Journal of Criminal Law and Criminology

Volume 85
Issue 3 *Winter*

Article 3

Winter 1995

The Relevance of Matching DNA: Is the Window Half Open Or Half Shut

D.H. Kaye

0091-4169/95/8503-0676
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY
Vol. 85, No. 3
Printed in U.S.A.

CRIMINOLOGY

THE RELEVANCE OF "MATCHING" DNA: IS THE WINDOW HALF OPEN OR HALF SHUT?

D.H. KAYE*

The basic concepts are always the hardest. This is particularly true in the study of evidence. Relevant evidence is evidence that alters the probability of a fact that matters, and relevant evidence generally is admissible unless it is too prejudicial.[1] Despite this seemingly simple formulation, questions of relevance relating to DNA evidence—even to the most elementary concept of a "match" between two DNA samples— can be confusing.

Well established methods of molecular biology permit laboratories to compare DNA from a crime scene to DNA from a defendant.[2] If the DNA in these samples is similar, the match usually is powerful evidence that the incriminating DNA came from the defendant. To describe the incriminating effect of the resemblance, scientists may use numbers. The numbers seen most frequently in criminal cases are

* Regents' Professor, Arizona State University, College of Law, Box 877906, Tempe AZ 85287-7906 (602 965-2922; k@asu.edu). I am grateful to Bruce Budowle, Jay Koehler, William Thompson, Bruce Weir, and especially David Balding for comments on a draft of this paper.

[1] *See* FED. R. EVID. 401 ("'Relevant evidence' means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."); FED. R. EVID. 402 ("All relevant evidence is admissible, except as otherwise provided . . . ."); FED. R. EVID. 403 ("Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.").

[2] The DNA at the crime scene may come from the defendant's blood, semen, or saliva. In other cases, the victim's DNA may be found on the defendant's property. *See, e.g.*, State v. Bible, 858 P.2d 1152 (Ariz. 1993); People v. Castro, 545 N.Y.S.2d 985 (Sup. Ct. 1989). Although the discussion here explicitly refers to only the former situation, the analysis applies to both methods of linking a defendant to a crime.

numbers referred to as the match frequency. These are estimates of the relative frequency of other people whose DNA also would match the crime scene DNA according to the criteria the laboratory uses for declaring a match.[3]

In court, the most successful objection to these numbers has been that the standard computational method does not account adequately for "population structure."[4] Arguments targeting the criteria for declaring a match have not fared as well.[5] In a recent article on the "DNA War," Professor William C. Thompson sought to correct this situation.[6] Already a veteran of a few skirmishes in the war,[7] Professor Thompson believes that when a defendant argues that "the match criteria should have been drawn in a narrower manner that excludes him," the finding of a match "could properly be excluded under Federal Rule 403."[8] At the very least, he believes that the court should

---

3 Some courts seem to require such testimony. *E.g.*, Nelson v. State, 628 A.2d 69, 76 (Del. 1993) (finding trial court's exclusion of match frequency "inherently inconsistent" with its admission of testimony of a match, because "without the necessary statistical calculations, the evidence of the match was 'meaningless' to the jury."); State v. Brown, 470 N.W.2d 30, 31-33 (Iowa 1991) (ruling expert testimony that "the likelihood of a person matching in all four fragments . . . would be one in several billion" admissible, since "[w]ithout statistical evidence, the ultimate results of DNA testing would become a matter of speculation."); State v. Vandebogart, 616 A.2d 483, 494 (N.H. 1992) ("A match is virtually meaningless without a statistical probability expressing the frequency with which a match could occur."). It would not, however, be "meaningless" to inform the jury that two samples match and that this match makes it more probable, in an amount that is not precisely known, that the DNA in the samples comes from the same person. Also, when all estimates of the frequency are in the millionths or billionths, it would not be meaningless to inform the jury that there is a match that is known to be extremely rare in the general population. D.H. Kaye, *The Forensic Debut of the National Research Council's DNA Report: Population Structure, Ceiling Frequencies, and the Need for Numbers*, 34 JURIMETRICS J. 369, 380-82 (1994).

4 United States v. Jakobetz, 747 F. Supp. 250, 263 (D. Vt. 1990) ("substructure is arguably the weakest link of the DNA profiling chain"), *aff'd*, 955 F.2d 786 (2d Cir.), *cert. denied*, 113 U.S. 104 (1992). *See also* COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, NAT'L RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE 79 (1992) [hereinafter NRC REPORT] ("whether actual populations have significant substructure for the loci used" is "the key question"). "Substructure" refers to the presence of subgroups that have distinctive DNA patterns and that tend to mate among themselves. For the view that courts sometimes have erred in excluding the estimates on this ground, see D.H. Kaye, *DNA Evidence: Probability, Population Genetics, and the Courts*, 7 HARV. J.L. & TECH. 101 (1993).

5 *See* Kaye, *supra* note 4; William C. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the "DNA War,"* 84 J. CRIM. L. & CRIMINOLOGY 22, 50-51 (1993) ("Only one court has held that the dispute over the FBI's matching procedure precludes the admissibility of the test under *Frye*.").

6 Thompson, *supra* note 5.

7 *See, e.g.*, People v. Simpson, No. BA097211 (Los Angeles County Superior Court, Motion to Exclude DNA Evidence, Oct. 4, 1994); State v. Anderson, No. CR 46255 (Dist. Ct., Bernalillo County, Memorandum in Opposition to the Introduction of the FBI's DNA Evidence, Apr. 3, 1990).

8 Thompson, *supra* note 5, at 60. Professor Thompson recommends exclusion only

allow expert testimony to the jury that the match criteria are too inclusive.[9] In contrast, I have argued that:

> Disputes over the strictness of particular windows or *the* optimal match window—when there is no such thing—may confuse and perplex the jury when it considers the probative value of a match. As a result, a court has discretion to exclude this testimony.[10]

This Article shows that Professor Thompson's call for exclusion of DNA evidence or for expert testimony on the definition of a match is not responsive to the concern that animates and underlies his proposal. If anything, a more radical response—rejecting the very notion of a simple "match"—may be in order. To reach these conclusions, Parts I and II describe the basic ideas of "matching" and "binning" that are central to the usual analysis of DNA evidence. Part III outlines a theory of probative value that illuminates the limitations of match-binning. With these fundamentals in place, Parts IV and V address Professor Thompson's concern and demonstrate that, depending on circumstances yet to be established, it could warrant a modification in the way that DNA evidence is presented, though not necessarily the one that Professor Thompson endorses.

The analysis demands a careful definition of probative value and a clear understanding of the matching process and its statistical properties. These can be obtained only in the currency of some mathematical notation and concepts. This price, however, is well worth paying. At the most practical level, the exercise promises to improve the presentation of DNA evidence. In addition, it illuminates the workings of a theory of probative value and inductive proof that evidence scholars have been propounding for some time.[11]

---

when "the defendant makes a strong showing prior to trial that the observed discrepancies are unlikely to have arisen if the 'matching' prints have a common source." *Id.*

9 *Id.* at 60 n.171.

10 Kaye, *supra* note 4, at 114-15 (footnotes omitted).

11 Although this theory has deep roots in the works of logicians, philosophers, psychologists, and statisticians, much of the exposition and development in the legal literature has come from the University of Michigan Law School. *See, e.g.*, Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439 (1986), *reprinted in* PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM 61 (Peter Tillers & Eric D. Green eds., 1988) [hereinafter Tillis & Green]; Richard Friedman, *Character Impeachment Evidence: Psycho-Bayesian [!?] Analysis and a Proposed Overhaul*, 38 UCLA L. REV. 637 (1991); Richard D. Friedman, *A Close Look at Probative Value*, 66 B.U. L. REV. 733 (1986). The "Michigan School," however, emphasizes Bayes' rule for combining new information with prior probabilities of hypotheses. The approach used in this Article focuses on the concept of relative likelihood which can be embraced without a commitment to Bayesian reasoning in general. *See, e.g.*, D.H. Kaye, *Introduction: What is Bayesianism?*, *in* Tillers & Green, *supra*, at 1. There are critics of the Michigan School, but they do not always offer clear alternatives. *See, e.g.*, Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 Nw. U. L. REV. 604, 627-28 (1994) (sketching a "theory of juridical evidence" that "reduces to the proposition that a disinterested fact finder reconstructs the past based

I. MATCHING

The most common form of DNA analysis in criminal cases utilizes four or five "single locus VNTR probes" to produce a "multilocus genotype," or, more simply, a "DNA profile." Technicians use techniques of molecular biology to excise fragments of chromosomes[12] that begin and end with certain sequences of DNA base pairs from samples found in blood, semen, or other material containing sufficient DNA.[13] The beginning and ending sequences are chosen so that the material they bracket tends to vary in size from person to person.

The lengths of the DNA fragments are measured by seeing how far they move, relative to DNA fragments of known lengths, through a slab of gelatinous material, under the pull of an electric charge. The procedure is known as electrophoresis. In a given time, shorter fragments (with low molecular weight) migrate farther down the electrophoretic gel than longer fragments (with high molecular weight).[14] Fragments of approximately the same size will stop at more or less the same point, coalescing to form a spot or band. By measuring the locations of the bands it is possible to determine the *approximate* lengths of the fragments.[15] The prevailing method of agarose gel electrophoresis of VNTR fragments is not sensitive enough to distinguish between

---

on all the observational inputs available at the moment of judging").

12 A chromosome is a tightly coiled molecule of DNA packed in protein. Each parent supplies one of 23 different chromosomes, so the human genome consists of 46 chromosomes arranged into 23 pairs.

13 For present purposes, DNA may be thought of as a book whose letters are "base pairs." Bacterial enzymes are used to cut the DNA "text" into fragments. A given "restriction enzyme" binds to DNA when it encounters a certain short sequence of DNA base pairs and cleaves the DNA at a specific site. For example, the Hae III enzyme acts like a scissors cutting the text . . .GGCC. . .GGCC. . . to yield the fragment CC. . .GG. "Digesting" DNA with such an enzyme usually produces fragments ranging from several hundred to several thousand base pairs in length.

14 The molecular weight of a compound is equal to the total mass of its constituent atoms. Since DNA fragments all have pretty much the same mix of atoms, a fragment that has twice the length of another also has about twice the molecular weight.

15 In many instances, the bands are sharp and clear; sometimes, they are fuzzy. Indeed, in some cases there are disputes over the presence or absence of bands, since laboratory artifacts sometimes produce missing bands or spurious bands. *See* David J. Balding & Peter Donnelly, *How Convincing is DNA Evidence?*, 368 NATURE 285, 286 (1994); William C. Thompson & Simon Ford, *The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints*, *in* FORENSIC DNA TECHNOLOGY 93, 138 (M.A. Farley & J.J. Harrington eds., 1991).

This Article limits itself to the problem of determining the size of the match window. To keep the focus sharply on this issue, it considers only cases in which there is no doubt as to the existence and measurements of the bands on the autoradiogram. When they are present, uncertainties created by missing or extra bands and deviations from objective procedures for ascertaining the positions of bands can degrade the probative value of a declared match. *Cf.* D.H. Kaye, *Comment: Uncertainty in DNA Profile Evidence*, 6 STAT. SCI. 196, 196-97 (1991) (noting the potential importance of these factors).

fragments that are extremely close in size. Measurement error creeps in.

The likely extent of this measurement error is found through reproducibility studies. For example, a laboratory may compare the fragment lengths in DNA obtained from vaginal swabs with that of DNA in blood taken from the same woman.[16] Such studies enable the laboratory to choose a "window" wide enough to be likely to result in a match when two samples come from the same source. For instance, studies conducted by the South Carolina Law Enforcement Division DNA laboratory found that corresponding bands never differed by more than ±2.8% of their average length.[17] The FBI reports that the biggest difference observed in its laboratory is less than ±2.5%.[18]

Once a laboratory establishes a numerical matching rule, technicians can compare the crime scene fragment lengths to fragment lengths from a suspect. Suppose that a window of width ±2.5% is chosen.[19] If a visual match is declared, and if all pairs of corresponding bands in the two profiles differ by no more than ±2.5%, then the two profiles are said to match.

The match window does not have to be any particular size. Big windows are likely to result in the declaration of a match when the DNA samples come from the same person. In this case, the procedure is highly sensitive—it has a large probability of declaring a match when the samples have a common source. Along with a high sensitivity, a suitable window should yield a high specificity. Specificity refers to the probability that the procedure will not declare a match when the samples do *not* come from the same person. A wide window enhances sensitivity; a narrow window favors specificity. When the sensitivity is high, DNA from the same sources almost always match and there are relatively few false negatives.[20] When the specificity is high,

---

16 Bruce Budowle et al., *Fixed-Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci, for Use in Forensic Comparisons*, 48 AM. J. HUM. GENETICS 841 (1991). Since epithelial cells from the swabs and blood cells from the same person contain identical DNA, the measurements would be the same in the absence of measurement error. An alternative corresponding to rape cases would be to compare semen and blood samples from the same man.

17 B.S. Weir & B.S. Gaut, *Matching and Binning DNA Fragments in Forensic Science*, 34 JURIMETRICS J. 9 (1993).

18 Budowle et al., *supra* note 16, at 844.

19 Because the speed at which fragments move through the gel varies over different regions of the gel, the absolute size of the experimental error tends to be larger for larger bands. Hence, the window is a percentage of the molecular weight or size of a fragment. In light of the studies cited above, the FBI uses a window of ±2.5%; the South Carolina laboratory uses ±2.8%.

20 For the purposes of this Article, a false negative is a failure to declare a match between two samples of DNA from the same person. One group of researchers found that a match window of ±1.2% with three probes produced false negatives nearly 20% of the

DNA from different people almost always fail to match and there are few false positives.[21] Thus, statistical reasoning gives the operating characteristics of the matching procedure,[22] but scientific values do not dictate the choice of a unique match window.[23]

Once a reproducibility study shows that a window has reasonable specificity, the evidentiary analysis should focus on ensuring that the testimony about the match fairly describes the probative value of the laboratory results. Will testimony about the properties of different windows aid in this task? Are there situations where the difficulty of expressing the probative value makes it better to exclude the evidence of a match? To answer these questions, it is necessary to analyze plausible ways to express the probative value of the "match." At least three possibilities exist: proportions, similarity likelihood ratios, and match-binning likelihood ratios. Parts II and III describe these measures. Part IV applies them to expose the flaws in Professor Thompson's argument for admitting testimony about the desirability of smaller

---

time. The rates would be larger for four or five probes, since each probe creates more opportunities for a false negative to occur. I.W. Evett et al., *An Illustration of the Advantages of Efficient Statistical Methods for RFLP Analysis in Forensic Science*, 52 AM. J. HUM. GENETICS 498, 502 (1993).

21 For present purposes, a false positive is a declaration of a match between DNA from different people.

22 It is possible to argue that some minimum values of the sensitivity and specificity of the test are needed to establish scientific validity. The need for scientific validity (under the rubric of "evidentiary reliability") is emphasized in Daubert v. Merrill Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993). The first two federal appellate cases applying the *Daubert* standard of "good [scientific] grounds" to DNA evidence are United States v. Bonds, 12 F.3d 540 (6th Cir. 1993) (affirming the admission of a match on the record developed in the extensive pre-trial hearing in United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991)), and United States v. Chischilly, 30 F.3d 1144 (9th Cir. 1994) (affirming the admission of a match and a "conservative" estimate of the genotype frequency among Navajos).

The need to demonstrate scientific validity under *Daubert* may require reproducibility studies to establish that a laboratory is capable of declaring a match for DNA samples from the same person. *See* Thompson, *supra* note 5, at 56-57. But the minimum error rates that make a clinical test acceptable to scientists (or to anyone else) depend on the costs of false positives and false negatives in light of the uses to which the test is put. Low specificity, for example, may be acceptable in a test, like a pap smear, that is used to screen for certain conditions whose presence can be confirmed by more definitive (and more expensive) tests. Rather than ask the abstract and unanswerable question of what values of sensitivity and specificity science demands, it seems more profitable for courts to inquire directly into the helpfulness of the test to the jury. In other words, the "science" requirement of *Daubert* should apply to the determination of whether the analytic procedure is capable of performing as warranted by scientists and technicians—whether it is indeed scientific—but not to whether the advertised or known level of performance is satisfactory for legal purposes.

23 *See* Kaye, *supra* note 4, at 114-15 n.67; Kathryn Roeder, *Rejoinder*, 9 STAT. SCI. 267, 275 (1994) ("the 'objective' match criterion . . . is, in fact, simply an arbitrary rule."); *see also infra* note 43.

match windows. Part V describes an alternative approach that may be better tailored to the possible problem.

II. BINNING

The conventional procedure for presenting incriminating DNA evidence in this country entails not just the declaration of a match, but also an estimate of the relative frequency with which other people in some population would match the crime scene DNA. This matching proportion of the population, designated as some number P, plainly has some connection to the probative value of the match.[24] If P approaches 1, then so many innocent people would match that the match proves little; if P approaches zero, then it may prove a great deal.[25]

It is possible to estimate P from a random sample of people from a suitable reference population. For example, consider a case in which a woman is kidnapped, assaulted, and robbed at a rest stop on an interstate highway by a truck driver who is white,[26] and in which the DNA comparison is limited to a single band.[27] With DNA from a suitable random sample of Caucasians, P could be estimated on the basis of the proportions of bands that fall into predefined intervals of molecular weights. Figure 1 shows a hypothetical distribution of band weights over ten intervals, or fixed "bins,"[28] numbered zero through

---

24 *See supra* note 4.

25 Often, P is interpreted as the probability that a randomly selected, or innocent person will have DNA that matches the crime scene sample. If so, it is akin to what statisticians usually call a "p-value" or "significance probability"—the probability under the "null hypothesis" of observing data within a region (the match window) that (when small enough) prompts rejection of the "null hypothesis." Under the null hypothesis that someone other than the defendant is the source of the crime scene DNA (and assuming that suspects are not selected because of factors related to their DNA profiles and that no laboratory errors occur), the probability of declaring a match is just the proportion of people in the population who have profiles that satisfy the match criteria.

26 *See* United States v. Jakobetz, 955 F.2d 786 (2d Cir.), *cert. denied*, 113 U.S. 104 (1992).

27 Computing P when more than one band is compared in the two samples has proved intensely controversial. For reviews, see, for example, Kaye, *supra* note 4; Kathryn Roeder, *DNA Fingerprinting: A Review of the Controversy*, 9 STAT. SCI. 222 (1994); Thompson, *supra* note 5. The issue here, however, is not the details of combining the information from the analyses at each band, but what use to make of even undeniably accurate estimates of P, either for a single band or for many bands. For convenience, the text only considers the simplified, one-band comparison.

28 A variation uses "floating" bins. *See, e.g.*, Keith L. Monson & Bruce Budowle, *A Comparison of the Fixed Bin Method with the Floating Bin and Direct Count Methods: Effect of VNTR Profile Frequency Estimation and Reference Population*, 38 J. FORENSIC SCI. 1037 (1993); B.S. Weir, *Independence of VNTR Alleles Defined as Floating Bins*, 51 AM. J. HUM. GENETICS 992 (1992). Because the differences do not affect the results of the analysis here, the discussion is limited to fixed bins.

nine.[29]

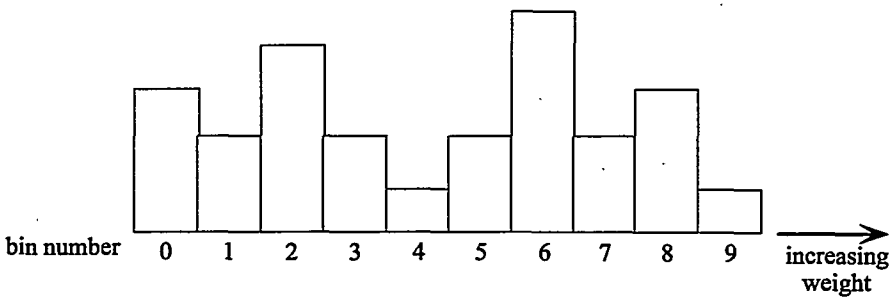

FIGURE 1
HISTOGRAM OF BAND WEIGHTS.

The height of each column is proportional to the fraction of bands that lie within each "bin," which are 12%, 8%, 16%, 8%, 4%, 8%, 20%, 8%, 12%, and 4%, respectively. No bands lie outside the numbered intervals.

Suppose that the crime scene DNA band is in the center of bin 0 and that the bins are much wider than the match window.[30] Then an estimated twelve percent of the bands from the Caucasian population would "match." Under the simplified scenario of examining a randomly chosen single band, P=.12. In this hypothetical case, the prosecution might offer testimony that the samples match and that such matches would be seen in no more than twelve percent of the Caucasian population.

III. PROBATIVE VALUE

The best developed and most plausible theory of probative value

[29] The choice of ten bins is illustrative; in practice, more may be employed. *See, e.g.*, John S. Waye & Ronald M. Fourney, *Forensic DNA Typing of Highly Polymorphic VNTR Loci*, *in* 3 FORENSIC SCIENCE HANDBOOK 358, 385 (Richard Saferstein ed., 1993). To avoid values of P that underestimate the proportion of matches that would be obtained in the population, the width of the bins must be such that no bin is smaller than the match window. In practice, considerably larger bins often are used, arguably leading to values of P that are too large and thus understate the rarity of a match. *But see* Thompson, *supra* note 5, at 67 (reporting that only Cellmark Diagnostics uses bins larger than its "match standards").

[30] If the crime sample band were close to the boundary of two bins, then it ordinarily would be safe to use the larger of the two bin heights—12% in this example. An extremely cautious approach would be to use the sum of the heights—12 + 8 = 20%. *See* NRC REPORT, *supra* note 4.

articulated by legal scholars builds on a statistical concept known as the likelihood ratio, or, more generally, the likelihood function.[31] The idea is elegantly simple and exceedingly powerful. Evidence has probative value with respect to competing hypotheses about the process that generated the evidence. With DNA measurements there are only two hypotheses: S, that the data arose from a common source of DNA, and -S, that they did not.[32] To the extent that the data are more probable under S than -S, they support S over -S. To measure this degree of support, the "likelihood" of S is defined as being proportional to the conditional probability of the data given that S is true.[33] A convenient measure of probative value is the ratio of the likelihoods for the two hypotheses:[34]

$$L = \frac{Likelihood\ (S)}{Likelihood\ (\text{-}S)} = \frac{Pr(data|S)}{Pr\ (data/\text{-}S)} \qquad (1)$$

As long as L is greater than 1, the DNA data support the hypothesis S. When L is less than 1, they support -S. When L equals 1, they are expected to arise as frequently when one hypothesis is true as when the other is true. Hence, they provide no reason to favor one hypothesis over the other and are irrelevant to choosing between them. Thus, the more the value of L departs from unity, the greater the probative value of the data.

Despite its mathematical form, equation (1) is ambiguous: How should the "data" be represented? At the most detailed level, DNA data consist of two sets of numbers denoting the approximate molecular weights of the fragments from the two samples. The useful information in these two profiles involves both the observed weights of each pair of corresponding fragments and the similarity in each pair of measurements. For instance, suppose that a single-locus probe produces a band that seems to be 2328 base pairs long for the crime scene sample and a corresponding band 2352 base pairs long for the defendant's DNA,[35] and that this is the only band to consider. The

---

31 *See supra* note 11.

32 The hypothesis -S can be thought of as the disjunction of hypotheses about everyone else in the relevant population. If all the people who might be the source of the crime scene DNA are labelled 1 through n (with the defendant being designated as person 1), and if the defendant, who is the source of the DNA being compared to the crime scene DNA, is not the source of the latter sample, then either person 2 is the source, person 3 is, and so on, through person n.

33 *See* A.W.F. EDWARDS, LIKELIHOOD: AN ACCOUNT OF THE STATISTICAL CONCEPT OF LIKELIHOOD AND ITS APPLICATION TO SCIENTIFIC INFERENCE (1972).

34 *See, e.g.*, Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977). For other possible expressions of probative value, see Friedman, *supra* note 11; D.H. Kaye, *Comment: Quantifying Probative Value*, 66 B.U. L. REV. 761 (1986).

35 *See* Thompson, *supra* note 5, at 58 (figure 3).

measurements are each 12 base pairs away from their mean of 2340 base pairs. The probative value of this pair of measurements depends on how rare bands in this region are, as well as the measured difference of 24 base pairs (1.03%). This difference, (d), together with the mean weight x, may be substituted for the expression "data" in (1) to yield the "similarity likelihood ratio":

$$L = \frac{Pr(x, d|S)}{Pr(x, d|\text{-}S)}. \qquad (2)$$

For a particular mean fragment weight x, the closer d is to 0, the better the corresponding bands match, and the greater the value of L. For a given separation d, the rarer the measurement weights are in the population, the greater the value of L.[36]

The match-binning approach does not make full use of d and x. "Pure" match-binning merely asks whether d is inside or outside the predefined match window.[37] This offers a simpler, but cruder repre-

---

[36] On computing L, see David Jarjoura et al., *Likelihood Ratios for Deoxyribonucleic Acid (DNA) Typing in Criminal Cases*, 39 J. FORENSIC SCI. 64 (1994); Kaye, *supra* note 4, at 162 n.254; Roeder, *supra* note 27. While grossly oversimplified, this exposition is intended to capture the features of L that are important to the analysis of a claim that a match window is too broad.

[37] In more technical jargon, instead of treating d as the continuous variable that it is, match-binning converts d into a binary variable, throwing away information. In practice, however, the declaration of a match is not so black-and-white. For one thing, human judgment may enter into the determination of whether a faint band actually is present or whether seemingly extra bands are laboratory artifacts. *See supra* note 15. Furthermore, even if the full set of bands to consider is clear, and even if they meet the statistical criteria for a match, subjective judgment may be involved in deciding whether these bands "really" match. Realizing that an electrophoretic artifact known as band-shifting sometimes results in all the bands being shifted in more or less the same proportion from their true values, an examiner may decide that two profiles match when all the differences between the samples are in the same direction and close to the edge of the match window, but not when the differences go in both directions. The former situation is consistent with a very close match distorted by band-shifting; the latter is not. To the extent that examiners rely on the actual size of the differences and not just the fact that the differences fall within the preset window in declaring matches, the match-binning procedure does use some of the information about the degree of matching.

It is difficult to model the effect of these subjective aspects on probative value. For the purpose of analyzing the relevance of testimony about match windows, this Article assumes that there are no ambiguities in the existence of bands. As for the second, subjective phase of matching, since examiners presumably screen out the cases in which matches are weak more often than they discard excellent matches, one might argue that their review can only enhance the probative value of a purely statistical match. Even if the subjective judgments were entirely arbitrary—made, for example, by flipping a coin to decide whether a statistically tolerable match will be deemed a match—the subjective component still could not lower the probative value. Consequently, the type of match-binning described and analyzed in this Article—involving purely statistical matching—may understate the probative value of those matches that arise in practice. *But see infra* note 46. On the other hand, if Professor Thompson is correct in his claims that examiners frequently declare matches that do not satisfy statistical criteria, then the analysis applicable to purely statistical match-

sentation of the "data" in (1). The fragments either match (m=1) or they do not (m=0)—there are no shades of gray:

$$L = \frac{Pr(m|S)}{Pr(m|\text{-}S)}. \qquad (3)$$

This likelihood ratio is relatively easy to compute. Suppose, as in the example of Part II, that there is a match with a bin frequency of P=.12. The denominator is the probability of declaring a match when -S is true.[38] When the suspect and the source of the crime scene DNA are unrelated, this probability cannot exceed a generous estimate[39] of the proportion of the Caucasian population with a matching band:[40]

$$L = \frac{Pr(m|S)}{P}. \qquad (4)$$

The numerator is the probability of a match given S. It depends on the size of the match window. It approaches one (1) as the match

---

ing could overstate the probative value of the matches that arise in practice.

38 *But cf.* David J. Balding et al., *Comment: Some Causes for Concern about DNA Profiles*, 9 STAT. SCI. 248, 249 (1994) ("[O]ne needs to consider each possible culprit and to assess the probability that they would match the crime profile, given that the defendant does."); B.S. Weir, Conditional Genotypic Frequencies in Forensic Analysis (Oct. 19, 1993) (unpublished manuscript focusing on "[t]he conditional probability that the accused suspect has the matching profile, given that the perpetrator of the crime has the profile and that these people are not the same").

39 *See supra* note 29.

40 Balding et al., *supra* note 38, at 249, contend that one must consider the likelihoods of each alternative hypothesis in note 32 *supra*, and that "[s]tatistical analyses of forensic databases are not directly relevant to an assessment of the conditional probabilities" that determine these likelihoods. They argue that the very fact that the defendant matches makes it more probable that relatives of the defendant and others in the same ethnic subpopulation also would match. Because of these "positive correlations," they conclude that using the population proportion is "both false and detrimental to the defence." *Id.* So too, Balding and Nichols maintain that the conditional probability differs from the estimated population proportion because "[i]f the suspect is innocent then we have observed two copies of a profile which is believed to be extremely rare in the database population," and "this observation makes more likely the possibility that the profile is more common in the suspect population than in the database population." David J. Balding & Richard A. Nichols, *DNA Profile Match Probability Calculation*, 64 FORENSIC SCI. INT'L 125, 128 (1994). Thus, these researchers propose using the proportions for the defendant's ethnic subpopulation rather than the broad racial population in estimating the likelihoods. Balding et al., *supra* note 38, at 249. This Article does not pursue these ideas—not because they are unimportant, but because the topic is the objection that the match window is too wide rather than the choice of the reference population or subpopulation.

Second, the use of the proportion in the database for the denominator of the likelihood ratio ignores the possibility of laboratory error. This possibility is important, but it is another distinct problem. The complaint that P is misleading in a way that justifies testimony about the size of match windows does not turn on the risk of such error. For a review of possible ways to ensure that jurors do not ignore this consideration, see Kaye, *supra* note 4, at 156-58. The issue also is addressed by several commentators in Roeder, *supra* note 27.

window grows very large. The FBI, for example, uses a match window that is slightly larger than the biggest difference observed in its laboratory for the same sample measured twice.[41] Some critics have complained that this window amounts to plus-or-minus four standard errors of measurement.[42] This, they say, is too large.[43] Although the figure of four standard errors may be misleading,[44] there are indications that the errors follow the ubiquitous normal curve.[45] Since 99.994% of the area under a normal curve lies within four standard errors to either side of its center, an examiner who always declares a match for a band when the band from the crime sample and the corresponding band from the defendant's sample are within

---

41 Budowle, *supra* note 16, at 844. Specifically, the FBI requires that bands be separated by no more than ±2.5% of their mean molecular weight to declare a match. *See supra* note 19. It also requires the examiner to say that profiles that meet this criteria look like matches.

42 *E.g.*, Seymour Geisser, *Some Statistical Issues in Medicine and Forensics*, 87 J. AM. STAT. ASS'N 607, 609 (1992). The standard error is the standard deviation of the difference between two measurements; it indicates the variability of the differences. For a sample batch of differences, it is computed by subtracting each difference from the mean difference, squaring this deviation from the mean, adding up all the squares, dividing by the size of the sample, and extracting the square root. In other words, it is the root mean square deviation. If all the differences are identical, the deviations from the mean are all zero, and the standard error will be zero. If all lie at two polar extremes, all the deviations are large, and the standard error will be large.

The standard error of DNA molecular weight measurements is not really a constant percentage of the true weight. *See* I.W. Evett et al., *An Efficient Statistical Procedure for Interpreting DNA Single-Locus Profiling Data in Crime Cases*, 32 J. FORENSIC SCI. SOC'Y 307, 316 (1992) (table 2). In the interest of brevity, this Article ignores this complication.

43 Geisser, *supra* note 42 ("an extraordinarily wide net"). Some experts in United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991), *aff'd sub nom.*, United States v. Bond, 12 F.3d 540 (6th Cir. 1993), claimed that the FBI's window is too wide because "the level of risk for false positives (matches)" exceeds that which is "generally accepted as reliable in the research and clinical communities for similar applications of this technology." Thompson, *supra* note 5, at 48-49 nn.122-23 (quoting Report of Professor Peter D'Eustachio, *An Evaluation of the FBI's Environmental Insult Validation Study and the FBI's Quantitative Matching Criteria*, at 16, *in Yee*, 129 F.R.D. 629). Since these communities have no similar applications of the technology, however, this version of the criticism seems odd. Unlike a diagnostician or (perhaps) a research scientist, the forensic expert need not and should not choose between two hypotheses. Rather, the forensic expert faces the unique challenge of presenting the scientific evidence bearing on the hypotheses so that the jury will appreciate its actual probative value. *See* Kaye, *supra* note 4.

44 The ±4 figure pertains to the standard error derived from K12 cell line (control DNA) measurements. These involve fresh DNA that produces less variable measurements than forensic samples, reducing this size of the standard error, and making it appear that a given difference represents a larger number of standard errors. Thus, Eric S. Lander, *Invited Editorial: Research on DNA Typing Catching Up with Courtroom Application*, 48 AM. J. HUM. GENETICS 819, 820 (1991), reports that the Bureau's laboratory has a standard error of about 1.5% of the molecular weight of their mean. If so, the ±2.5% match window corresponds to only ±1.7 standard errors.

45 Donald A. Berry et al., *Statistical Inference in Crime Investigations Using Deoxyribonucleic Acid Profiling*, 41 APPLIED STAT. 499 (1992).

the ±2.5% window will be correct in 99.994% of the cases when S is true:[46] $\Pr(m=1|S) = .99994$. More generally, if $A(w)$ stands for the area within a match window that extends a distance $\pm w$ about the center of the normal error curve for duplicate measurements, the match-binning likelihood ratio becomes:

$$L = A(w)/P. \quad (5)$$

Two salient points emerge. First, there is some tension between the similarity likelihood ratio (2) and the match-binning ratio (5). The former attends to the precise degree of matching; the latter pretends that all "matches" are equally revealing. In a case in which the value obtained from (2) is much less than that obtained with (3), the match-binning likelihood ratio can give an inflated impression of probative value.[47]

Second, even within the confines of (5), P is incomplete because it relates solely to the denominator of L. A juror who implicitly thinks in terms of likelihood ratios, and who is not apprised of $A(w)$, might treat $A(w)$ as if its value were 1, and weigh the evidence as if its probative value were given by the following "P only" likelihood ratio:

$$L = 1/P. \quad (6)$$

Comparing (5) and (6), if $A(w)$ were dramatically less than 1,[48] then presenting only the value of P could lead a jury to give the fact of a match more weight than it deserves. If this is the argument for excluding match-binning evidence, however, it has little force. As Part IV explains, the match windows are wide enough to ensure that matches are probative.

---

[46] This applies to a single band. For most people, eight or more bands are compared. According to Bonferroni's inequality, the probability of at least one mismatch in eight comparisons is, at most, 8 x .00006 = .00048. Therefore, the probability of eight matches cannot be any less than 1 - .00048 = .99952. Examiners who rely on their impressions of band-shifting to exclude some statistical matches, *see supra* note 37, may lower the probability of declaring a match when the DNA samples originate from the same individual.

[47] The results of matching with a subjective element could have a likelihood ratio closer to that given by (2). *See supra* note 37.

[48] If $A(w)$ were less than P, then the likelihood ratio (5) would indicate that the evidence does not support the hypothesis of a common source. *Cf.* Thompson, *supra* note 5, at 57. To some extent, this reinforces the conclusion of Part I. Unless the laboratory can demonstrate that it can reliably declare matches for DNA from the same source, the evidence should be inadmissible. However, in terms of logical relevance as opposed to the more demanding *Daubert* standard, what counts is not whether the chance of declaring a match when S holds close to 1, but whether it is substantially larger than the chance under -S. If the DNA bands are extremely rare, then the latter quantity can be quite small, so that even modestly reliable DNA tests may produce highly probative matches.

IV. THOMPSON'S ARGUMENT

Professor Thompson agrees that in many cases the demand for narrower windows would not assist the jury. His argument for admitting testimony about the width of the match window is terse. He cautiously suggests that it "might well be appropriate" to exclude testimony of a match "if the defendant makes a strong showing prior to trial that the observed discrepancies are unlikely to have arisen if the 'matching' prints have a common source."[49] This showing, he thinks, would support a defense argument that "the match criteria should have been drawn in a narrower manner."[50] Noting the inability of "lawyers and jurors to deal with complex scientific and statistical arguments,"[51] Thompson concludes that Rule 403 "could properly" justify exclusion of the evidence.[52] But if testimony as to a match is to be admitted, he insists that reasonable people could not differ on the relevance of testimony favoring a smaller window that would not produce a match.[53]

His argument, however, is not conclusive. In more detail, the argument splits into two branches, as follows:

I When the numerator of the appropriate likelihood ratio is substantially less than 1, *P* is a misleading measure of the probative value of a match.

IIa Expert testimony that a smaller window would exclude the defendant might prevent the jury from being misled by the discrepancy between $1/P$ and the more appropriate likelihood ratio.

IIIa Therefore, courts should allow expert testimony that a smaller match window would exclude the defendant.

IIb Alternatively, explaining to a jury that the numerator of the appropriate likelihood ratio is substantially less than 1 would be too hard or ineffective.

IIIb Therefore, courts should exclude evidence of the match and of the match frequency *P*.

This formulation discloses an ambiguity. Which likelihood ratio—(2) or (5)—must be substantially less than 1? If (I) refers to (5), then for the hypothetical case outlined in Part II, the likelihood ratio is $L = .99994/.12 = 8.327$. A juror who implicitly uses the oversimplified "*P* only" ratio (6) would think that $L = 1/.12 = 8.333$—hardly a mistake that cries out for redress. This is because the condition in (I) that might produce a misleading picture is not satisfied—the numera-

49 Thompson, *supra* note 5, at 60.

50 *Id.*

51 *Id.* at 60 n.171. *See also id.* at n.170.

52 *Id.* at 60 & n.171.

53 *Id.* at 60 n.171.

tor is not substantially less than 1. Instead, it is all but identical to 1.

Suppose, however, that the condition in (I) were fulfilled—that a laboratory used a window amounting to ±1 standard errors instead of ±4. Then the numerator $A(w)$ becomes .6827, a value that is substantially less than 1. It follows that $L = .6827/.12 = 5.689$ instead of the "P only" value of 8.333.[54] In this event, the argument still fails, because (IIa) is false. Expert testimony that a smaller window would exculpate the defendant would not prevent the jury from being misled. The expert concerned with the match-binning likelihood ratio would have to testify that the laboratory should use a larger window—exactly the opposite of the testimony that has been heard to date.[55]

So if the focus is on match-binning proportions, the argument founders at the outset. If Professor Thompson was writing about the match-binning likelihood ratio, he would be correct in identifying $\Pr(m|S)$ as a possible source of cognitive error. Instead of justifying smaller match windows that would exonerate the defendant, however, a small $\Pr(m|S)$ would motivate larger windows that would continue to inculpate the defendant.

Thus, Thompson must not be concerned with the match-binning likelihood ratio. Instead, his argument may be that when the numerator $\Pr(x,d|S)$ of the similarity likelihood ratio (2) is "quite small,"[56] P is misleading and either justifies testimony that the match window is too large or warrants exclusion of the match altogether.

But even under this interpretation, what is in question is the full likelihood ratio, not just its numerator, and the event of an observed distance d between two bands is mathematically very different from the existence or non-existence of a match. Whereas the numerator of the match-binning likelihood ratio may be calculated as an area $A(w)$ under a part of a normal curve, the numerator of the similarity likeli-

---

54 This assumes that the number of bins remains unchanged. With a smaller match window, more bins with smaller bin widths would be appropriate. This change, however, could only reduce P (since fewer bands would fall into most of the smaller bins). Reducing P would raise both the correct value of the match-binning likelihood ratio, $A(w)/P$, and the perceived value, $1/P$. The illustrative numbers would change, but the general picture would not.

55 *See* United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991), *aff'd sub nom.*, United States v. Bonds, 12 F.3d 540 (6th Cir. 1993); *supra* text accompanying note 42.

56 Thompson writes: "In formal terms, the defendant should make a showing that *p(D/H1)* is quite small." Thompson, *supra* note 5, at 60 n.168. The expression *p(D/H1)*, in the notation used here, would be $\Pr(d|S)$. The text of his article imposes the condition in (I) that "the *observed discrepancies* are unlikely to have arisen if the 'matching' prints have a common source." *Id.* at 60 (emphasis added). The "observed discrepancies" are abbreviated as d in equation (2) and D in Thompson's expression for the difference likelihood ratio. *Id.* at 57 n.163. His definition of the likelihood ratio does not refer explicitly to the location of the "matching" bands, presumably to make the expression easier to read.

hood ratio is more like the area of a thin strip under the curve located a distance d from the peak.[57] Being a thin strip under a low-lying curve rather than a broad region under the entire match window $\pm w$, the numerator $\Pr(x,d|S)$ always will be small.[58] Likewise, the denominator in (2) is not the proportion P of matches within a fixed, overly broad window, as it was in (5). Now, the denominator is $\Pr(x,d|\text{-}S)$, the probability (actually, the probability density) of seeing two bands that represent the same DNA fragment separated by d and centered at x, and it changes as x or d changes. Just as the numerator can be much smaller than 1, this denominator can be much smaller than P. P, after all, is the proportion of matches estimated with a bin that is bigger than the match window, which, in turn, is bigger than the small difference d. $\Pr(x,d|\text{-}S)$ will involve but a small slice of this big bin.

Consequently, the "substantially less than 1" or "quite small" condition of (I) is neither necessary nor sufficient to establish that the "P only" likelihood ratio $1/P$ is too large. In principle at least, the numerator can be much less than 1, but the probability of seeing the same degree of matching with an innocent suspect also may be much less than P. Therefore, a small numerator in (2) does not establish that (2) is substantially less than (5). In short, it is not obvious that Thompson's criterion accurately picks out those cases that could involve the cognitive error at issue.[59] Once again, the argument for ex-

---

[57] This "thin strip" description is an approximation. Because d is a continuous variable, (2) is the ratio of probability densities rather than probabilities, and the numerator is given by the height of the normal curve at d rather than an area beneath the curve.

[58] Even two bands that are very close will produce a small value of $\Pr(x,d|S)$. For instance, discrepancies of just about one standard error occur well under 1/3 of the time for normally distributed errors. One third is the approximate probability of discrepancies of one *or more* standard errors. The quantity in the similarity likelihood ratio (2) is not this tail-end probability, but rather the probability density for exactly one standard error. *See supra* note 57. This density might be given by the height of the normal curve at one standard error, which is 0.24. The height of the curve for two perfectly aligned bands—the best result one could hope for when the samples are from the same source—is less than 0.40.

Of course, the normal curve is merely a model that fits the observed distribution of errors in repeated measurements. A particular reproducibility study might show that more than 40% of the replicates are perfectly aligned, or that fewer are perfectly aligned.

[59] One study using one version of the difference likelihood ratio with data on three of Lifecodes' probes reports that, on average, the combination of bands at each of the three loci produce monotonically increasing ratios with decreasing values of d. The study finds that at a separation of three standard errors, one probe gives an average likelihood ratio below 1 (supportive of -S). At four standard errors, the average ratio for each probe is much less than 1. Jarjoura et al., *supra* note 36, at 70 (table 1). The extent to which this study indicates that focusing on the numerator of the difference likelihood ratio might identify the troublesome cases is unclear. Being averages, these values do not pertain to all cases. Moreover, they are averages for a ratio rather than just the numerator. If Professor Thompson wished to test his criterion empirically, he could identify those cases in which the difference likelihood ratio is substantially less than $1/P$ in a data set obtained by mea-

pert testimony about the desirable size of match windows falters at the first step.

V. Reformulating the Argument and Solving the Problem

When inspected, Professor Thompson's argument about match windows is less than airtight. Nevertheless, the underlying concern that P is an imprecise proxy for probative value could be valid. Part III explained that the match-binning likelihood ratio pretends that all matches within an arbitrary window are equally revealing. Sometimes the more precise similarity likelihood ratio (2) exceeds the match-binning ratios. Other times, the match-binning ratios exceed the similarity likelihood ratio. If the disparity is serious, relying solely on P as a measure of probative value could be misleading. But instead of identifying these cases by looking solely to $\Pr(x,d|S)$, as Thompson recommends, it is better to look to suitable computations of the similarity likelihood ratio. This captures the spirit of Thompson's analysis and avoids the initial misstep.

If this reformulation is correct, the pivotal question becomes what to do about the possible discrepancies between the best measure of probative value (2) and the measure (6) that currently is put before a jury. Clearly, the correct response is not testimony about the statistical properties or "scientific" acceptability of particular match windows. The objection, at its root, is to the entire enterprise of binary matching.[60]

The most radical reform would be a rule that barred prosecution testimony about match-binning proportions, but admitted testimony about likelihood ratios. While prosecutors have been advised to rely on the similarity likelihood ratio instead of P,[61] unless the match-bin-

---

suring many fragments twice and then see how well these discrepancies correspond to "quite small" values for $\Pr(x,d|S)$.

[60] Of course, actual practice departs from the model of purely binary matching. *See supra* note 37.

[61] *See, e.g.*, Donald A. Berry, *DNA, Statistics, and the Simpson Case*, Chance, Fall 1994, at 9, 10; Donald A. Berry, *Inferences Using DNA Profiling in Forensic Identification and Paternity Cases*, 6 Stat. Sci. 175 (1991); J. Buckleton et al., *A Continuous Model for Interpreting the Positions of Bands in DNA Locus-Specific Work*, 31 J. Forensic Sci. Soc'y 353 (1991); Bernard Devlin et al., *Forensic Inference from DNA Fingerprints*, 87 J. Am. Stat. Ass'n 337 (1992); Evett et al., *supra* note 42. Evett favors characterizing the numerical value with descriptions like "very strong evidence of identity." *Id.* Jarjoura et al., *supra* note 36, propose that the expert also provides the jury with the proportion of times that likelihood ratios as large or larger than the one at bar have been obtained from different persons in large databases. According to Evett et al., *supra* note 42, at 325 (Table 7), likelihood ratios exceeding 1000 almost always arise for repeated measurements of the same DNA and almost never occur in comparing DNA from different people in a database. *See also* I.W. Evett, *DNA Statistics: Putting the Problems Into Perspective*, 33 Jurimetrics J. 139 (1992) (exhaustive pairing of 1500 Caucasians tested at three loci, generating over one million between-person comparisons

ning ratios are typically far from the mark, compelling them to do so may amount to overkill.[62] The American criminal justice system generally relies on the parties to pick and choose from the available evidence and to build the case that they think is best. When the value of the match-binning ratio exceeds that of the similarity likelihood ratio, but the prosecution prefers to rely on the simpler and perhaps more easily comprehended approach, the defense can hardly object that the prosecution has understated the strength of its own evidence.

In contrast, a court might permit the prosecution to introduce *P*, and leave it to the defense to argue that the better measure of probative value paints a different picture. Such adherence to the usual system of "free proof"—the analog of the economist's "free market"—is reasonable as long as two conditions hold. First, the identification of the cases in which *P* is misleading must not be prohibitively difficult. This consideration counsels requiring the government to disclose to the defense in advance of trial its best estimate of the similarity likelihood ratio.[63]

Second, the free proof system would fail if jurors would be unduly influenced by *P* even when the defense counters it with cross-examination and testimony about the similarity likelihood ratio. If reasonable defense efforts would not prompt the jury to appreciate the prosecutor's fallacy in relying on *P*, then the prosecution does not risk being shown to be acting unfairly by intentionally or unwittingly presenting evidence that overstates its case. Professor Thompson implies that this type of "market failure" is a distinct possibility. He characterizes the effort to present the similarity likelihood ratio as "a formidable task" that "arguably would impose an unfair burden on a criminal defendant."[64] The point is indeed arguable, since it is not clear how juries would react to a combination of *P* and the similarity likelihood ratio.[65] The general philosophy of most modern evidence codes is to allow the jury to hear the evidence unless it is clear that prejudicial effect

---

intended to simulate cases of false accusations, gave only nine likelihood ratios greater than 1, and most of these were below 100).

62 *See* Evett et al., *supra* note 42, at 325 (Table 7).

63 *See* Paul C. Giannelli, *Criminal Discovery, Scientific Evidence, and DNA*, 44 VAND. L. REV. 791 (1991) (emphasizing the importance of discoverable and meaningful expert reports).

64 Thompson, *supra* note 5, at 60. More recently, he has suggested that likelihood ratios should not be used unless they "accurately take into account" uncertainties in the examiner's decisions as to whether a band is present and where it lies on the gel. William C. Thompson, Comment, 9 STAT. SCI. 263, 264 (1994). Even if this view were accepted, it would not apply in the category of cases that is the subject of this Article. *See supra* note 15.

65 One might buttress the argument slightly by recognizing that if the specific likelihood ratio (2) is going to be presented anyway, the jury learns little or nothing useful from *P*. What makes (2) superior to (5) and (6), after all, is that it makes use of more of the data.

outweighs probative value, but few appellate courts would find an abuse of discretion if a trial judge limited the jury to receiving the expert description of import of the DNA profiles in the form of a likelihood ratio.

Other reforms lie between these extremes. The forensic scientist could include the likelihood ratio in a written report accessible to the prosecution and the defense. If the ratio is markedly less than $1/P$, the prosecution could be barred from introducing P without contemporaneous testimony during the direct examination of a prosecution expert to the following effect: (a) it would be wrong to rely on P alone in weighing the DNA evidence; (b) it is important to compare the chance that comparable DNA profiles would arise when the defendant is the source to the chance when the defendant is not the source; and (c) this comparison shows that in this instance P exaggerates the worth of the DNA evidence.[66] This policy of full disclosure reinforces the norm of fair use of the evidence by the government without depriving the prosecution of the opportunity to place an appropriately computed value of P before the jury.

Ultimately, even this solution seems unsatisfactory. If clear and contemporaneous testimony about the similarity likelihood ratio succeeds in ensuring that the limitations on P are explained in an effective way, what is the point of the initial testimony about P? If the jury does not rely on P, then the time and argument expended on it is a waste of resources—in economic terms, a dead weight loss. On the other hand, if the jury does use P to gauge the probative value of the DNA profiles, it makes a potentially serious cognitive error; by definition, the case is one in which $1/P$ is much higher than the similarity ratio (2).

In this way, the result, in broad outline, might resemble Professor Thompson's. In a subset of cases match-binning proportions should be excluded, but the DNA profile evidence should be admissible when explained in terms of a suitable likelihood ratio.[67] The subset of cases may differ from those identified by Professor Thompson, but it consists of precisely those in which the form of prejudice he identifies is a serious risk. The response to this risk is neither outright exclusion of DNA evidence nor admission of nonresponsive testimony about the properties of match windows.[68] If the problem is sufficiently wide-

66 Other limitations on P that might justify similar, cautionary testimony in the interest of full disclosure are described in Kaye, *supra* note 4, and Thompson, *supra* note 5.

67 For discussion of how a likelihood ratio might be presented or used, see Berry, *supra* note 61; Kaye, *supra* note 4; Richard Lempert, *Comment: Theory and Practice in DNA Fingerprinting*, 9 STAT. SCI. 255, 258 (1994); Roeder, *supra* note 27; *supra* note 61.

68 Whether any special rules are required to handle the subset of cases in which match-

spread, the best solution may be to use the similarity likelihood ratio to more fairly present the DNA data.[69]

binning frequencies are misleading is not clear. Without an empirical analysis of the rate at which the similarity likelihood ratio (2) is substantially less than match-binning likelihood ratios, the size of this subset—and hence the suggestion that there is a widespread problem that warrants a procedural solution—is speculative. Before demanding that all laboratories compute similarity likelihood ratios to isolate the cases in which there would be a discrepancy, empirical analysis to ascertain the extent of the problem would seem advisable.

69 Some statisticians may question whether existing methods for computing (2) are adequate. If such doubts are justified, then implementation of the approach recommended in this Article would be premature, but it still would be reasonable to use the cruder match-binning likelihood ratio (3) to measure probative value. This second best likelihood ratio could stand in for the presently unavailable ratio (2) in the approach that is outlined in this Article.