The Interpretation of DNA Evidence A Case Study in Probabilities An Educational Module

Prepared by

David H. Kaye Associate Dean for Research and Distinguished Professor The Pennsylvania State University (Penn State Law)

For

Committee on Preparing the Next Generation of Policy Makers for Science-Based Decisions Committee on Science, Technology, and Law

The National Academies of SCIENCES • ENGINEERING • MEDICINE

June 2016

The copyright in this module is owned by the author of the module, and may be used subject to the terms of the <u>Creative Commons Attribution-NonCommercial 4.0 International Public License</u>. By using or further adapting the educational module, you agree to comply with the terms of the license. The educational module is solely the product of the author and others that have added modifications and is not necessarily endorsed or adopted by the National Academy of Sciences, Engineering, and Medicine or the sponsors of this activity.

Contents

Preface
Introduction
1. A Laboratory Report5
2. The Basis for the Reported Frequencies
2.1 Sampling Error: Estimating allele and single-locus genotype frequencies by counting
their occurrence in samples
2.2 Modeling Uncertainty: Estimating single-locus genotype frequencies from allele
frequencies 11
2.3 Estimating multilocus genotype frequencies with the independence assumption 13
3. The Final Laboratory Report: The Value of Conservatism
4. A Probability Primer
4.1 Probability Mass19
4.2 Conditional Probability
4.3 Transposing Conditional Probabilities: Bayes' Rule
5. The Expert Testimony at Trial25
6. Letter Introduced to Support a Petition for a Writ of Habeas Corpus
7. Majority Opinion in <i>Brown v. Farwell,</i> 525 F.3d 787 (9th Cir. 2008) ⁷⁶
8. Supreme Court Opinion in McDaniel v. Brown, 558 U.S. 120 (2010)
9. The Transposition Fallacy in <i>Brown</i>
9.1 The Fallacy Exposed 389.2 The Fallacy Corrected (for the hypothesis of an unrelated
perpetrator)
9.3 The Fallacy Corrected (for the Hypothesis of a Brother)
10. The Seven Principles Revisited 45
Further Reading
Answers to Test Your Understanding Questions 47

Examination and Extensior	۱ Questions	51
---------------------------	-------------	----

PREFACE

This module is designed to enable readers with only the most general knowledge of biology, probability, and statistics to understand the major scientific aspects of a case in which a leading DNA analyst presented "critical testimony that was later proved to be inaccurate, misleading, [and] faulty."¹ It is the only case in which the U.S. Supreme Court has inquired into the substance of DNA testimony. When it did so, the Court in *McDaniel v. Brown*² identified "two inaccuracies" in the testimony, namely, "equating random match probability with source probability, and an underestimate of the likelihood that one of [the defendant's] brothers would also match the DNA left at the scene."³

This case study presents the most pertinent laboratory reports, testimony, and opinions in the case along with questions about the scientific basis for the statements in them. Explanations of some of the basics of forensic DNA analysis and the probabilities that forensic scientists or technicians present in such cases are included. Nevertheless, this module will not make you an expert in DNA evidence, probability, and statistics. That is not its purpose. It is designed instead to illustrate the potential strengths, weaknesses, and limitations of probability computations; to supply a basic vocabulary and conceptual framework for reasoning about DNA evidence; and to illustrate through a detailed study of one case how a grasp of correct reasoning with conditional probabilities is necessary to apply evidentiary and constitutional principles of fairness to testimony about DNA and other forensic-science evidence.

After reading and discussing in class the case materials (the laboratory reports, excerpts from the trial transcript) and the explanations of the science and the probabilistic interpretation of the scientific findings, and after answering the Test Your Understanding questions before class, you should be able to:

- Explain how DNA typing helps identify individuals.
- State how and why frequencies of DNA alleles, single-locus, and multilocus profiles are estimated.
- Describe the sources of uncertainties that result from sampling allele frequencies and combining them to estimate profile frequencies.
- Interpret confidence intervals for estimated frequencies and state their limitations.
- Define the concept of mathematical probability.
- Perform elementary computations for the probabilities of events such as a DNA match to a randomly selected individual.
- Articulate the difference between independent and dependent events.
- Recognize when a conditional probability is improperly transposed.
- Define likelihood ratios for pairs of hypotheses about the source of DNA evidence and compute their magnitudes.

¹ Brown v. Farwell, 525 F.3d 787, 789 (9th Cir. 2008), rev'd sub nom. *McDaniel v.* Brown, 558 U.S. 120 (2010).

² 558 U.S. 120 (2010).

³ Id. at 129–130.

- Use the likelihood ratio to measure the probative value of DNA matches and other forensic scientific test results.
- Use Bayes' rule to deduce a posterior probability from a likelihood ratio and a prior probability.
- Apply likelihood ratios and Bayes' rule to assess the impact of the improper transposition of conditional probabilities in *McDaniel v. Brown* and other cases.

In addition, the module has broader goals. It illustrates general features of scientific and statistical reasoning and proof that are important to the best use of DNA and other forensic-science test results. The main themes are how uncertainty can be quantified with probability and the basis and limitations of statistical and probabilistic models. The following principles should emerge from examination and discussion of the materials in this case study:

- Principle 1. Uncertainty arises in choosing mathematical models of phenomena.
- Principle 2. Models rest on assumptions. Checking whether the results are highly sensitive to changes in the assumptions is advisable.
- Principle 3. Uncertainty arises when generalizing from samples to populations. One should ask: (a) How were samples chosen? (b) Is the method likely to produce biased samples? (c) What is the plausible range of statistical error (variation arising from "the luck of the draw")? (d) Is the sample drawn from the population of interest or from a good proxy for it?
- Principle 4. Uncertainty can be quantified with probabilities.
- Principle 5. The probability of evidence indicating that a disputed event did (or did not) occur does not generally equal the probability of that event, as assessed in the light of the evidence.
- Principle 6. Bayes' rule offers one logical way to use probabilities about the evidence of an event to determine the probability of the event itself.
- Principle 7. The strength, or weight, of evidence can be quantified by comparing the probability of the evidence given one party's hypothesis to the probability of the evidence given an alternative hypothesis.

INTRODUCTION

On July 20, 1994, Marshall Smith, the District Attorney for the County of Elko, Nevada, filed a criminal information charging Troy Don Brown with the rape of a 9-year-old girl. Troy, then 23 years old, worked at a mine outside of Carlin, Nevada, and lived in the same trailer park as the little girl. He had grown up in a Mormon family, with two older and two younger brothers, on a small ranch in Loa, Utah. Troy insisted that he was innocent even when the police lied to him about finding his fingerprints in the little girl's bedroom. As the trial opened, the district attorney announced that "[t]his case is about an unimaginably brutal rape and sodomy and attempted murder of a child." The child, he told the jurors,

Doesn't know who [sexually assaulted her, strangled her], and her left for dead. The police went out to [her] bedroom where this occurred, and found a fingerprint. That fingerprint does not match the defendant. But we have another type of fingerprint, and that is called DNA. Located on a pair of bloody panties that [she] had been wearing was semen—sperm. Each one of us has a specific genetic pattern. The DNA is found in sperm, and it's found in blood. DNA was extracted from that sperm found on those child's bloody panties and was compared to the DNA extracted from the defendant's blood. The likelihood of those same characteristics occurring in those two samples is not 1 out of 100,000; it's not 1 out of 500,000; it's not 1 out of 2 million; it's 1 out of 3 million. One out of 3 million will share those characteristics. The characteristics the DNA found in the semen is exact, and compared to the DNA in that blood is 1 in 3 million. They match.

Jurors wept as they returned their guilty verdict. Troy maintained his innocence when the judge offered leniency during sentencing if only he would accept responsibility. Instead, Troy appealed and brought a series of actions for postconviction relief.

As the prosecutor's opening statement made plain, DNA evidence—the new "fingerprinting"—was pivotal. Renee Romero, a criminalist in the Washoe County Sheriff's Laboratory and a member of the American Academy of Forensic Sciences and the FBI's select Technical Working Group on DNA Analysis Methods, convincingly testified to a "very conservative" probability of "99.99967 percent . . . that the DNA found in the panties is the same as the DNA found in the defendant's blood."⁴ Twelve years later, a federal district court concluded that her testimony was fallacious. It held that these facts (and the failure of defense counsel to counter the critical DNA evidence effectively) deprived Brown of due process of law. The Court of Appeals for the Ninth Circuit agreed that the testimony about probability relating to the DNA match should not have been admitted and that without the DNA match, no reasonable jury could have convicted. However, the Supreme Court unanimously rejected this reasoning. It held that the issue of whether admitting inaccurate statements of probability was itself a violation of due process was not properly before the lower federal courts and that the undisputed existence of a DNA match, combined with other evidence in the case, supported the conviction.

The materials that follow focus on the laboratory findings in *Brown*, the underlying scientific and statistical principles, and how the findings were used in the courtroom and in postconviction proceedings. Sections 1 through 3 present and then explain the laboratory reports in the case. They define such key terms in genetics and forensic DNA science as *DNA*, *allele, chromosome, locus,* and *genotype,* and they indicate how and why the frequencies of identifying DNA types are estimated. Discussion questions are intended to identify uncertainties that are not always incorporated in probabilities computed by laboratories. More broadly, these sections raise questions about the relationship between scientific models and the phenomena they are intended to illuminate. They ask how science and law should respond

⁴ According to one true-crime story writer, Renee Romero, who is now the director of the Washoe County crime laboratory, "could easily have passed for a model in a glamour magazine rather than a criminalist in the Sheriff's Department." Robert Scott, *Rope Burns* 28 (2001).

to the fact that scientific and statistical models are always inexact simplifications of a complex reality.

Section 4 is a quick primer on probability, with emphasis on conditional probability and Bayes' rule as a way to update prior information and to move from statements of the probability of evidence under different hypotheses or models to statements of probabilities for the hypotheses and models themselves. The reasoning it describes is helpful in understanding what went wrong in the prosecution of Troy Brown.⁵

Sections 5 through 8 show how the laboratory findings were presented during the trial and how the federal district court, the Court of Appeals, and the Supreme Court responded to an expert's criticisms of the trial testimony.

Section 9 assesses these criticisms and the courts' reasoning about the problems with the probabilities in *Brown*. It explains the nature and varying severity of a commonly encountered fallacy in reasoning about probabilistic and statistical proof. It applies two tools—likelihood ratios and Bayes' rule—to reason more soundly about the probative value of DNA and other evidence of identity. It illustrates how clear thinking about the questions that probability computations can help answer is important in forensic science and legal proof.

Section 10 recapitulates the seven principles for thinking about uncertainty, statistical modeling and estimation, and forensic inference as they applied in *McDaniel v. Brown*. These principles and the discussion of the testimony and argument in *Brown* remain important even though the particular tests used to produce the DNA profiles have been superseded by improved technologies and a different and more extensive set of loci and even though the technology now in common use (capillary electrophoresis of short-tandem repeats) could well be replaced in the near future with "next-generation sequencing."

Rather than become mired in the details of a specific technology for DNA profiling, we will only examine the underlying genetic and statistical issues and ideas reflected in the reports and testimony in *Brown*. These matters apply to all DNA profiling, and, in significant part, to other forms of forensic identification as well. As one forensic DNA scientist reflected, "[d]espite all of these technical advances, the single most important advance in forensic genetic thinking is the realization that the scientist should address the probability of the evidence."⁶

⁵ The notion that mathematical probabilities meaningfully can be assigned to hypotheses as opposed to only the outcomes of random processes is not without controversy. See, e.g., David H. Kaye & David A. Freedman, Reference Guide on Statistics, in *Reference Manual on Scientific Evidence* 211 (Federal Judicial Center, 3d ed. 2011); Paul Weirich, The Bayesian Decision-theoretic Approach to Statistics, in *Philosophy of Statistics* 33 (Dov M. Gabbay et al. eds. 2011). This module makes use of the "subjective probability" framework simply to illuminate the relationship between an expert's description of probabilities for DNA matches and the probabilities that an expert, judge, or jury willing to envision a mathematical probability should assign to the claim that a given individual is the source of the matching DNA.

⁶ Angel Carracedo, Forensic Genetics: History, in *Forensic Biology* 19, 22 (Max M. Houck ed. 2015).

1. A Laboratory Report

The Washoe County Sheriff's laboratory prepared three written reports. Before delving into their contents, some relevant biological principles and terminology might be worth noting. You probably know these facts already, but Box 1 summarizes some key terms and ideas.

Box 1: Some Definitions of Forensic *Genetics*

- 1. *Deoxyribonucleic acid* (DNA) is a molecule that contains four distinct nucleotide bases (abbreviated A, T, C, and G) arrayed inside the famous double helix that is the "backbone" of the molecule. An A on one strand binds to a T on the other, and a C pairs with a G. A DNA sequence is the order of these nucleotides as one moves along the backbone. Because of the pairing rule, a single letter can denote a pair of bases—an A can be written for A:T, for example.
- 2. The vast bulk of the DNA in a human cell is crammed into the cell nucleus, coiled around proteins in structures called *chromosomes*. Humans normally have 46 chromosomes. Fortyfour come in pairs, with one member inherited from the father and one from the mother. The other two chromosomes are the sex chromosomes. Women have two copies of the X chromosome; men have one X (inherited from their mothers) and one Y (inherited from their fathers).
- 3. A *locus* is a physical location of a gene or other forensically interesting sequence of base pairs. DNA sequences in a *gene* specify the order of the basic building blocks of proteins. Other DNA sequences have different functions—or none at all.
- 4. From the standpoint of forensic DNA profiling, all that matters is that the sequences at designated locations have detectable differences across the population of possible perpetrators of crimes. These DNA variations (in either the order of the base pairs or the length of the sequence) are known as *alleles*. The more the DNA loci vary across individuals, the more useful the alleles at these loci will be in discriminating among individuals.
- 5. A *single-locus genotype* is the pair of alleles at a single locus. A *multilocus genotype* is the set of pairs at more than one locus. In forensic genetics, the term *profile* is more appropriate because the DNA variants are usually not the different forms of genes but other types of DNA sequences.
- 6. *Population genetics* is the study of the occurrence of particular alleles and genotypes in population groups and subgroups.
- 7. A *random match* would be a match between the DNA of an individual drawn at random from the general population (or a subset of it) and the crime-scene DNA (also called the trace evidence or questioned sample).

The first laboratory report was limited to one locus—the physical location of a gene named HLA DQA on the sixth largest human chromosome.⁷ The laboratory's initial report on the DQA gene is not reproduced below because the same data are in the second report. This second report extended the analysis to another five genetic loci that encode other protein products. These loci are abbreviated as LDLR, GYPA, HBGG, D7S8, and GC. To see the legal implications of the report, it is not necessary to know any particular details about these loci. It is enough to recognize that a "3,4" individual has a type 3 allele on one chromosome and a type 4 allele on the other; an "AA" individual has a type A allele on the both chromosomes, and so on.

The report was presented in a preliminary hearing but not at the trial. To make it easier to understand, material has been added in brackets. Take a few minutes to read it and think about how strongly it points to Troy Brown as the rapist. The questions we will consider relate to how the Sheriff's laboratory produced the numbers at end of the report.

[SECOND REPORT] WASHOE COUNTY SHERIFF'S OFFICE FORENSIC SCIENCE DIVISION

6/06/1994

LABORATORY NUMBER:	L0277-94-A (SUPPLEMENTAL REPORT)
SUSPECT:	BROWN, TROY
VICTIM:	[Jane Doe]
DATE OF SUBMISSION:	02/08/94
OFFENSE:	SEXUAL ASSAULT

Received from WCSO EVIDENCE SECTION, on 02/03/94

The submitted items were identified as: 1. Blood standard from [Doe, Jane] 2. Blood Standard from Brown, Troy 3. #14 T1 panties from [Doe, Jane] 4. #14 T1 and T2 pajamas 5. R03692 vaginal swabs from [Doe, Jane] 6. R03692 gauze pad

RESULTS OF EXAMINATION

The above items of evidence (3-6) were differentially extracted [a chemical procedure was used to separate out the DNA in semen] and semen was identified on all of the items. DNA was isolated from the above items and subjected to PCR amplification [a chemical process for making many identical copies of a DNA molecule]. Examination at six different genetic loci gave the following results:

 $^{^{7}}$ The DQA gene, which encodes for a protein known as DQ α , is part of a family of genes called the human leukocyte antigen (HLA) complex. HLA plays an integral role in the immune system.

[TABLE 1]

	HLA	LDLR	GYPA	HBGG	D7S8	GC
	DQα					
[DOE]	3,4	AB	BB	AB	AA	CC
BROWN	3,3	AB	AA	BB	AB	AA
EPITHELIAL	3,4	AB	BB	AB	AA	CC
[female]						
PANTIES						
SPERM	3,3	AB	AA	BB	AB	AA
PANTIES						
EPITHELIAL	3,4					
[female]						
PAJAMAS						
SPERM	3,3					
PAJAMAS						

No results were obtained on the remaining items of evidence at this time. Based on the above results Troy Brown is consistent with being the source of the sperm fraction from the stain on the panties and the pajamas. The frequency of occurrence of a 3,3 is 1 in 22 in the Caucasian population, 1 in 110 in the Black population and 1 in 4 in the Hispanic population. The frequency of occurrence of the pattern from all six markers obtained for the standard from Troy Brown and the sperm fraction from the panties is 1 in 18,900 in the Caucasian population, 1 in 2,460,000 in the Black population and 1 in 4,800 in the Hispanic population.

The statistics were calculated using the population data available through Roche Molecular Systems.

/s/ ____

ANALYST RENEE L. ROMERO, CRIMINALIST

Now that this impressive report is before us, let's consider how precise the numbers in it really are. Psychologists have found that exact numbers are especially persuasive,⁸ but the number of digits in a laboratory report does not always indicate more precision. Whether stated or not, uncertainties inhere in even the best empirical estimates. Expounding slightly on Principles 1 (uncertainty in modeling) and 3 (uncertainty due to sampling), some common sources of uncertainties in estimated quantities are imprecision in individual measurements, sampling error in selecting units for measurements, imperfections in the assumptions used to

⁸ Alexandra Jerez-Fernandez et al., Show Me the Numbers: Precision as a Cue to Others' Confidence, 25 Psychol. Sci. 633 (2014).

construct a statistical model for evaluating measurements, and incomplete knowledge of the parameters of such models. Let's see if we can identify some of these in the laboratory report.

2. The Basis for the Reported Frequencies

The bottom line of the report is that "[t]he frequency of occurrence of the pattern from all six markers obtained for the standard from Troy Brown and the sperm fraction from the panties is 1 in 18,900 in the Caucasian population, 1 in 2,460,000 in the Black population and 1 in 4,800 in the Hispanic population." How exact are these estimates, why should we believe them, and what do they prove? The report is conspicuously silent about where they came from. It merely states, "The statistics were calculated using the population data available through Roche Molecular Systems." It reveals nothing about the *how* the analyst used "the population data" to compute anything. To dig deeper, we need to start with the data.

2.1 Sampling Error: Estimating allele and single-locus genotype frequencies by counting their occurrence in samples

Look back at the column for the DQA locus in Table 1. The sperm cells on the panties as well as the cells in a sample of Troy's blood exhibited the same genotype, (3,3). What does this match prove? If the only allele in the population were the one designated as "3", then it would prove exactly nothing. The (3,3) genotype would be equally likely to arise if Troy were the source than it would be if any other man were the source. But research shows that there are other DQA alleles in the population. According to the Sheriff's laboratory, "The frequency of occurrence of a 3,3 is 1 in 22 in the Caucasian population, 1 in 110 in the Black population and 1 in 4 in the Hispanic population."

Discussion Question 2.1. Roche Molecular Systems must have obtained these numbers from samples of people in these three population groups. What is required to make such numbers reliable and valid?

Some terms and ideas about sampling methods may help answer this question. Statisticians distinguish between the numerical characteristics of a *population* and the corresponding ones of a *sample*. The former are *parameters*; the latter are *statistics*. For example, the proportion of all students enrolled at a university who are over 6 feet tall is a population parameter. All the enrolled students constitute the population; the proportion in that population is the parameter.

Suppose you wanted to know the proportion of these tall students in the whole university. The university does not record its students' heights.⁹ In theory, you could measure the height of every student to find the population proportion directly. But that would be time-

⁹ The Department of Motor Vehicles includes this information on driver's licenses, but let's assume that you do not have access to its records.

consuming and tedious. Instead, you could find a sample of students, ascertain the proportion in the sample, and treat that sample statistic as an estimate of the population proportion. However, even if your height measurements are perfectly accurate, the estimate is all but certain to be wrong—because you cannot guarantee that any sample is perfectly representative of the entire population. The difference between the unknown, "true" population value and the sample-based estimate is *sampling error*.

What factors cause sampling error and therefore might reduce your confidence in the accuracy of your estimate? In the student-height case, it might not be a good idea to go the basketball courts at the gym to draw a sample. Using only that location might introduce a bias toward greater heights. In statistics, *bias* is a systematic departure between the quantity as measured or estimated and its true value.

Drawing units strictly at random from the population of interest eliminates bias in the way the sampled items are obtained. Moreover, well-defined statistical models are available to describe the properties of statistics computed from such samples. For example, if you were to take a *simple random sample* of students—one in which every enrolled student had an equal probability of being in the sample—the sample proportion would be a "random variable" that follows the "normal distribution"-the bell-shaped curve shown in every elementary statistics textbook.¹⁰ We won't go into all the details, but statisticians typically compute an interval to express the precision of the estimate. The most common interval estimate is called a *confidence* interval, but the meaning of "confidence" is tricky. It is not necessarily the probability that the true value lies within the particular interval, but rather the approximate proportion of a great many confidence intervals—one for each random sample—that would include the population value, whatever it happens to be. About 95% of all 95% confidence intervals for the average height of all students will cover that unknown height. To capture the true value that often, the interval must be wider than it would be to cover the true value in, say, 60% of all sample intervals. A scientist who wants to generalize from a single sample will be correct more often by making a less precise claim (higher confidence, less precision) than by making a sharper claim (less confidence, more precision). Of course, high confidence is desirable, and for a high degree of confidence, a narrow interval suggests that the sample estimate should be pretty good, whereas a wide interval indicates that the estimate is not very reliable.¹¹

For a given degree of confidence, the width of the interval depends on two things: the number of students in the sample (*n*), and the extent of the variation in heights in the population (expressed as the standard deviation *s*). If the size of the random sample of students is small compared to the size of the entire student body, then the width is directly proportional to *s* and inversely proportional to \sqrt{n} . Thus, small samples may be adequate for learning about a relatively homogeneous population, and increasing the sample size yields diminishing returns. For example, quadrupling the sample size will improve the precision of the estimate by only a factor of 2. To improve the precision by a factor of 5, the sample would have to be 25 times larger.

¹⁰ Click on the link in the sentence for an example. The sample statistic is only approximately normal. The approximation is good when the sample size is very small compared to the population size.

¹¹ See David H. Kaye & David A. Freedman, Reference Guide on Statistics, in *Reference Manual on Scientific Evidence* 211 (Federal Judicial Center, 3d ed. 2011).

TEST YOUR UNDERSTANDING

2.1. You measure the heights of every 25th student leaving the student union between the hours of 1 and 3 p.m. (a) Is this simple random sampling? If not, is it likely to yield a sample that is comparable to ones that would come from simple random sampling? (b) Is the sample drawn from the right population? If not, is the group from which the sample comes a good proxy for the population of interest?

2.2. A simple random sample of 100 students yields a mean height of 67 inches and a 95% confidence interval of 67 \pm 8 inches. A second simple random sample of 400 students yields a mean height of 69 inches. You would expect the 95% confidence interval for this sample to be closest to (a) 67 \pm 8; (b) 69 \pm 8; (c) 67 \pm 4; (d) 69 \pm 4.

2.3. If the 95% confidence interval were 70 \pm 2, the 99% interval could be (a) 70 \pm 3, or (b) 70 \pm 1.

2.4. Measurements of height might not be exactly correct. Does the confidence interval for the sample described in Question 2.1 account for this measurement error?

Suppose that the Sheriff's laboratory report had included the confidence interval for frequencies of the DQA allele (3,3) in the three population groups, based on the simple random sampling model. Would the expansiveness of the interval be a good indication of the sampling error in the reported frequencies of "1 in 22 in the Caucasian population, 1 in 110 in the Black population and 1 in 4 in the Hispanic population"? One concern would be the realism of the statistical model of simple random sampling. As Principle 1 indicates, choosing the right model is important. Did Roche use simple random sampling to pick people to type? Or did it use "convenience" or "grab-bag" samples from blood banks, parentage testing laboratories, or perhaps its own employees to obtain DNA for what forensic DNA testers call "population databases"? If it used convenience sampling, we have to ask whether the results would be comparable to simple random samples. The issue is similar to the sampling students at the convenient location of the student union in Question 2.1(b). In some cases, convenience samples, are likely to be very similar to random samples. For example, to determine the proportion of oxygen in the air in a room, it is not necessary to use a random sampling method to decide on exactly where in the room the sample of air to be tested should come from. Trapping some air from anywhere in the room will do. In DNA cases, some experts claimed that the failure to obtain random samples was a fatal flaw in DNA probabilities, and some courts seemed to agree, but this view did not prevail in the long run.

One reason convenience sampling came to be accepted was that different convenience samples gave roughly similar results. For instance, a study conducted by FBI scientists at the time of the *Brown* case reported that among the individuals providing DNA samples, the (3,3) DQA genotype occurred with the frequencies of 0.007 (145 African-Americans), 0.061 (148

Caucasians), 0.032 (94 Southeastern Hispanics), and 0.042 (94 Southwestern Hispanics).¹² Thus, 9 of the 148 Caucasians in the FBI sample were type (3,3).¹³

Discussion Question 2.2. How close are FBI and the Roche estimates of the proportion of Caucasians who are type (3,3)? What factors might account for the difference?

In sum, the report of the DQA findings illustrates the first three principles listed in the preface. Uncertainty arises in choosing mathematical models of phenomena, herein using the model of a normal random variable for the sample proportions of DQA alleles, even though the samples are not quite random. Uncertainty also arises when generalizing from samples to populations. One should ask whether the method by which Roche acquired "population data" is likely to produce representative samples, and whether the plausible range of statistical error (variation arising from "the luck of the draw") with this method is small. In this instance, later study largely replicated the report of the DQA allele frequency.

2.2 Modeling Uncertainty: Estimating single-locus genotype frequencies from allele frequencies

Estimating the frequencies of genotypes such as (3,3) by sampling from a large population and counting how many individuals have the genotype does not always work well. If the genotype is very rare, as the genotypes now in use are, there is a good chance it will not be seen even once in any sample of realistic size. If a large haystack contains only a few needles, one cannot expect to find any of them by inspecting only a cubic centimeter of the hay. This is the "zero-numerator problem" in statistics. It is impossible to specify with precision the frequency of an event that is too rare to show up in the vast majority of samples. One solution is to devise a model with which to calculate the unknown, rare frequency from more easily measured quantities. Each DQA allele is sufficiently common to allow the allele frequencies to be ascertained with reasonable accuracy in a moderately sized sample. The question then becomes how to derive a DNA profile frequency from the estimated allele frequencies. Population-genetics models address this question. The simplest model posits a single "randomly mating" population. Here, random mating is a term of art. It does not mean that individuals choose their mates at random, but only that the selections are unrelated to the alleles of the DQA gene. At first blush, it seems plausible that people's tissue types (which is what the gene helps determine) have no bearing on their choice of reproductive partners. Whether the random-mating model is reasonable is informed further by sociological knowledge about how people choose their reproductive partners and by genetic studies of allele frequencies in different groups.

 $^{^{12}}$ Bruce Budowle et al., Validation and Population Studies of the Loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQ α Using a Multiplex Amplification and Typing Procedure, 40 J. Forensic Sci. 45 (1995).

 $^{^{13}}$ 9/148 = 0.0608 = 6.1%.

In a randomly mating population, any individual's alleles are a random draw from the gene pool. The FBI found that 21.6% of the DQA alleles in the Caucasians tested were type 3. More generally, we can use the symbol p to stand for the proportion represented by an allele. Here, p = 0.216, so the probability of an offspring with allele 3 from both the mother and the father is $P(3,3) = p^2 = (0.216)^2 = 0.0467 = 4.7\%$. The estimate from the random-mating model can be compared to the directly observed proportion of Caucasians who are type (3,3). That number is 0.061 = 6.1%. The model estimate is not bad.¹⁴

The simple multiplication for the probability of the HLA genotype is an example of a general rule of probabilities: For any two independent events *A* and *B*, the probability of both *A* and *B* is the product of the probability of *A* and the probability of *B*—that is, P(A,B) = P(A) P(B). For instance, if two fair, cubical dice are rolled simultaneously, the probability that die number 1 will come up 3 and that die number 2 will come up 5 is P(3,5) = P(3) P(5) = (1/6)(1/6) = 1/36. If six dice are rolled, the probability of any particular string of numbers is the product of the individual probabilities. For example, $P(3,1,4,6,4,5,2) = (1/6)^6 = 0.000214$.

But what is an independent event? Two events are *independent* if the outcome of one has no bearing on or correlation to the outcome of the other. For example, if there are 50 marbles stamped A and 50 marbles stamped B in an urn, and one is drawn at random, the probability that it is stamped A is $50/100 = \frac{1}{2}$. The probability it is stamped B is the same. If we replace the marble and draw again at random, the probabilities do not change: $P(A) = P(B) = 50/100 = \frac{1}{2}$. The type of marble on one draw and the type on the other draw are independent. Hence, the probability of drawing an A twice is $P(A, A) = P(A) P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

TEST YOUR UNDERSTANDING

2.5. You turn over the card from the top of a well-shuffled deck of 20 cards (the ace through the 10 of diamonds, and the ace through the 10 of clubs), replace it, shuffle again, and turn over the top card again. (a) What is the probability that the first card was a 10? (b) That the second card was a 10? (c) That both cards were 10s? (d) Did you treat the two drawings as independent events?

2.6. You do not replace a card after turning it over. (a) What is the probability that the first two cards that are turned over are 10s? (b) The first three? (c) Is the identity of the card turned over the second and third times independent of the identity of the previous card?

2.7. Which of the following pairs of events are statistically independent: (a) hair color and eye color; (b) height and weight; (c) red blood type and IQ score; (d) whether a new car is blue or green and its sticker price?

¹⁴ The FBI researchers performed comparisons for all DQA genotype frequencies and concluded that "[t]he distribution of the genotype frequencies for . . . HLA-DQa meet HWE." Budowle et al., supra note 12, at 53. (HWE is short for *Hardy-Weinberg equilibrium*, the technical term for the allele and single-locus genotype proportions expected in a randomly mating population.)

2.3 Estimating multilocus genotype frequencies with the independence assumption

The HLA DQA match is not the only genetic information pointing to Troy as the source of the sperm. The Sheriff's laboratory reported that "[t]he frequency of occurrence of the pattern from all six markers obtained for the standard from Troy Brown and the sperm fraction from the panties is 1 in 18,900 in the Caucasian population, 1 in 2,460,000 in the Black population and 1 in 4,800 in the Hispanic population," but it did not say how it arrived at these numbers. It simply listed (in Table 1) the following matching genotypes at all six loci:

	HLA DQα	LDLR	GYPA	HBGG	D7S8	GC
BROWN	3,3	AB	AA	BB	AB	AA
SPERM	3,3	AB	AA	BB	AB	AA
PANTIES						

Presumably, the laboratory used the random-mating model to compute single-locus frequencies as the product of the allele frequencies. As we just saw, P(3,3) would be $p(3) \times p(3)$. When the single-locus genotype is a pair of two different alleles, such as Brown's LDLR type (A,B), the genotype frequency is $2 \times p(A) \times p(B)$. (The factor of 2 is present because there are two equally likely ways for Brown to be type AB.)¹⁵

After calculating the frequencies of the single-locus genotypes in this way, the laboratory had to combine them to obtain the multilocus frequency. Some of the single-locus genotypes in the case are common, occurring in over 50% of the individuals tested. Others are rarer, occurring in only 5 or 6% of the sample. And, just as a string of six heads in six consecutive tosses of a coin is rarer than a head on a single toss, a six-locus genotype is rarer than a single-locus genotype. But how rare? Again, a solution is to devise and apply a population-genetics model—this time one that expresses the relationship between (1) the frequencies of the single-locus genotypes that can be estimated reasonably with small-to-moderate samples and (2) the multilocus frequency that is just too small to measure directly.

The Sheriff's office used the random-mating model and assumed that the different genes are inherited separately rather than in pairs or other combinations.¹⁶ Under these conditions, not only are two alleles at one locus statistically independent, but the single-locus genotypes are all independent of one another. Using these assumptions, the laboratory arrived at the figure of 1/18,900 for the multilocus profile in Caucasians.

Consequently, there are two sources of uncertainty in this number. Beyond the uncertainty in the estimates of the allele frequencies, there might be departures from random mating in real populations. Indeed, several population geneticists emphasized this possibility around the time of Brown's trial. We will see that this uncertainty is manageable. Its magnitude

¹⁵ He could have inherited the A from his mother with probability p(A) and the B from his father, with probability p(B). This joint event has probability p(A) p(B). Alternatively, he could have inherited the B maternally and the A paternally. This joint event has probability p(B) p(A). The probability that he turned out to be type AB one way or another is thus p(A) p(B) + p(B) p(A) = 2 p(A) p(B).

¹⁶ This assumption would be false if any pair of loci were close by one another on a chromosome.

can be estimated, or a more complicated model can be used. As this point, it is enough to note that throughout science and statistics, simplified models are used to represent a more complex reality. Physicists speak of frictionless surfaces, perfectly elastic collisions, and the gravitational interactions of only two bodies in a universe populated with inconceivably many bodies. If the assumptions used to construct the model are reasonably close to the actual situation, then they are justified. ("Reasonably close" means that realistic departures from the assumptions will produce similar results.) As two statisticians famously wrote, "all models are wrong, but some are useful."¹⁷

Discussion Question 2.3. How could one go about establishing that the assumption that the six loci that the laboratory typed are independent is reasonably accurate? (Remember Principles 1 and 2.)

To point toward an answer to this question, we can consider a slightly more familiar situation—tossing two pairs of dice. You toss each pair 1,000 times and record the numbers showing on each die each time. To check whether a 1,1 on Pair 1 is independent of, say, a 3,6 on Pair 2, you could see if the proportion of the outcomes (1,1) is roughly the product of the proportion of 1,1 outcomes for Pair 1 and the proportion of 3,6 outcomes for Pair 2. If P[(1,1) & (3,6)] is very different from $P(1,1) \times P(3,6)$, then either you have a strange sample or the independence assumption does not work well for these dice. Notice, however, that if you have six pairs of dice, this test will be less useful because there are so many possibilities for the six pairs of numbers. Even so, checking at least some of them for independence could be reassuring (or disturbing). What would be an analogous empirical test for independence of genotypes?

The uncertainties in the final, multilocus genotype frequencies in the Sheriff's report are not listed in the report, but that does not mean they are not present. The allele frequencies are not known exactly. In addition to this data uncertainty, the formula for combining them rests on independence assumptions (modeling uncertainty) that must be defended on theoretical grounds and with some assistance from statistical tests. Regrettably, there are no simple rules for determining whether, given these uncertainties, the conclusions are in the right ballpark. But the questions that must be addressed to come to a sound assessment are well known, and the laboratory made no effort in the report to identify any limitations in its measurements and computations of frequencies.

¹⁷ George E. P. Box & Norman R. Draper, Empirical Model-Building and Response Surfaces 424 (1987). The full quotation is

The fact that the polynomial is an approximation does not necessarily detract from its usefulness because all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.

3. The Final Laboratory Report: The Value of Conservatism

Figures such as 1 in 18,900 in the first two reports are impressive, but District Attorney Smith did not introduce those reports at trial. Perhaps he was reluctant to invite arguments about what then was a new forensic technology (PCR-based testing). Instead, he had Renee Romero testify solely to a third report. Relying only on tests for the additional loci listed there that had been in use across the country for several years, he was able to open his case with the powerful claim that "it's 1 out of 3 million."

This figure pertained to five "variable number tandem repeat" (VNTR) loci. The DNA sequences at these loci do not code for proteins, and variations in the number of repeats at these loci do not seem to have any health consequences that would affect the number of children who could pass along the VNTR alleles. Because these DNA variants confer no selective advantage or disadvantage, one would expect random mutations to spread and persist in the population. Over the generations, the result should be populations with many alleles, each occurring with low frequencies.

Beyond naming the five VNTR loci that were studied, however, the report contained little information. It did not list the alleles found at any locus, and it provided only a cursory statement that "the frequency of this [unspecified] matching pattern is rarer than 1 in 3,000,000 in the Caucasian, Black, and Hispanic populations." This report was introduced into evidence at the trial during the analyst's testimony. Again, let's look at it and then decipher it further.

[THIRD REPORT] WASHOE COUNTY SHERIFF'S OFFICE FORENSIC SCIENCE DIVISION

9/15/1994

LABORATORY NUMBER: L0277-94-1 SUSPECT: BROWN, TROY VICTIM: [DOE, JANE] DATE OF SUBMISSION: 02/08/94 OFFENSE: SEXUAL ASSAULT

The submitted items were identified as: Blood standard from [Jane Doe] Blood Standard from Troy Brown #14 T-1 Panties from [Jane Doe]

RESULTS OF EXAMINATION:

High molecular weight DNA was isolated from the above items and processed with the Hae III restriction enzyme. DNA fragments for five polymorphic loci were measured: D2S44 (YNII24), D4S139 (PII30), D10S28 (TBQ7), D14S13 (CMM101), and D17S79 (V1). Examination of the DNA fragments from the sperm fraction from the panties shows that each fragment matches the comparison sample from Troy Brown.

While this analysis does not constitute an absolute identification of the suspect Troy Brown as being the source of the sperm fraction from the panties, the frequency of this matching pattern is rarer than 1 in 3,000,000 in the Caucasian, Black, and Hispanic populations. This statistic was calculated using the "Ceiling Principle" as recommended in the 1992 National Research Council Report. This method of statistical analysis assumes an arbitrary minimum frequency of 10% for each DNA fragment and results in conservative estimates.

/s/ ______ANALYST RENEE L. ROMERO, CRIMINALIST

According to this report, in all three major population groups, no more than one person in three million would have the incriminating five-locus VNTR profile. Again, there is no statement of the uncertainty associated with the estimate. The estimated frequency is small because there are 10 multiplications for the five pairs of uncommon VNTR alleles. To be "conservative," the laboratory "arbitrarily" did not use any factor smaller than 1/10 in this chain of multiplications. This modification (and another one not stated in the report, but required by the "Ceiling Principle") are a response to modeling uncertainty.

The introduction of the "ceiling principle" constitutes a remarkable chapter in the history of science.¹⁸ Advocated as a compromise by a bitterly divided National Research Council (NRC) committee, it was immediately criticized as neither a ceiling nor a principle, and a nonsolution to a nonproblem. The raucous reception in the scientific and legal communities particularly among population geneticists—prompted the NRC to appoint a second committee, which concluded that the proposed modifications to the independence model for multilocus profile frequencies in major population groups were excessive.¹⁹

What motivated the invention of the "ceiling principle" was the fear that the model of random mating was oversimplified as a result of "population substructure." The major, censustype population groups of African Americans, Asian Americans, Caucasian Americans, and Native Americans are not simple randomly mating groups; rather, they are composites of ethnically diverse subpopulations that tend to mate most often among themselves. The first NRC committee explained that

For example, a person who has one allele that is common among Italians is more likely to be of Italian descent and is thus more likely to carry additional alleles that are common among Italians. The true genotype frequency [in the Italian subpopulation] is

¹⁸ See David H. Kaye, *The Double Helix and the Law of Evidence* (2010) [cited as *Double Helix*].

¹⁹ The author of this module was a member of the second committee.

thus higher than would be predicted by applying the multiplication rule and using the average frequency in the entire [Caucasian] population.²⁰

Fearing that such population structure could have a major effect on profile frequency estimates, this first NRC committee "strongly" recommended a new research program to acquire "[r]andom samples of 100 persons . . . drawn from each of 15-20 populations, each representing a group relatively homogeneous genetically."²¹ The "ceiling frequency" for an allele would be the largest frequency encountered in any of these samples. Moreover, to be particularly cautious, if this largest observed frequency was less than 5%, the ceiling frequency would be rounded up to that number. These allele ceiling frequencies then would be multiplied to provide what was believed to be an upper bound for the genotype frequency.²² Because the allele frequencies would come from many different population—structured or otherwise. It would treat everyone not as members of any particular ethnic group, but rather as if they have a fluctuating ethnicity that changed as each new locus was considered. For the first VNTR locus, an individual might be Cuban; for the second, Apache; for the third, Nepalese; and so on.

But the random samples of such smaller populations did not yet exist.²³ The report therefore recommended a temporary expedient—using the highest frequency in any of the major population groups in the United States (Caucasians, African-Americans, Hispanics, Asians, and Native Americans), or 10% (if the frequencies were less than 10%). This interim ceiling method is what the Sheriff's laboratory used. It relaxes the assumption of random mating within the general population (but still assumes random mating in subpopulations) and thus can be thought of as a check on how sensitive the computational model is to certain assumptions (Principle 2).

²⁰ Committee on DNA Technology in Forensic Science, National Research Council, *DNA Technology in Forensic Science* 79 (1992). On the other hand, multiplying the allele frequencies for all Caucasians with the independence assumptions in place to estimate the multilocus genotype frequencies for all Caucasians would tend to favor defendants. Using the average allele frequencies rather than accounting for population structure by computing genotype frequencies within each randomly mating subpopulation and combining them to obtain an estimate for the overarching population usually leads to overestimates of the population-wide genotype frequencies. In other words, ignoring population structure already is somewhat "conservative" in cases such as *Brown*, in which there is no reason to focus on suspects in any one ethic subpopulation. See David H. Kaye, DNA Evidence: Probability, Population Genetics and the Courts, 7 Harv. J. L. & Tech. 101 (1993).

²¹ Committee on DNA Technology in Forensic Science, supra note 19, at 83.

²² The computation also requires a factor of two for every locus at which an individual has two distinct alleles. This is because there are two, equally likely ways for an individual to have a pair of alleles *A* and *B* at a locus. The person could have inherited the *A* from the father and the *B* from the mother, or vice versa. In addition, with VNTR loci, it was not always clear whether observing only one allele *A* at a locus meant that the individual's genotype was *AA* (in case which the probability would be the square of the proportion *p* of *A* alleles in the population) or whether the genotype at the locus was *A* plus some other allele that was not detected. The work-around was to be "conservative"—to assume that the true genotype was *AX*, where *X* is some allele that has the maximum conceivable frequency of 1, resulting in the estimate of 2*p* for the frequency at that locus. Following this approach, Romero should have multiplied the allele frequencies by a factor of 2⁵ = 32 for the five-locus VNTR-profile frequency estimate.

²³ These groups the committee listed as examples of the subpopulations to be sampled. Committee on DNA Technology in Forensic Science, supra note 20, at 84.

This is not the place to examine the scientific merits of or the perceived need for either version of the "ceiling principle." But its use in *Brown* and other cases of that era exemplifies one response to modeling uncertainty—bending over backwards to avoid one type of error. This is a perennial issue for scientific evidence—when is it appropriate for the forensic scientist to be conservative as opposed to being as accurate as possible? Indeed, the question is not confined to forensic science. For example, in a death penalty case in which the IQ of the defendant was a critical consideration, the Justices debated whether a relatively rigorous 95% confidence interval for assessing measurement error was appropriate.²⁴ Outside of litigation, one might ask whether climatologists should provide conservative estimates of global warming as a response to uncertainties in their knowledge and modeling.

Discussion Question 3.1. Should the Sheriff's laboratory have used the allegedly "conservative" interim ceiling method rather than the simple multiplication method it had used in the first two reports?

In thinking about this question, consider the observation of one of the world's leading forensic scientists:

There is a widespread view that the expert must quote an assessment of the evidence which is conservative in the sense of erring so as to favour the defendant. This idea seems partly based on the view that, by understating the evidence, the scientist will give himself a more comfortable time in the witness box. On the contrary, once we concede the principle that it is necessary to be conservative we put ourselves in the situation which invites the defence to show that there is another method which is still more conservative than ours. And there always is — unless we opt out completely and say the evidence has no value.²⁵

Do you agree? Does the fact that the prosecution must prove guilt beyond a reasonable doubt militate in favor of scientific conservatism in presenting its experts' findings in criminal cases?

Discussion Question 3.2. A single-nucleotide polymorphism (SNP) is a substitution of one nucleotide base for another in a DNA sequence that occurs in more than 1% of the general population. For example, the sequences ATATTGCT and ACATTGCT display a SNP at the second base (shown in bold). SNPs occur throughout the human genome—about once in every 300 nucleotide base pairs. Suppose a test for 20 SNPs detected which variant (SNP allele) is present

²⁴ Hall v. Florida, 134 S.Ct. 1986 (2014).

 ²⁵ Ian W. Evett, Expert Evidence and Forensic Misconceptions of the Nature of Exact Science. 36 Sci. & Just.
 118, 120 (1996).

at each SNP locus. Under what conditions would it be reasonable to compute the frequency of this SNP profile as the product of the individual allele frequencies?

Now that we understand the origins of the numbers for DNA profile frequencies in *Brown*, it is time consider a second set of issues. These issues pertain to the presentation of even perfectly precise and valid frequencies or probabilities. They arise because of the ease with which certain kinds of probabilities are misconstrued. The next section sketches the framework of the mathematical theory of probability so that later sections can examine how it was applied in *Brown*.

4. A Probability Primer

4.1 Probability Mass

As originally formalized, probabilities are numbers assigned to all possible events (observations or outcomes of an experiment). Mathematically, the events are represented by sets, and the set of all possible events (the "sample space") can be denoted *S*. Three axioms circumscribe the ways in which the numbers can be assigned and allow us to think of probability as mass. We have a large quantity of mass to distribute across the events in the sample space. The following set of axioms is standard:

- First, we have to put either some probability mass on each event, or none at all. Hence, the probability of every event is at least zero. More formally, for every event A, P(A) ≥ 0.
- Second, we have to use up all the probability mass: The probability mass that collectively covers all the events is P(S) = 1, but we can put different fractions of it on the various events if we wish.
- Finally, the amounts must be such that if any two events A and B are mutually exclusive or disjoint—meaning that if one occurs, the other does not—then the probability that one *or* the other occurs is the probability that one occurs plus the probability that the other one does: P(A or B) = P(A) + P(B).

For example, if a cubical die must come to rest with one of the numbers 1 through 6 on top when tossed, then those six events comprise the sample space *S* of all possible outcomes. The probabilities assigned to these events (and combinations of them, such as an odd number (1, 3, or 5)) turning up, must obey the three axioms. For a fair die, the probability mass is evenly distributed: P(1) = P(2) = ... = P(6) = 1/6. For the fun of it, you might verify that the axioms hold for this uniform distribution. Do all event probabilities have non-negative values? Yes. Do the probabilities sum to 1? Yes. Does P(odd) = P(1) + P(3) + P(5)? Yes. And so on.

These axioms and the many theorems that follow from them constitute a rich mathematical structure. We described it as applying to events, but other interpretations are possible. One that is particularly fruitful in law and forensic science allows probabilities to be

assigned to propositions. A proposition that *must* be true—a logical truth such as "there are twelve eggs in a basket that has a dozen eggs in it"—has a probability of one. A proposition that has no chance of being true—a logical contradiction such as "someone who always tells the truth sometimes lies"—has a probability of zero. Other propositions—which is to say, all those encountered in the empirical sciences and in the courtroom—have probabilities larger than zero and smaller than one.

The mathematical rules that follow from the axioms allow one to calculate the probabilities of complicated events or propositions from the probabilities of simpler ones. For example, if A denotes the event that the die will turn up with a one showing, then the chance that the toss will *not* yield a one is the remaining probability mass: 1 - P(A) = 5/6.²⁶ Likewise, if any proposition H has probability P(H) attached to it, then its negation $\sim H$ must have the rest of the probability mass: $P(\sim H) = 1 - P(H)$.

TEST YOUR UNDERSTANDING

4.1. The victim of a face-to-face robbery perpetrated by a man with a knife is called to the police station and shown similar photographs of six men fitting the general description he gave of the robber. One of the six is the prime suspect, but the police do not tell him this. He identifies the suspect as the robber. Suppose that the victim had been determined to pick a photograph from the photo spread and simply guessed without paying the slightest attention to the pictures. What is the probability that he would have picked out the suspect? The probability that he would have guessed wrong?

4.2. The witness in the previous problem is quantitatively inclined. He tells the police that he cannot be sure that anyone in the photo spread is the robber, but he thinks the following are the chances that each man in the six photographs is the robber are 0.1, 0.2, 0.4, 0.05, 0.15, and 0.05—and that there is a chance of 0.1 that the robber is not in the six photos. Do these numbers meet the requirements for mathematical probabilities?

4.3. The witness says that the probability that either the man in the first photo or the one in the second committed the robbery is 0.5. When asked which one is more likely, he states that the first has probability 0.2 and the second has probability 0.3. Do these numbers meet the requirements for mathematical probabilities?

4.2 Conditional Probability

²⁶ This result follows from the axioms. The probability of an event that is certain is 1. For example, because a football team either will or will not win the next game it plays to completion, $P(A \text{ or } \sim A) = 1$, where A designates a victory for the team. Axiom 3 stipulates that $P(A \text{ or } \sim A) = P(A) + P(\sim A)$. Therefore, $P(A) + P(\sim A) = 1$, and $P(\sim A) = 1 - P(A)$.

A probability such as P(A) = P(1) = 1/6 for the toss of a die is an *unconditional probability* of a single event or proposition A. For any two events A and B, P(A & B) is the *joint probability* of A and B. For example, the probability of a 1 on the first toss and a 6 on a second, independent toss of a fair die is P(A & B) = P(AB) = P(1,6) = 1/36.²⁷

Another category of probabilities are conditional. They will prove particularly important for the testimony in the *Brown* case. The *conditional probability* of *A* given *B*, or P(A|B) for short, is the probability that event *A* will occur *given that B* occurs. For example, the conditional probability of the outcome of a toss of the fair die being a one (the event *A*) given that the outcome was an odd number (B = 1, 3, or 5) is P(A|B) = 1/3. More formally, the conditional probability P(A|B) is defined by its relationship to the joint probability: P(A & B) = P(B) P(A|B). In our example, the probability of a 1 and an odd number is the probability of an odd number times the probability of a 1, given that the number is odd. If you prefer, you can swap *A* and *B*, and write P(B & A) = P(A) P(B|A).

Conditional probability is related to the concept of independence that, we saw, was used in computing the frequencies in the first two laboratory reports. Events *A* and *B* are independent if and only if P(A & B) = P(A) P(B). Comparing this definition to the last two equations, events *A* and *B* are independent when P(A|B) = P(A) and P(B|A) = P(B). In other words, the occurrence of *B* has no bearing on whether *A* occurs, and vice versa—learning that *B* occurs supplies no information on whether *A* does, and vice versa. Test Your Understanding questions 2.5, 2.6, and 2.7 further illustrate these ideas.

4.3 Transposing Conditional Probabilities: Bayes' Rule

What is the relationship between P(B|A) and P(A|B)? They are obviously different (although they could have the same numerical values in some special cases). The probability that a toss of a fair die produces a 1 (event A) given that it produces an odd number (event B) is P(A|B) = 1/3. Exactly one of the three equally likely odd numbers is a 1. But the probability that the toss produces an odd number given that it produces a 1 is P(B|A) = 1. If a 1 appears, then an odd number surely has appeared. This is an example of Principle 5 at work—the probability of evidence indicating that a disputed event did (or did not) occur does not generally equal the probability of that event, as assessed in the light of the evidence.

Although P(A|B) is not generally equal to P(B|A), it is not difficult to use the expressions for conditional probabilities in the preceding section to transpose the events or propositions Aand B in a conditional probability expression. The formula is the most compact if odds are used in place of probabilities. Odds are just another way to express chances, but whereas probabilities are defined on a scale of zero to one, odds go from zero (cannot happen) to infinity (must happen). If the probability of an event is, say, 1/6 (one chance out of six), then the odds are 1 to 5 (one chance compared to five, which can be written as 1:5, 1/5, or 0.2).²⁸

 $^{^{27}}$ For two independent tosses of a fair die, there is exactly one way for the event (1,6) to occur, and this event is 1 of 36 equally likely events.

²⁸ More generally, Odds(A) = P(A) / [1 - P(A)]. Therefore, when P(A) = 0 (A is impossible), Odds(A) = 0. As P(A) approaches 1 (certainty), Odds(A) grow infinitely large.

We can compute the odds of A occurring given that B does from the odds of A (without knowing whether B has occurred) and the ratio of two conditional probabilities:

$$Odds(A|B) = \frac{P(B|A)}{P(B| \sim A)} Odds(A)$$

This equation is known as Bayes' rule.²⁹ It will be immensely useful in understanding the opinions in *Brown*, for it offers one logical way to use probabilities about the evidence of an event to determine the probability of the event itself.

The term on the far right of the equation is known is the *prior odds* because it denotes the odds of event *A* prior to learning about event *B*. Now, unless *B* is entirely independent of A, its occurrence supplies information about whether or not *A* occurs. The rule specifies how to update the odds in light of this information—just multiply the prior odds by the middle term, which is called the *likelihood ratio*. The result is the *posterior odds*. We can summarize the rule in these six words: *posterior odds* = *likelihood ratio* × *prior odds*, or even more compactly, *posterior odds* = *LR* × *prior odds*, where *LR* is short for "likelihood ratio."

To make the meaning of this equation more concrete, let us return to tossing the fair die once. At the outset, we know that the probability of a 1 is 1/6. So the prior odds are Odds(A) = Odds(1) = 1:5 = 1/5. Now, we learn that the outcome of the toss (*B*) is an odd number. So we calculate the likelihood ratio *LR*. The numerator is the probability of *B* given *A*, and for the roll of the die, it is the probability that the number is odd given that it is a 1. Thus, the numerator is P(B|A) = 1. To find the denominator, we need to think about what happens for ~*A*. If the die is not showing a 1, it is a 2, 3, 4, 5, or 6. The probability that the number is odd, given that it is not 1, therefore is 2/5. (There are two odd numbers among the five possibilities that are left when ~*A* occurs.) So the denominator, P(B|~A) = 2/5, and the ratio is 1 / (2/5) = 5/2. In other words, *B* is 2.5 times more likely to occur when *A* does than when *A* does not. That *B* occurred therefore raises the prior odds of *A* from Odds(A) = 1/5 to the posterior odds, $Odds(A|B) = 5/2 \times 1/5 = 1/2$. These updated odds correspond to a posterior probability P(A|B) = 1/(2+1) = 1/3—exactly what we found earlier.

In sum, before we received the evidence that the top number on the die was odd, we had considerable doubt that a one had appeared. Our probability for this event A was only 1/6, just under 17%. After processing the evidence *B*, we are more certain that A occurred— although we are still not convinced. The prior probability of 17% has been updated to the posterior probability of 1/3, just under 34%. The changing probability thus measures the degree of certainty about the truth of the hypothesis that A occurred. This is a manifestation of Principle 4 (uncertainty can be quantified with probabilities) as well as Principle 6 (Bayes' rule

$$P(^{A}|B) = P(^{A}) P(B|^{A}) / P(B).$$

Dividing the last equation into the one above it gives the form of Bayes' rule presented in the text.

²⁹ The rule is easily derived. By definition, P(A & B) = P(B) P(A|B) and P(B & A) = P(A) P(B|A). But the joint event A & B is identical to B & A. Hence, P(B) P(A|B) = P(A) P(B|A). Dividing both sides by P(B) gives P(A|B) = P(A) P(B|A) / P(B|A).

But we can substitute ~A for A in this equation to obtain

shows how to go from the probability of evidence about the event to the probability of the event itself).

The example also illustrates Principle 7—that the strength, or weight, of evidence can be quantified by comparing the probability of the evidence given one hypothesis to the probability of the evidence given an alternative hypothesis. The shift in the odds of *A* resulting from the evidence *B* was given by the ratio of (1) the conditional probability for evidence given the hypothesis that *A* occurred to (2) the conditional probability for evidence given the alternative hypothesis that *A* did not occur. Here, this likelihood ratio was LR = P(B|A) / P(A|B)= 1 / (2/5) = 2.5. To be clear about what is the evidence and what are hypotheses, we can write

$$LR = \frac{P(E|H)}{P(E|\sim H)}$$

The probability of evidence conditioned on a hypothesis is called the *likelihood* for the hypothesis,³⁰ and the ratio of this probability for one hypothesis as opposed to a competing hypothesis is the likelihood ratio for those hypotheses.

The idea in Principle 7 fits the intuitive observation in Sections 1 and 2 that the rarer the matching DNA profile, the stronger the evidence of identity. If the suspect is the source of the crime-scene DNA, then the laboratory should report a match with a probability near 1. If the suspect is not the source and the true source is not genetically related to the suspect, then the probability that the laboratory will find a match is the proportion p of the relevant population with the crime-scene DNA profile. The likelihood ratio is thus LR = 1/p. A rare profile has a small p—which makes the LR large. For example, if only one in a million individuals share the profile, LR is one million, but if one in five people have the profile, LR is only 5. The match is relevant evidence in both cases—it changes the odds that the suspect is the source—but its strength varies with the frequency of the profile.

Bayes' rule is also a handy tool because it tells us how to transpose the *E* and the *H* in a conditional probability. For example, a clinical test for a disease can be performed on subjects known to have the disease and on other subjects known to be disease-free. Thus, the PSA test for prostate cancer has been shown to give a positive result (to indicate that the cancer is present) about 70% of the time when the cancer is in fact present and about 10% of the time when it is not.³¹ Letting *H* be the hypothesis that a patient has cancer, and letting +*E* denote positive evidence from the PSA test, the conditional probabilities are P(+E|H) = 0.7 and P(+E|~H) = 0.1.

But the patient and physician want to know whether the patient has the cancer. The probability of cancer given a positive test result is P(H|+E). Bayes' rule teaches us that we cannot simply use 0.7, or 70%, as an estimate of this probability (Principle 5). We have to consider not only this value of 0.7 for P(+E|H), but also the likelihood for the alternative hypothesis, $P(+E|\sim H) = 0.1$, and the prior odds. Suppose, then, that the age-adjusted frequency

³⁰ Technically, the likelihood is only proportional to the conditional probability. Because the proportionality constant cancels out when forming a likelihood ratio, we can ignore it.

³¹ Carvell T. Nguyen & Michael W. Kattan, Prediction Models in Prostate Cancer Diagnosis, in *Prostate Cancer Diagnosis: PSA, Biopsy and Beyond* 85, 86 (J. Stephen Jones ed., 2012).

of prostate cancer in similar patients is 1 in 11. This statistic suggests the prior odds are 1 to 10. Bayes' rule instructs us to multiply these odds by 0.7 / 0.1 = 7. The high PSA level in this patient has raised the odds from 1/10 to 7/10. Odds of 7 to 10 are the same as a probability of 7/(10 + 7) = 0.41. With a 41% probability of cancer, a biopsy might well be advisable.

TEST YOUR UNDERSTANDING

4.4. According to the World Health Organization, a rapid test for the presence of the Ebola virus in blood has a sensitivity of 92% and a sensitivity of 85%.³² That is, the probability of positive evidence from the test when applied to blood that contains the virus (a true positive) is P(+E|H) = 0.92, and the probability of negative evidence from the test when applied to blood that does not contain the virus is P(-E|~H) = 0.85. (a) What is the probability of a false positive, P(+E|~H)? (b) What is the likelihood ratio for H based on a positive result? (c) Suppose that considering the prevalence of the disease in the relevant population and the symptoms presented by the patient, the clinician believes that the probability that the patient has Ebola is 20%. What are the odds of H prior to learning the test result? (d) The test is positive for Ebola. What are the odds now? (e) The posterior probability? (f) Suppose the prior probability, as assessed by the clinician, are 60% rather than 20%. What are the prior odds now? (g) The posterior odds? (h) The posterior probability?

4.5. FBI criminalists examine "questioned" hairs recovered for crime scenes and "known" hairs from suspects for gross and microscopic physical similarities that would indicate whether the questioned hairs could have come from the suspects. In one study, researchers compared the criminalists' judgments to DNA tests on the hairs (which, we may assume, are certain to be positive (show matching DNA sequences) when the hairs have a common source and are certain to be negative (show different DNA sequences) when the hairs come from different individuals. A portion of the results are shown in the following table:³³

	MtDNA –	MtDNA +
Criminalists –	17	0
Criminalists +	9	69

(a) What is the true positive probability P(+E|H)? (b) What is the false positive probability $P(+E|\sim H)$? (c) What is the likelihood ratio? (d) How do the judgments of the criminalists compare to the flip of a coin? To the Ebola test?

http://apps.who.int/iris/bitstream/10665/175554/2/WHO_EVD_HIS_EMP_15.2_annex_eng.pdf.

³² WHO, Selection and Use of Ebola In Vitro Diagnostic (IVD) Assay Annex 1. Predictive Values of Ebola RDTs and Implications for Decision-makers 13 (2015),

³³ David H. Kaye, Ultracrepidarianism in Forensic Science: The Hair Evidence Debacle, 72 Wash. & Lee L. Rev. Online, 227, 234 tbl. 1 (2015) (analyzing table 2 of Max M. Houck & Bruce Budowle, Correlation of Microscopic and Mitochondrial DNA Hair Comparisons, 47 J. Forensic Sci. 1, 3 (2002)).

Having seen how probabilities for evidence can be computed under different hypotheses and how those conditional probabilities can be inverted or transposed with knowledge of the prior probability, let's look at the expert testimony about the probabilities of evidence and hypotheses in Troy Brown's trial. We will see that the prosecution neglected Principle 5, but that how much this logical fallacy mattered is debatable.

5. The Expert Testimony at Trial

From the direct examination of Renee Romero:

Q: Ma'am, what is your occupation? A: I'm a criminalist with the Washoe County Sheriff's Office. Forensic Crime Lab. Q: How long have you been doing that? A: Five and a half years. Q: Would you please describe to the jury, your educational background? A: I have a Bachelor's in Chemistry with the fulfillment of a Bachelor's in Forensic Science. And my thesis project is currently being reviewed for Master's in Cellular Molecular Biology. Q: What training have you had specifically in DNA analysis? A: I spent a month at the FBI, training specifically in DNA techniques, and have also spent time in a training course at Roche Molecular Systems in Alameda, California. And part of my graduate work was working on population studies in the DNA laboratory. Q: And are you a member of any professional organizations or associations related to that field? A: I'm a provisional member of the American Academy of Forensic Scientists, and I'm also a member of TWGDAM, that's an acronym --T-W-G-D-A-M -- and it stands for the Technical Working Group on DNA Analysis Methods. Q: And what is that group responsible for? A: We meet back at the FBI twice a year, and we discuss quality control and quality assurance methods with the DNA technique. O: Would you describe what DNA is and then how you can make comparisons of DNA. *** Q: All right. Did you have an occasion to examine some pieces of evidence that were submitted to you concerning the defendant, Troy Brown? A: Yes, I did. Q: All right. Would you please describe to the jury, then, once you had those stains, what did you do with them? A: From the very beginning? Q: Please.

A: Once I had these stains in my hands, I extracted the DNA, the first test I ran was a PCR test. And the second test I ran was an RFLP test. Q: All right. With reference to the RFLP test, what testing did vou do? A: I did the testing that I just described here - the RFLP testing, extracted the DNA from these stains and ran them out on the gel, blotted them, and examined them using five probes. Q: Now, what does that mean when you used five probes? A: I examined five different areas of that DNA. I have that membrane that contains the DNA which remains single stranded. First I attached one probe to it, read those results, stripped that probe off, attached another probe to it, stripped that probe off, attached another probe to it, etc., and read those results. 0: What was the reason for continuing to follow through with additional probes? A: Each probe provides more discriminating value. Your first probe, your statistic may be one in a few hundred. And your second probe it jumps up to thousands, and with each probe your discrimination level gets tighter and tighter. Q: And would you please explain what you mean by discrimination level. A: The frequency of occurrence of that DNA pattern gets smaller and smaller. Q: In other words, the -a person -two persons having that same genetic code decreases. A: Correct. Q: Why is important to do one probe after the other. In other words, what is your goal? A: The goal is to make it more discriminating. Q: All right. That -A: We want to decrease the amount of people - the pool of peoplethat could have contributed to that stain. With each probe we use we're throwing out a certain amount of people. *** Q: Now, in terms of the probe matches that you just previously described, what is the frequency that you would expect to find that match in the White, Black, Hispanic population? A: We determine the frequency using a statistical calculation called the ceiling principle. The ceiling principle is from the National Research Council, which was formed under the National Academy of Scientists - came out with a report on DNA technology because there was some controversy over the statistical application. They came out with this conservative method of the ceiling principle. I utilized the ceiling principle in this case, in the frequency, to answer your question, it was one in 3 million. Q: So that means that only one in 3 million people will share the same genetic code.

A: Approximately, yes.

Q: Now, for my benefit, we're looking at a one in 3 million statistic, is there another way to show that statistic? In other words, what - let's say 100 percent - what is the likelihood that the DNA found in the panties is the same as the DNA found in the defendant's blood? A: Paternity testing uses percentages. 0: 0kav. A: Not the way forensics likes to look at it. We prefer the one in 3 million. Q: I understand that, but for just another way to look at it, what would that percentage be? A: It would be 99.99967 percent. That's what -MR. SMITH: May I go to the blackboard, Judge? THE COURT: Well, we'll pull it out for you and you want her to write on it? MR. SMITH: No, I'd like to write on it. THE COURT: Well, I don't think you're a witness. I'm not going to let you write on it now. MR. SMITH: All right. THE COURT: If you want her to write on it, she can write on it. By MR. SMITH (continuing): Q: Okay. If - Ms. Romero, if you'd write it down, please. A: Okav. THE COURT: Is there a - could you make the decimal a little bigger for us older -By MR. SMITH (continuing): Q: So, okay. So, if you would do, also, put 100 percent on top of that, if you would, please, with the corresponding number of zeros after the decimal point. And if you would please, then, subtract the lower number from the higher number. A: I don't think this is right. Just a minute. Q: All right. So, if you put a little line under the 99 there and a minus – subtraction indication. All right. All right. Thank you. If you'd just take - resume the stand. [The blackboard looked roughly as follows:

RFLP

100.000000 - 99.999967% .000033]

Q: So, the – would it be fair to say, based on that that the chances that the DNA found in the panties – the semen in the panties – and the blood sample, the likelihood that it is not Troy Brown would be .000033?

A: I'd prefer to refer to it as the one in 3 million.

Q: All right. But from a mathematical standpoint, would that be inaccurate? A: Repeat the question, please. Q: Would it be fair, then, to say that with that mathematical calculation there, that the likelihood that the DNA extracted from the semen in the panties and the DNA extracted from the blood that the likelihood that it's not Troy Brown, that it's not a match is .000033? MR. LOCKIE: Your Honor, I'm going to object on relevance. The witness is testifying that it's not scientifically valid in her opinion. So it's not relevant. THE COURT: Well, I don't know that -MR. LOCKIE: That's just a subtraction problem. THE COURT: Let's go back. I don't think that's what she said. I don't think that's what she said. Let's go back a step and find out. I don't think that's what she said. By MR. SMITH (continuing): Q: Now, I understand that - and what I'm trying to do is make this into a percentage where I can understand it. And so I recognize that as far as your testing, you would prefer to have it as a one in 3 million, but just as another way of looking at it, would it be inaccurate to state it that way? A: It's not inaccurate, no. Q: All right. Then in response to my question, would the likelihood that the semen from the DNA found in the panties and the blood from Troy Brown, that it's not the same, would it be the chances that they are not a match would be .000033? A: Yes. That's the way the math comes out. Q: All right. THE COURT: Let's make sure. It's the same thing - it's the same math just expressed differently. Is that correct? THE WITNESS: Yes. Exactly, your Honor.

Discussion Question 5.1. Let H_T be the hypothesis that Troy is the source of the semen and H_U be the hypothesis that someone genetically unrelated to Troy is. Let +*E* be the fact of the fivelocus match. Has Romero testified that the probability of the evidence +*E* given the hypothesis that an unrelated person is the source is $P(+E|H_U) = 1/3,000,000 = 0.000033\%$? Has she testified that the probability that someone else is the source, given the match, is $P(H_U|+E) = 0.000033\%$? That $P(H_T|+E) = 99.999967\%$? Out of court, Romero maintained that "the prosecutor . . . led her through a complicated mathematical journey aimed at reducing Brown's guilt to a percentage. . . . Romero found herself agreeing with the district attorney's math, but not in how he was applying it. She says the judge stepped in to clarify the matter . . ." Ed Pearce, DNA Expert Says Evidence Was Solid, KOLO8 News Now, "Rape Conviction Overturned," May 7, 2008, http://www.kolotv.com/home/headlines/18712759.html. Did the answer to the judge's question clarify the matter?

On recross examination:

Q: Does that statistical probability change with brothers? A: Yes. Q: How does it change? A: With a brother, there would be some genetic relationship. They have a 25 percent chance of sharing both alleles – both bands, and 50 percent chance of sharing one band. THE COURT: Well, I - can you help us with that. What does that mean? Is there any way to help us with that more than what you've just said? THE WITNESS: Yes. THE COURT: Would you please. THE WITNESS: In the National Research Council report, the one I spoke about earlier, the committee wrote the report under the auspices of the National Academy of Sciences, they addressed that problem. in their discussion about useful data-banking DNA, and they have a formula in there, and they state that they could share -25 percent could share two bands and 50 percent chance to share one band. And they have a formula in there that they use utilize the frequency of occurrence of the alleles. And in this case that turns out to be one in 6,500. Meaning those two adults would have to mate and produce offspring 6,500 times to come up with that pattern again.

* * *

On redirect examination:

Q: Yeah. When - what was the percentage again? A: It was one in 6,500. Q: So, in practical terms that means that the - would it be fair to say that the parents would have to have 6,000 - bear, actually bear 6,000 offspring? A: Um-hum. Q: For them to have 2 children, who are not identical twins, share the exact genetic code. A: That's how that formula comes out, yes. MR. SMITH: That's all I have. THE COURT: Mr. Lockie?

On recross examination:

Q: Would it necessarily have to be the 6,000th one that was identical?
A: No.
Q: All right. It could be the second one, couldn't it? Just as easily as it could be the 6,000th, then, right?
A: Sure.

MR. LOCKIE: All right. Thank you. I have no further questions.

Discussion Question 5.2. Romero agreed that "the parents would have to . . . actually bear 6,000 offspring . . . for them to have 2 children . . . share the exact genetic code." Suppose that you are tossing a die that has a probability of 1/6 of showing a 1 on each try. Would you have to flip it six times before observing a 1?

Discussion Question 5.3. If the probability that two parents would have one child followed by the next child with the same DNA identification profile is 1/6,000, is the probability that *at least* two of five brothers have the same profile higher, lower, or the same as 1/6,000? Notice that there is more than one possible pair of brothers to consider. How many pairs are there?

On redirect examination:

Q: So you indicated that - what is the percentage - we're looking at one to 100 percent, one in 6,000. A: One in 6,500? Q: Yes. All right. And if you would put a 100 - 100 above that with the: corresponding number of zeros, with a - I think you need another zero there. A: I don't think so. Q: Oh, I'm sorry, that's a percent sign. And a - subtraction and a line. And would you subtract the lower number from the higher number, please. All right. So, would it be fair, then, to say that the likelihood of the parents having one child, and then the very next child having the same genetic code would be .02 percent? A: Yes.

[The blackboard now looked roughly as follows:

RFLP 100.000000 <u>- 99.999967%</u> <u>.000033</u>

Brother 10000 <u>99.982</u>% .02]

MR. SMITH: Thank you. That's all I have.

On recross examination:

Q: Does that change at all with two brothers? A: No.

6. Letter Introduced to Support a Petition for a Writ of Habeas Corpus

UNIVERSITY OF CALIFORNIA, IRVINE LAURENCE D. MUELLER Department of Ecology & Evolutionary Biology Irvine, CA. 92697-2525

28 February 2006

Mr. Paul Turner Law Offices of the Federal Public Defender 411 E. Bonneville Ave., Suite 250 Las Vegas, Nevada 89101

RE: Brown v. Farwell

Dear Mr. Turner:

* * *

On page 343 of the trial transcript Ms. Romero states that there is a 25% chance of two brothers sharing both alleles at a single locus in common. This conclusion is only correct if both parents are heterozygotes and share at most one allele in common. For other possible parental pairs the probability of two sibs matching could be 50% or 100%. Thus, in this portion of Ms. Romero's testimony she has chosen a special case which suggests that sibs have the lowest chance of matching that is biologically possible.

On page 344 Ms. Romero utilizes the assumption that two siblings have a 25% chance of matching at a single locus to estimate that the chances of two siblings matching at five of the loci used in this case would be 1 in 6500. Even if we assume that 25% is the proper number to use in this calculation the chance of two brothers matching is $(0.25)^5 = 1$ in 1024 not 1 in 6500. As before the error made here by Ms. Romero tends to suggest that the chance of two brothers matching is actually much less than it really is.

I have been told that defendant Brown may in fact have 2 brothers that were living in the near vicinity of the crime and two additional brothers that lived a greater distance away. With these additional facts we can address the more relevant question about siblings which is: what is the chance that any one of the two (four) brothers would match the evidence profile? Even if we use the least likely chance of two brothers matching at one locus, 25%, the chance that one or more brothers would match the evidence is 1 in 512. With four brothers this total goes up to 1 in 256. The calculations in the preceding two paragraphs used the smallest probability that two sibs will have matching profiles. As mentioned before this probability could be as high as 100%. The second National Research council report on DNA typing gave a formula for computing the chance of matching profiles when the profiles of the parents are unknown. Using this formula and the FBI Caucasian database the chance of a single sib matching defendant Brown's profile is 1 in 263. The chance that among two brothers one or more would match is 1 in 132 and the chance that among four brothers one or more would match is 1 in 66.

In the course of questioning Ms. Romero about the statistical meaning of the DNA match Prosecutor Smith says (pg. 338) "the likelihood that it is not, Troy Brown would be 0.000033?" The accepted meaning of these statistics is that they represent the chance that a single person chosen at random from the suspect population would match the evidence. The phrase above suggests that the probability is that chance of anyone other than the defendant matching the evidence sample. This is not correct since in the population at large there are very many people not just one and the chance of finding a match to any one of these people is much greater (more likely) than the chance of finding a match to just a single person. This erroneous phraseology is in fact so common it has been given a special name, the prosecutor's fallacy.

* * *

Sincerely, /s/ Laurence D. Mueller

7. Majority Opinion in *Brown v. Farwell*, 525 F.3d 787 (9th Cir. 2008)³⁴

WARDLAW, Circuit Judge:

In the early morning of January 29, 1994, Jane Doe, a nine-year-old girl, was sexually assaulted in the bedroom of her trailer home in Carlin, Nevada. At the time, Jane was home alone with her four-year-old sister while their mother, Pam, was drinking at a bar and their step-father . . . was working the night shift. . . .

* * *

... Pam accompanied Troy to Peacock Bar where they had a drink. Troy was clearly drunk, but "he behaved like a gentleman and made no sexual advances toward" Pam. Pam stated that the last time she saw Troy was between 11 p.m. and midnight. However, one bartender stated that Troy left the bar no later than 12:20 a.m., and another bartender stated that she believed she saw Troy at the bar at 1:30 a.m. Between midnight and 12:30 a.m., Jane called Pam at the bar and explained that some man was at the trailer looking for Pam and had

³⁴ Judge Diarmuid Fionntain O'Scannlain dissented. Various citations as well as all footnotes and headings are omitted from the opinion of Judges Kim McLane Wardlaw and Michael D. Hawkins reproduced here.

hurt her. When Pam arrived home, she found Jane covered in blood from the waist down and called 911. A police officer and the paramedics responded...

Later at the hospital, vaginal and anal penetration were confirmed. Jane had bruises on her neck and scratches on her face. A vaginal smear was taken because sperm was present. Debris was collected from her teeth because she stated she had bitten the assailant's hands. Jane described the assailant to the police that night as follows:

[H]e did not wear a hat and had blonde or sandy-colored hair which was curly at the bottom and thinning on top; she thought he had a small moustache; he was wearing dark jeans, a black jacket with "a zipper for sure," a western type shirt, boots, and a watch which scraped her face.... [She] stated that the assailant smelled like cologne but that it was an "awful smell" ... [like] "beer or puke or something."

That night, Troy was wearing a cowboy hat, dark jeans, a black satin jacket with an orange and yellow CG's logo on the back, and boots. Two witnesses testified that, at 1:05 a.m. that morning, they saw near Jane's trailer a man wearing a cowboy hat, dark jeans, and a black satin jacket with a bright green emblem on the back that looked like a skull or bandit.

Troy stated that he had been drinking steadily that night and, while walking home to his trailer located ten trailers away from Jane's, had vomited several times, soiling his pants and shirt. When he arrived home, Troy's brother, Travis, awoke from sleeping on the couch. Travis stated that it was 1:32 a.m. when he awoke and that he did not see any traces of blood in the house. Troy washed his clothes as soon as he returned home because he was leaving that day to go to Utah for a week and all of his clothes were already packed. When a police officer arrived at 5 a.m. to question Troy, he saw no blood on Troy or his boots; he also checked Troy's hands, which did not have any evidence of bite marks.

Jane also stated that she fell asleep with a night light on, but that the man who assaulted her must have turned it off because it was off when the man left. Troy's fingerprints were not found anywhere in Jane's trailer, and the one fingerprint found on the night light did not match Troy's....

At trial, Respondents presented the testimony of DNA expert Renee Romero of the Washoe County Sheriff's Office Crime Lab. Romero testified that, among other things, there was a 99.99967 percent chance that Troy was the assailant.

The jury found Troy guilty of two counts of sexual assault on a child under the age of fourteen . . . and one count of abuse or neglect of a child resulting in substantial bodily harm Troy appealed to the Nevada Supreme Court, [which] vacated the third charge and remanded for resentencing on the second sexual assault count. The trial court resentenced Troy to life with the possibility of parole after ten years on both sexual assault counts, to run consecutively. Troy again appealed to the Nevada Supreme Court, which rejected his appeal. Troy next filed a state petition for post-conviction relief, and, after holding an evidentiary hearing, the state courts denied relief.

On February 6, 2004, Troy filed his federal petition for writ of habeas corpus . . . , arguing, inter alia, violations of due process and ineffective assistance of counsel. Judge Pro permitted Troy to expand the record, admitting, among other things, an uncontested report discrediting Romero's testimony by Dr. Laurence Mueller (the "Mueller Report"), a professor of Ecology and Evolutionary Biology at the University of California, Irvine.

The district court granted Troy's petition. First, the district court concluded that, in light of the Mueller Report, Romero's testimony was unreliable. Absent that testimony, no rational trier of fact could conclude beyond a reasonable doubt that Troy was guilty of each and every element of the offenses with which he was charged. The district court also concluded that Troy's attorney's failure to diligently defend against Respondents' DNA testimony, as well as his failure to investigate the alibi of . . . a potential suspect, amounted to ineffective assistance of counsel. Respondents timely appealed. . . .

Troy asserts that there was insufficient evidence to convict him. His argument rests on the admission of Romero's later discredited testimony regarding the DNA evidence, which was introduced without rebuttal at trial. Respondents have conceded that absent introduction of Romero's DNA evidence, the remaining evidence is insufficient to sustain Troy's conviction....

We agree with the district court that the Nevada Supreme Court's decision was both "contrary to" and "an unreasonable application of" *Jackson v. Virginia*, 443 U.S. 307 (1979). In *Jackson*, the Supreme Court held that a conviction must be upheld if, "after viewing the evidence in the light most favorable to the prosecution, any rational trier of fact could have found the essential elements of the crime beyond a reasonable doubt." Id. at 319....

The district court also correctly concluded that the Nevada Supreme Court's decision was "an unreasonable application of" *Jackson* because, in light of the Mueller Report, no rational trier of fact could have found Troy guilty beyond a reasonable doubt on the evidence presented at trial. The validity and accuracy of the Mueller Report went unchallenged by Respondents, and the district court found it to be credible....

The Mueller Report indicates that Romero's testimony was unreliable for two main reasons. First, Romero testified that there was a 99.99967 percent chance that Troy's DNA was the same as the DNA discovered in Jane's underwear—or, in other words, that the science demonstrated a near 100 percent chance of Troy's guilt. This assertion was incorrect, as it falls directly into what has become known as the "prosecutor's fallacy."...

Here, Romero initially testified that Troy's DNA matched the DNA found in Jane's underwear, and that 1 in 3,000,000 people randomly selected from the population would also match the DNA found in Jane's underwear (random match probability). After the prosecutor pressed her to put this another way, Romero testified that there was a 99.99967 percent chance that the DNA found in Jane's underwear was from Troy's blood (source probability). This testimony was misleading, as it improperly conflated random match probability with source probability. In fact, the former testimony (1 in 3,000,000) is the probability of a match between an innocent person selected randomly from the population; this is not the same as the probability that Troy's DNA was the same as the DNA found in Jane's underwear, which would prove his guilt. Statistically, the probability of guilt given a DNA match is based on a complicated formula known as Bayes's Theorem, and the 1 in 3,000,000 probability described by Romero is but one of the factors in this formula. Significantly, another factor is the strength of the non-DNA evidence. Here, Romero improperly conflated random match and source probability, an error that is especially profound given the weakness of the remaining evidence against Troy. In sum, Romero's testimony that Troy was 99.99967 percent likely to be guilty was based on her scientifically flawed DNA analysis, which means that Troy was most probably convicted based on the jury's consideration of false, but highly persuasive, evidence.

Second, Romero inaccurately minimized the likelihood that Troy's DNA would match one of his four brothers' DNA, thus underestimating the likelihood that one of Troy's brothers could have been the perpetrator. She testified that there was a 25 percent chance of two brothers sharing both alleles at one locus, and, using that figure, a 1/6500 chance that one of Troy's brothers would match Troy's DNA at all five loci. The Mueller Report indicated that Romero's calculation was incorrect, as the correct figure is 1/1024. More importantly, Romero's testimony is misleading because it presented the narrowest interpretation of the DNA evidence. Had Romero accounted for Troy's four brothers, two of whom lived in Carlin and two of whom lived in neighboring Utah, the chance that Troy's DNA would match at least one of his four brothers' DNA can increase to 1/66— almost one hundred times the probability asserted by Romero.... Again, Respondents introduced nothing to contradict the findings of the Mueller Report.

A federal court on habeas may exclude evidence admitted in the state court if the evidence "rendered [the] trial so fundamentally unfair as to violate federal due process." *Butcher v. Marquez*, 758 F.2d 373, 378 (9th Cir. 1985). We agree with the district court that Romero's testimony was unreliable, as it was inaccurate and ignored logical implications about Troy's four brothers, each of whom lived in the general vicinity. Admission of this unreliable testimony most certainly rendered the trial fundamentally unfair, as even Respondents concede that "[t]here was insufficient evidence to convict the Defendant unless the DNA evidence established his guilt." Thus, the admission of Romero's unreliable and misleading testimony violated Troy's due process rights

... The conflicts in the evidence are simply too stark for any rational trier of fact to believe that Troy was the assailant beyond a reasonable doubt, an essential element of any sexual assault charge. This conclusion is confirmed by Respondents' own concessions. Therefore, the Nevada Supreme Court's decision was "an unreasonable application of" *Jackson*.

Because we affirm the district court's grant of Troy Brown's habeas petition on due process grounds, we need not reach his arguments regarding ineffective assistance of counsel. The district court's grant of Troy's petition for writ of habeas corpus and reversal of his conviction is AFFIRMED. Respondents shall retry Troy within 180 days or shall release him from custody.

8. Supreme Court Opinion in McDaniel v. Brown, 558 U.S. 120 (2010)

PER CURIAM.35

Respondent . . . filed this federal habeas petition, claiming there was insufficient evidence to convict him on the sexual assault charges and that the Nevada Supreme Court's rejection of his claim was both contrary to, and an unreasonable application of, *Jackson*. He did not bring a typical *Jackson* claim, however. Rather than argue that the totality of the evidence admitted against him at trial was constitutionally insufficient, he argued that some of the evidence should be excluded from the *Jackson* analysis. In particular, he argued that Romero's testimony related to the DNA evidence was inaccurate and unreliable in two primary respects: Romero mischaracterized the random match probability and misstated the probability of a DNA match among his brothers. Absent that testimony, he contended, there was insufficient evidence to convict him. . . .

* * *

We granted certiorari to consider two questions: the proper standard of review for a *Jackson* claim on federal habeas, and whether such a claim may rely upon evidence outside the trial record that goes to the reliability of trial evidence...

The Mueller Report does not dispute Romero's opinion that only 1 in 3,000,000 people would have the same DNA profile as the rapist. Mueller correctly points out, however, that some of Romero's testimony—as well as the prosecutor's argument—suggested that the

³⁵ Justice Thomas, joined by Justice Scalia, wrote separately to "disagree with the Court's decision to complicate its analysis with an extensive discussion of the Mueller Report." As with the Court of Appeals opinion, citations, footnotes, and headings are omitted without further notice.

evidence also established that there was only a .000033% chance that respondent was innocent.

Looking at Romero's testimony as a whole, though, she also indicated that she was merely accepting the mathematical equivalence between 1 in 3,000,000 and the percentage figure. At the end of the colloquy about percentages, she answered affirmatively the court's question whether the percentage was "the same math just expressed differently." She pointed out that the probability a brother would match was greater than the random match probability, which also indicated to the jury that the random match probability is not the same as the likelihood that someone other than Troy was the source of the DNA.

The Mueller Report identifies a second error in Romero's testimony: her estimate of the probability that one or more of Troy's brothers' DNA would match. Romero testified there was a 1 in 6,500 (or .02%) probability that one brother would share the same DNA with another. When asked whether "that change[s] at all with two brothers," she answered no. According to Mueller, Romero's analysis was misleading in two respects. First, she used an assumption regarding the parents under which siblings have the lowest chance of matching that is biologically possible, but even under this stingy assumption she reported the chance of two brothers matching (1 in 6,500) as much lower than it is (1 in 1,024 under her assumption). Second, using the assumptions Mueller finds more appropriate, the probability of a single sibling matching respondent is 1 in 263, the probability that among two brothers one or more would match is 1 in 132, and among four brothers it is 1 in 66.

In sum, the two inaccuracies upon which this case turns are testimony equating random match probability with source probability, and an underestimate of the likelihood that one of Troy's brothers would also match the DNA left at the scene.

Although we granted certiorari to review respondent's *Jackson* claim, the parties now agree that the Court of Appeals' resolution of his claim under *Jackson* was in error. Indeed, respondent argues the Court of Appeals did not decide his case under *Jackson* at all, but instead resolved the question whether admission of Romero's inaccurate testimony rendered his trial fundamentally unfair and then applied *Jackson* to determine whether that error was harmless . . .

Respondent no longer argues it was proper for the District Court to admit the Mueller Report for the purpose of evaluating his *Jackson* claim, and concedes the "purpose of a *Jackson* analysis is to determine whether the jury acted in a rational manner in returning a guilty verdict based on the evidence before it, not whether improper evidence violated due process." There has been no suggestion that the evidence adduced at trial was insufficient to convict unless some of it was excluded. Respondent's concession thus disposes of his *Jackson* claim. . . .

... Even if we set that concession aside, however, and assume that the Court of Appeals could have considered the Mueller Report in the context of a *Jackson* claim, the court made an egregious error in concluding the Nevada Supreme Court's rejection of respondent's insufficiency-of-the-evidence claim "involved an unreasonable application of ... clearly established Federal law."

Even if the Court of Appeals could have considered it, the Mueller Report provided no warrant for entirely excluding the DNA evidence or Romero's testimony from that court's consideration. The Report did not contest that the DNA evidence matched Troy. That DNA evidence remains powerful inculpatory evidence even though the State concedes Romero overstated its probative value by failing to dispel the prosecutor's fallacy. And Mueller's claim that Romero used faulty assumptions and underestimated the probability of a DNA match between brothers indicates that two experts do not agree with one another, not that Romero's estimates were unreliable. Mueller's opinion that "the chance that among four brothers one or more would match is 1 in 66," is substantially different from Romero's estimate of a 1 in 6,500 chance that one brother would match. But even if Romero's estimate is wrong, our confidence in the jury verdict is not undermined. First, the estimate that is more pertinent to this case is 1 in 132—the probability of a match among two brothers—because two of Troy's four brothers lived in Utah. Second, although Jane Doe mentioned Trent as her assailant, and Travis lived in a nearby trailer, the evidence indicates that both (unlike Troy) were sober and went to bed early on the night of the crime. Even under Mueller's odds, a rational jury could consider the DNA evidence to be powerful evidence of guilt....

Resolution of the *Jackson* claim does not end our consideration of this case because respondent asks us to affirm on an alternative ground. He contends the two errors "in describing the statistical meaning" of the DNA evidence rendered his trial fundamentally unfair and denied him due process of law. Because the Ninth Circuit held that "the admission of Romero's unreliable and misleading testimony violated [respondent's] due process rights," and in respondent's view merely applied *Jackson* (erroneously) to determine whether that error was harmless, he asks us to affirm the judgment below on the basis of what he calls his "DNA due process" claim.

As respondent acknowledges, in order to prevail on this claim, he would have to show that the state court's adjudication of the claim was "contrary to, or involved an unreasonable application of, clearly established Federal law." 28 U.S.C. § 2254(d)(1). The clearly established law he points us to is *Manson v. Brathwaite*, 432 U.S. 98, 114 (1977), in which we held that when the police have used a suggestive eyewitness identification procedure, "reliability is the linchpin in determining" whether an eyewitness identification may be admissible, with reliability determined according to factors set out in *Neil v. Biggers*, 409 U.S. 188 (1972). Respondent argues that the admission of the inaccurate DNA testimony violated *Brathwaite* because the testimony was "identification testimony," 432 U.S., at 114, was "unnecessarily suggestive," id., at 113, and was unreliable.

Respondent has forfeited this claim, which he makes for the very first time in his brief on the merits in this Court. Respondent did not present his new "DNA due process" claim in his federal habeas petition, but instead consistently argued that Romero's testimony should be excluded from the *Jackson* analysis simply because it was "unreliable" and that the due process violation occurred because the remaining evidence was insufficient to convict. In the Ninth Circuit, too, respondent presented only his *Jackson* claim, and it is, at the least, unclear whether respondent presented his newly minted due process claim in the state courts. Recognizing that his *Jackson* claim cannot prevail, respondent tries to rewrite his federal habeas petition. His attempt comes too late, however, and he cannot now start over.

[A]mple DNA and non-DNA evidence in the record adduced at trial supported the jury's guilty verdict under *Jackson*, and we reject respondent's last minute attempt to recast his claim under *Brathwaite*. The Court of Appeals did not consider, however, the ineffective-assistance claims on which the District Court also granted respondent habeas relief. Accordingly, the judgment of the Court of Appeals is reversed, and the case is remanded for further proceedings consistent with this opinion.

It is so ordered.

9. The Transposition Fallacy in Brown

9.1 The Fallacy Exposed

The Supreme Court described the testimony in *Brown* as "a classic example of erroneously equating source probability with random match probability." Let's spell out the error. The Sheriff's laboratory found a profile with an uncontested frequency of 1/3,000,000. Of course, only one man in the population was the source of the semen. On the hypothesis that this man was *not* Troy Brown and not closely related to him $(H_{\sim T\&\sim R})$, the probability of the evidence that Troy has matching DNA (+*E*) is $P(+E|H_{\sim T\&\sim R}) = 1/3,000,000 = 0.000033\%$. Yet, Romero agreed with the prosecutor's suggestion that "the likelihood that . . . it's not Troy Brown . . . is .000033." The prosecutor and the expert are transposing. They are interpreting the probability of the evidence given the hypothesis that the semen is not Troy's and not a close relative's, $P(+E|H_{\sim T\&\sim R})$, as the probability of the hypothesis that the semen is neither Troy's nor a close relative's given the evidence of a matching DNA evidence, $P(H_{\sim T\&\sim R}|+E)$.

As noted in section 4.3, P(A|B) does not generally equal P(B|A). Examples abound:

- The probability of a 6 on a toss of a die given that an even number of dots turns up is 1/3, but the probability of an even number given a 6 is 1.
- The probability of the elevated PSA level given prostate cancer was 0.70, but the probability of prostate cancer given the PSA level was only 0.41.
- The probability that a person selected at random from the U.S. population is a Supreme Court justice, given that this individual is a woman, is virtually zero; however, the probability that the individual is a woman given that she is a justice is (as of this writing) 3/8.
- The Zika virus (ZIKV) infects humans only through bites from a mosquito that is not found in Canada. If validation studies establish that a blood test for the virus gives positive results with probability 0.95 when administered to infected individuals, but the test is used on a Canadian who never has been outside the country and who has not had sexual intercourse with an infected person, then it would be foolish to regard 0.95 as the probability that the virus is in this individual's blood—even though that is what the test indicates: $P(+E|H_{ZIKV}) = 0.95 \neq P(H_{ZIKV}|+E)$.

As Principle 5 states, we cannot just assume that $P(+E|H_{T\&\sim R}) = P(H_{T\&\sim R}|+E)$.

9.2 The Fallacy Corrected (for the hypothesis of an unrelated perpetrator)

The Court of Appeals went further. Not only did the majority of the three-judge panel observe that transposition is a statistical fallacy, but it insisted that

Statistically, the probability of [a source] given a DNA match is based on a complicated formula known as Bayes's Theorem, . . . and the 1 in 3,000,000 probability . . . is but one of the factors in this formula. Significantly, another factor is the strength of the non-DNA

evidence. Here, Romero['s] error . . . is especially profound given the weakness of the remaining evidence against Troy.

But Bayes' rule is not a "complicated formula." We derived it in section 4.3 in a short footnote. It involves nothing more complex than one multiplication and one division. For the test for the Zika infection, the formula, *posterior odds* = $LR \times prior odds$, becomes

$$Odds(H_{ZIKAV}) = [P(+E|H_{ZIKV}) \div P(+E|H_{ZIKV})] \times Odds(H_{ZIKV}).$$

The likelihood ratio is the bracketed term in the middle. It states how many times more probable the positive test result (+*E*) is when the test is applied to a patient who is infected than when it is applied to a patient who is virus-free. Previously, we stated that the test detected the virus when it was present with probability 0.95. Suppose that the validation studies for the test also establish that it gives a positive reading when the patient is virus-free with a smaller probability, 0.0475. Dividing these quantities to find the likelihood ratio indicates the probative value of the positive test result. The likelihood ratio is $P(+E|H_{ZIKV}) \div P(+E|H_{\sim ZIKV}) = .95/.0475 = 20$, which means that the positive finding is more probative than was the positive PSA test whose likelihood ratio was only 7.

For a likelihood ratio of 20, Bayes' rule simply states that

$$Odds(H_{ZIKV}) = 20 \times Odds(H_{ZIKV})$$

If the patient with flu-like symptoms lived in Brazil, maybe the prior odds for a Zika infection could be something like 1 to 1,000. Let's hope it is lower than this (as it would be in Canada!), but 1:1000 will do for an illustration. The odds, taking the positive test result into account, are $20 \times 1:1000 = 20:1000$, or 1:50.

Now let's apply this reasoning to *Brown* to see if Judge Wardlaw's opinion about the "especially profound" effect of "the weakness of the remaining evidence against Troy" is tenable. As it applies in *Brown*, Bayes' rule means that

$$Odds(H_{T}) = [P(+E|H_{T}) \div P(+E|H_{T})] \times Odds(H_{T}).$$

$$\uparrow \qquad \uparrow \qquad \uparrow$$

$$posterior odds = likelihood ratio \times prior odds.$$

The prior odds are the odds in favor of the prosecution's hypothesis H_T that Troy is the source of the semen before considering the DNA match +*E*. The opinions acknowledge that the nongenetic evidence was less than compelling. There were discrepancies in witnesses' description of when Troy left the Peacock Lounge, the clothing worn by the man walking near the trailer, the fingerprint on the night lamp, and identifications made by the little girl. For a start, let's say that the nongenetic evidence gives rise to odds of about 1:1. These prior odds

correspond to a prior probability of one-half: Troy's being the source of the semen, based on what we know so far, is as likely to be true as it is to be false.

As with the test for ZIKV (and the examples in Section 4.3), the likelihood ratio measures of the strength of the positive test finding, here the five-locus VNTR match. The numerator is the probability of the match given the prosecution's hypothesis. If Troy is in fact the source, the DNA samples should match: $P(+E|H_T) = 1$. If Troy is not the source and an unrelated individual is, then the probability of a match is Romero's uncontested random match probability: $P(+E|H_{\sim T}) = 1/3,000,000$. Dividing 1 by 1/3,000,000 yields 3,000,000 for the likelihood ratio—far better than the ZIKV test. Principle 7 thus applies here—the strength, or weight, of evidence can be quantified by comparing the probability of the evidence given one party's hypothesis to the probability of the evidence given an alternative hypothesis.

The likelihood ratio in *Brown* swamps the prior odds. When we multiply it by the prior odds of 1:1, we obtain 3,000,000:1. Notwithstanding the weakness in the nongenetic evidence, the posterior odds that Troy is the source are overwhelming. Conversely, the odds that he is not the source are a minuscule 1:3,000,000.

This a particularly interesting figure. Odds of 1:3,000,000 correspond to a probability of 1/(3,000,000 + 1) = 0.000033—exactly the number that Ms. Romero wrote on the blackboard and that she agreed was "the likelihood that . . . it's not Troy Brown"! It appears that the transposition was not fallacious after all.

Let's think about this. The equality between the transposed conditional probabilities holds for the specific numbers we have plugged into Bayes' rule. Will it hold for other numbers? And, will the finding that the weakness in the nongenetic evidence had no real effect on the Bayesian result that the Ninth Circuit presented as the proper way to reason about the DNA finding change if different, but still plausible odds and likelihood ratios are used? These questions are related to Principle 2—we should check whether the results are highly sensitive to changes in the assumptions.

Discussion Question 9.1. Suppose that the prior odds of 1:1 are too favorable to the prosecution—that the other evidence in the case is not merely equivocal but strongly supports the hypothesis that someone unrelated to Troy attacked the girl. Let's say that the prior odds on Troy are only 1:1000. (a) What are posterior odds that Troy is the source given the reported VNTR match? (b) What is the posterior probability of this hypothesis? (c) Was the Ninth Circuit correct in stating that the weakness in the nongenetic evidence had an "especially profound effect" on the Bayesian result? (d) What does this tell us about the belated "DNA due process claim" that the Supreme Court did not reach in *Brown*?

TEST YOUR UNDERSTANDING

9.1. Suppose Troy had become a suspect as a result of a search of the state DNA database, which contained his DNA profile along with those of 3,000,000 other people. (a) Without considering any further evidence (or the lack of it) against these people, what are the prior odds that Troy is the source? (b) For a likelihood ratio of 3,000,000, what are the posterior odds? (c) What is the posterior probability? (d) Is the likelihood ratio 3,000,000 if the identification of Troy resulted from a database trawl?

9.3 The Fallacy Corrected (for the Hypothesis of a Brother)

The calculation of the likelihood ratio in the previous section rested in large part on the fact that $P(+E|H_{-T}) = 1/3,000,000$. But strictly speaking, the frequency of 1/3,000,000 that Romero testified to translates into the probability of the positive DNA test given that the semen originated from *neither* Troy *nor* a close relative. In symbols, 1/3,000,000 is really $P(+E|H_{-T\& -R})$. Because close relatives are more likely to have the same DNA profile than are pairs of unrelated individuals, the probability of the evidence—the match to Troy—could be larger than the supposed "ceiling" of 1/3,000,000. For example, if Troy had had a twin brother in Carlin on the night of the assault, then a false match to Troy could have resulted *either* from an unrelated person *or* the twin having been the source. If the probabilities of these two source hypotheses (conditional on Troy's not being the source) were, say, 99.9% and 0.1% respectively, then, to the nearest decimal place, $P(+E|H_{-T})$ would be (.999 × 1/3,000,000) + (.001 × 1) = 0.001 = 0.1%.³⁶ The match to Troy would still be improbable if Troy were not the source, but not nearly so improbable as 1/3,000,000 = 0.00003%.

Troy had no identical twin brother, but he did have four brothers. If one of them was the rapist and if Troy and this brother happened to share the same five-locus VNTR profile, the DNA test would falsely implicate Troy. Although Troy did not want his lawyer to raise this possibility, defense counsel elicited testimony that there were five brothers in the family— Trent, Troy, and Travis (living in Carlin), and two younger ones (still living in Loa)—"all bear[ing] a similar resemblance." District Attorney Smith responded by having Renee Romero testify that

$$P(+E | H_{\sim T}) = P(H_{\sim R}) P(+E | H_{\sim R}) + P(H_{Twin}) P(+E | H_{Twin}).$$

³⁶ More generally, partitioning the population of possible perpetrators into three classes—Troy, all unrelated people, and the twin—the probability of the evidence under the hypothesis that Troy is not the source is the probability of the evidence under the two remaining hypotheses weighted by the probability of each of hypothesis:

Analogously, one can say that the probability of snow tomorrow is (a) the probability that it will snow given that it will be a sunny day times the probability that it will be sunny plus (b) the probability that it will snow given that it will be a cloudy day times the probability that it will be cloudy.

the probability of any pair of brothers having the same DNA profile would be "one in 6,500" and then subtract this number from 1 to arrive at 99.982% (in contrast to the 99.99967% for the source probability of an unrelated person). This arithmetic raises the same questions about the transposition fallacy presented in the previous section. Indeed, the brief of the National Association of Criminal Defense Lawyers as amicus curiae transposes when it characterizes 1/6500 as "the chance that a brother could have been the source of the DNA."³⁷ The correct statement is that 1/6500 is the calculated probability that *if* a specific brother were the source, *then* Troy's DNA also would match. The number is an evidence probability, not a source probability.

Regardless of whether the sibling match probability that Romero proffered was transposed into a source probability, Mueller argued that the reported figure of 1/6500 was not what the jury really needed to know. His letter maintained that "the chance of two brothers matching" would be at least 1/1024.³⁸ And even that, he opined, was too small, for "the more relevant question about siblings . . . is what is the chance that any one of the two (four) brothers would match the evidence profile?" His answer to this question was that for Troy's two brothers in Carlin (Travis, and Trent), "the chance that one or more brothers would match the evidence is 1 in 512. With four brothers this total goes up to 1 in 256."

These numbers might seem a little mysterious, but they are readily derived with the probability axioms. The situation is akin to tossing a die that has a probability p of 1/1024 of showing a one (a match to a brother) and 1 - p = 1 - 1/1024 of some face bearing a different number of dots (a nonmatch) on each independent toss. On the first toss, the probability of two nonmatches in a row is $(1 - p) (1 - p) = (1 - 1/1024) \times (1 - 1/1024)$, which is approximately 1 - 1/512. This is the probability of no matches to Troy for two brothers that flows from Romero's testimony). But P(at least one match) + P(no matches) = 1. Therefore, P(at least one) = 1 - P(none) = 1 - (1 - 1/512) = 1/512.

³⁷ Brief of the National Association of Criminal Defense Lawyers as Amicus Curiae in Support of Respondent, McDaniel v. Brown, No. 08-559, July 24, 2009, at 16.

³⁸ If the father has two distinct alleles A_1 and B_1 at a locus (is heterozygous), and the mother has two distinct alleles A_2 and B_2 at that locus, then there are four equally probable pairs of alleles that a son could have: A_1B_1 , A_2B_1 , A_1B_2 , and A_2B_2 . Suppose that the first son is A_1B_1 . Then the probability that a second son also is A_1B_1 is 1/4. The same result applies to every other possible pair. Hence, the probability that two specific sons have the same genotype at one locus is 1/4, as Romero testified. Of course, if a parent had two copies of the same allele at a locus, the probability of sharing a genotype would be higher, as Mueller pointed out. Because the two alleles that a child inherits at any locus is independent of the genotypes at every other locus, the probability that a pair of brothers inherit the same alleles at *n* loci from two parents who have no alleles in common is $(1/4)^n$. For five loci, this probability is 1/1024.

³⁹ The binomial approximation $(1 - p)^n \approx 1 - np$ works well here.

TEST YOUR UNDERSTANDING

9.2. Verify that the probability of a match to *at least one* of Troy's *four* brothers is 1/256 if the probability of a match for each brother is p = 1/1024. (a) What is the probability *P*(no matches) to all four brothers? Hint: For p = 1/256, $(1 - p)^4 \approx 1 - 4p$. (b) What is the probability at least one match? Hint: *P*(at least 1 match) + *P*(no matches) = 1.

The previous computations are based on a sibling match probability of 1/1024. By using a larger probability, 1/263, that accounts for scenarios in which Troy's parents have some of the same alleles, Mueller arrived at "[t]he chance that among two brothers one or more would match is 1 in 132 and the chance that among four brothers one or more would match is 1 in 66.

TEST YOUR UNDERSTANDING

9.3. Verify that 1/66 is the probability of a match to at least one of any of Troy's four brothers if the probability of a match for each brother is p = 1/263.

The Ninth Circuit was deeply impressed by the 1/66 figure. Judge Wardlaw observed that `"two of [Troy's brothers] lived in Carlin and two . . . lived in neighboring Utah," and that "the chance that Troy's DNA would match at least one of his four brothers' DNA [was] almost one hundred times the probability asserted by Romero." An amicus brief submitted by no less than "20 Scholars of Forensic Evidence" reiterated that "Ms. Romero's statements suggested it was up to 100 times less probable that there would be a fraternal match than in fact was the case" and insisted that "the probability that at least one of the [four] brothers will have the matching profile . . . is an important factor for jurors to consider when assessing the likelihood that someone other than respondent could be the source of the semen stain."⁴⁰

Yet, the Supreme Court settled on the smaller figure of "1 in 132" for "the probability of a match among two brothers" as "more pertinent . . . because two of Troy's four brothers lived in Utah." A law review comment on the Ninth Circuit's opinion had expressed this idea more fully:

The Ninth Circuit described the Tenth Circuit state of Utah as "neighboring" and the family ranch in Loa where the two younger brothers apparently lived with their parents as lying within "the general vicinity." Yet, the driving distance from Loa to Carlin is over

⁴⁰ Brief of 20 Scholars of Forensic Evidence as Amici Curiae Supporting Respondents, McDaniel v. Brown, No. 08-559, at 24.

440 miles. The relevance of 1 in 66 therefore rests on such speculations as a 13-year-old brother sneaking away from home to go hundreds of miles to another state.⁴¹

Moreover, the Supreme Court noted that "although Jane Doe mentioned Trent as her assailant, and Travis lived in a nearby trailer, the evidence indicates that both (unlike Troy) were sober and went to bed early on the night of the crime. Even under Mueller's odds, a rational jury could consider the DNA evidence to be powerful evidence of guilt."⁴²

From a Bayesian perspective, the argument about the improbability of any of the brothers being a source leans on nongenetic evidence—locale, age, sobriety, and so on—and refers to a posterior probability ascertained in the light of the DNA evidence *and* the additional evidence. As the Supreme Court concluded, even for the many probabilities floated in Mueller's letter, the verdict may have been rational given all the evidence, but which of the various probabilities is logically and legally most appropriate?

Discussion Question 9.2. What is the best statistic for indicating the probative value of the DNA profile match to Troy as opposed to his brothers? The single sibling match probability (1/263)? The probability of a match to one or two brothers (1/132)? Or the probability of a match to one, two, three or four brothers (1/66)?

In considering this last question, it may help to think about the situation for defendants with no close relatives. Imagine that Troy had been an only child and his parents were deceased. His DNA profile is the same as that of the semen. The datum is 3,000,000 times more probable under the prosecution hypothesis that Troy is the source than under the defense hypothesis that some unrelated person is. As with the diagnostic tests for prostate cancer and ZIKAV test, which had likelihood ratios of 7 and 20, respectively, this quantity expresses how strongly the test result supports the prosecution hypothesis compared to the defense hypothesis. Would it make sense to argue that the relevant probability to consider in assessing probative value is not the random match probability, but rather the probability that one or more unrelated people would have matching profiles? If the latter, disjunctive probability is not so relevant, then was Romero correct to present only the probability for a match to one randomly selected sibling? Consider these questions about the possibility of a matching brother: What is the prosecution hypothesis about brothers? The alternative hypothesis? The resulting likelihood ratio?

⁴¹ Id. at 30.

⁴² David H. Kaye, False, But Highly Persuasive: How Wrong Were the Probability Estimates in *McDaniel v. Brown*?, 108 Mich. L. Rev. First Impressions 1, 6 (2009).

Discussion Question 9.3. In *Brown* should the testimony have been that the probability of the match is (1) millions of times greater when testing an actual source than when testing unrelated people, and (2) about 263 times greater when testing an actual source than when testing brothers? Or is there yet another way that you think DNA evidence in a case like *Brown* should be presented?

10. The Seven Principles Revisited

At the outset, this module offered seven principles as sources of guidance in understanding DNA evidence. They apply in all forensic-science disciplines. As such, it is worth reviewing how they informed our understanding of the DNA evidence in *Brown*.

- Principle 1. Uncertainty arises in choosing mathematical models of phenomena. We saw this in the selection of the random-mating model for estimating the frequencies of combinations of DNA alleles at and across loci. It also arose in using the model of simple random sampling to estimate individual allele frequencies from small samples of individuals.
- Principle 2. Models rest on assumptions. Checking whether the results are highly sensitive to changes in the assumptions is advisable. Performing such checks suggested that the simple random-mating model was reasonable for estimating random-match probabilities. In addition, checking the Bayesian model of the impact of new evidence on a range of prior odds indicated that the court of appeals may have overstated the importance of the fact that the non-DNA evidence in the case was not compelling.
- Principle 3. Uncertainty arises when generalizing from samples to populations. One should ask: (a) How were samples chosen? Is the method likely to produce biased samples? (b) What is the plausible range of statistical error (variation arising from "the luck of the draw")? The state made no effort to account for sampling error in Brown. The letter that Brown submitted seeking federal habeas corpus relief did not raise this point, but sampling error probably was not a major source of uncertainty in the case.
- Principle 4. Uncertainty can be quantified with probabilities. We saw that both sides used probabilities (at various stages of the litigation). The district attorney argued that the probabilities were such that Brown's guilt was virtually certain. In securing a writ of habeas corpus, the federal public defender's office contended that the DNA evidence left open a substantial probability that an untested brother might have been the rapist.

- Principle 5. The probability of evidence indicating that a disputed event did (or did not) occur does not generally equal the probability of that event, as assessed in the light of the evidence. This principle cautions against naive transposition of an evidence probability into a source probability, as occurred in *Brown*. The prosecution characterized the probability of the DNA evidence given that the defendant was not the source of the semen as the probability that the defendant was the source given the DNA evidence. Although this transposition may not have departed greatly from the correct inversion according to Bayes' rule, it could be argued that it was psychologically prejudicial in that it may have invited the jury to overlook the other evidence in the case and the possibility that a brother was responsible.
- Principle 6. Bayes' rule offers one logical way to use probabilities about the evidence of an event to determine the probability of the event itself. We used this principle repeatedly to discuss the posterior probability that the defendant was the source in light of the DNA evidence and other evidence in the case, and to underscore the distinction between the probability of evidence given some hypothesis and the probability of that hypothesis given the evidence.
- Principle 7. The strength, or weight, of evidence can be guantified by comparing the probability of the evidence given one party's hypothesis to the probability of the evidence given an alternative hypothesis. This likelihood-ratio principle finds favor with many forensic scientists and statisticians but is not widely used in U.S. courts as a way to present forensic evidence. But whether it is used to grade evidence numerically for judges or juries or not, it has considerable value in understanding which probabilities are logically appropriate for presentation. Moreover, the probative value of all circumstantial evidence—even that which cannot be assigned precise probabilities—can be understood in terms of broad ranges for the likelihood ratio. For example, evidence that is much more probable under the prosecution's hypothesis than under the defense's is highly probative. Evidence that is equally probable under both side's theories is irrelevant—it does not support one's party's theory compared to the other party's view of the matter. Evidence that is slightly less probable under the prosecution hypothesis than it is under an alternative for the defense undercuts the prosecution's theory, at least a little. Considering the relative probability of the evidence under competing hypotheses can improve assessments of its relevance and weight.

Further Reading

Books

- Colin Aitken, Paul Roberts & Graham Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings* (2010).
- David J. Balding & Christopher D. Steele, *Weight-of-Evidence for Forensic DNA Profiles* (2nd ed. 2015).

John M. Butler, Advanced Topics in Forensic DNA Typing: Methodology (2011).
John M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation (2015)
David H. Kaye, David E. Bernstein & Jennifer L. Mnookin, The New Wigmore, A Treatise on Evidence: Expert Evidence (2d ed. 2011).
David H. Kaye, The Double Helix and the Law of Evidence (2010).

Forensic DNA Evidence Interpretation (John S. Buckleton et al. eds., 2d ed. 2016)

Roberto Puch-Solis, Paul Roberts, Susan Pope & Colin Aitken, Assessing the Probative Value of DNA Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses (2012).

Book Chapters and Articles

David H. Kaye & David A. Freedman, Reference Guide on Statistics, in *Reference Manual on Scientific Evidence* 211 (Federal Judicial Center, 3d ed. 2011).

- David H. Kaye & George Sensabaugh, Reference Guide on DNA Evidence, in *Reference Manual* on Scientific Evidence 129 (Federal Judicial Center, 3d ed. 2011).
- David H. Kaye, False, But Highly Persuasive: How Wrong Were the Probability Estimates in *McDaniel v. Brown*?, 108 Mich. L. Rev. First Impressions 1 (2009).

Richard O. Lempert, Modeling Relevance, 75 Mich. L. Rev. 1021 (1977).

Websites

David H. Kaye, Forensic Science, Statistics, and the Law, <u>http://for-sci-law.blogspot.com/</u>. NFSTC, President's DNA Initiative, <u>http://projects.nfstc.org/pdi/</u>. NIST, Short Tandem Repeat DNA Internet DataBase, <u>http://www.cstl.nist.gov/strbase/</u>.

Answers to Test Your Understanding Questions

2.1.

Part (a) asks about the fit between the model (simple random sampling) and the sampling method employed, which implicates Principles 1 (choosing models), 2 (assumptions in models), and 3(a) (sampling error). The sampling plan calls for systematic sampling at one location and time. A rigorous simple random sampling plan would start with the list of all students and use a chance mechanism (such as a random number generator) to pick the ones to contact. As a result, we must examine how closely the systematic sample can be expected to resemble a random one. The protocol does limit selection bias. For example, it would protect against a researcher's perhaps unconscious inclination to approach students of one gender. Nevertheless, being person 25, 50, 75, and so on, could be related to height. What if the big football players all had lunch together and came out of the building at the same time? Part (b) asks about the difference between the group samples and the population of interest, which invokes Principle 3(d) (right population). The population of students who

frequent the student union may or may not be a good proxy for the population of all registered students.

2.2.

This is a more technical question about the effect of sample size on the precision of an estimate. It relates to Principles 3(a) and 3(c) (sampling error). The best answer is (d). The width of a confidence interval is inversely proportional to the square root of the sample size. The second sample of 400 is four times larger than the first sample of 100, so the interval width should be half of that for the first sample. Instead of ± 8 we expect ± 4 . The center of the interval (the mean) for the larger sample was 68. So we should have a confidence interval of about 68 \pm 4.

2.3.

This is a technical question about the relationship between "confidence" and precision. Principles 3(a) and 3(c) are again involved. To achieve higher confidence is to have intervals for repeated samples that would cover the true value more often. That is done by enlarging the intervals. (It is easier to capture a butterfly with a big net than a small one.) Hence, the answer is (a), which has a larger interval.

2.4.

No. The confidence interval reflects only sampling error—the uncertainty in estimation arising from the luck of draw, that is, from random sample-to-sample variation. Some samples might have an unrepresentative group of tall students; others might be composed of an unusual proportion of short students. Errors in measuring how tall each student is adds to the total uncertainty. In some samples, an unusual proportion of the errors of measurement might suggest that the sample contains students who are taller (or shorter) than the students in the sample really are. The calculation of the confidence interval ignores this source of variability. Another way to say this is that the statistical model for simple random sampling assumes that there is no measurement error (Principle 2—Models rest on assumptions.)

2.5.

(a) There are two 10s among the 20 cards. Shuffling makes the probability of a 10 being the top card 2/20 = 1/10. (b) Replacing and reshuffling makes the second draw just like the first. Again, the probability is 1/10. (c) Shuffling ensures that the two events are independent, so the probability of the joint event (the two 10s) is the product (1/10)(1/10) = 1/100. (d) Yes.

2.6.

(a) As before, the probability of a 10 on first trial is 1/10. But if the card is not replaced there are only nine 10s in the 19 remaining cards, so the probability of a second 10 (given the first 10) is 9/19. Let's write this is symbols. $P(10_1) = 1/20$ and $P(10_2 \text{ given } 10_1) = 9/19$. We want to know $P(10_1 \& 10_2)$. The special multiplication rule for independent event does not apply because the card that came to the top on the first shuffle

influences which card comes to the top on the second shuffle. A more general multiplication rule for probabilities is P(A & B) = P(A) P(B|A), where the vertical line stands for "given." For example, the probability that tomorrow will be both a cloudy day and a rainy day is the probability that it will be cloudy times the probability that it will be rainy if it is cloudy. We conclude that $P(10_1 \& 10_2) = (1/20) (9/19) = 9/380$. (b) If 10s were removed on the first trials, the probability of a 10 on the third trial is zero. It is impossible to draw three 10s because there are only two in the abbreviated deck. We can write $P(10_1 \& 10_2 \& 10_3) = 0$. (c) No.

2.7.

(a) Dependent. A person with blue eyes is more likely to have blond hair than is a person with brown eyes. (b) Dependent. Taller people tend to be heavier than shorter ones. (c) Probably independent. It is doubtful that people with type O blood, for instance, perform better on IQ tests. (d) Independent. The sticker prices for ordinary green and ordinary blue cars are the same. (Some metallic colors might command a little higher price.)

4.1.

(a) P(suspect) = 1/6. (b) P(~suspect) = 1 - P(suspect) = 5/6.

4.2.

No. The sample space is $S = \{Man 1, Man 2, Man 3, Man 4, Man 5, Man 6, Other Man\}$. P(S) has to equal 1. Here, the sum of the "probabilities" of the mutually exclusive events that comprise S is only 0.95. Either the witness is not using numbers in the sense of mathematical probability, or he has made a mistake in his subjective estimates of the probabilities.

4.3.

Yes. They satisfy the three axioms.

4.4.

(a) Ignoring the possibility of an "inconclusive" test result, there are only two possible outcomes for a test of an Ebola-free patient, namely, +E and -E. Therefore, $P(-E|\sim H) + P(+E|\sim H) = 1$, and $P(+E|H) = 1 - P(+E^{-}|H) = 1 - 0.85 = 0.15$. (b) $LR = P(+E|H) / P(+E|\sim H) = 92/15 = 6.1$. (c) Odds(H) = 20 to 80 = 1 to 4 = 1/4. (d) Odds(H|+E) = LR Odds(H) = 6.1/4. (e) P(H|+E) = 6.1/(6.1 + 4) = 0.60 = 60%. (f) Odds(H) = 60 to 40 = 3/2. (g) Odds(H|+E) = LR Odds(H) = 6.1(3/2) = 18.3/2. (h) P(H|+E) = 18.3/(18.3+2) = 0.90 = 90%.

4.5.

(a) Because the mtDNA test definitively establishes whether *H* is true, P(+E|H) can be estimated as the proportion of cases with a positive mtDNA result for which the criminalists offered a positive judgment. There were 0 + 69 cases in which H was true, and the criminalists declared an association in all of them. So P(+E|H) = 1. Their sensitivity was very high. (b) On the other hand, of the 17 + 9 cases in which ~*H* was

true, the criminalists found an association in 9 of them. So $P(+E|^{H}) = 9/26$. (c) $LR = P(+E|H) / P(+E|^{H}) = 1 / (9/26) = 26/9 = 2.9$. (d) If we used a head on a coin toss to mean that the hairs had a common source, the likelihood ratio would be one. A head would have the same 0.5 probability when the hairs were from the suspect as when they were from a different individual. The criminalists are better than that but not as good as the Ebola in vitro diagnostic test: 1 (coin) ≤ 2.9 (criminalist) ≤ 6.1 (Ebola IVD test).

9.1.

(a) We are not supposed to consider any other evidence in forming the prior odds, so it might seem that every individual in the database and Troy are equally probable to be the source. That would make the prior odds on Troy 1 to 3,000,000. However, it is hard to believe that everyone in the database could have committed the crime. Some may have been in prison, incapacitated, or far from Carlin, at the time. Nevertheless let us bend over backwards and assume that the prior odds are 1/3,000,000. (b) The posterior odds then would be $LR \times (1/3,000,000) = 3,000,000/3,000,000 = 1$. (c) Posterior odds of 1 (to 1) correspond to a probability of ½. So the likelihood ratio of 3,000,000 does not swamp the prior odds. A posterior probability of only 50% is not nearly enough for a conviction. Now the weakness in the other evidence would be critical. (d) The answer is "no," but whether the LR is larger or smaller than in the one-suspect case is not obvious. The effect of a database trawl has sparked tremendous controversy in the statistics literature and in court. The dominant view is that, if anything, the database match is more probative than the match in the one-suspect case.⁸⁵ In the one-suspect case, the DNA evidence is simply that the suspect matches, but for all we know, an individual in the large database could have been the source. After the database trawl, we have excluded all of those 3,000,000 people as possible sources. The probability mass that had been placed on them has to go somewhere, and some of it flows to the matching suspect. Consequently, starting from the same prior odds in a one-suspect case as in a database-trawl case, we should have more, not less, confidence that the matching suspect is the true source for the database-trawl match. Inasmuch as the ratio of the posterior odds to the prior odds is larger, the likelihood ratio must have been larger.⁴³

9.2.

(a) $P(\text{no matches}) = (1-p)^4 \approx 1 - 4p = 1 - 4/1024 = 1 - 1/256$. (b) P(at least 1 match) = 1 - P(no matches) = 1/256.

9.3.

(a) $P(\text{no matches}) = (1-p)^4 \approx 1 - 4p = 1 - 4/263$. (b) $P(\text{at least 1 match}) = 1 - P(\text{no matches}) = 4/263 = 1/65.575 \approx 1/66$.

⁴³ A more precise statement would be that the Bayes factor is larger. In simple situations, the Bayes factor equals the likelihood ratio.

Examination and Extension Questions

Section 1

- 1. In your own words, explain the difference between a locus and an allele.
- 2. Although human beings normally have two pairs of homologous chromosomes, an extra chromosome number 18 occurs in about 1 out of every 2,500 pregnancies in the United States and disrupts the normal pattern of development in significant ways that can be life-threatening, even before birth. Trisomy 18, or Edwards syndrome, as the condition is called is thus present 1 in about 6,000 live births in the United States. In the case of an individual with this trisomy, how many distinct alleles might be detected at a locus on chromosome 18?
- 3. The presence of distinct alleles at a locus is known as heterozygosity; having the same allele on the homologous chromosomes is known as homozygosity. According to Table 1 of the Washoe County Sheriff's Laboratory report, the victim (Doe) is homozygous at how many loci?
- 4. In an action for child support, Mother (M) names Alleged Father (AF) as the father of her child (C). A paternity test reveals the following DQA types: M is (1,3), C is (3,3), and AF is (2,3). Is the test positive or negative for paternity?

Section 2.1

- 1. In *Garcia v. Tyson Foods, Inc.*, 890 F.Supp.2d 1273 (D. Kan. 2012), "Dr. Radwin testified that his study was conducted within a confidence interval of 95—that is 'if I did this study over and over again, 95 out of a hundred times I would expect to get an average between that interval." Id. at 1285. Is Dr. Radwin correct? If not, why not?
- 2. In *United States v. Nelson*, 419 F.Supp.2d 891 (E.D. La. 2006), the federal district judge wrote that "his 2004 [IQ] score of 67, with a confidence interval of 64-72, means that there is a 95% chance that Nelson's actual IQ is between 64 and 72." Id. at 896. Is this interpretation correct? If not, why not?
- 3. In *United States v. Gaines*, 979 F.Supp. 1429 (S.D. Fla. 1997), "Dr. Tracey concurred with the recommendation in the 1996 NRC Report that the jury also be advised of the 'confidence level,' or margin of error of this frequency. That is, although in any given population of African Americans, the most likely frequency for the genetic profile found in this case is 1 in 6.1 million, in some populations, it could be a frequent as 1 in 600,000, or as rare as 1 in 61 million." Id. at 1441. What problems are there with these statements?
- 4. Box 1 noted that women have two X chromosomes (and no Y chromosomes) and that a male inherits an X from his mother and a Y from his father. For most of the Y chromosome, the DNA is inherited as a single block. Thus, aside from infrequent mutations within the block, the "short-tandem repeat" (STR) loci on the Y chromosome are the same for all members of a paternal lineage—brothers, sons, fathers, grandfathers, and so on. In *Commonwealth v. Lally*, 46 N.E.3d 41 (2016), "the defendant argued that DNA results from the Y—STR testing were erroneously admitted without a

'confidence interval' A "confidence interval" adjusts . . . to account for sampling errors and identical profiles being passed through a paternal line" Id. at 50. Suppose that a population database of 100 Y-STRs does not include a Y-STR profile that matches the (male) Y-STR profile derived from a crime-scene and that the upper limit of a 95% CI for the profile's population frequency is 0.03, 3%. Does the adjustment from 0 to 3% account for "identical profiles being passed through a parental line"?

Sections 2.2–3

- Which of the following pairs of items are likely to be independent: (a) fingerprint minutiae and intelligence; (b) height and gender; (c) income and life expectancy; (d) shoe size and vocabulary of elementary school students?
- 2. More Y-STR loci in a profile (see previous problem) increase the power to discriminate among different paternal lineages, but the DNA sequences at all the loci are copied together and move from father to son as connected units. Are these alleles statistically independent of one another, so that the individual Y-STR allele frequencies can be multiplied to arrive at the multilocus Y-STR frequency?

Section 4

- 1. A test for deception has a true positive probability of P(+E|H) = 0.8 and a true negative probability of $P(-E|^{\sim}H) = 0.8$. Applied to a suspect who denies her involvement in a robbery, the test is positive for deception. What is the likelihood ratio for H?
- In an action for child support, Mother (M) names Alleged Father (AF) as the father of her child (C). A paternity test reveals the following DQα types: M is (1,3), C is (1,2), This implies that M transmitted the 1 to the child, and the biological father, whoever he may be, supplied the 2. AF is type (2,3).

(a) Does the test indicate that AF is a possible biological father (BF) of C?
(b) The DQA gene lies on chromosome 6. Half of AF's sperm will carry the chromosome with the 2, and the other half will have a 3. Let the hypothesis that AF is the BF be H. What is the value of P(C receives 2 | H)?

(c) If AF is not the BF, and if M does not mate in some manner associated with a man's DQA type, what is $P(C \text{ receives } 2 \mid {}^{\sim}H)$? Assume that the proportion of type 2 alleles in the relevant population is 20%.

(d) What is the likelihood ratio for H?

(e) If the prior odds for *H* are 1:1, what are the posterior odds of *H*?

(f) The posterior probability?

Section 9

1. In *People v. Axell*, 1 Cal.Rptr.2d 411, 415 (Ct. App. 1991), the California Court of Appeal wrote that "Subsequently, Cellmark reported that the frequency of that DNA banding pattern in the Hispanic population is approximately 1 in 6 billion. Appellant is part Hispanic. Simply put, Cellmark's analysis meant that the chance that anyone else but

appellant left the unknown hairs at the scene of the crime is 6 billion to 1." Has the court transposed a conditional probability? If this were the testimony at trial, would it have been a violation of due process?

- 2. In People v. Wilson, 136 P.3d 864 (Cal. 2006), the California Supreme Court wrote that (a) "Experts calculate the odds . . . that a random person from the relevant population would have a similar match."; (b) "the prosecution expert calculated the odds of an Hispanic and a Caucasian being the donor of a certain semen sample," Are the two statements of the nature of the statistical testimony equivalent? If not, which one is correct, and what is the relationship, if any, between the two?
- 3. In People v. Pike, No. 1–12–2626, 2016 IL App (1st) 122626, 2016 WL 359117 (III. App. Ct. Jan. 27, 2016), the Illinois Appellate Court wrote that "We hold that the admission of expert testimony of the 50% inclusion probability statistic in this case was error because the statistic was irrelevant. The probability of inclusion of 50% of the population as a potential contributor to the mixed DNA profile on the gun did not tend to make defendant's identification more or less probable, and so as a whole the DNA expert testimony in this case was irrelevant." Discuss this conclusion using the likelihood ratio as a measure of the probative value of the evidence that 50% of the population has a DNA profile consistent with the alleles detected in the DNA on the gun.