Washington and Lee Law Review

Volume 47 | Issue 4 Article 3

Fall 9-1-1990

Primer For The Nonmathematically Inclined On Mathematical Evidence In Criminal Cases: People V. Collins And Beyond

David McCord

Recommended Citation

David McCord, *Primer For The Nonmathematically Inclined On Mathematical Evidence In Criminal Cases: People V. Collins And Beyond*, 47 Wash. & Lee L. Rev. 741 (1990).

Available at: https://scholarlycommons.law.wlu.edu/wlulr/vol47/iss4/3

A PRIMER FOR THE NONMATHEMATICALLY INCLINED ON MATHEMATICAL EVIDENCE IN CRIMINAL CASES: *PEOPLE V. COLLINS* AND BEYOND

DAVID MCCORD*

Mathematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him.[1]

INTRODUCTION

An old hymn tells us, "Time, like an ever-rolling stream, bears all its sons away."[2] This religious verity states an indisputable truth concerning legal fact-finding processes: every event about which people litigate has been "borne away" by time, so that what happened can never be exactly reproduced, but only imperfectly reconstructed through the residue that the event left in the world—witnesses' memories, physical evidence, documents, and the like. Thus, a factfinding process can never lead to an absolutely certain conclusion. As Justice Harlan succinctly put it, "[I]n a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what *probably* happened."[3] The legal system straightforwardly recognizes this inherent uncertainty through the system's formulation of the burdens of persuasion: the burden in a criminal case is not "beyond any doubt," but "beyond a reasonable doubt;"[4]

---

* Professor of Law, Drake Law School; B.A. 1975, Illinois Wesleyan University; J.D. 1978, Harvard Law School. This research was supported by a stipend from the Drake Law School Endowment Trust, for which the author is deeply grateful. The author wishes to thank his Drake Law School colleagues, Professors James Adams and Stanley Ingber, Stuart Klugman, Professor of Actuarial Science at Drake, and Professor David Kaye of Arizona State University College of Law for their insightful suggestions for improvements. The author also wishes to thank his research assistants, Stephanie Sarcone and Michael Gabor, and his secretary, Karla Westberg, for their diligent assistance. Finally, thanks go to Professors Leonard Jaffee and Richard Wright for their valuable insights offered by telephone. Any errors, of course, are solely the responsibility of the author.

1. People v. Collins, 68 Cal. 2d 319, 319, 438 P.2d 33, 33, 66 Cal. Rptr. 497, 497 (1968).

2. "O God Our Help in Ages Past," The Methodist Hymnal, No. 28, stanza 5 (1966).

3. In re Winship, 397 U.S. 358, 370 (1970) (Harlan, J., concurring).

4. Prior to the seventeenth century in England, juries were often instructed to find guilt in a criminal case only if there was *no* doubt. But with the advent in the 1600s and 1700s of legal rules which excluded evidence that earlier would have been admissible, the system recognized that absolute certainty had become impossible (indeed, it had never been possible in the first place). By the late 1700s, the "beyond a reasonable doubt" standard had displaced the "no doubt" standard. Morano, *A Reexamination of the Development of the Reasonable Doubt Rule*, 55 B.U.L. REV. 507 (1975).

the burden in civil cases is not "to a certainty from the evidence," but "by a preponderance of the evidence." Accordingly, the legal factfinding process is quite aptly described as "a social invention for deciding between disputed alternatives under conditions of uncertainty."[5]

But the legal system is not alone in its interest in "alternatives under conditions of uncertainty": mathematics, through the subdiscipline of probability, has developed a coherent and intellectually powerful system for addressing this very subject. It was inevitable that the legal system would confront the question whether mathematical conceptions correctly and helpfully describe the kind of uncertainty that exists in legal factfinding processes. The legal system began to confront that question in earnest about twenty years ago, continues to address it on a daily basis today, and will undoubtedly be called upon to deal with it in the foreseeable future. The question is persistent because statistical data from which probabilities can be generated is available on an ever-increasing number of issues in our technological society. And the question is a monumental one for the law because if the law concedes that mathematical concepts correctly describe legal uncertainty, then two conclusions seem to follow: first, that the burdens of persuasion are amenable to mathematical expression; and second, that mathematical evidence is an acceptable and perhaps even preferred mode of proof.

Not surprisingly, this monumental question has generated a large body of case law and legal scholarship. Unfortunately, the legal scholarship has consisted mainly of articles written by mathematical sophisticates, and much of the work is impenetrable[6] to many law students, evidence professors, lawyers, and judges, who are by nature and training nonmathematically inclined. Yet the issue is so fundamental and timely that the nonmathematically inclined person involved or interested in the legal system should have some accessible entree into the debate. Thus, the goal of this article is to explain in terms understandable to the nonmathematically inclined those issues raised by the intersection of legal proof and mathematics, the arguments on both sides of each issue from a theoretical perspective, how those issues and arguments have been dealt with in the case law, and what valuable insights social scientists have to offer. Although the article does not advocate a particular resolution of most of the major theoretical issues (the field is already rife with competing theories), the author hopes that a thorough yet understandable explanation of the area will be helpful to those "in the trenches."[7] The article is limited to a discussion of the issues as they arise

---

5. Saks & Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 L. & SOC'Y REV. 123, 123 (1980-81).

6. At least they are "impenetrable" given the amount of time and energy that most lawyers and judges have to devote to the task. That they are not *totally* impenetrable to the non-mathematically inclined is evidenced by this Article, whose author falls squarely into that category.

7. This is not the first such attempt, but should still be helpful. The earliest piece along these lines, Broun & Kelly, *Playing the Percentages and the Law of Evidence*, 1970 U. ILL. L. REV. 23, was published before most of the pertinent case opinions and scholarship came into

in criminal cases, where the stakes are the highest. Much of the discussion, however, is equally applicable to civil cases.

The article attempts to accomplish these goals in four parts. Part I provides some basic building blocks necessary to initially understand the topic. In Part II, the article re-examines the landmark mathematical evidence case, *People v. Collins*,[8] which must be understood completely because it marks the real beginning of the intersection between the law and mathematics, and because it continues to be considered by evidence texts, courts, and scholars as the preeminent case on the subject. Next, in Part III, the article gives an overview of the development of scholarship and case law since *Collins*. The article culminates in Part IV by analyzing the five kinds of mathematical evidence that arise in criminal cases and applying the scholarship, case law, and social science research which address their appropriateness as proof.

I. The Basic Building Blocks

A. The Mathematical Toolkit

Happily for the nonmathematically inclined, the basic toolkit necessary for understanding mathematical issues in the proof process consists of only six items: the relationships among "data," "statistics," and "probability;" three theories of probability—classical, frequency and subjective; and two mathematical rules, the product rule for independent events and Bayes' Theorem. More happily, the relationships among "data," "statistics," and "probability," the three theories of probability, and the product rule for independent events are quite commonsensical and easily understandable. Even Bayes' Theorem, while embodied in a mathematically imposing equation, is simple to understand in principle.

Let us begin by examining the relationships among "data," "statistics," and "probability." "Data" are simply observations about the world.[9] "Statistics" are numerically presented summaries of data conveying information regarding the chosen subject.[10] A "probability" is a calculation of the mathematical likelihood of some statement about the world being true, where

---

existence. The next such effort, Braun, *Quantitative Analysis and the Law: Probability Theory as a Tool of Evidence in Criminal Trials*, 1982 Utah L. Rev. 41, is helpful in explaining theories of probability and their applications, but is not comprehensive in its coverage of case law and scholarship. The succeeding article, Jaffee, *Of Probativity and Probability: Statistics, Scientific Evidence, and the Calculus of Chance at Trial*, 46 U. Pitt. L. Rev. 925 (1985), includes a great deal of thought-provoking analysis, but is a book-length, mathematically complex treatment. Finally, Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials—Part II*, 27 Jurimetrics J. 160 (1987), while providing numerous incisive concepts, again is not comprehensive in its coverage of case law and scholarship. None of these pieces incorporates social science research.

8. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

9. Webster's New World Dictionary 374 (College ed. 1968).

10. *Id.* at 1425.

it is unknown whether that statement is true.[11] The relationships among these three concepts have four important aspects. First, neither statistics nor probabilities can be formed without data. Second, probabilities are often based on statistics, *e.g.*, one can form a probability based on the statistic that a certain die has rolled a "6" twenty times in 100 throws. Third, probabilities can be based on data that is not numerical (and thus not statistical) in nature: *e.g.*, one may believe that, based on events, there is a forty percent probability that the democracy movement will prevail in China by the year 2000. Hereinafter in this article, probabilities based on statistical data will be referred to as "statistically-based probabilities," while those not based on statistical data will be referred to as "non-statistically-based probabilities." Fourth, while a statistic may be used to form a probability, it need not be so used.[12] A simple example will illustrate this point. Suppose that a population of 100 persons is given blood tests, and it is determined that only two of them have factor X in his or her blood. The statement, "One-fiftieth of this population has factor X in his or her blood," is a statistic. It has validity, and may be helpful in a criminal case, without ever being made part of a probability calculation. For example, suppose that a crime has been committed, the previously mentioned 100 people are the only possible suspects, and blood left by the perpetrator at the scene is found to contain factor X. As a matter of statistics, we can say, "Ninety-eight percent of the suspect population could not have left such a blood trace." This is not a probability because it does not concern the *likelihood* of any given person's having committed the crime. It speaks in terms of *certitude*, not likelihood: "Only two percent of the population could have left such a trace; the other ninety-eight percent could not have."

There is a more subtle relationship between statistics and probability that is disguised by the somewhat unreal example set forth in the preceding paragraph, where the entire universe at issue was known and tested. In most real-world situations, the entire universe is unknown—instead, only some *portion* ("sample") of the universe is known and tested. The question then becomes: To what extent can the statistical data derived from the sample be extrapolated as fairly representative of the occurrence of the sampled characteristic in some larger segment of the universe (the "population parameter")? Obviously, it must be acknowledged that there is some possibility that the statistics derived from the sample do not accurately reflect the occurrence of the sampled characteristic in the population parameter because the possibility always exists that the sample is not representative of the population parameter. Statisticians have developed sophisticated techniques to estimate the likelihood that the statistics derived from the sample do not accurately reflect the occurrence of the sampled characteristic throughout the population parameter. One common sense principle is fundamental, though: the larger

11. *Id.* at 1160.

12. Special thanks go to Professors Kaye and Klugman for illuminating these distinctions for the author.

the sample, the less likelihood of error. For example, in a population of one million suspects the statistic derived from testing 100 individuals for the presence of blood factor X obviously has a greater probability of error when extrapolated over the entire population than a statistic derived from testing 10,000 individuals. In either case, the statement, "Two percent of the suspect population has blood factor X," is implicitly based on a probability calculation that the rate of occurrence in the sampled segment continues to occur proportionately in the unsampled segment. Any statistical statement based upon less than full testing, then, *does* make a probabilistic statement about the world: there is a probability that this proportion continues throughout the universe at issue. But, importantly, in the criminal law context this implicit statistical probability that the statistic is accurate does *not* make a statement regarding a probability related to guilt of any suspect. The statistic simply has a simultaneous inclusive and eliminative effect on whether the defendant is within the *class* of possible culprits, *e.g.*, "The defendant has blood factor X, which we believe only two percent of the population has, based upon sampling of the population with the results of the sample extrapolated to continue over the whole population." Once we know there are 100 people in the entire population, and the additional non-statistical information that factor X blood was left by the perpetrator, and that two percent of the population has factor X blood, we can calculate probabilities relating to random selections of blood factor X carriers (one in fifty of selecting a person with factor X blood; one in two of selecting the culprit from between the two carriers of factor X blood). The distinction between statistics and probability is absolutely fundamental in understanding both the scholarly literature and the case law.

The next necessary explanation concerns the three theories of probability.[13] The classical theory was the first to be developed. The theory grew out of a study of gambling and game problems that began in the mid-seventeenth century.[14] This theory calculates probability based on the number of possible outcomes. For example, the probability of rolling a "6" with a fair die under this theory is one-sixth since there are six possible outcomes. The classical theory is quite limited in application because its results are rationally acceptable only when one is willing to assume that each of the possible outcomes is equally likely. In non-game-like real-world situations,

13. For a discussion of the classical, frequency, and subjective theories, see Braun, *supra* note 7, at 45-55; Cullison, *Probability Analysis of Judicial Fact-finding: A Preliminary Outline of the Subjective Approach*, 1 U. TOL. L. REV. 538, 545-59 (1969). More elaborate taxonomies of probability theories are available, but are unnecessary and unhelpful for purposes of this Article because the classical, frequency, and subjective theories are the ones that arise in practice. *See* J. WIGMORE, 1A WIGMORE ON EVIDENCE 37.6 n.5 (P. Tillers ed. 1983) (noting that "any division of probability theory into types is not without its difficulties" and going on to discuss several proposed taxonomies). *See also* Kaye, *What Is Bayesianism? A Guide For The Perplexed*, 28 JURIMETRICS J. 161, 164-67 (1988) (identifying seven types of probability).

14. *See* Cullison, *supra* note 13, at 545-46. This theory is also known as the "*a priori*" and is usually identified with LaPlace. *Id.* at 546.

where outcomes are very rarely equally likely, the classical theory does not lead to rationally acceptable results. For example, a purely classical approach to the question, "What is the probability that the high temperature will be over eighty degrees in Des Moines, Iowa, next January 1?," would be fifty percent since there are only two possible outcomes—eighty degrees or less and over eighty degrees. Of course, one could get more elaborate. Assuming the hottest it ever gets in Des Moines is 110 degrees, and the coldest is thirty degrees below zero, there are 141 possible temperatures, and the classical theory would calculate the probability of the temperature exceeding eighty degrees to be thirty out of 140 (there are thirty possibilities over eighty degrees, and 111 below it). Either way, real-world experience tells us that the probability calculation is way off—there is practically no chance of such a temperature in Des Moines on that date.

The limited number of possible applications of the classical theory led to the development of the frequency theory of probability in the last half of the nineteenth century.[15] The frequency model bases its predictions on the relative frequency with which particular outcomes are expected to result from repeated trials.[16] The statistical information needed to calculate this relative frequency can be obtained by counting within a known sample (*e.g.*, sixty of the 100 marbles in this bag are black and forty are red) or by repeated tests of an unknown sample (*e.g.*, out of 100 draws from this bag which contains an unknown proportion of black and red marbles, I drew black sixty times). Either derivation of the statistics leads to the same frequency probability: over a long number of draws the probability is that black will be drawn three-fifths of the time. Note, however, that in the real world of criminal cases the sample is never specifically known, and thus the second derivation of the statistic—sampling of the population—is the only way a valid statistic can be generated. The obvious advantage of the frequency theory over the classical theory is that it can be applied to real-world situations where the possible outcomes are not equally likely. Of course, when the possible outcomes *are* equally likely, the classical and frequency theories result in the same probability: the probability under either theory of a fair die rolling a "6" is one-sixth.

The third probability theory—the subjective model—had its origins in the work of James Bernoulli in the early eighteenth century,[17] and was popularized by an influential work published by Professor Leonard Savage in 1950.[18] Savage demonstrated that given a few simple and relatively noncontroversial assumptions, a probability subjectively developed by an individual can still meaningfully be called a "probability." Such a probability can, if so desired, be used as a meaningful input into Bayes' Theorem.

15. Cullison, *supra* note 13, at 548.

16. Brilmayer & Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116, 138 (1978).

17. *See* J. BERNOULLI, ARS CONJECTANDI (1713).

18. L. SAVAGE, FOUNDATIONS OF STATISTICS (1950). The subjective theory is also known as the "personalistic" theory.

Subjective probability theory is obviously the only one of the three theories of probability that can accommodate purely non-statistically-based probabilities, such as the probability that the democracy movement in China will triumph before the year 2000.

Let us now turn to the two mathematical rules necessary to complete the toolkit. The product rule for independent events (hereinafter "the product rule") is used in calculating the probability of two or more independent events occurring together, and simply states that the probability of the occurrence of both is the product of the probabilities of each, *i.e.*, the individual probabilities are simply multiplied together. It can also be used to figure a statistic regarding the coincidence of two or more independent characteristics. Events or characteristics are "independent" when the incidence of occurrence of one is neither increased nor decreased by the occurrence of any of the others.[19] For example, suppose that the question is "What is the probability that a randomly chosen person from the adult population of the United States will be a lefthanded woman?" Suppose that frequency data shows that one-half of the adult population are women, and one-tenth of adults are lefthanded. Since handedness is not a function of gender, *i.e.*, a woman is no more likely to be lefthanded than a man, the two characteristics would be independent. Thus, the product rule for independent events would tell us that there is a one-twentieth chance of randomly selecting a lefthanded woman. But suppose the question is, "What is the probability that a randomly chosen person from the adult population of the United States will be a woman lawyer?" Assume that the frequency data shows that one-half the adult population are women, and one-two-hundredth are lawyers. We might initially be tempted to use the product rule for independent events to calculate a probability of one-four-hundredth. But this would be wrong because we know that the two characteristics are not independent: if a chosen person is a woman, she is much less likely to be a lawyer than a person who is a man. Accordingly, another equation must be used to account for this situation. There is a relatively simple formula for calculating the probability of all of two or more mutually dependent events occurring. This formula multiplies the probability of the first event's occurring times the probability of the second event's occurring, where it is known that the first event occurs.[20] But this formula is rarely, if ever, used in cases or discussed by commentators. Instead, a mathematically complex formula known as Bayes' Theorem, which provides more information than the product rule for mutually dependent events, sometimes shows up in case law, and has often been the focus of commentators.[21]

---

19. *See* Cullison, *supra* note 13, at 541-42.

20. In mathematical notation the formulation is written $P(A + B) = P(A) \times P(B/A)$, where $P(A + B)$ represents the probability of the occurrence of both the first event (A) and the second event (B), $P(A)$ represents the probability of the first event (A), and $P(B/A)$ represents the frequency with which the second event (B) occurs out of all the cases in which the first event (A) occurs. Cullison, *supra* note 13, at 541-42.

21. For a derivation of Bayes' Theorem from the product rule for mutually dependent

Bayes' Theorem[22] is legally significant because it offers a means by which a probability estimate concerning a matter at issue can be revised in light of a new piece of probabilistic evidence,[23] *i.e.*, how a "prior probability" can be revised in light of new data to reach a "posterior probability." Suppose that a person has been selected at random from the adult population of the United States, and the question is what is the probability that this person is a lawyer. Suppose our "prior probability," based upon the proportion of lawyers in the population, is one-two-hundredth. Suppose we now learn that the person selected is a woman, and evidence is introduced that only one-fifteenth of all lawyers are women. How do we revise our "prior probability" of one-two-hundredth in light of this new evidence? Bayes' Theorem tells us: the "posterior" probability is 1/1500.[24] This illustration uses a frequency theory-derived figure for the prior probability, a frequency theory-derived figure for the new evidence, and results in a frequency theory posterior probability. But the Theorem works equally well using a subjective theory-derived figure for one of the inputs. For example, suppose that a defendant is charged with murder and, based upon the evidence thus far presented, the jury believes that there is a fifty percent likelihood that he is guilty. This is a subjective probability figure. Suppose further that the prosecution introduces new evidence of a palm print on a knife left at the murder scene, accompanied by expert testimony that the defendant's palm print is consistent

---

events, see Cullison, *supra* note 13, at 544-45. The "more" provided by Bayes' Theorem is that it permits the calculation of P(A/B), not merely P(A + B).

22. There are various ways to write Bayes' Theorem. A simple formulation is as follows:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

Where P(A/B) represents the frequency with which A occurs out of all cases in which B occurs (the "posterior probability"); P(B/A) represents the frequency with which B occurs out of all cases in which A occurs; P(B) represents the probability of B occurring; and P(A) represents the probability of A occurring (the "prior probability").

A more complicated and more often used version is as follows:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(A)P(B/A) + P(\text{not } A)\, P(B/\text{not } A)}$$

The only new terms here are P(not-A), which represents the probability of A's non-occurrence; and P(B/not-A), which represents the frequency with which B occurs out of all the cases in which A does not occur.

23. Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489, 502 (1970).

24. Using the simple formulation of Bayes' Theorem from note 22, the calculation is as follows: P(L/W) stands for the probability that the person is a lawyer given that she is a woman; P(W/L) stands for the probability that the person is a woman given that she is a lawyer; P(L) stands for the probability that the person is a lawyer; and P(W) stands for the probability that the person is a woman. Solving the equation to determine $P(L/W) = 1/15 \times 1/200 \div 1/2 = 1/1500$.

It should be noted that while nobody doubts the validity of Bayes' Theorem from a mathematical standpoint, statisticians are as divided as legal scholars on the question whether Bayes' Theorem is an appropriate device with which to manipulate real-world statistical data.

with the palm print on the knife, and that only about one of every 1000 people in the population could leave such a print. This is a frequency probability figure. By applying Bayes' Theorem, the jurors' fifty percent prior probability estimate of guilt should be adjusted because of the new probabilistic evidence up to a 99.9 percent posterior probability.[25] Here, the Theorem has combined a subjectively-derived prior probability with a frequency-derived new piece of evidence to arrive at a subjective posterior probability. The Theorem can accommodate any combination of subjective or frequency-derived inputs as the prior or new-evidence probabilities, but it is important to remember that if either the prior probability or the new evidence figures are subjectively-derived, then the posterior probability will be subjective.

With this simple mathematical toolkit the reader has the equipment necessary to understand the mathematics of virtually every reported criminal case involving mathematical evidence and every scholarly writing on the subject. The reader cannot understand the possible legal ramifications of the applications of these tools to the proof process, however, without clearly understanding the basic contentions between the two competing schools of thought.

B. The Mathematical Probabilist and Anti-Mathematical Probabilist Positions

Those who favor the use of probabilities in the proof process (hereinafter "mathematical probabilists" or "probabilists") and those who oppose the use of probabilities (hereinafter "anti-mathematical probabilists" or "anti-probabilists") have fundamental disagreements.[26] While both camps agree

25. This example is a famous one originated by Finkelstein & Fairley, *supra* note 23, at 496-501. If the reader is disconcerted by the concept of the jury forming a belief in the likelihood of guilt before all of the evidence has been presented, the reader is not alone. Indeed, this has been one of the most hotly debated topics concerning the application of Bayes' Theorem to criminal cases, and will be discussed *infra* in the text accompanying notes 285-88.

26. Not everyone involved in the debate, of course, falls strictly into one camp or the other. *See, e.g., infra* text accompanying notes 38-41 for a difference of opinion within the probabilist camp between the "utilitarian" probabilists and other probabilists. Some scholars see these schools of thought as not so much in conflict as simply speaking to different aspects of the proof process. *See, e.g.,* Allen, *A Reconceptualization of Civil Trials*, 66 B.U.L. REV. 401 (1986); Schum, *Probability and the Processes of Discovery, Proof, and Choice*, 66 B.U.L. REV. 825 (1986) (a particularly good introduction to the whole mathematical evidence debate); Tillers, *Mapping Inferential Domains*, 66 B.U.L. REV. 883 (1986). Perhaps not surprisingly the most avid probabilists are not lawyers, but rather persons like Professor Stephen Fienberg, Professor of Statistics and Social Science at Carnegie Mellon University, and Professor Mark Schervish, Professor of Social and Decision Sciences at Carnegie Mellon University. *See* Fienberg & Schervish, *The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decision-Making*, 66 B.U.L. REV. 771 (1986), probably the most pro-probabilist writing to be found in the literature. At the other end of the spectrum, seemingly the most vehement anti-probabilist is Professor Leonard Jaffee. *See* Jaffee, *Prior Probability—A Black Hole in the Mathematician's View of the Sufficiency and Weight of Evidence*, 9 CARDOZO L. REV. 967 (1988); Jaffee, *supra* note 7, at 925.

on the starting point—that absolute truth regarding a past event can never be known, and thus that the judicial system can only strive to decide, in the words of Justice Harlan, "what *probably* happened,"[27]—the camps split regarding what "probably" means in the context of the proof process. There are two main points of contention, with opposing positions on each contention having a momentous impact on what the camps view as the value of the elements of the mathematical toolkit to the proof process, or the meaning of the burden of persuasion.

1. The First Point of Contention—Is All Evidence Inherently Probabilistic?

The first point of contention is as follows. Probabilists, believing that the legal proof process can be described in the language of mathematical probability, believe that *all* evidence is in essence probabilistic. They argue that, because every item of evidence is inherently less than absolutely true, each item's evidentiary value can be expressed in terms of the subjective probability of its being true. This leads to the further conclusion that the line of reasoning from proof to conclusion is subjectively probabilistic as well.[28] Anti-probabilists respond that all evidence is *not* inherently probabilistic and that the reasoning from proof to conclusion in the trial process is not subjectively probabilistic.[29] Rather, each piece of evidence goes some distance, depending on its weight, toward establishing or refuting one party's explanation or some alternative to it:

> This sort of [probabilistic] analysis, however, misconstrues how juries reason when they consider "particularistic" evidence such as witness testimony. Their actual reasoning is quite different from the sort of statistical reasoning espoused by the probabilists, and thus undermines the claim that all evidence is ultimately statistical.
>
> There are numerous forms of inductive inference in addition to statistical syllogisms. The type that most accurately characterizes how the jury reasons when it considers "particularistic" evidence is inference to the best explanation. . . . With inference to the best explanation, the jury considers alternative hypotheses that would explain the available evidence, and chooses the most plausible explanation.[30]

---

27. *See supra* text accompanying note 3.

28. *See, e.g.*, Ball, *The Moment of Truth: Probability Theory and Standards of Proof*, 14 VAND. L. REV. 807, 812-13 (1961); Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1023 (1977); Saks & Kidd, *supra* note 5, at 151-53; Williams, *A Short Rejoinder*, 1980 CRIM. L. REV. 103, 103 [hereinafter Williams, *Rejoinder*]; Williams, *The Mathematics of Proof-I*, 1979 CRIM. L. REV. 297, 299-300.

29. *See, e.g.*, Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 IND. L.J. 1, 16-17 (1982); Jaffee, *supra* note 7, at 1017.

30. Note, *Gambling on the Truth: The Use of Purely Statistical Evidence as a Basis for Civil Liability*, 22 COLUM. J. L. & SOC. PROBS. 31, 53 (1988) [hereinafter Note, *Gambling on the Truth*].

This aspect of the debate between the camps is persistent and pervasive.

The ramifications of each camp's positions on this point to their views of the value of the items in the mathematical toolkit is straightforward. Regarding *data, statistics, and probabilities*, mathematical probabilists, believing as they do that all evidence is ultimately probabilistic, are firm believers not only in the use of empirically sampled data, empirical statistics, and empirical probabilities, but also in the use of non-empirically sampled data to form non-empirical probabilities. As to the *theories of probability*, probabilists recognize that neither the classical nor frequency theories alone serve their purpose of assigning a probability to the defendant's guilt. The classical theory alone is unsuitable because in the real world the assumption of equal probability of results is almost always unjustified. The frequency theory alone is inappropriate because it speaks only to expected results over repeated trials. Only subjective probability can be used to legitimately form a probability with respect to a unique past event such as whether the defendant committed the crime.[31] Subjective probabilities are appropriate, argue the mathematical probabilists, because they are the mathematical way of stating beliefs. Regarding the two *rules* in the mathematical toolkit, probabilists recognize that the product rule is helpful in manipulating statistics and in forming classical or frequency probabilities where the statistics are mutually independent. The product rule is recognized by probabilists to be inappropriate, however, in the formation of most subjective probabilities where the inputs, which are usually quite varied (*e.g.*, the probability regarding the truth of eyewitness testimony, the probability of truth of proof of motive, and the probability of truth of blood factor analysis) cannot be shown to be mutually independent. Mathematical probabilists instead are enamored of Bayes' Theorem for use with their subjective probabilities because it can be used to meld different varieties of evidence without requiring that they be mutually independent.[32]

As to the anti-probabilist position, regarding *data, statistics, and probabilities*, the one point on which they agree with the probabilists is with respect to the use of empirically sampled data and empirical statistics. As long as such data and statistics are used only to eliminate possible suspects and are not used to form probabilities, the logic of the anti-probabilist position supports the use of such empirical results because they help to

---

31. M. FINKELSTEIN, QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS 62-65 (1978); Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34, 41-47 (1979); Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1348 (1971); Note, *Gambling on the Truth*, *supra* note 30, at 38.

32. *See, e.g.*, Kaye, *Paradoxes, Gedanken Experiments and the Burden of Proof: A Response To Dr. Cohen's Reply*, 1981 ARIZ. ST. L.J. 635, 638-39; Lempert, *supra* note 28, at 1023. One commentator who attended the Boston University Symposium on probabilities in the proof process concluded that "Bayesian Theory is indeed an orthodoxy [among evidence law scholars]". Edwards, *Summing Up: The Society of Bayesian Trial Lawyers*, 66 B.U.L. REV. 937 (1986).

negate alternative explanations (*i.e.*, the guilt of someone other than the defendant) inconsistent with the guilt of the defendant.[33] Anti-probabilists denounce the use of non-empirically sampled data to form non-empirical probabilities, however, because anti-probabilists deny that all evidence is inherently probabilistic, and that the reasoning process from proof to conclusion is subjectively probabilistic.[34] With respect to the *theories of probability*, anti-probabilists reject the use of probabilities based upon the classical and frequency theories for the same reasons that probabilists recognize that they cannot rely solely on such probabilities—the classical theory is usually inappropriate in real-world situations where outcomes are not equally likely, and the frequency theory speaks only to expected results over repeated trials.[35] Anti-probabilists deny that subjective probability is appropriate in the proof process because the only belief it leads to is a belief in what *might have* happened, a legally insignificant matter, since what the legal system calls for is a belief in what *actually* happened.[36] With respect to the two *rules* in the toolkit, anti-probabilists do not deny that the product rule can properly be used to calculate a combined statistical occurrence where the underlying statistics are mutually independent. Anti-probabilists do, however, vigorously oppose the use of Bayes' Theorem for several reasons, the most fundamental of which is that multiplication is inappropriate as applied to the proof process where mutually independent pieces of evidence and inferences are involved. To anti-probabilists each piece of evidence and each inference is ultimately an all-or-nothing proposition. Either the jury believes the evidence or inference (because it is the most plausible alternative) or does not believe it (because it is not the most plausible alternative). If the jury believes the evidence or inference, it can then proceed to examine the truth of other items dependent upon it, whereas if the jury does not believe the evidence or inference, the reasoning chain stops.[37]

2. The Second Point of Contention—Do Mathematical Probabilist Concepts Describe the Optimal Operation of the Trial Process as It Operates Through the Burden of Persuasion?

Some probabilists limit their justification of the value of mathematical probability concepts in the proof process to that set forth with respect to the first point of contention outlined above: mathematical probability concepts correctly and helpfully describe the nature and weight of evidence, and

33. Telephone Conversation with Professor Richard Wright, Associate Professor of Law, ITT Chicago-Kent School of Law (Feb. 1990).

34. *See supra* note 29 and authorities cited therein.

35. *See, e.g.*, Note, *Gambling on the Truth*, *supra* note 30, at 38.

36. Jaffee, *supra* note 7, at 934, 936; Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001, 1054, 1060 (1988).

37. Cohen, *The Logic Of Proof*, 1980 CRIM. L. REV. 91, 93; Schum, *supra* note 26, at 857; Shafer, *The Construction Of Probability Arguments*, 66 B.U.L. REV. 799, 803 (1986); Wright, *supra* note 36, at 1061.

the line of reasoning from evidence to conclusion.[38] Other probabilists go further, and argue that mathematical probability concepts describe the optimal operation of the trial process. This latter group, to whom we will refer as "utilitarian probabilists," argues that an optimally operating legal system will maximize the number of correct verdicts (*i.e.*, verdicts that reflect what actually happened) while minimizing the number of incorrect verdicts.[39] They then contend that mathematical probability concepts exactly describe this optimal operation: in a civil case a plaintiff should win if and only if the plaintiff has proven that the probability that her proof warrants recovery is greater than fifty percent;[40] in a criminal case the prosecution should win if and only if the prosecution has proven that the probability that its proof warrants conviction is greater than whatever probability figure should be assigned to the "beyond a reasonable doubt" standard (a number which even utilitarian probabilists hesitate to assign, but which is generally agreed to lie somewhere above ninety percent, and is expressed merely as something like "extremely probable."[41])

---

38. *See, e.g.*, Kaye, *supra* note 31; Lempert, *supra* note 28.

39. *See, e.g.*, Ball, *supra* note 28, at 816-817; Brook, *The Use of Statistical Evidence of Identification in Civil Litigation: Well-Worn Hypotheticals, Real Cases, and Controversy*, 29 St. Louis U.L.J. 293 (1985); Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 Harv. L. Rev. 530, 532 (1989); Winter, *The Jury and the Risk of Nonpersuasion*, 5 L. & Soc'y Rev. 335, 337 (1971).

40. *See, e.g.*, M. Finkelstein, *supra* note 31, at 65; Brook, *supra* note 39, at 296; Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. Rev. 385, 390-91 (1985) (this is Professor Neil Cohen, a probabilist, who is not to be confused with Professor L. Jonathan Cohen, a vehement anti-probabilist); Eggleston, *The Probability Debate*, 1980 Crim. L. Rev. 678, 680; McBaine, *Burden of Proof: Degrees of Belief*, 32 Calif. L. Rev. 242, 246, 262 (1944).

41. The utilitarian probabilists have difficulty with the "beyond a reasonable doubt" standard. *See, e.g.*, Brook, *supra* note 39, at 309 ("[T]he goal of accuracy always has had to contend with other factors in the criminal context."); Kaplan, *Decision Theory and The Factfinding Process*, 20 Stan. L. Rev. 1065 (1968) ("Probably the most important reason why we do not attempt to express reasonable doubt in terms of quantitative odds, however, is that in any rational system of utilities (or disutilities) that determine the necessary probability of guilt will vary with the crime for which the defendant is being tried, and indeed with the particular defendant." *Id.* at 1073.); Tribe, *supra* note 31 (although Tribe is often viewed as an anti-probabilist because he reached the conclusion that probabilistic evidence should not be admissible in criminal cases, he reached that conclusion not because he disagreed with the precepts of the probabilist position but instead because he believed the evidence was unfairly prejudicial). Professor Daniel Shaviro is a utilitarian probabilist who is comfortable applying the mathematical approach in the criminal context. Shaviro asserts that:

> Perhaps the most obvious purpose of the trial system is to reach accurate verdicts—for example, to convict criminal defendants when and only when they have committed the unlawful acts of which they are accused. Apart from any practical benefits of deciding cases accurately (such as improving deterrence), the accuracy of verdicts has moral implications. Consider a criminal trial in which the defendant, accused of rape, defends on the basis of consent, and the presence or absence of consent would be unambiguous if all the facts were known. A false conviction will punish an innocent man, perhaps severely, for a heinous crime. A false acquittal may cause the rape

For anti-probabilists, the "probability" of finding liability in a civil case or of finding guilt in a criminal case sufficient to meet the burden of persuasion can only be arrived at by the process of eliminating alternative explanations inconsistent with the guilt of the defendant.[42] To look at the long run of cases, they argue, results in callously sacrificing a certain percentage of defendants who have not been proven to be culpable.[43] Only if all plausible alternative explanations other than guilt have been negated can the juror find guilt beyond a reasonable doubt.[44] Thus, according to the anti-probabilists, the required juror mindset for conviction is not a subjective probability that the defendant's guilt is "extremely probable," but rather, "No explanation other than the defendant's guilt appearing plausible to me, I *actually believe* that the defendant is guilty (even though I recognize the possibility that I may be wrong)." In support of their position, anti-probabilists can point to the fact that no jurisdiction phrases its burden of persuasion in criminal cases in probabilistic terms, and in fact some courts state that jurors must be convinced of the defendant's guilt "to a moral certainty," which means "that the inference of guilt is the only one that can fairly and reasonably be drawn from the facts, and that the evidence excludes beyond a reasonable doubt every reasonable hypothesis of innocence."[45]

The difference between the utilitarian probabilist and anti-probabilist positions with respect to the burden of persuasion is best demonstrated by an illustration. We will use a variation on a well-known hypothetical proposed

---

victim unjust humiliation, in addition to other harms.

Sheviro, *supra* note 39, at 532; *see also* Kaye, *supra* note 31, at 40 ("Surely it is not some defect in probability theory that restrains us from instructing jurors that they should convict so long as they are, say, at least ninety-five percent certain that the defendant is guilty. A much simpler explanation is that we would prefer not to advertise the fact that we are willing to sacrifice one innocent person in order to secure the conviction of nineteen guilty ones."); Williams, *supra* note 28, at 306-07 (arguing that while theoretically jurors should be instructed on the degree of probability that the law counts as beyond a reasonable doubt, "[t]he answer, [why they are not], I suppose, is that there is no consensus upon this probability, and that the question is theoretical so long as we find it almost impossible to state forensic probability in mathematical terms.").

42. *See, e.g.*, Wright, *supra* note 36, at 1065; Note, *Gambling on the Truth*, *supra* note 30, at 49 (although this Note does not deal with criminal cases, it is, to the shame of all the professional scholars who have written in this area, by far the most concise and understandable explication of the debate between the two camps).

43. *See, e.g.*, Callen, *supra* note 29, at 9.

44. Brilmayer & Kornhauser, *supra* note 16, at 144; Callen, *Second-Order Considerations, Weight, Sufficiency and Schema Theory: A Comment on Professor Brilmayer's Theory*, 66 B.U.L. Rev. 715, 727 (1986); Cohen, *The Role of Evidential Weight in Criminal Proof*, 66 B.U.L. Rev. 635, 648 (1986); Cohen, *supra* note 37, at 91-92; Jaffee, *supra* note 26, at 1001; Jaffee, *supra* note 7, at 939, 947, 1045; Note, *Gambling on the Truth*, *supra* note 30, at 45.

45. People v. Johnson, 140 A.D.2d 626, 528 N.Y.S.2d 666 (1988); *see also* People v. Collins, 68 Cal. 2d 319, 332, 438 P.2d 33, 41, 66 Cal. Rptr. 497, 505 (1968) ("In essence this argument of the prosecutor was calculated to persuade the jury to convict defendants whether or not they were convinced of their guilt to a moral certainty and beyond a reasonable doubt.").

by Professor Nesson: the prison guard murder.[46] Suppose it is known that ninety-nine of 100 prisoners in an enclosed yard collaborated in the murder of a prison guard while the one hundredth prisoner did not collaborate or participate. Suppose that one of the 100 prisoners is picked at random to be tried for the murder. No evidence is presented other than the statistical information. If this imposing statistic—ninety-nine of 100 are guilty—would be sufficient for a utilitarian probabilist to form a subjective belief that it is "extremely probable" that the defendant is guilty (whether a utilitarian probabilist *would* find this statistic to be sufficient to create such a subjective probability is discussed in the next paragraph), the utilitarian probabilist would hold the evidence sufficient for conviction. If each of the 100 possible defendants were tried, the utilitarian probabilist would argue that each should be convicted despite our certain knowledge that this would result in one false conviction.[47] An anti-probabilist, on the other hand, would argue that a directed verdict would be in order for each and every one of the 100 defendants, since as to each there is a distinct possibility of non-guilt that has not been negated by the prosecution.[48]

Utilitarian probabilists sometimes try to soften the mechanistic nature of their position by arguing that a high probability based on the classical or frequency theories does not necessarily translate into a sufficient subjective probability to support a finding of guilt. They argue that the very absence of other evidence which could have been presented by the prosecution may prevent the formation of a subjective probability sufficient to support a finding of guilt.[49] Anti-probabilists contend that this attempt by the utilitarian probabilists to ameliorate their position to eliminate its most obvious objec-

---

46. This hypothetical (using the figure that twenty-four of twenty-five were guilty) was developed by Professor Charles Nesson in Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1192-93 (1979). In the civil context the two famous hypotheticals are the rodeo gatecrasher developed by Professor L. Jonathan Cohen, in THE PROBABLE AND THE PROVABLE 75 (1977), and the Blue Bus case created by Professor Laurence Tribe in Tribe, *supra* note 31, at 1340-41.

47. C. ALLEN, LEGAL DUTIES AND OTHER ESSAYS IN JURISPRUDENCE 286-87 (ed. 1977):

> I dare say some sentimentalists would assent to the proposition that it is better that a thousand, or even a million, guilty persons should escape than that one innocent person should suffer; but no sensible and practical person would accept such a view. For it is obvious that if our ratio is extended indefinitely, there comes a point when the whole system of justice has broken down and society is in a state of chaos. In short, it is only when there is a reasonable and uniform probability of guilty persons being detected and convicted that we can allow humane doubt to prevail over security.

Id.

48. Professor Glanville Williams, who is a probabilist on most points, cannot bring himself to make the final step to probabilizing the burden of persuasion. Williams states "The true reason why the proof fails in the gatecrasher case and the Blue Bus case is that it does not sufficiently mark out the defendant from others. No doubt, we are illogical in this." Williams, *supra* note 28, at 305. In a later continuation of this thought Williams acknowledges that even in a civil case a mathematical probability of 999 out of 1000 of liability would not be sufficient to support a verdict for the plaintiff. Williams, *Rejoinder*, *supra* note 28, at 106.

49. *See, e.g.*, Cohen, *supra* note 40, at 399; Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 106; Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U.L. REV. 439, 457 (1986).

tionable aspect is unconvincing.[50] Professor Wright has summarized the anti-probabilist position in the context of another well-known hypothetical put forth by L. Jonathan Cohen—the rodeo gatecrasher (501 of 1000 spectators are known to have snuck into a rodeo without paying):

> This [probabilist] response clearly is inadequate. First, it does not explain why the jury is not even allowed to consider the naked statistics, rather than being allowed to consider and possibly discount the statistics. Second, it does not explain why the plaintiff, rather than the defendant, is being charged with failure to supply other types of evidence. Third, it at least implicitly admits that other types of evidence are more probative than naked statistics; otherwise why insist on more than the naked statistics? Fourth, it fails to explain why there is no liability even when the objective probability is much higher than fifty percent—for example when only fifty of the 1000 spectators paid for their tickets. In this situation, even when the objective ninety-five percent probability is discounted, the subjective probability almost certainly is greater than fifty percent. Fifth, it fails to address the hypothetical as Cohen presented it, which assumes that no other evidence is available.[51]

The debate regarding the burden of persuasion[52] perhaps consumes more pages in the scholarly debate than any other single aspect of the war between the probabilists and anti-probabilists[53] (with the applicability of Bayes' Theorem to the proof process running neck-and-neck).[54]

3. Ramifications for Relevance

Although probabilists usually take the relevance of statistics and probabilities for granted and focus instead on the burden of persuasion issue and the applicability of Bayes' Theorem, in fact one of the stronger arrows in the quiver of the probabilists is the language of the modern rule of relevance: "having any tendency to make the existence of any fact that is of consequence . . . *more probable or less probable* than it would be without the evidence."[55]

---

50. *See, e.g.*, Brilmayer, *Second-Order Evidence and Bayesian Logic*, 66 B.U.L. Rev. 673, 676-77 (1986).

51. Wright, *supra* note 36, at 1055-56.

52. Special thanks go to Professor Kaye for his help in sorting out the camps in this debate (which is not to say that the author has ended up with an explanation in which Professor Kaye completely concurs).

53. The war shows no signs of abating: "The whole topic remains fascinating, and despite the amount of ink that has been spilled, I fear that more remains to be said." Letter from Professor Kaye to David McCord (May 15, 1990).

54. The author should disclose that he finds the anti-probabilist position to be correct because he finds convincing the anti-probabilist positions with respect to the definition of relevance and to how the burden of persuasion is met.

55. Fed. R. Evid. 401 (emphasis added). Green, *Foreward*, 66 B.U.L. Rev. 377, 377 (1986).

On the infrequent occasion when probabilists do address relevance they simply point to the language of this rule.[56]

Anti-probabilists argue that the material issue in any case is what *actually* happened and that, thus, probabilities are irrelevant because they speak only to the immaterial issue of what *might have* happened.[57] Anti-probabilists argue that the word "probable" in the modern definition of relevance does not refer to mathematical concepts of probability:

> Too, "probable" can mean "worthy of being believed true" as well as it can mean "likely." So, evidence that tends to make a fact "probable" is "evidence that affords ground for belief." Whether a particular "fact" is worthy of being believed true, then, depends upon whether the items evidencing it are, collectively, worthy of being believed to "add up" to the "fact.". . . "Probative" not only does not mean "statistically significant;" it means, as a term of art in Evidence Law, "capable in logic and experience of proving."[58]

Thus, while the modern rule of relevance does indeed speak in terms of making "probable," the anti-probabilists argue that term is not meant in a mathematical sense.

---

56. *See, e.g.*, Brook, *supra* note 39, at 321; Friedman, *A Close Look at Probative Value*, 66 B.U.L. REV. 733, 734-35 (1986); Lempert, *supra* note 28, at 1025-26.

Sometimes probabilists get carried away with their argument however. L. Jonathan Cohen posed a hypothetical in which the issue was whether a man was likely to survive to age seventy and evidence was offered that men with six letter names had a higher survival rate to that age than other men. Cohen concluded that the evidence would not be relevant because there was no causal connection between the number of letters in a name and the age to which a person would survive. Cohen, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 ARIZ. ST. L.J. 627, 633. For the probabilists, Professor Kaye argued "[W]ere we actually confronted with a finite population in which substantially more men with six letter names became septigenerians, then the evidence would be relevant. . . .[T]here is no legal requirement of causal as opposed to purely statistical association." Kaye, *supra* note 32, at 639, 640. Kaye was taken to task for this assertion by Professor Jaffee in Jaffee, *supra* note 7, at 1021. The anti-probabilists clearly have the better of this argument since one cannot imagine a court admitting statistical evidence that was obviously a product of pure happenstance. Professor Kaye refuses to budge on this point, however:

> I can imagine a court admitting statistical evidence that was obviously a product of pure happenstance. A murder by stabbing occurs on a yacht with 10 people on board. Seven are female and six of them have blond hair. A person on a fishing boat in the area says that she saw a blond woman toss a knife overboard. Are you really so sure that the court would sustain an objection to proof by the woman so identified that five other women on board have blond hair? Or would you argue that hair color is causally relevant (unlike some "antiprobabilists") or that the number of women is not a statistic? Or suppose that all passengers were killed instantly in a plane crash. Quite by happenstance, all those with long names were more than 50 years old, and everyone with short names was less than 50 years old. For some good reason, the information that permitted this to be determined directly is no longer available, and it is important to decide whether a passenger lived more than 50 years. Why are you so sure that the name length—a matter of pure happenstance—would be inadmissible?

Letter from Professor Kaye to David McCord (May 15, 1990).

57. *See supra* text accompanying note 36.

58. Jaffee, *supra* note 7, at 1059.

So far we have examined the mathematical toolkit, and the competing schools of thought regarding mathematical evidence in the criminal proof process. But before the reader can confidently plunge into the deep waters of case law and commentary, one more introductory explanation is necessary.

C. The Five Categories of Mathematical Evidence

The scholarly literature and case law are not organized along the lines of categories of mathematical evidence. Indeed, this is perhaps the most serious obstacle to the non-mathematically inclined's attempts to understand those writings because the mathematically sophisticated authors often make broad statements about "probabilistic evidence" which in fact only apply to some kinds of mathematical evidence. A close analysis of the case law demonstrates that there are five categories of mathematical evidence[59] that arise in the discussions of mathematics in the proof process. These categories will be briefly explained here, illuminated further in Parts II and III through an analysis of the landmark *Collins* case and what has happened since, and analyzed in detail in Part IV.

The first category will be denominated "empirical statistics." It consists of use of statistics based on empirical sampled data to eliminate possible culprits, while simultaneously *not* eliminating the defendant as the culprit. The key characteristic of this category is that it involves no probability calculation (although it provides data upon which the second category—probabilities of a random match—can readily be based). Our earlier example in explaining the meaning of "statistics" will suffice here: two of 100 suspects have blood factor X, and blood of the perpetrator containing factor X was found at the scene. Statistically, without formulating a probability, we can say that the ninety-eight percent of the suspect population without blood factor X is eliminated as suspects. If statistics exist with respect to more than one factor, and the occurrence of the factors is mutually independent, then the product rule for independent events can be used to multiply the statistical frequencies together to reach a new statistical incidence of occurrence.

The second category consists of what will be termed "probabilities of a random match." Any empirical statistic can be used to form such a probability. For example, if two of 100 suspects have factor X blood and the blood of the perpetrator containing factor X was found at the scene, we can generate several probabilities. Using only the statistics that the suspect population contains one and only one perpetrator, and knowing that the suspect population is 100, we can generate the probability that a random choice from the population will be the culprit: one-one hundredth (using the

59. Statistics are clearly "facts" and thus there is no difficulty in characterizing them as "evidence." Probabilities are more of a way of thinking about the significance of evidence and thus more easily characterizable as "arguments" than as "evidence." Nonetheless, most courts and commentators have characterized probabilities as "evidence" and thus we will consider probabilities as "evidence" for purposes of this Article as well.

classical theory since there is an equal likelihood of each person's being selected). Using the same statistics and the frequency theory, we generate the probability that over the long run of random choices, we would choose the culprit one-one hundredth of the time. Using the additional information from the hypothetical, we can generate both classical and frequency probabilities regarding selecting at random a person with factor X blood (one-fiftieth), and selecting at random the culprit from a pool consisting only of the two persons with factor X blood (one-half). Note that these probabilities (like all classical and frequency probabilities) do not even purport to make a statement concerning what any particular random choice would *actually* turn up. Each *actual* choice will either turn up the characteristic at issue (a "probability" of "1," after the fact) or not turn it up (a "probability" of "0," after the fact).

The remaining categories of mathematical evidence all involve *subjective* probabilities of the defendant's *guilt*. The third category is called "non-empirical probabilities of guilt incorporating empirical statistics without Bayes' Theorem." In this category, empirical statistics form part of the basis for formation of a subjective probability, but not via use of Bayes' Theorem (*e.g.*, "Based upon my assessment of the probability of truth of the eyewitness testimony plus the statistics regarding blood factors, I believe that the defendant's guilt is extremely probable.") The fourth category is called "non-empirical probabilities of guilt developed without empirical statistics and without Bayes' Theorem." This category is composed of instances where the prosecutor, in line with the mathematical probabilist tenet that all evidence is inherently probabilistic, probabilizes non-empirically sampled data, and then further relies on the probabilist position that the burden of persuasion is probabilistic by arguing that the subjective probability of guilt based upon the non-empirically sampled data is sufficient to convict, all without using Bayes' Theorem. The quintessential example of this category is *People v. Collins*.[60] The fifth category is entitled "non-empirical probabilities of guilt incorporating empirical statistics via Bayes' Theorem," (*e.g.*, "Based upon my evaluation of the probability of truth of the eyewitness testimony, as combined with the blood factor evidence's probability via Bayes' Theorem, I believe that defendant's guilt is extremely probable.") These five categories exhaust the academic debate and the case law regarding mathematical evidence in criminal cases. A discussion of these categories may exhaust the reader as well, but it is necessary to gain a firm understanding of the subject matter. First, though, let us make sure that we understand the most famous of all mathematical evidence cases, *People v. Collins*.[61]

II. UNDERSTANDING THE BEDROCK CASE: *PEOPLE V. COLLINS* RE-EXAMINED

A. *The Forerunners of* Collins

While *People v. Collins* was not the first case involving issues of mathematical probability in the proof process, more than two decades have

60. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).
61. *Id.*

passed since it was decided, and numerous cases involving mathematical evidence have been the subject of appellate opinions, *Collins* nonetheless remains the best-known and most influential case addressing the intersection between mathematical probabilities and the criminal law. Commentators have not used pale adjectives to describe their reaction to the prosecution's approach in the case, including "bizarre,"[62] a "flagrant" misuse of statistical evidence,[63] and "notorious."[64] The *Collins* opinion has been hailed as a "seminal judicial contribution" to the debate regarding probability in the proof process.[65] In short, legal commentators have found *Collins* "irresistible."[66]

Although *Collins* was not the earliest case to raise probability issues, a brief review of its three case law predecessors shows that they did not have the ingredients necessary for staying power because either they did not confront the mathematical issues head-on, or if they did, did not confront them fully. The earliest and most significant of the three cases is the 1915 New York Court of Appeals case of *People v. Risley*.[67] Defendant Risley, a lawyer, was convicted of offering into evidence as genuine a forged and fraudulently altered document. The prosecution's theory was that Risley had typed in the key words on the document on a typewriter at his office. Specimens of the typing of a machine at his office were compared with the allegedly inserted words and the prosecution's experts claimed that there were eleven matching characteristics.[68] The prosecution then called a mathematics professor who assumed probabilities for each of the defects, and then used the product rule to reach the conclusion that the probability of those defects being reproduced by a typewriter other than the defendant's was one in four billion.[69] *Risley* is the first, but certainly not the last, case

62. Finkelstein & Fairley, *supra* note 23, at 496.

63. Note, *Admissibility of Mathematical Evidence in Criminal Trials*, 21 AM. CRIM. L. REV. 55, 61 (1983).

64. Cohen, *supra* note 40, at 388; Gerjuoy, *The Relevance of Probability Theory to Problems of Relevance*, 18 JURIMETRICS J. 1, 25 (1977).

65. Green, *supra* note 55, at 377.

66. Gerjuoy, *supra* note 64, at 25.

67. 214 N.Y. 75, 108 N.E. 200 (1915). An earlier American case had involved probabilistic evidence, but the admissibility of that evidence was never reviewed by an appellate court. That case was the *Howland Will* case, discussed in 4 AM. L. REV. 625 (1870). There the authenticity of a signature on a will was at issue. Opponents of the document claimed that the signature had been traced by someone other than the alleged testator. Noted Harvard mathematician Benjamin Pierce testified for the opponents. He examined 50 specimens of the decedent's genuine signature and concluded that the downward strokes coincided about one-fifth of the time, yet the signature on the proffered will showed 30 such coincidences with the signature on the source from which the opponents claimed that the signature had been traced. Assuming that each downward stroke was independent of the others, and using the product rule for independent events, Pierce calculated that the probability of such coincidence occurring naturally was one-fifth multiplied by itself 30 times, that is, about one in 931 millions of millions of millions.

68. People v. Risley, 214 N.Y. 75, 83, 108 N.E. 200, 202 (1915).

69. *Id.* at 85, 108 N.E. at 202.

showing how easily prosecutors can fall into the trap of (or can consciously resort to as a gambit) the following sequence: (1) create claimed empirical statistics which are illegitimate because they are not based upon a valid empirical sampling (the first category described above); (2) use this statistic to calculate the probability of a random match (the second category described above); and (3) argue that this probability, which is invariably infinitesimal, requires the jury to come to an extremely high subjective probability of guilt (the third category described above), when in fact the evidence falls into the fourth category described above—a non-empirical probability of guilt developed without empirical statistics and without Bayes' Theorem. Ostensibly, the evidence in *Risley* only shows that statistically one could eliminate most of the typewriters in the world, but not defendant's, as having produced the words in issue, and, thus, the evidence falls into the empirical statistics category. Clearly, though, the statistic was a spurious one, and the prosecution sought to use it to convince the jury that the statistical unlikelihood of the occurrence of more than one such typewriter, combined with the fact that Risley owned such a typewriter, should convince them that the subjective probability of Risley's guilt closely approached one.

The New York Court of Appeals found the evidence to have been erroneously admitted for two reasons. First, it held that the professor was not qualified as a typewriting expert, and thus his estimates of the probability were without foundation.[70] In mathematical terms, this meant the court believed that there were no competent statistics upon which to base a probability. As part and parcel of this discussion, the court made some observations that seemed to be the beginning of a head-on attack on mathematical probabilities in the proof process:

> These factors and many others which we cannot foresee, and which, in all likelihood, are beyond the possibility of any human being to ascertain, would enter into any calculation of this character. The statement of the witness was not based upon actual observed data, but was simply speculative, and an attempt to make inferences deduced from a general theory in no way connected with the matter under consideration.[71]

Here is the germ of an argument that subjective probability theory does not correctly model the proof process and thus that probabilistic evidence is irrelevant, a position adhered to by anti-mathematical probabilists to this day. The court in *Risley*, however, did not really follow up on these observations to the point of making them full-blown arguments, leaving the opinion on these points subject to being overshadowed by *Collins*.

The second basis for reversal in *Risley* was that probability was an inappropriate tool in the proof process because probability theory speaks only to future events, not to past events: "The fact to be established in this

70. *Id.* at 85, 108 N.E. at 203.
71. *Id.*

case was not the probability of a future event, but whether an occurrence asserted by the people to have happened had actually taken place."[72] Here again is a germ of an idea that is not fully developed. From a purely mathematical standpoint the court is wrong because probability theory speaks equally to future events, which are by their nature unknown, and unknown past events. For example, the probability of selecting a woman lawyer at random from the adult population of the United States is the same for a random selection that has already been made (with the characteristics of the selected person being concealed from the person performing the probability calculation), as for a random selection that has not yet been made. From a legal standpoint, however, the court's perception that a past event, which either definitely occurred or definitely did not occur (in mathematical terms, a posterior "probability" of "1" and "0," respectively), cannot be proven or disproven by probabilistic evidence showing a likelihood of occurrence between "1" and "0" has continued to be a tenet of the anti-probabilist position. Again though, the *Collins* case developed this argument more fully, and has overshadowed *Risley.*

The issue of probabilistic evidence in criminal cases lay dormant for half a century after *Risley.* Then, in 1966, two appellate courts were confronted with the issue. In an Arkansas case, *Miller v. State,*[73] the defendant was charged with burglary. A soil sample taken from where the perpetrator had fallen near the scene was compared by an expert with soil stains on the defendant's clothes. The expert assumed that there was a one-tenth probability of the same color soil appearing at another location, a one-one-hundredth probability of the same texture of soil appearing, and a one-one-thousandth probability of the same density of soil appearing. Then, assuming that those characteristics were independent, and applying the product rule for independent events, the expert testified that there was a one in a million chance that the defendant's clothing had become soiled in some location other than the burglary scene.[74] The Arkansas Supreme Court held that the expert's testimony had been erroneously admitted because the assumed frequencies were without foundation[75] *i.e.*, the statistics on which the probability calculations were based were simply made up out of thin air. In a New Mexico case, *State v. Sneed,*[76] the defendant was charged with murdering his parents. The prosecution sought to prove that the defendant had purchased a .22 caliber revolver on the day prior to the murders using the name "Robert Crosset," an alias that he had used twice in other cities during the week before the deaths. The gun sales register kept by the pawn shop where the gun had been purchased indicated the buyer as "Robert Crosset" of "Box 210, Las Cruces" and that the buyer was 5'9" tall and

72. *Id.* at 86, 108 N.E. at 203.
73. 240 Ark. 340, 399 S.W.2d 268 (1966).
74. Miller v. State, 240 Ark. 340, 343, 399 S.W.2d 268, 270 (1966).
75. *Id.*
76. 76 N.M. 349, 414 P.2d 858 (1966).

had brown eyes and brown hair. The defendant matched this description, but the pawn shop sales clerk was unable to make a positive identification at trial.[77] To bolster the inference that the purchaser of the gun had been the defendant, the prosecution called a mathematics professor who, on the basis of review of some western states' phone books, estimated that the chance of the name "Robert Crosset" appearing at random was one in thirty million, and then, using the pawn shop's handgun sales register, estimated that the probability of a gun purchaser having a combination of brown hair and brown eyes and a height between 5'8" and 5'10" was one-eighth, and testified that the probability of two people at random choosing the same post office box from a bank of 1000 was one-one-thousandth.[78] The professor used the product rule for independent events to multiply these three figures together and arrived at the probability of one in 240 billion that all of these characteristics would be randomly associated with one person.[79] The New Mexico Supreme Court found reversible error in the admission of the professor's testimony because there was a lack of foundation concerning the frequency estimates for both the name "Robert Crosset" and the physical characteristics.[80] Both *Miller* and *Sneed* illustrate, like *Risley*, a prosecutor's attempt to use illegitimate evidence appearing to fall into the empirical statistics category to support an infinitesimal probability of a random match, and then to argue that the defendant's guilt was proven by an extremely high subjective probability. The *Miller* and *Sneed* cases are less significant than *Risley* because they did not attempt to confront head-on the issue of whether mathematically probabilistic concepts have a place in the proof process. Rather, they simply held that the statistics on which the probabilities were based were without proper support.[81] Thus the field remained open for the California Supreme Court in the *Collins* case to make

77. Some of this information about the case does not appear as part of the appellate opinion. It is to be found in Note, *Evidence: Admission of Mathematical Probability Statistics Held Erroneous for Want of Demonstration of Validity*, 1967 DUKE L.J. 665, 666, which cites to the brief-in-chief of the defendant in *Sneed*, pp. 8-11.

78. Neither the appellate opinion nor the note in the Duke Law Journal, *supra* note 77, gives any indication how that particular post office box number was significant. There was seemingly no showing that the defendant had a habit of choosing that particular post office box number, which means that multiplying by the one-one-thousandth figure simply and arbitrarily increases the final probability.

79. State v. Sneed, 76 N.M. 349, 352-53, 414 P.2d 858, 860-61 (1966).

80. *Id.* at 353-54, 414 P.2d at 861-62. The court does not mention the probability figure regarding the post office box choice at all.

81. It is interesting to note that in a sense the *Sneed* case is a progeny of *Collins*, even though the California Supreme Court opinion in *Collins* was rendered two years after the New Mexico Supreme Court decision in *Sneed*. In fact, the *Collins* case was tried first and obtained nationwide media exposure. This prompted the prosecutor in the *Sneed* case to call the prosecutor in the *Collins* case to discuss the possible use of probability evidence in the *Sneed* case. The *Collins* prosecutor recommended as an expert witness his brother-in-law, who was in fact used by the *Sneed* prosecutor as his expert. This interesting tidbit is contained in a letter from Raymond Sinetar, prosecutor in the *Collins* case, to the author, August 26, 1989.

the first effort to tackle head-on the fundamental issues of the intersection of mathematics and the proof process.

B. Collins: *The Facts*

The basic facts of the *Collins* case are plainly outlined in the California Supreme Court opinion. Yet the question which inevitably intrigues a reader—"What prompted the prosecutor to resort to such an unusual mode of proof and argument?"—is not illuminated by the opinion. It is helpful, therefore, to examine how the case developed from the perspective of the prosecuting authorities by examining the facts provided in the People's brief for the case, and the recollections of the prosecutor (who still remembers the case well after two decades).

On June 18, 1964, the police in the San Pedro area of the city of Los Angeles received a report of a robbery. Mrs. Juanita Brooks reported that she had been shopping and was walking home along an alley in that area of the city. She was pulling a wicker basket containing groceries, with her purse on top. She was using a cane. As she stooped to pick up an empty carton she was suddenly pushed to the ground by a person whom she neither saw nor heard approach. Although stunned, she managed to look up and saw a young woman running from the scene whom she described as weighing about 145 pounds, wearing something dark, and having hair between dark blond and light blond. Mrs. Brooks quickly discovered that her purse, which had contained between thirty-five to forty dollars, was missing.[82]

The police found another eyewitness, John Bass, who lived on the street at the end of the alley and who had been watering the lawn in front of his house at the time of the incident. Attracted by the crying and screaming coming from the alley, he looked in that direction and saw a woman run out of the alleyway and enter a yellow automobile parked across the street from him. He described the woman as Caucasian, slightly over five feet tall, of ordinary build, with her hair in a dark blond ponytail and wearing dark clothing. The car took off immediately and passed within six feet of Bass. He described the driver as a black male having a mustache and beard.[83]

The San Pedro area of Los Angeles had a strong neighborhood feel to it, and the police in the area believed that they had a good reading on who most of the criminally-inclined inhabitants were.[84] The police immediately focused their suspicion regarding the Brooks robbery on Malcolm and Janet Collins, an interracial couple in the area who drove a yellow car, and who were on the police's list of suspicious characters. Within a day or two of June 18, Malcolm had been brought in for a lineup and identified by Mr. Bass.[85] The police noted that Malcolm was not wearing a beard, which made

82. People v. Collins, 68 Cal. 2d 319, 321, 438 P.2d 33, 34, 66 Cal. Rptr. 497, 498 (1968).

83. *Id.*

84. Conversation with Raymond Sinetar, prosecutor in *Collins* (April 12, 1989).

85. Brief for Respondent at 10, *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

Bass' identification less than certain, but also fueled the police's suspicion because it appeared that Malcolm had changed his appearance immediately after the robbery. Evidently the police did not feel confident enough about the evidence to make an arrest at this point and Malcolm was released while the police sought more evidence. On June 22, the police interviewed both Malcolm and Janet Collins. Janet claimed that she had been at her job as a housemaid in San Pedro from about eight o'clock the morning of the robbery until one o'clock that afternoon. The police obviously did not believe this alibi because later that afternoon the same officer who had conducted the interview began surveillance of the Collinses, and for some undisclosed reason called for other officers to meet him at the Collinses' home. As uniformed officers approached the Collinses' front door, Malcolm ran out the back door and was eventually discovered hiding in a neighbor's closet. Undoubtedly, the police viewed this as one more strong fact indicating guilt. Both Malcolm and Janet were arrested and detained for forty-eight hours, during which time the police observed another fact which they believed corroborative of guilt. The officer who had interviewed Janet on the morning of June 22 noticed that after her arrest her hair appeared both shorter and darker, leading to the conclusion that immediately after the June 22 interview she had cut and dyed her hair to befuddle witnesses.[86]

Further investigation unearthed other circumstantial evidence against the Collinses. On June 19, the day after the robbery, Malcolm had paid thirty-five dollars in traffic fines, approximately the same amount that had been taken from Mrs. Brooks.[87] Further, their alibi fell apart when Janet's employer stated that Malcolm had picked Janet up in a yellow car at about 11:30 in the morning the day of the robbery, which would have given time for the Collinses to have arrived at the location of the robbery in time to commit it.[88] Additionally, the eyewitness Bass identified Janet from a police photograph that had been taken on June 22.[89]

Finally, on July 9, 1964, the defendants were arrested and charged. While in custody both Janet and Malcolm talked to the police and, although neither defendant confessed or expressly made damaging admissions, the whole tone of the conversation evidenced a strong consciousness of guilt on both their parts.[90] Further, they told inconsistent stories about how the money Malcolm had used to pay the traffic fines had been obtained. Malcolm stated that he had won it gambling, while Janet stated that it had come from her earnings, even though at the time of the couple's marriage seventeen days earlier on June 2, 1964, they had only twelve dollars to their name, part of which had been spent on a trip to Tijuana, and Janet's earnings

86. *Id.*

87. *Collins*, 68 Cal. 2d at 323, 438 P.2d at 35, 66 Cal. Rptr. at 499.

88. *Id.* at 322, 438 P.2d at 34, 66 Cal. Rptr. at 498.

89. Brief for Respondent at 8, *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

90. *Collins*, 68 Cal. 2d at 324, 438 P.2d at 36, 66 Cal. Rptr. at 500.

since the marriage were not more than twelve dollars a week.[91] The final nail in the Collinses' coffin for the prosecuting authorities was that before the trial of the Brooks robbery, Malcolm was arrested for two other robberies—one of a gas station and one of a restaurant—that occurred two months after the Brooks robbery.[92]

The Brooks robbery case came to trial in December, 1964. The file was assigned to Raymond Sinetar, a Los Angeles County assistant district attorney.[93] Because the file appeared to be a "completely routine felony from all appearances," the prosecutor did not get the file much in advance of trial, perhaps the day before, or perhaps even the same day. His discussions with the police officer witnesses, however, plus the facts in the file, soon left no doubt in his mind that the Collinses were the culprits.

As the prosecutor put his elderly eyewitnesses on the stand, however, it quickly became apparent that their identifications of the Collinses were "shaky." In fact, Mrs. Brooks had never seen the driver of the car at all, while Mr. Bass' identification of Malcolm was shaken by his uncertainty of identification at a prior lineup.[94] One night toward the end of the prosecution's case-in-chief Mr. Sinetar lay in bed convinced of the guilt of the Collinses, yet also convinced that his case was not nearly as strong as he would have liked. It was clear to him that the best argument for the

---

91. Brief for Respondent at 22, *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

92. Malcolm was tried on these charges after the trial of the Brooks robbery and convicted of both robberies. These convictions were sustained on appeal and Malcolm was imprisoned for these crimes. Prosecuting attorney Sinetar cannot recall for sure whether he knew that Malcolm had allegedly committed these other robberies at the time the Brooks robbery came to trial, but he thinks that he most likely would have. Telephone conversation of the author with Raymond Sinetar, prosecutor in *Collins* (July 19, 1989).

93. The information that follows was kindly provided by Mr. Sinetar during a phone conversation with the author on April 12, 1989. Mr. Sinetar still remembers the case vividly because it turned into such a *cause celebre*. Mr. Sinetar went on to have a distinguished career. He was with the Los Angeles County District Attorney's Office for a total of about fifteen years, leaving in 1979 to become the chief deputy district attorney for Ventura County. He is currently a judicial attorney for the California Court of Appeals.

Unfortunately, the author was unable to track down the public defenders who represented the Collinses. The author did manage, however, to locate the attorneys on each side who handled the appeal. Nicholas C. Yost, who represented the state, is now a specialist in environmental and natural resources law in a Washington, D.C., law firm. He recalls the case as, "personally, an immense and stimulating intellectual challenge; societally, it went to the heart of how cases are proved." Telephone conversation Nicolas C. Yost (March 15, 1989). Rex K. DeGeorge, who represented Malcolm Collins on appeal, practices law in Beverly Hills. At the time, he was an associate in an insurance defense law firm. He read about the case, was intrigued, and told the partners he would like to volunteer to handle the case on appeal, working on it after hours. They refused permission, but he took the case anyway and was fired. He says, "That was the best thing that could have happened to me. I opened my own practice, and was quite successful." He added an interesting twist to the facts by noting that he remembers finding out that Malcolm and Janet were not even married; their marriage had been void for some reason that he cannot now recall. Telephone conversation with Rex K. DeGeorge (March 13, 1989).

94. *Collins*, 68 Cal. 2d at 325, 438 P.2d at 36, 66 Cal. Rptr. at 500.

prosecution focused on the "discrete bits of circumstantial evidence" such as the hair color of the assailant, the color of the car, the interracial nature of the couple, and the hair on the face of the getaway driver. As he lay in bed pondering how to communicate the impact of those bits of circumstantial evidence to the jury, the idea occurred to him to "funnel these subjective words into a quantitative analysis."

Acting upon that inspiration the next morning, he telephoned a local college to contact a mathematics instructor and eventually was put in touch with a Mr. Martinez, who taught basic math courses. During a brief "five minute" phone call, Martinez agreed to come to court that day to testify about the application of the product rule. The prosecutor then went to work attempting to decide on ballpark figures to use as statistics for the various characteristics on which he intended to focus. He did this by asking the clerical staff in his office for their estimates of the proportion of occurrences of blond hair, ponytails, interracial couples, yellow cars, and facial hair on black men. Having obtained these estimates, he decided upon the figures that he would use at trial that day and made up a chart.

Martinez did indeed appear at the trial that day. The first time the prosecutor had ever met him was when he walked into the courtroom. During Martinez's brief testimony, he first identified himself and his qualifications and then explained the product rule using a simple hypothetical involving the probability of rolling a certain number two times in a row with a die. The prosecutor then showed him the chart on which were listed the famous six factors and the probabilities estimated for them by the prosecutor (as per his conversations with members of the clerical staff), namely: partly yellow automobile one-tenth; man with mustache one-fourth; girl with ponytail one-tenth; girl with blond hair one-third; Negro man with beard one-tenth; and interracial couple in car one-one-thousandth.[95] The prosecutor attempted to make clear that Martinez was in no way responsible for the assigned statistics by the following question and answer sequence:

> Question: "Your specialty does not equip you, I suppose, to give us some probability of such things as a yellow car as contrasted with any other kind of car, does it?. . . I appreciate the fact that you can't assign a probability for a car being yellow as contrasted to some other car, can you?
> Answer: "No, I couldn't."[96]

The prosecutor also attempted to make clear that the statistics were not to be taken as givens by the jury but were "conservative estimates" used in a "hypothetical way" to illustrate the application of the product rule:

> Question: Mr. Martinez, if I haven't made it clear before, may I make it clear now, that the purpose of this diagram and illustration is only for illustrative purposes to illustrate the proposition that you have already indicated is true that, namely, when you have separate

95. *Id.* at 325-26 n.10, 438 P.2d at 37 n.10, 66 Cal. Rptr. at 501 n.10.
96. *Id.* at 328-29 n.13, 438 P.2d at 38 n.13, 66 Cal. Rptr. at 502 n.13.

> and independent factors, that to arrive at what the likelihood would be of their concurring, it is the product of each of their probabilities. That is the only purpose for this diagram.[97]

The trial judge also attempted to bring this home to the jury, instructing that the testimony "has only been received for the purpose of illustrating the mathematical probabilities of various matters, the possibilities for them occurring or re-occurring."[98] The defense objected to Martinez's testimony on the grounds that it was immaterial, invaded the province of the jury, and was based on unfounded assumptions.[99] The Collinses then presented their defense. Each of them testified that they had not been in the vicinity of the crime, but instead had gone directly from Janet's work to visit with friends.[100] Malcolm also testified, as did other defense witnesses, that he had shaved off his beard sixteen days before the Brooks robbery.[101]

When the time arrived for summation the prosecutor again attempted to make clear that the individual probabilities assigned to the six characteristics were not ironclad, but merely illustrative:

> "You can put your own estimate on it, it doesn't matter. You might, for your own purpose, put in your own figures, but the process is the same. . . . I think lastly, as I say, these are my own ideas, my own estimates, I have tried to make them conservative,. . . Again, it doesn't matter that I put down this figure. I would invite counsel to use his own figures or you to use your own, just taking these six figures alone, and I have ignored all the rest, just for illustrative purposes the chances of their being any other couple which fits the description of the Collinses on that occasion is at least one in twelve million . . ."[102]

The prosecutor ended his mathematical argument with a flourish, arguing that "the chances of anyone else besides these defendants being there, . . . having every similarity, . . . is somewhat like one in a billion."[103] These modes of proof and argument were so original that the case was reported by a local newspaper, from whence it was picked up by a wire service, and eventually reported twice in *Time* magazine.[104] The Collinses were both convicted and Malcolm appealed, although Janet did not. Malcolm's case worked its way to the California Supreme Court, where a little over three years later the court delivered its famous opinion.

---

97. Brief for Respondent at 45-46, *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

98. *Id.* at 42.

99. *Collins*, 68 Cal. 2d at 326, 438 P.2d at 37, 66 Cal. Rptr. at 501.

100. *Id.*

101. *Id.* at 323 n.5, 438 P.2d at 35 n.5, 66 Cal. Rptr. at 499 n.5.

102. Brief for Respondents at 47, *People v. Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

103. *Collins*, 68 Cal. 2d at 326, 438 P.2d at 37, 66 Cal. Rptr. at 501.

104. April 26, 1968, at 41; Jan. 8, 1965, at 42.

C. *Collins*: The Opinion

The California Supreme Court reversed Malcolm Collins' conviction, finding the prosecution's approach invalid for six reasons.[105] We need to examine the court's opinion from two perspectives: (1) what does it say regarding the respective virtues of the tenets of the mathematical probabilist and anti-mathematical probabilist positions?; and (2) what does it say about any of the five categories of mathematical evidence?

1. Lack of Foundation for the Six Statistics

At first glance, *Collins* appears to be an exact analog of *Risley*, *Miller*, and *Sneed*—a prosecution attempt to use illegitimate empirical statistics to support a spurious argument that there was a high subjective probability of guilt based upon those statistics. The California Supreme Court dealt with the prosecution's proof in the same manner as the *Risley*, *Miller* and *Sneed* courts before it—the evidence was improper because of a lack of foundation for the statistics.[106] Commentators, even those who are mathematical probabilists, have agreed.[107] Yet the *Collins* prosecutor's efforts are distinguishable from those of the prosecutors in the earlier cases. In those earlier cases, the prosecutors attempted to convince the jury that made-up statistics were in fact *empirical*; in *Collins* the prosecutor *told* the jury that the statistics were *not* empirically based, and urged them to develop their own. The prosecutor adhered to both the basic probabilistic tenet that all evidence is probabilistic, as is the reasoning from proof to conclusion, and to the utilitarian probabilist position that the burden of persuasion is probabilistic. The court's rejection of the prosecution's efforts means that it did not agree: (1) that all evidence is probabilistic; (2) that the burden of persuasion is probabilistic (although the court is more explicit regarding this point later in the opinion); and (3) that the fourth category discussed above—non-empirical probabilities of guilt developed without empirical statistics and without Bayes' Theorem—is an appropriate one in a criminal case.

2. Mutual Dependance of the Six Factors

The second reason for reversal was that there was no showing of mutual independence of the six factors which would legitimate the use of the product rule.[108] Indeed, it is obvious to most readers of the opinion that at least two of the characteristics are not independent, namely, "man with mustache"

---

105. Malcolm was not retried on the charge. According to the prosecutor's recollection, since Malcolm had been convicted and sentenced for two other robberies in a subsequent trial, which convictions were not reversed on appeal, it would have served no purpose to have retried him on the Brooks robbery, because even if convicted he would not have stood to serve any more time. Conversation with Raymond Sinetar (April 12, 1989).

106. *Collins*, 68 Cal. 2d at 327, 438 P.2d at 38-39, 66 Cal. Rptr. at 502.

107. *See, e.g.*, Finkelstein & Fairley, *supra* note 23 (the first scholarly proponents of the use of mathematics in the proof process).

108. *Collins*, 68 Cal. 2d at 327, 438 P.2d at 38-39, 66 Cal. Rptr. at 503.

and "Negro man with beard," since the occurrence of mustaches and beards are related to each other. Further, other possible dependencies exist. For example, it is possible that "girls with blond hair" wear their hair in ponytails more frequently than do "girls" with other color hair. Similarly, it is possible that an interracial couple in a car is more likely to be found in a yellow automobile than are non-interracial couples. Further, it is possible that "Negro men with beards" prefer the company of "girls with blond hair" as opposed to "girls" with other color hair. Thus the court's conclusion that the testimony was defective because independence of the factors was not shown has struck virtually every reader of the opinion as correct. There are three points to be made here, however, before we leave this topic. First, it should be noted how easily the prosecution could have avoided the most obvious dependency, that is, the one between "man with mustache" and "Negro man with beard." The prosecutor could simply have combined the two factors and estimated a probability of a "Negro man with mustache and a beard." His failure to do so is probably reflective of the haste with which the demonstration was conceived and presented. Second, aside from the obvious dependency between mustaches and beards which we know from common experience are correlated, our intuition regarding the likelihood of the other possible dependencies is much weaker. We have no strong feeling that, for example, "Negro men with beards" prefer yellow automobiles in a higher proportion than other groups in the population, or that "girls with blond hair" wear ponytails with substantially greater frequency than "girls" with other color hair. In short, while it is true that the prosecutor did not prove a lack of mutual dependence, our intuition regarding most of the factors is that they are likely not to be mutually dependent or, if they are, only to a small extent. Third, and most importantly, to the extent that there is a possibility of mutual dependency, the significance of this fact should be minimal under the subjective theory of probability. Non-empirical probabilities are acknowledged to be only ballpark common sense estimates anyway. They result in a final probability using the product rule that is not exact, but only suggestive of a likelihood. Thus, by rejecting the prosecution's approach, the court emphasized that probabilizing non-empirically sampled data is inappropriate in the proof process.

3. Defects in "the Entire Enterprise"

a. What Was "the Entire Enterprise" Upon Which the Prosecution Embarked?

The California Supreme Court stated that even after assuming away the lack of foundation defect and the mutual dependence defect, "the entire enterprise upon which the prosecution embarked . . . was gravely misguided"[109] in two ways. But before we can fully understand the court on those two

109. *Id.* at 329-31, 438 P.2d at 39-40, 66 Cal. Rptr. at 503-04.

points, we must be clear regarding what exactly *was* the enterprise upon which the prosecution embarked. The prosecutor was not clear and, unfortunately, as will be shown, neither was the court. But before we can see how both the prosecutor and the court were muddled in their thinking, we must try to identify exactly what enterprises *were available* for the prosecution to embark upon. From a mathematical standpoint, there are three separate and distinct "enterprises" upon which the prosecution might have embarked.

First, the prosecution could have attempted to calculate the probability of a random match, *i.e.*, that a randomly chosen couple from a population that included the Collinses would have all six characteristics. Assuming that the statistical occurrence of each factor was independent of the others, the product rule would tell us that all six characteristics would appear in one couple in twelve million. The probability of a random match that naturally follows, using the classical theory, is that there is a one in twelve million chance that any couple picked at random from the population would manifest those six characteristics.[110] The second enterprise upon which the prosecution might have embarked was the following: given that we know there is one couple in the population, namely the Collinses, who possess all six characteristics, what is the probability that, excluding the Collinses from the population, there is *another* couple which duplicates those characteristics? This will be referred to as the "probability of duplication." Surprisingly to the non-mathematically inclined, this calculation, which stated in prose seems little different than the probability of a random match is, from a mathematical standpoint, entirely different. Further, the "probability of duplication" calculation is much more complex. The complicated formula for calculating the "probability of duplication" is set forth in the Appendix to the Collins opinion,[111] and is discussed at length in the scholarly literature.[112] This lengthy calculation does not need to be repeated here, although a brief comment will be made later concerning the population figure that the court chose for its calculation.[113] It is sufficient for our purposes to note two things. First, the probability of a random match and the probability of duplication are quite distinct mathematically. Second, unless the suspect population is quite small, the probability of duplication will be dramatically higher (*i.e.*, more likely) than the probability of a random match.[114] The

110. *See, e.g.*, Finkelstein & Fairley, *supra* note 23. "Because the court was dealing with an existing, finite population, the frequency with which couples with the identifying characteristics may be found in that population is identical to the probability of selecting one at random." *Id.* at 493.

111. *Collins*, 68 Cal. 2d at 333-35, 438 P.2d at 42-43, 66 Cal. Rptr. at 506-07.

112. *See, e.g.*, R. EGGLESTON, EVIDENCE, PROOF AND PROBABILITY 241-47 (2d ed. 1983). Charrow & Smith, *A Conversation About "A Conversation About Collins"*, 64 GEO. L.J. 669 (1976); Cullison, *Identification by Probabilities, and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision)*, 6 HOUS. L. REV. 471, 495-98 (1969); Fairley & Mosteller, *A Conversation About Collins*, 41 U. CHI. L. REV. 242, 248-52 (1974); Finkelstein & Fairley, *supra* note 23, at 492-95.

113. *See infra* note 127.

114. For example, using a suspect population of twelve million couples, the court calculated

third enterprise upon which the prosecution might have embarked is to calculate whether the Collinses were in fact the guilty couple. This will be referred to as the "probability of guilt." This calculation differs from both the probability of a random match and the probability of duplication. The probability of guilt can be mathematically modeled, but is even more complex than the probability of duplication. Professor Alan Cullison undertook in an article shortly after *Collins* to explain in mathematical terms the derivation and application of the formula needed to engage in the probability of guilt calculation.[115] The explanation consumes twelve pages and, despite Professor Cullison's lucidity, is virtually incomprehensible to one not possessing a high level of mathematical sophistication.[116] Again, the mathematical formula need not be repeated here since the key fact to be noted is that the calculation differs both from the probability of a random match and the probability of duplication.[117]

Given these three possible enterprises, exactly which enterprise did the prosecution embark upon in *Collins*? The answer seems to be "some of each." As has been shown, the prosecutor had set sail across very deep mathematical waters. Given the spur-of-the-moment nature of the prosecutor's consultation with the mathematician, we can be confident that the prosecutor did not understand the difference among the three probability calculations, and thus it is not surprising that he was not clear regarding exactly what the evidence was intended to show. According to the court, the prosecutor sought to convince the jury "that there was but one chance in twelve million that any [other] couple possessed the distinctive characteristics of the defendants."[118] While this could be read as either a probability of a random match argument or as a probability of duplication argument, the latter seems to be more true to the prosecutor's intentions. Similarly, the prosecution's argument that "the chances of anyone else besides these defendants being there, having every similarity, is somewhat like one in one billion"[119] could be read as either of those arguments, but again seems to be better read as a probability of duplication argument. According to the court, the prosecutor also sought to convince the jury that "it was to be inferred that there could be but one chance in twelve million that the defendants were innocent,"[120] which clearly appears to be a probability of guilt argument. In summary, while the prosecutor was somewhat muddled,

---

the probability of duplication as 41% compared with the one in twelve million figure suggested by the prosecution, which is in practical effect zero percent. *Collins*, 68 Cal. 2d at 334-35, 438 P.2d at 42-43, 66 Cal. Rptr. at 506-07.

115. Cullison, *supra* note 112.

116. *Id.* at 484-95.

117. For an attempt to calculate the probability of guilt on the facts in *Collins*, see Kingston, *Probability & Legal Proceedings*, 57 J. CRIM. L. CRIMINOLOGY & POL. SCI. 93 (1966), which reaches the conclusion that in the long run about five percent of the conclusions of guilt in cases exactly like *Collins* could be expected to be in error. *Id.* at 97.

118. *Collins*, 68 Cal. 2d at 325, 438 P.2d at 37, 66 Cal. Rptr. at 501.

119. *Id.*

120. *Id.*

probably the best characterization of his argument is that it was primarily a probability of duplication argument, with the probability of duplication being so small that it could be disregarded and the jury could conclude that the probability of guilt was certain.[121]

Did the court, then, correctly characterize the prosecutor's "enterprise", which it found to be "greatly misguided"? While the court eventually did correctly characterize the prosecution's argument, the beginning of its analysis is questionable. Examine the opening paragraph of the discussion:

> We now turn to the second fundamental error caused by the probability testimony. Quite apart from our foregoing objections to the specific technique employed by the prosecution to estimate the probability in question, we think that the entire enterprise upon which the prosecution embarked, *and which was directed to the objective of measuring the likelihood of a random couple possessing the characteristics allegedly distinguishing the robbers,* was gravely misguided. *At best, it might yield an estimate as to how infrequently bearded Negroes drive yellow cars in the company of blond females with ponytails.*[122] (Emphasis added.)

There are two questionable aspects to this paragraph. First, the court identifies the enterprise upon which the prosecution had embarked as "measuring the likelihood of a random couple possessing the characteristics allegedly distinguishing the robbers." As we have just pointed out, the prosecution was not attempting primarily to use a probability of a random match argument, but rather was attempting to use probability of duplication and probability of guilt arguments. Second, the court indicates that "an estimate as to how infrequently bearded Negroes drive yellow cars in the company of blond females with ponytails" is dramatically different from the calculation "measuring the likelihood of a random couple possessing the characteristics allegedly distinguishing the robbers." In fact, as was pointed out earlier,[123] these two figures are little more than restatements of each other: once the frequency of occurrence of the six characteristics and their mutual independence is assumed, the probability of a random match naturally follows. Surprisingly, these two weaknesses in the opinion have never been pointed out despite the voluminous commentary and numerous case cites to *Collins.*

After this shaky beginning, the court succeeded in righting itself in the next paragraph when it noted that one of the reasons that it doubted the prosecution's approach could furnish guidance on the crucial issue of determining which of the admittedly few such couples could have been guilty of

---

121. *See also* Cullison, *supra* note 112. "The prosecutor was neither clear nor consistent in identifying his exact purpose in bringing probability theory into the *Collins* case." *Id.* at 477.

122. *Collins*, 68 Cal. 2d at 329-30, 438 P.2d at 39-40, 66 Cal. Rptr. at 503-04.

123. *See supra* note 110 and accompanying text.

committing the robbery was that the probability did not prove that "only *one* couple possessing those distinctive characteristics could be found in the entire Los Angeles area."[124] Thus, the court realized that the thrust of the prosecution's argument relied on the low probability of duplication from which the prosecutor extrapolated an astronomically high probability of guilt. The court correctly recognized that, even assuming away the foundation and mutual dependence problems, the most that the prosecution's formula "could *ever* yield would be a measure of the probability that a random couple would possess the distinctive features in question."[125] The court then referred to its Appendix, which showed that "the prosecution's figures actually imply a likelihood of over 40 percent that the Collinses could be 'duplicated' by at least *one other couple who might equally have committed the San Pedro robbery.*"[126] Having caught the prosecution's drift, the court proceeded to find two defects in the approach.[127]

---

124. *Collins*, 68 Cal.2d at 330, 438 P.2d at 40, 66 Cal. Rptr. at 504.

125. *Id.*

126. *Id.*

127. While the mathematical computations performed by the *Collins* court in the Appendix are indisputably correct, there is some considerable question whether the suspect population figure, set by the court at 12 million, is the proper figure to have been used. Apparently the court selected that figure simply by assuming that since the probability of a random couple's possessing those six characteristics was one in 12 million, a population of 12 million sample couples was logical. Yet in fact there is no special reason for selecting a suspect population of 12 million couples just because the probability of a random couple's possessing the six characteristics is one in 12 million. *See* Fairley & Mosteller, *supra* note 112, at 251. The object should not have been simply to take the number of couples which would seem to be required to generate one couple possessing all six characteristics, but rather to find the sample population best representing the number of couples who might have committed the crime. Then it could be determined out of that population what the likelihood was that more than one couple possessing those characteristics existed. Accordingly, 12 million couples seems highly arbitrary. Yet what the correct suspect population figure should be is highly problematical. On the one hand, one could argue that the 12 million figure is far too high for the very reason that the entire "coupled" population of the Los Angeles area, even including visitors who did not reside in that area, must have been much smaller than 12 million in 1964. One could argue for an even smaller suspect population on the theory that some proportion of the "coupled" population of the Los Angeles metropolitan area on the day of the crime could not physically have been at the scene at the time the crime occurred. On the other hand, one could argue that a suspect population of 12 million couples understates the correct size because theoretically *any* couple in the United States (or the United States and Mexico, or the *whole world*) could have been in Los Angeles on the day of the crime. Yet given what the court was doing in the Appendix, which was basically playing along with the prosecution's approach, it seems that the most logical suspect population would have been the population from which the prosecution had calculated the probabilities for the six characteristics. The prosecution never made clear what was the population from which those estimates were made, but implicitly it seems that it must have been roughly the population of the Los Angeles metropolitan area (presumably including a generous addition for couples who did not reside there but who were physically there that day). This population could have been only roughly determined, but would certainly have been several times less than 12 million couples. Decreasing the size of the suspect population to a more realistic level, say to something like two million, would have dramatically decreased the probability of duplication.

b. Traditionally Nonprobabilistic Evidence Cannot Be "Probabilized" Via the Subjective Theory

As a third basis for reversal, the court stated that: "[T]he prosecution's theory of probability rested on the assumption that the witnesses called by the People had conclusively established that the guilty couple possessed the precise characteristics relied on by the prosecution. But no mathematical formula could ever establish beyond a reasonable doubt that the prosecution's witnesses correctly observed and accurately described the distinctive features which were employed to link defendant to the crime."[128] Here the court was at its most vigorous in rejecting the probabilist principle that all evidence, and the burden of persuasion, is probabilistic: "[T]he likelihood of human error or of falsification obviously cannot be quantified; that likelihood must therefore be excluded from any effort to assign a *number* to the probability of guilt or innocence."[129]

c. The Probabilistic Evidence Was Insufficient, Even if Proper Foundation and Mutual Dependence Are Assumed

As a fourth basis for reversal, the court noted that even if the prosecution had used the figure obtained from the product rule for its mathematically proper purpose, which was to demonstrate the low probability of a random match, that probability "could furnish the jury with absolutely no guidance on the crucial issue: *Of the admittedly few such couples, which one, if any, was guilty of committing this robbery*?. . . Urging that the Collinses be convicted on the basis of evidence which logically establishes no more than this seems as indefensible as arguing for the conviction of X on the ground that a witness saw either X or X's twin commit the crime."[130] It is clear that the court found the evidence to be insufficient, but insufficient in what respect? The most obvious reading is that the evidence was insufficient to support a conviction, *i.e.*, an argument going to the weight of the evidence, not its admissibility. But this reading is bedeviled by the fact that the issue, as framed by the parties and the court, was one of admissibility, not sufficiency. If the court had really been conducting a review of the sufficiency of the evidence, it would have looked at all the evidence in the case, including an eyewitness identification, and would likely have found sufficient evidence to support the conviction. The less obvious reading is that the court found the evidence to have zero probative value, and thus insufficient evidentiary worth to even be relevant. This reading, while not without difficulty, is probably the preferable one because it comports with the admissibility issue before the court.

---

128. *Collins*, 68 Cal. 2d at 330, 438 P.2d at 40, 66 Cal. Rptr. at 504.

129. *Id.* It is possible to read the opinion on this point, as does Professor Kaye, as not rejecting the assigning of a subjective probability, but rather rejecting an attempt to derive such a probability from a mathematical formula. Letter from Professor Kaye to David McCord (May 15, 1990). The author believes that the *Collins* opinion sweeps more broadly than that.

130. *Id.* at 330-31, 438 P.2d at 40-41, 66 Cal. Rptr. at 504-505.

4. Can the "Beyond a Reasonable Doubt" Standard Be Quantified?

The fifth basis for reversal found by the court was the prosecution's attempted probabilization of the burden of persuasion. The court characterized the prosecutor's summation to the jury as follows:

> [T]he prosecutor told the jurors that the traditional idea of proof beyond a reasonable doubt represented "the most hackneyed, stereotyped, trite, misunderstood concept in criminal law." He sought to reconcile the jury to the risk that, under his "new math" approach to criminal jurisprudence, "on some rare occasion . . . an innocent person may be convicted." "Without taking that risk," the prosecution continued, "life would be intolerable because there would be immunity for the Collinses, for people who chose not to be employed to go down and push old ladies down and take their money and be immune because how could we ever be sure they are the ones who did it?"[131]

The court concluded, "In essence this argument of the prosecutor was calculated to persuade the jury to convict defendants whether or not they were convinced of their guilt to a moral certainty and beyond a reasonable doubt."[132] A clearer rejection of the utilitarian probabilist position that the burden of persuasion is probabilistic cannot be imagined. Instead, the court, by using the concept of elimination of doubt "to a moral certainty," clasped to its bosom the anti-probabilist position that all plausible explanations inconsistent with guilt must be negated by the prosecution before guilt beyond a reasonable doubt can be found.

5. The Mathematical Arguments Were Unfairly Prejudicial

The sixth basis for reversal was the court's conclusion that not only did the mathematical probabilities have no probative value, they caused unfair prejudice. Most of the unfair prejudice found by the court inhered in the prosecution's misuse of the "probability of a random match" with the "probability of duplication" and "probability of guilt," coupled with the fact that the defense attorney had no advance warning to prepare to meet the evidence.[133] But even though the court asserted that it was making no generally applicable "appraisal of the proper applications of mathematical techniques in the proof of facts,"[134] the court went on to intimate its feeling that juries may generally not be competent to deal with such evidence: "[W]e have strong feelings that such [mathematical] applications, particularly in a criminal case, must be critically examined in view of the substantial unfairness to a defendant which may result from ill conceived techniques with which

---

131. *Id.* at 331-32, 438 P.2d at 41, 66 Cal. Rptr. at 505.
132. *Id.*
133. *Id.* at 330-31, 438 P.2d at 40-41, 66 Cal. Rptr. at 505.
134. *Id.* at 331, 438 P.2d at 41, 66 Cal. Rptr. at 505.

the trier of fact is not technically equipped to cope."[135] This position again strongly supports the anti-probabilist side of the debate.

D. *Why* Collins *Has Had Staying Power*

There are at least six reasons why *Collins* became and has remained the preeminent case regarding mathematical probability in the proof process. First, the facts are vivid and memorable.[136] Second, *Collins* was the first case in which the prosecution used subjective probabilities based on acknowledgedly non-empirical data. This meant that the stakes were high—if the prosecutor's technique in *Collins* had passed scrutiny, one imagines that many other prosecutors would have tried the technique. Third, the case was decided by the California Supreme Court as opposed to the supreme court of some smaller and less influential state. Fourth, the *Collins* court was the first to tackle the issue of mathematics in the proof process head-on. The court's elaborate mathematical Appendix in particular demonstrated the willingness of the court to engage the probabilists on their own turf. Fifth, the *Collins* case has struck most readers as powerfully reasoned and correctly decided. For a case addressing an issue that was both fundamental and novel, *Collins* is a remarkably sophisticated opinion. Sixth and finally, *Collins* was decided at a propitious moment in the history of evidence scholarship. As Professor Richard Lempert has noted, by the mid-1960s the great common-law evidence authorities such as Wigmore, Maguire, and McCormick had finished their work, but no new significant evidence scholarship efforts were being undertaken.[137] Interest in the field was awakened, however, by the drafting and promulgation of the Federal Rules of Evidence shortly thereafter, which led to a transformation of evidence scholarship "from a field concerned with the articulation of rules to a field concerned with the process of proof."[138] Professor Lempert adds that "[N]owhere is the concern for proof more central than in that body of scholarship which seeks to build on or criticize mathematical models as modes of proof or as a means of understanding trial processes."[139] Indeed, there is a bit of a "chicken and egg" problem here about whether *Collins* had a significant role in providing impetus to this wave of evidence scholarship, or whether *Collins* merely rode the wave of interest in a topic that was already of concern. In any event, *Collins* is generally identified with the inception of this trend in evidence scholarship.[140] This is evidenced by the preeminence of *Collins* in law school

---

135. *Id.*

136. In informal conversations with colleagues the author has discovered that even professors whose areas of expertise are far removed from evidence law at least vaguely remember "that crazy case involving the woman with the pony tail, the black man and the yellow car.

137. Lempert, *supra* note 49, at 439.

138. *Id.*

139. *Id.* at 440.

140. *See, e.g.*, Green, *supra* note 55, at 377. "After *Collins*, judges and scholars questioned whether mathematical proof should be admitted under *any* circumstances, even assuming that

evidence texts, which provide the introduction for most lawyers and attorneys to the issues of mathematical probabilities in the proof process.[141]

III. An Overview of Case Law and Scholarship Since *Collins*

The substance of the case law and scholarship since *Collins* has already been largely revealed in the earlier summary of the probabilist and anti-probabilist positions, and will be examined further in the analysis included in Part IV. It will be helpful, however, to gain an overview of the case law and scholarship in the two decades subsequent to *Collins* to illuminate the trends that have developed over that time which have led to the current state of case law and scholarship.

The case law in the first decade following *Collins* is easily reviewed since only one significant criminal case involving probabilistic issues was decided during that period, and that case had already been in the judicial pipeline before the *Collins* decision was rendered.[142] Either prosecutors during that decade did not think of using probabilistic techniques, or defendants did not seek to challenge such efforts on appeal—both of which seem unlikely—or *Collins* had a powerful general deterrent effect. On the scholarly front, however, *Collins* proved to be anything but a deterrent: it spawned an outburst of scholarly writing, two pieces of which are well known even today.[143] The first noteworthy article spawned by *Collins*, by Michael Finkelstein and William Fairley, was entitled "A Bayesian Approach To Identification Evidence."[144] The first part of the article examined the *Collins* case

---

the particular defects in the proof in that case could be avoided. In the ensuing debate, numerous blue buses have run untold numbers of near-sighted elderly ladies off the road, hundreds of alleged gate crashers have been collared, dozens of murderous prisoners have been brought to justice, and countless articles, books and opinions have been written on the subject." *Id.* at 377-78.

141. This author has 14 evidence texts on his shelf, 11 of which contain a substantial number of judicial opinions and three of which do not. Of the 11 texts that include a substantial number of opinions, *Collins* is used as a principal case in seven of them, as a note case in two of them, and not mentioned in the other two. Of the three books that have virtually no opinions in them, *Collins* is mentioned in a footnote in two of them and not mentioned in the other one.

142. State v. Coolidge, 109 N.H. 403, 260 A.2d 547 (1969), *rev'd on other grounds*, 403 U.S. 443 (1971). For a discussion of this case, see *infra* notes 232-35 and accompanying text.

143. *Collins* is one of those unusual cases that generated scholarly commentary even before the California Supreme Court opinion was rendered. *See* Stoebuck, *Relevancy and the Theory of Probability*, 51 Iowa L. Rev. 849, 859-61 (1966); Kingston, *supra* note 117, at 93; Note, *Criminal Law: Mathematical Probabilities Misapplied to Circumstantial Evidence*, 50 Minn. L. Rev. 745 (1966). None of these articles attained lasting prominence. Immediately after the publication of the opinion, two articles (in addition to the two that became famous) were published, and both are quite helpful. One, Broun & Kelly, *supra* note 7, at 23, was the first attempt to survey the intersection between mathematical probability and the proof process. The second, Cullison, *supra* note 112, was a scintillating mathematical tour de force which analyzed the technical mathematical aspects of *Collins* with a sure-handedness that cannot be improved upon.

144. Finkelstein & Fairley, *supra* note 23.

and concluded that the California Supreme Court had been correct in its decision,[145] but it was the second portion of the article which caught the fancy of legal scholars. There, Finkelstein and Fairley advocated a "modest" use of Bayes' Theorem to apprise the jury of the proper weight to be accorded to trace evidence left by a defendant at a crime scene. They argued that where trace evidence such as palm prints, blood, hair, etc. is found at a crime scene, and it is established that the trace matches a trace that could have been left by the defendant, but not by a very large percentage of persons in the general populace, the jury should be presented with a chart illustrating how the small probability of a random person having left the trace at the scene should affect the jury's assessment of the defendant's guilt, formed on the basis of non-trace evidence.[146] This argument became famous for two reasons. First, it was in the vanguard in relating Bayes' Theorem to the trial process[147] and was the first to take a position in favor of using it.[148] Second, it provoked the other famous article that was spawned by *Collins*: Lawrence Tribe's "Trial By Mathematics: Precision and Ritual In the Legal Process."[149] In fact, Tribe's article became even more prominent and influential than Finkelstein and Fairley's piece. Tribe's article has been characterized by scholars as "brilliant and enormously helpful,"[150] "justly celebrated,"[151] and so powerful that it "almost shattered"[152] the movement to advocate the use of probabilistic proof at trial. Courts have been even more taken with Tribe's article than scholars: while telling salvos have been launched at Tribe's arguments by scholars, and numerous other works of scholarship have been published regarding mathematical evidence in the proof process during the ensuing two decades, Tribe's article continues to be virtually the only scholarly piece ever cited by courts when they deal with mathematical issues.[153]

Tribe's article was as skeptical about the advisability of the use of Bayes' Theorem in the proof process as Finkelstein and Fairley's article was hopeful. Many of Tribe's numerous objections to the use of Bayes' Theorem at trial were easily generalizable to *all* probabilistic modes of proof, whether they involved Bayes' Theorem or not. Ironically, while Tribe might thus be viewed

145. *Id.* at 489-96.

146. *Id.* at 500-02.

147. The concept had been broached earlier in two articles, Kaplan, *supra* note 41, at 1084-91, and Ball, *supra* note 28, at 807.

148. Ironically, *Collins*, which served to spark the Bayesian debate, did not involve any use of Bayes' Theorem.

149. Tribe, *supra* note 31.

150. J. WIGMORE, *supra* note 13, at 37.6.

151. Lempert, *supra* note 49, at 441-42.

152. Tillers, *supra* note 26, at 884.

153. *Collins* and the Finkelstein and Fairley/Tribe debate directly spawned some additional scholarship of lesser significance, including Finkelstein and Fairley's rebuttal, *A Comment on "Trial by Mathematics,"* 84 HARV. L. REV. 1801 (1971), and Tribe's surrebuttal, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810 (1971). *See also, e.g.*, Charrow & Smith, *supra* note 112, at 669; Fairley & Mosteller, *supra* note 112, at 242.

as an anti-probabilist, in fact he tacitly acknowledged that the proof process could be correctly described in probabilistic terms, even though he believed that jurors would only be confused by injecting probabilities. Tribe did, though, vigorously oppose the utilitarian strand of probabilism, arguing that it exacted too great a cost in terms of forcing acknowledgement by the legal system that it is willing to sacrifice an occasional innocent defendant.[154]

But while the force of *Collins* and Tribe's article seemingly deterred prosecutors from using mathematical techniques in criminal cases, the use of statistics and probabilities in civil cases was growing apace. In particular, litigation under anti-discrimination statutes such as the Equal Pay Act of 1963,[155] the Civil Rights Act of 1964,[156] and the Voting Rights Act of 1965[157] typically involved statistical and probabilistic evidence.[158] Further, in the mid-1970s both evidence scholars and non-legal scholars interested in the concept of inference and proof began to focus on the proof process. This interest produced three important books in the mid-1970s, all of which dealt with the intersection between probabilities and the proof process. The first of these, Professor Glenn Shafer's *A Mathematical Theory Of Evidence*,[159] in 1976, contended that the proof process could be modeled in mathematical terms, but that the appropriate terms were *not* the traditional probabilistic tools (the frequency theory, the subjective theory, the product rule for independent events, and Bayes' Theorem). Shafer's proposed alternative system is rich and complex, but his basic complaint with traditional probabilistic tools as applied to the proof process is the "additivity property" that undergirds those tools, *i.e.*, that the probability of something happening and the probability of that same thing not happening must equal one.[160] Shafer contended that a more flexible system is necessary for expressing beliefs in light of evidence. For example, the system may wish to distinguish between both "disbelief" and "lack of belief," and between "disproof" and "lack of proof."[161] While it is difficult to pigeonhole Shafer in either the probabilist or anti-probabilist camp (since his work does not directly address the tenets over which those camps do battle), it seems fair to say that his position tends toward the anti-probabilist side.

154. Tribe, *supra* note 31, at 1358 *et seq.*

155. 29 U.S.C. 206(d) (1988).

156. 42 U.S.C. 2000e-2000e-17 (1988).

157. 42 U.S.C. 1971, 1973-1973bb(a) (1988).

158. For a discussion of the influence of litigation under these statutes on the development of the intersection between mathematical probabilities and the law, see Cohen, *supra* note 40, at 387.

159. While at first glance the title of Shafer's book would lead one to believe that it was written from a legal perspective, in fact Shafer is a statistician, not a lawyer, and the book is not primarily concerned with the legal concept of proof, but with concept as it is applied by statisticians. Nonetheless, Shafer's work is taken seriously by evidence scholars, as is evidenced by Shafer's participation in the Boston University Symposium. *See* Shafer, *supra* note 37.

160. For an excellent overview of Shafer's system, see Schum, *supra* note 26, at 847.

161. *Id.*

The second influential book concerning proof processes published in the mid-1970s was L. Jonathan Cohen's *The Probable And The Provable*, in 1977. Like Shafer, Cohen was concerned about the "additivity property," and found other deficiencies in the traditional probabilistic system (Cohen calls this system the "Pascalian" system) when it is used to describe the legal proof process. He compiled a list of six "anomalies and paradoxes" which he claimed would inevitably emerge from any attempt to explain the process of judicial proof in "Pascalian" terms.[162] Cohen proposed an alternative system of "inductive" probability which he traced back to Sir Francis Bacon, and thus termed "Baconian."[163] Cohen's system, although complicated, in essence stands foursquare in favor of the anti-probabilist tenet that the burden of persuasion can only be met through the process of eliminating alternative explanations inconsistent with guilt. Cohen's work has been the lightning rod which has attracted much of the scholarship in the area.[164]

As a counterpoint to Shafer and Cohen, the third influential book published in the mid-1970s concerning the proof process was vigorous in its support of traditional probabilistic concepts in the proof process. This book was Michael Finkelstein's *Quantitative Methods In Law: Studies In The Application Of Mathematical Probability And Statistics To Legal Problems*, published in 1978. Finkelstein's book provoked two book reviews which have become well known, one of which was quite critical of his approach,[165] and the other of which was more complimentary.[166]

The year 1978 heralded more than the publication of Finkelstein's book: after a ten year drought following *Collins*, appellate opinions in criminal cases involving probabilistic modes of proof began to appear again. Increases in scientific knowledge concerning blood and hair characteristics during the

---

162. L. COHEN, THE PROBABLE AND THE PROVABLE 49-120 (1977). For a good overview of Cohen's system, see Schum, *supra* note 26, at 853.

163. L. COHEN, *supra* note 162, at 124-41.

164. Cohen first proposed his ideas in THE PROBABLE AND THE PROVABLE in 1977. Some of his subsequent defenses include: Cohen, *supra* note 44, at 635; Cohen, *supra* note 56, at 627; Cohen, *supra* note 37, at 91. The attacks on Cohen are virtually too many to list, although representative ones include: R. EGGLESTON, *supra* note 112, at 34-49; Fienberg, *Misunderstanding, Beyond a Reasonable Doubt*, 66 B.U.L. REV. 651 (1986); Kaye, *supra* note 49, at 101; Williams, *supra* note 28, at 297; Williams, *The Mathematics Of Proof—II*, 1979 CRIM. L. REV. 340. In fact, at times the debate has become quite scathing. Consider this exchange between two civilized Englishmen. Professor Williams, in *The Mathematics of Proof—II*, *supra*, at 354, states in a review of Cohen's book that: "In almost all other respects, so far as they are of legal interest, he has arrived at wrong conclusions (or, at best, of right conclusions adulterated by misleading elements), on account of what seem to me to be palpable errors of reasoning." Professor Cohen responded, "In sum, I should have expected worthier and more pointed objections from Professor Williams." Cohen, *supra* note 37, at 103. Professor Williams rejoined, characterizing one of Cohen's footnotes as "rather disgraceful", chastising Cohen for a lack of "grace" to admit his mistakes, and concluding "it is difficult to avoid the feeling that Mr. Cohen is immune against criticism because he is determined never to admit an error." Williams, *Rejoinder*, *supra* note 28, at 106-07.

165. Brilmayer & Kornhauser, *supra* note 16.

166. Kaye, Book Review, *Naked Statistical Evidence*, 89 YALE L.J. 601 (1980).

decade following *Collins* emboldened prosecutors to use mathematics to demonstrate that the defendant was one of a very small percentage of the population who could have left an incriminating trace at the scene. With regard to blood, by the mid-1970s scientists had amassed data showing that many blood factors existed besides the traditional ABO typing system factors, and that these new factors were mutually independent. With respect to hair identification, while comparison evidence that a hair from a known source "was consistent with" a hair from an unknown source had been admitted since at least 1882,[167] by the mid-1970s statistical data had been developed upon which experts were willing to opine regarding the likelihood of hairs from different individuals being consistent with each other.[168] Statistics derived from blood trace evidence and hair comparisons began to appear in appellate opinions in 1978 and have continued to appear ever since.[169] Scholarly literature addressed specifically to these kinds of evidence soon followed.[170] Prosecutors on occasion used more idiosyncratic kinds of trace evidence statistics as well, such as those relating to bitemarks and fiber matches. Sometimes prosecutors offered empirical statistics merely for their inclusive/exclusive effect, and on other occasions to calculate the probability of a random match. Often defendants argued that the statistical evidence was implicitly used for the still further purpose of calculating a subjective probability regarding the defendant's guilt. Prosecutors did in fact *explicitly* use such subjective probabilities in the area of sex crimes where a child was born allegedly as a result of sex acts with which the defendant was charged, and blood factor evidence was used to calculate the defendant's "probability

---

167. Knoll v. State, 55 Wis. 249, 12 N.W. 369 (1882), is apparently the first case at the appellate level.

168. Gaudette & Keeping, *An Attempt at Determining Probabilities in Human Scalp Hair Comparison*, 19 J. FORENSIC SCI. 599 (1974); Gaudette, *Probabilities and Human Pubic Hair Comparisons*, 21 J. FORENSIC SCI. 514 (1976).

169. The first cases involving blood factors are *State v. Carlson*, 267 N.W.2d 170 (Minn. 1978) (holding the testimony inadmissible as unfairly prejudicial); and *State v. Rolls*, 389 A.2d 824 (Me. 1978) (holding the testimony properly admitted). The first hair comparison case involving the Royal Canadian Mounted Police probability figures is the Minnesota case just referred to, State v. Carlson, 267 N.W.2d 170 (Minn. 1978) (holding the testimony improperly admitted because it was unfairly prejudicial). Also in 1978 a probability based upon bite mark comparisons was held to have been properly admitted in State v. Garrison, 120 Ariz. 255, 585 P.2d 563 (1978). For a further discussion of this case, see *infra* notes 236-39 and accompanying text.

170. *See* Barnett & Ogle, *Probabilities and Human Hair Comparison*, 27 J. FORENSIC SCI. 272 (1982); Imwinkelried, *Forensic Hair Analysis: The Case Against the Under-Employment of Scientific Evidence*, 39 WASH. & LEE L. REV. 41 (1982); Jonakait, *Will Blood Tell? Genetic Markers in Criminal Cases*, 31 EMORY L.J. 833 (1982); Jonakait, *When Blood is Their Argument: Probabilities in Criminal Cases, Genetic Markers, and, Once Again, Bayes' Theorem*, 1983 U. ILL. L. REV. 369 [hereinafter Jonakait, *Probabilities*]; Miller, *Procedural Bias in Forensic Science Examinations of Human Hair*, 11 L. & HUM. BEHAV. 157 (1987); Note, *The Admissibility of Electrophoretic Methods of Genetic Marker Bloodstain Typing Under the Frye Standard*, 11 OKLA. CITY U.L. REV. 773 (1986); Note, *Splitting Hairs in Criminal Trials: Admissibility of Hair Comparison Probability Estimates*, 1984 ARIZ. ST. L.J. 521.

of paternity."[171] This use of blood factor evidence has likewise provoked scholarly literature.[172]

In 1980 social scientists made their entry into the debate concerning mathematics and the proof process. Two social scientists, Michael Saks and Robert Kidd, examined many studies which showed systematic irrational biases in human decisionmaking and concluded that the more structured and rational approach to decisionmaking provided by traditional probabilistic concepts could result in higher quality legal decisions.[173] These researchers

---

171. *See, e.g.*, People v. Alzoubi, 133 Ill. App. 3d 806, 89 Ill. Dec. 202, 479 N.E.2d 1208 (1985); Davis v. State, 476 N.E.2d 127 (Ind. Ct. App. 1985).

172. Ellman & Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. REV. 1131 (1979); Peterson, *A Few Things You Should Know About Paternity Tests (But Were Afraid to Ask)*, 22 SANTA CLARA L. REV. 667 (1982); Reisner & Bolk, *A Layman's Guide to the Use of Blood Group Analysis in Paternity Testing*, 20 J. FAM. L. 657 (1982); Shafer, *supra* note 37, at 810-814.

173. Saks & Kidd, *supra* note 5. Saks and Kidd did no research themselves, but they provided a thorough review of the social science literature then available regarding human decision making (principally that based on research of Amos Tversky and Daniel Kahneman). Saks and Kidd then cataloged some major "heuristic [simplifying] biases", all of which speak directly or indirectly to the usefulness of probabilistic evidence. The first heuristic—"representativeness"—arises when people attempt to connect two events by assessing the degree of similarity between them. While for many purposes this is a useful strategy (when probability is highly correlated with similarity), in many other cases it may lead to defects in judgment for any of three reasons. First, people seem to be oblivious to the important role played by the "prior probability" or "base rate" of the events occurring. *Id.* at 133. Second, people exhibit an "insensitivity to sample size," failing to recognize that larger samples are more likely than smaller samples to approximate the characteristics of the population from which they were drawn. *Id.* at 134. Third, the "illusion of validity" describes the tendency of people to make intuitive predictions by selecting the outcome that is more similar to their stereotype. For example, given a brief personality description, people rely on their stereotypes and go from the description, however meager, to the prediction. They do this even when informed ahead of time about this propensity. *Id.* at 135.

A second heuristic explained by Saks and Kidd is entitled "availability." This refers to the likelihood that people will judge the probability or frequency of an event based upon the ease with which they can recall instances or occurrences of similar events. One effect of this is that experiences which are bizarre or extreme are more likely to be remembered than commonplace events. *Id.* at 137, 139. Another effect is that expert witnesses who present statistical data seem to have less impact on the jury than does a person who reports a case study, relates a compelling personal experience, or offers anecdotal evidence. *Id.* at 137. Another availability heuristic is called "illusory correlation." Here a person's estimates of two events occurring together departs systematically from the evidence they actually experience. *Id.* at 139-40. People perceive stereotypical correlations even though there is no evidence present for them, and this prevents them from detecting relationships that are actually present.

A third heuristic pointed out by Saks and Kidd is entitled "adjustment and anchoring." This heuristic simply demonstrates that when making certain types of judgments that require revision, the revised result depends heavily upon the initial judgment and thus different initial values often lead to quite different final estimates. Another aspect of "adjustment and anchoring" are the "biases in the evaluation of conjunctive and disjunctive events." *Id.* at 140-41. Here, people tend to overestimate the probability of the occurrence of conjunctive events and to underestimate the probability of disjunctive events. *Id.* at 142.

A fourth heuristic is the tendency of people to undervalue probability data. Saks and Kidd did not even spare the sacred citadel erected by Professor Tribe in his "Trial by Mathematics,"[174] asserting that it contained "a Swiss cheese of assumptions about human behavior—in this case human decision-making processes—which are asserted as true simply because they fall within the wide reach of the merely plausible, not because any evidence is adduced on their behalf."[175]

The combination of the interest generated by the three books published in the mid-1970s, the resurgence of appellate cases involving mathematical issues, and the entrance by social scientists into the debate spurred continuing activity by legal scholars.[176] This scholarly interest culminated in a symposium at Boston University Law School in 1986 which spawned a book-length issue of the Boston University Law Review containing contributions from most of the leading scholars concerned with mathematics and the proof process.[177] Even this symposium, however, did not result in the final word on these issues, as is evidenced by continuing scholarship in the field.[178] Moreover,

---

assert:

> Research demonstrates, however, that people do not process probabilistic information well, that in the face of particularistic information, they cannot integrate the statistical and anecdotal evidence and consequently tend to ignore the *statistical* information. Intuitive, heuristic, human decision makers must dispense with certain information, and that tends strongly to be the quantitative information.

Id. at 149.

Having pointed out these failings of human decisionmaking, Saks and Kidd conclude that the legal system, like many other endeavors, should make use of mathematical tools as decision aids:

> It has been well established for some time now that when the same information is available to intuitive humans or a good mathematical model, the human's decisions are consistently less accurate. . . Even when mathematical tools are modeled after human decision processes, the copy works better than the original. . . The mathematical model of a person's own decision policies is more accurate than the person because it consistently applies the same logic, while the human decision maker fluctuates, being over-influenced by fortuitous, attention-catching pieces of information that vary from time to time, and processing a too-limited set of variables.

Id. at 146-47. Thus, they conclude that, "[E]xperts ought to be permitted to offer their data, their algorithms, and their Bayesian theorems," and more information regarding prior probability should be admitted. *Id.* at 148.

174. *See supra* note 31.

175. *See* Saks & Kidd, *supra* note 5, at 125.

176. This scholarship includes, but is not limited to, Brook, *supra* note 39; Braun, *supra* note 7; Brilmayer & Kornhauser, *supra* note 16; Callen, *supra* note 29; Cohen, *supra* note 37; Cohen, *supra* note 40; Eggleston, *supra* note 40; Jaffee, *supra* note 7; Kaye, *supra* note 32; Kaye, *supra* note 166; Kaye, *supra* note 31; Kaye, *supra* note 49; Nesson, *supra* note 46; Tyree, *Probability Theory and the Law of Evidence*, 1984 CRIM. L.J. 224; Williams, *supra* note 28; Williams, *supra* note 164; Williams, *Rejoinder*, *supra* note 28.

177. 66 B.U.L. REV. 377 (1986).

178. *See* Birmingham & Dunham, *An Evidentiary Value Reading of Naked Statistical Proofs*, 31 ST. LOUIS U.L.J. 797 (1987); Jaffee, *supra* note 26; Kaye, *supra* note 13; Kaye, *supra* note 7; Shaviro, *supra* note 39; Tillers & Schum, *Charting New Territory in Judicial Proof: Beyond Wigmore*, 9 CARDOZO L. REV. 907 (1988); Wright, *supra* note 36. A follow-up conference to the one held at Boston University in 1986 is set for March 24-26, 1991, at the Cardozo School of Law. The title of the conference is "Decision and Inference in Litigation."

the findings of three significant social science research projects concerning how factfinders deal with mathematical proof were published in the late 1980s.[179] Meanwhile, by the mid-1980s prosecutors had become emboldened to try more novel and *Collins*-like uses of probabilistic modes of proof. A limited number of such attempts have been reported in appellate opinions, but they are significant to the extent that they show that the urge still beats within prosecutors' breasts to prove guilt through the creative use of probabilities.[180] Two decades after *Collins*, then, there is a rich and complex mixture of elements that has emerged at the intersection of mathematics and the proof process. The last part of this article attempts to analyze the current status of the five categories of mathematical evidence through an exposition regarding the legal scholarship, case law, and social scientific research.

IV. Analyzing the Issues Concerning the Five Categories of Mathematical Evidence

A. *An Overview Of Categories and Their Interrelationships*

The five categories of mathematical evidence are: (1) empirical statistics; (2) probabilities of a random match (the classical theory analogue of category one); (3) non-empirical probabilities of guilt incorporating empirical statistics without Bayes' Theorem; (4) non-empirical probabilities of guilt developed without empirical statistics and without Bayes' Theorem; and (5) non-empirical probabilities of guilt incorporating empirical statistics via Bayes' Theorem. One obviously key distinction that separates category four from the others is that it is the only category that does not utilize empirical statistics. The line separating empirical statistics from non-empirical data is sometimes blurry, however. Some statistics, like bodily fluid "markers," are obviously empirical because they are based on extensive empirical data. Some data, like that used in *People v. Collins*, is obviously non-empirical. The line tends to blur in the middle, however, because the skimpier the empirical basis for the statistic, the less empirical the data looks. The rule of thumb applied here for categorizing these issues is to place into the empirical category those statistics with some plausible basis in empirical testing, and to place into the non-empirical category those which are clearly and simply made up (prosecutorial protests to the contrary notwithstanding). Once we

---

Participants include many of the persons whose work is cited in this article, including Ronald Allen, L. Jonathan Cohen, Michael Finkelstein, David Kaye, Richard Lempert, David Schum, Peter Tillers, Charles Nesson, Richard Friedman, Glenn Shafer, Lewis Kornhauser, and Lea Brilmayer.

179. *See* Faigman & Baglioni, *Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 L. & Hum. Behav. 1 (1988); Thompson & Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 L. & Hum. Behav. 167 (1987); Goodman, *Probabilistic Scientific Evidence: Jurors' Inferences* 1986 (Doctoral Dissertation at the University of Washington, available through UMI Dissertation Information Service).

180. For discussion of these cases, see *infra* notes 232-78 and accompanying text.

conclude that a statistic is empirical, it is easy to tell whether it falls into category one or another category, since category one is the only category where the statistic is not used to form a probability. Similarly, category five is easy to recognize because of the presence of Bayes' Theorem. Categories two and three, however, provide another difficult question of categorization. Both use probabilities formed from empirical statistics and neither uses Bayes' Theorem. Further, when prosecutors use empirical statistics to form probabilities they almost always argue that they are forming the probability of a random match rather than a probability of guilt, undoubtedly because courts are antipathetic towards probabilizing the "beyond a reasonable doubt" standard. The rule of thumb that will be used herein to distribute cases between categories two and three is as follows: If the probability of a random match is such that there are a fair number of suspects other than the defendant who have not been eliminated, the case will be discussed in category two. If, on the other hand, the probability of a random match is such that there is virtually no chance that any person other than the defendant in the suspect population would match, then the case will be discussed in category three.[181]

Significantly, the great majority of scholarly writings regarding mathematics in the proof process focus on the theoretical issues concerning the probabilist and anti-probabilist positions—particularly on hypotheticals that raise the issue of the sufficiency of naked statistics to support a verdict, and the application of Bayes' Theorem to the proof process—while virtually ignoring category one—empirical statistics not used probabilistically. But category one usages in fact constitute the bulk of the case law. Thus, to a great extent the interaction between the scholarship and case law surrounding *Collins* are akin to the proverbial two ships passing in the night. One final note is in order: the cases reviewed will be limited almost exclusively to those that cite *Collins*. While other appellate opinions exist which address mathematical issues, the cases citing *Collins* provide a fair cross-section of the case law, and include virtually all of the more exotic uses of mathematical evidence.

---

181. Two social science researchers recently noted, "No studies have systematically investigated individuals' differential use of varying degrees of statistical information." Faigman & Baglioni, *supra* note 179, at 14. They went on to hypothesize, "Typically, researchers have used modest ratios such as 80/20 or 70/30. It may be that individuals process probability figures by giving some weight to extreme figures and little or no weight to modest figures, but do not discriminate between the two in any refined manner." *Id.* These researchers were apparently unaware of the doctoral thesis research conducted by Goodman, *supra* note 179, who *did* systematically test reactions to differing levels of statistics. Goodman presented subjects with the same nonmathematical evidence, but varied the inclusive/eliminative hard statistic among subjects at four levels: 10%, 5%, 1%, and one-tenth of 1%. *Id.* at 103-04. Her findings at least partially supported Faigman and Baglioni's hypothesis: the subjects seemed not to differentiate regarding the weight to be assigned to the various levels. *Id.* at 133-35. Even Goodman's one-tenth of 1% figure does not seem "extreme" however, since in a suspect population of 100,000 (which is what she used), it still leaves 100 suspects. Thus, Faigman and Baglioni's hunch, shared by some courts, that jurors can be bowled over by extreme probabilities, remains an unproven intuition.

B. Category One: Empirical Statistics

This category consists of evidence based on empirically derived statistics which are not used to form probabilities, but merely to show what portion of the population is excluded as the culprit while the defendant is not. In practice there is only one kind of evidence that prosecutors have used in this manner: bodily fluid "marker" evidence. Such evidence usually eliminates upwards of ninety-eight percent of the subject population. But assuming a sizeable population in the area where the crime occurred, such evidence leaves hundreds or even thousands of suspects other than the defendant. This evidence raises three questions which will be discussed separately below: (1) are the statistics valid; (2) are they relevant; (3) are they unfairly prejudicial?

As to validity, for statistics to have any inclusive/eliminative effect they must, of course, be valid statistics to begin with. "Validity" here means that the statistics were obtained through an appropriate sampling technique and that enough empirical data was obtained so that the generalizability of the occurrence of the characteristic at issue over the entire population at issue seems reasonable, *i.e.*, that the there is a good probability that the rate of occurrence within the sampled portion of the population continues at the same proportion through the unsampled portion of the population. Further, if two statistics are to be multiplied via the product rule to obtain a new statistic regarding the coincidence of these two statistics, then there must be a valid showing that the two factors are mutually independent.

Until relatively recently the only "marker" in human blood and semen that prosecutors found helpful as evidence was the type of fluid found in the ABO system. Such evidence has been common in criminal cases for decades, but has only slight probative value because each of the four types under that system is possessed by so many people that a showing of a match between the defendant and the trace found at the scene (or the victim and the trace found on the defendant), while eliminating a substantial portion of the population, leaves a significant proportion of the population remaining as possible culprits. In recent decades, however, scientists have discovered an ever-increasing number of genetically controlled proteins and enzymes in human cells (primarily blood and semen cells). Further, scientists have studied various ethnic groups, both in this country and other countries, and have cataloged the frequency of occurrence of many of these blood factors. They have also sought to determine whether these factors' frequencies are independent of each other in their occurrence, and have concluded that with rare exceptions they are mutually independent.[182] These scientific advances were quickly embraced by prosecutors as mechanisms by which to dramati-

182. Jonakait, *Probabilities, supra* note 170, at 375-77. However, studies have so far been limited to tests for "pairwise independence." Three or more factors can be independent when considered in pairs, yet still not be independent when considered in other combinations. Letter from Professor Kaye to David McCord (May 15, 1990).

cally narrow the population of possible suspects, sometimes to less than one percent of the population of a particular ethnic group.

Defendants have challenged the scientific validity of such evidence. In some jurisdictions which apply the *Frye*[183] general acceptance standard the evidence initially encountered some difficulties, primarily regarding whether the scientific test (usually electrophoresis) used to test for genetic markers was sufficiently reliable.[184] Perhaps surprisingly, no court seems to have undertaken a serious analysis whether the empirical database concerning the occurrence and mutual independence of the particular factors were the result of a valid and sufficient sampling procedure. Numerous courts have held, without seemingly much inquiry into the validity of this database, that there is a good foundation for the statistical rates of occurrence of blood factors.[185] Many others assume a good foundation for the figures without even discussing the issue.[186] Even those courts that hold such evidence inadmissible for unfair prejudice acknowledge the good foundation for use of these statistics.[187] Some scholars have been less willing than most courts to accept the validity of the empirical data underlying blood factor evidence. In one of the first academic forays into the topic, Professor Jonakait expressed substantial skepticism about the validity of the database even though he came to the conclusion that the database was good enough:

> Available research thus indicates that population frequency figures cannot be treated as precise and that great care must be taken in the interpretation of the products using the frequencies. This is especially true when only a handful of markers is identified, or if one of the markers is a rare phenotype. Nevertheless, data now exist which form a reasonable basis on which to ascribe genetic marker frequencies. The situation, therefore, is not like *Collins*. The probabilities are not fanciful creations; instead they have a foundation in the real world and may be verified scientifically. Because blood genetic markers are statistically independent from each other, the

183. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

184. *See, e.g.*, People v. Reilly, 196 Cal. App. 3d 1127, 242 Cal. Rptr. 496 (1987); People v. Brown, 40 Cal. 3d 512, 709 P.2d 440, 220 Cal. Rptr. 637 (1985) (but the evidence was found to pass the *Frye* test two years later); People v. Harbold, 124 Ill. App. 3d 363, 464 N.E.2d 734 (1984); People v. Young, 418 Mich. 1, 340 N.W.2d 805 (1983).

185. *See, e.g.*, People v. Yorba, 209 Cal. App. 3d 1017, 257 Cal. Rptr. 641 (1989); Martinez v. State, 549 So. 2d 694 (Fla. Dist. Ct. App. 1989); State v. Stukey, 242 Kan. 204, 747 P.2d 137 (1987); Commonwealth v. Gomes, 403 Mass. 258, 526 N.E.2d 1270 (1988); State v. Woodall, 385 S.E.2d 253 (W.Va. 1989).

186. *See, e.g.*, United States v. Gwaltney, 790 F.2d 1378 (9th Cir. 1986), *cert. denied*, 479 U.S. 1104 (1987); People v. Poggi, 45 Cal. 3d 306, 753 P.2d 1082, 246 Cal. Rptr. 886 (1988), *cert. denied*, 109 S. Ct. 3261 (1989); People v. Morris, 199 Cal. App. 3d 377, 245 Cal. Rptr. 52 (1988); Graham v. State, 168 Ga. App. 23, 308 S.E.2d 413 (1983); People v. Prewitt, 160 Ill. App. 3d 942, 515 N.E.2d 977 (1987); State v. Rolls, 389 A.2d 824 (Me. 1978).

187. People v. Harbold, 124 Ill. App. 3d 363, 464 N.E.2d 734 (1984); State v. Carlson, 267 N.W.2d 170 (Minn. 1978).

> criticisms about the use of the product rule as in *Collins* do not apply.[188]

Professor Jaffee goes much further in attacking the validity of the empirical database:

> Genetic marker statistics, theories, and methods have been developed around small, often unrandom and unsystematically obtained population samples, the more colorably valid studies having involved only a few northwest European "Caucasian" populations. And this will continue so, unless our privacies are to be severely invaded. Therefore, we must question the genetic frequency figure's reliability across various discrete populations. We do not know whether there is a fair match between the population relevant to any American case and those sampled. In fact, we can't ever *know* the relevant population.
>
> * * *
>
> The HLA data available respecting American Blacks are far fewer hence far less reliable. Many American Blacks are of "mixed stock," and in mixed-stock cases, genetic marker identification and frequency are difficult even to *estimate*.[189]

Despite these scholarly reservations, courts seem to be moving inexorably in the direction of accepting the underlying statistical data's validity without question.

As to relevance, prosecutors have a double-barreled argument regarding such evidence. First, and most obviously, it places the defendant in a relatively small class of persons who are not excluded from guilt. Second and less obviously, the prosecution counts on this evidence reinforcing other evidence in the case under the theory that the more pieces of evidence that point toward the defendant, the less likely it is that the congruence is coincidental. This is really a quite traditional usage of evidence which has simply been made more precise by new technology and information. The same theory underlies the admission of evidence in earlier times regarding ABO blood type factors, and, analogously, testimony that the defendant's hair is not inconsistent with a hair found at the scene; indeed, even non-scientific testimony, such as that the defendant drove a white car as did the culprit, or that the defendant was a Caucasian as was the culprit, is traditionally used to build the case brick-by-brick. Not surprisingly, then, bodily fluid "marker" evidence has never been held by a court to be irrelevant (although one court has held its probabilistic analogue—the probability of a random match—to be irrelevant).[190] Several courts have explicitly found such

188. Jonakait, *Probabilities*, *supra* note 170, at 381.

189. Jaffee, *supra* note 26, at 1025-26, 1049 (citations omitted).

190. *See infra* note 200 and accompanying text.

evidence to be relevant.[191] More often, courts have found the evidence to have been properly admitted without even discussing the question of relevance, which indicates that either the defendant did not raise the relevance objection or that the court was willing to assume relevance.[192] A further discussion of relevance is undertaken below in connection with the same issue in category two.

As to unfair prejudice, the defense argument is that such evidence "will be understood by the jury to be a quantification of the likelihood that the defendant, who shares that unique combination of blood characteristics, is guilty."[193] This language is derived from the only case to have held such evidence to be unfairly prejudicial (although, again, several other courts have held analogous probabilistic evidence in category two to be unfairly prejudicial).[194] On the other hand, most courts have found such evidence not to be unfairly prejudicial because the jury can understand the distinction between statistics which exclude a portion of the population, and a probability that the defendant is the guilty party out of the non-excluded portion of the population.[195] A discussion of the merits of the debate regarding unfair prejudice will be undertaken in connection with the closely related issue in category two.[196]

C. *Category Two: Probabilities of a Random Match—The Classical Probability Theory Analogue of Category One*

Using an inclusive/eliminative hard statistic and adopting an assumption that a random choice from a suspect population is equally likely to turn up any individual, the classical theory of probability allows us to form the probability of random match, that is, the probability that if we select a person at random from the population, that person will have characteristics that match the trace evidence connected with the crime. Thus, if we know statistically that only two percent of the suspect population has certain bodily

---

191. *See, e.g.* People v. Poggi, 45 Cal. 3d 306, 753 P.2d 1082, 246 Cal. Rptr. 886 (1988), *cert. denied*, 109 S. Ct. 3261 (1989); People v. Morris, 199 Cal. App. 3d 377, 245 Cal. Rptr. 52 (1988); Commonwealth v. Gomes, 403 Mass. 258, 526 N.E.2d 1270 (1988).

192. *See, e.g.*, United States v. Gwaltney, 790 F.2d 1378 (9th Cir. 1986), *cert. denied*, 479 U.S. 1104 (1987); People v. Yorba, 209 Cal. App. 3d 1017, 257 Cal. Rptr. 641 (1989); Graham v. State, 168 Ga. App. 23, 308 S.E.2d 413 (1983); People v. Redman, 135 Ill. App. 3d 534, 481 N.E.2d 1272 (1985); Davis v. State, 476 N.E.2d 127 (Ind. Ct. App. 1985); State v. Stukey, 242 Kan. 204, 747 P.2d 137 (1987); State v. Thompson, 503 A.2d 689 (Me. 1986); State v. Woodall, 385 S.E.2d 253 (W.Va. 1989); State v. Hartman, 145 Wis. 2d 1, 426 N.W.2d 320 (1988).

193. State v. Joon Kyu Kim, 398 N.W.2d 544, 548 (Minn. 1987). *See also* Tribe, *supra* note 31, at 1355.

194. *See infra* notes 199 and 209-10 and accompanying text.

195. *See, e.g.*, United States v. Gwaltney, 790 F.2d 1378 (9th Cir. 1986), *cert. denied*, 479 U.S. 1104; People v. Redman, 135 Ill. App. 3d 534, 481 N.E.2d 1272 (1985); People v. Alzoubi, 133 Ill. App. 3d 806, 479 N.E.2d 1208 (1985); Davis v. State, 476 N.E.2d 127 (Ind. Ct. App. 1985); State v. Woodall, 385 S.E.2d 253 (W.Va. 1989).

196. *See infra* notes 215-30 and accompanying text.

fluid "markers," we can state that the probability of randomly selecting a person from the suspect population with those bodily fluid "markers" is one-fiftieth. This second category consists of probabilities formed in such a manner (with the caveat that the probabilities must be such as to leave a fair number of suspects other than the defendant in the suspect population— if they do not, then they fall into category three). Two kinds of evidence— bodily fluid "markers" and hair comparisons—are found in this category. Each type of evidence raises issues of validity, relevance and unfair prejudice.[197]

With respect to bodily fluid "markers," most of the cases in this category involve sex offense prosecutions where the alleged victim became pregnant and delivered a child as a result of the allegedly illegal sexual activity. The prosecutor uses markers in the blood of the child and compares them with markers from the mother and the alleged father to calculate one or both of two probabilities. One probability is the probability of a random match calculation (known in the paternity area as the "probability of exclusion") which states what portion of the male population could have contributed genetic material containing the requisite factors. The second probability calculation is called the "paternity index" and is based on the fact that within a group of men genetically capable of fathering a child with certain genetic factors, some will be more genetically likely to have done so than others.[198] This probability cannot be calculated as to trace evidence left at the scene, since within the group of persons with such "markers," each is assumed equally likely to have left the trace. The paternity index, then, is a ratio that compares the alleged father's likelihood of producing the child's phenotypes with the likelihood of a randomly selected man doing so. Thus, the "paternity index" is more discriminating than the "probability of exclusion."

Courts have not questioned the validity of the underlying statistics in these paternity cases; indeed, the statistics are based on the same data as empirical statistics in category one. Of the few courts that have considered

---

197. Recall the dual import of inclusive/eliminative hard statistics—their inclusive/eliminative effect, and their tendency to show that the evidence in the case is not merely pointing coincidentally toward the defendant. When a prosecutor turns the hard statistics into the probability of a random match, the prosecutor is seeking to highlight the "not just a coincidence" effect of the evidence even further.

198. *See* Plemel v. Walter, 303 Or. 262, 269, 735 P.2d 1209, 1213-14 (1987) where the following example is given:

> For example, we noted above that Walter's phenotype for the ABO system was *A2B and that the child's father had to have one of the "B phenotypes". Walter could transmit to his child either his A2 or his B gene. The chance that he would transmit the B gene would be 50 percent. A man who had a B phenotype and two B genes (as opposed to a man with a B phenotype and one B gene and one O gene), however, would have a 100 percent chance of transmitting a B gene. Other things being equal, this man would be more likely to be the father of Plemel's child than would be Walter.*

Id. at 269, 735 P.2d at 1214.

the "probability of exclusion", all but one have found it to be relevant and not unfairly prejudicial.[199] One other case involving a blood trace not in the context of a paternity issue also found a probability of a random match to be unfairly prejudicial due to the confusion that might exist for the jury between that probability and the probability of the defendant's guilt.[200]

With respect to hair comparison evidence, testimony by experts that characteristics of two specimens of hair "were consistent with" each other had been used in courts for decades prior to *Collins*. It was not until the mid-1970s, however, that forensic scientists sought to establish any statistics concerning the occurrence of common characteristics among strands of hair. The Canadian Royal Mounted Police did two studies, one for scalp hair and one for pubic hair, in which researchers selected hairs from different people and then determined the probability of a match between hairs picked at random.[201] These studies determined that there was a one in 4500 probability of a random match between the scalp hair of two different persons and a one in 800 chance regarding pubic hair. Commentators have launched several telling blows at the validity of these figures.[202] These objections include the subjective nature of hair comparison, the fact that the examiners' subjective determinations may have been affected by their knowledge that no hairs should match, that the use of characteristics that are not significant in distinguishing the hairs of different people may have biased the results of the study, that the matches between pairs of hairs were not statistically independent, and that the study population was very small and not selected at random.[203] But two objections are particularly telling. First, the probability estimate was merely an average of the probability over all cases. Unlike bodily fluid "markers," the hair comparison figures failed to reflect the relative rarity or frequency of the combination of hair characteristics exhibited by one individual compared to the rest of the population.[204] The studies recognized that some types of hair were "common and featureless" and that they occurred with relatively more frequency in the population than hairs with more singular characteristics. Thus, as to a "common and featureless"

---

199. In the majority are People v. Alzoubi, 133 Ill. App. 3d 806, 479 N.E.2d 1208 (1985); Davis v. State, 476 N.E.2d 127 (Ind. Ct. App. 1985); State v. Jackson, 320 N.C. 452, 358 S.E.2d 679 (1987); State v. Hartman, 145 Wis. 2d 1, 426 N.W.2d 320 (1988). In the minority is State v. Boyd, 331 N.W.2d 480 (Minn. 1983) (too much danger of jury confusing probability of a random match with probability of guilt). See also a civil case, Commonwealth v. Beausoleil, 397 Mass. 206, 490 N.E.2d 788 (1986) ("probability of exclusion" while possibly relevant is of very little probative value because it does not pinpoint which person within the nonexcluded group of suspects is the father).

200. People v. Harbold, 124 Ill. App. 3d 363, 464 N.E.2d 734 (1984).

201. *See* Gaudette & Keeping, *supra* note 168; Gaudette, *supra* note 168.

202. *See, e.g.*, Barnett & Ogle, *supra* note 170; Miller, *supra* note 170; Robertson, *An Appraisal of the Use of Microscopic Data in the Examination of Human Head Hair*, 22 J. FORENSIC SCI. SOC'Y 390 (1982); Note, *Splitting Hairs in Criminal Trials: Admissibility of Hair Comparison Probability Estimates*, 1984 ARIZ. ST. L.J. 521 [hereinafter Note, *Splitting Hairs*].

203. These objections are summarized in Note, *Splitting Hairs*, *supra* note 202, at 529-33.

204. *Id.* at 534-35.

hair, the one in forty-five hundred or one in eight hundred figure would far overstate the rarity of the existence of other possible culprits. Second, the studies determined the probability of a match given that hairs are not from the same person, when the issue in criminal trials is almost always the logical converse, *i.e.*, with respect to hairs that have already been shown to match, what is the probability that the hairs came from different people?[205] While these two figures can be related through Bayes' Theorem, they are not the same. To calculate the probability that would be pertinent in a criminal case, one would also have to know the probability of two hairs matching given that they come from the same person, a figure not provided by the Royal Canadian Mounted Police studies.[206] Only if the chances of a match between hairs picked at random from the same person were greater than one in 4500 or one in 800 would the evidence tend to inculpate the defendant.[207] While one suspects that there is a substantially greater probability of a match between two hairs selected at random from the same person, forensic scientists admit that even hairs from the same person can exhibit quite a range of characteristics. For all of these reasons, the Royal Canadian Mounted Police studies are fatally flawed.

Four appellate decisions have considered the admissibility of expert testimony based upon the Canadian Royal Mounted Police studies. Surprisingly, none of the opinions seriously discussed the validity of the studies, although one opinion held that expert testimony referring to those studies had been improperly admitted because the witness knew little about the studies and thus could not provide a foundation from which the trial court could have concluded that the studies were valid.[208] One opinion, in fact, found that the probabilities were "based upon empirical scientific data of unquestioned validity."[209] That court, however, along with one other,[210] held that the probabilities were improperly admitted because of their potential for unfairly prejudicing the defendant by influencing the jury to equate the probability of a random match with the probability of guilt.

With respect to the relevance of evidence in categories one and two, then, while one court has held the evidence to be irrelevant, by far the bulk of authority is in favor of admission. As far as this author can tell, no anti-probabilist scholars would argue against the relevance of empirical statistics.[211]

---

205. Note, *Admissibility of Mathematical Evidence in Criminal Trials*, 21 AM. CRIM. L. REV. 55, 65-66 (1983).

206. *Id.* at 66.

207. *Id.*

208. United States v. Massey, 594 F.2d 676 (8th Cir. 1979).

209. State v. Carlson, 267 N.W.2d 170, 176 (1978). Lest we be too harsh on this court, note that none of the critiques of Gaudette & Keeping's work cited *supra* in note 168 had yet been published when this case was decided.

210. Brown v. State, 751 P.2d 1078 (Okla. Crim. App. 1988).

211. The closest is Professor Leonard Jaffee who characterizes eliminative evidence as requiring a "double negative inference" and contends that no number of such double negative inferences accumulated can combine with one another to substitute for "actualistic" proof. *See*

The bulk of authority clearly reaches the correct result as to category one. Inclusive/eliminative evidence, including empirical statistics, have long been recognized to be relevant because they eliminate a substantial number of possible culprits without exculpating the defendant.[212] Indeed, the one holding of *People v. Collins* that has been challenged effectively concerns this very issue. Recall that the fourth basis for a reversal in *Collins* was that even assuming the statistics had been valid, "[T]he most a mathematical computation could *ever* yield would be a measure of the probability that a random couple would possess the distinctive features in question. . . . Urging that the Collinses be convicted on the basis of evidence which logically establishes no more than this seems as indefensible as arguing for the conviction of X on the ground that a witness saw either X or X's twin commit the crime."[213] Recall also that the issue before the court was the admissibility of the probabilities and the prosecutor's argument based thereon, not the sufficiency of the evidence to convict. Thus, the excerpt just quoted can either be viewed as an unnecessary excursus on sufficiency, or as an implied holding that the probability of a random match was irrelevant. If the latter interpretation is what the court meant, then the court's own illustration shows the reasoning to be incorrect. While certainly the court is correct that eyewitness testimony that either X or X's twin committed a crime would not be *sufficient* alone to convict X of the crime, such testimony would certainly be relevant since it pares the possible suspect pool from the population of the world down to two. California courts in later cases involving empirical statistics have simply ignored this aspect of *Collins*.[214]

With respect to unfair prejudice regarding evidence in categories one and two, the majority of courts have rejected this objection, while a few have found too great a danger of the jury's confusing the empirical statistics or a probability of a random match based thereon with the probability of the defendant's guilt. One recent social science study has been conducted on this exact point,[215] and a second, broader study also reached some conclusions

---

Jaffee, *supra* note 7, at 1063-64. While Jaffee's illustration in the article just cited does not deal with hard statistical data, Professor Jaffee acknowledged in a conversation with the author that he would indeed extend his analysis to include inclusive/eliminative hard statistics.

212. Indeed, if such inclusive/eliminative hard statistics are deemed to be so precise as to narrow the category of suspects to one, the issue of probability does not even arise. Fingerprints are the one traditional kind of evidence viewed in this manner. Another recent contender for this honor is "DNA fingerprinting" which claims to be able to identify patterns of genetic characteristics that are unique to each individual. Such evidence has been approved by appellate courts in Florida. *See* Martinez v. State, 549 So. 2d 694 (Fla. Dist. App. 1989). *See also* Williams, *DNA Fingerprinting: A Revolutionary Technique in Forensic Science and Its Probable Impacts on Criminal Evidentiary Law*, 37 DRAKE L. REV. 1 (1987-88).

213. People v. Collins, 68 Cal. 2d 319, 331, 438 P.2d 33, 40-41, 66 Cal. Rptr. 497, 504-05 (1968). See discussion on this point *supra* note 130.

214. *See, e.g.*, People v. Poggi, 45 Cal. 3d 306, 753 P.2d 1082, 246 Cal. Rptr. 886 (1988), *cert. denied*, 109 S. Ct. 3261 (1989); People v. Brown, 40 Cal. 3d 512, 709 P.2d 440, 220 Cal. Rptr. 637 (1985); People v. Yorba, 209 Cal. App. 3d 1017, 257 Cal. Rptr. 641 (1989); People v. Morris, 199 Cal. App. 3d 377, 245 Cal. Rptr. 52 (1988).

215. Thompson & Schumann, *supra* note 179.

on this point.[216] The results of the first study (the "Thompson and Schumann" study) are illuminating. Thompson and Schumann's first experiment involved 144 volunteers from a pool of university students. Each was initially given the same non-mathematical evidence regarding the identity of a person who robbed a liquor store, *i.e.*, that the store clerk was able to describe the robber's height, weight, and clothing, but could not see his face or hair. The police apprehended the suspect near the store who matched the clerk's description but the suspect did not have the ski mask or the stolen money. The police found those items in a trash can near where the suspect was apprehended, however. At that point, the experiment subjects were asked to make an initial estimate of the probability of the suspect's guilt. Each of the subjects was then given mathematically-phrased expert testimony concerning hair comparison, with the evidence presented in one form to part of the subjects and in another form to the other subjects. The first form was an empirical statistic, *i.e.*, that in the city of one million people, two percent have hair that would be indistinguishable from that of the defendant and thus that there were approximately 20,000 people in the suspect population with hair consistent with that found inside the ski mask that was worn by the robber during the robbery. The other half of the subject pool was given testimony stating that there is "only a two percent chance a defendant's hair would be indistinguishable from that of the perpetrator if he were innocent. . .," *i.e.*, testimony in the form of probability of a random match. After reading the forensic evidence each of the subjects made a final judgment of the probability of the suspect's guilt. The researchers found that 13.2 percent of the subjects estimated the probability of guilt to be exactly ninety-eight percent, which is the probability obtained by subtracting the probability of a random match figure from one. The researchers concluded that these subjects were victims of "the Prosecutor's Fallacy," that is, that they had been induced to equate the probability of a random match with the probability of the suspect's guilt. On the other hand, 12.5 percent of the respondents did not change their probability of guilt before the forensic evidence to a higher probability afterwards, indicating that they gave no weight whatsoever to the forensic evidence. The researchers concluded that these subjects were victims of "the Defense Attorney's Fallacy," that is, that those subjects viewed the mutually supportive aspects of the non-mathematical evidence and the mathematical evidence as merely coincidental and having little or no probative value. These persons seemingly gave great weight to the proposition that the evidence only showed that the defendant was one of 20,000 persons who could have committed the crime. The characterization by the researchers of this reasoning as a "fallacy" betrays them as probabilists. Indeed, they claim that the evidence is really highly probative because Bayesian analysis so shows. An anti-probabilist, of course, would say that the line of reasoning which gives little weight to such evidence

---

216. Goodman, *supra* note 179.

is perfectly legitimate and praiseworthy and not a "fallacy" at all.[217] Without taking sides, for purposes of convenience in discussing this research we will adopt the name—"Defense Attorney's Fallacy"—used by the researchers. The researchers classified the remaining 74.3 percent of the subjects as victims of neither fallacy, because their final judgments of guilt were higher than their initial judgments but less than ninety-eight percent.[218] The incidence of occurrence of each fallacy was directly related to whether the subject had received the testimony in the form of empirical statistics or the probability of a random match. Of the nineteen persons who fell prey to the prosecutor's fallacy, sixteen of them had read the probability of a random match form of the evidence, while only three had read the empirical statistics form. Of the eighteen people who fell prey to the defense attorney's fallacy, only six had read the probability of a random match form of the evidence, while the remaining twelve had read the empirical statistics form.[219]

In their second experiment, Thompson and Schumann gave seventy-three undergraduate subjects a description of a murder case in which the killer's identity was unknown, but the victim was known to have wounded the killer with a knife. The police found some of the killer's blood at the crime scene and tests indicated that it was a rare type found in only one person in 100. While questioning the victim's neighbors, a detective noticed that one of them was wearing a bandage. Based on his impression of this man, the detective estimated the probability of his guilt to be ten percent. Later the detective received the information that this suspect had the same rare blood type as the blood of the killer found at the crime scene. The subject's task was to decide whether the detective should revise his estimate of the probability of the suspect's guilt in light of the new evidence, and if so, by how much. Subjects were given two brief arguments to read, one embodying the Prosecutor's Fallacy and the other the Defense Attorney's Fallacy. The

---

217. Thompson & Schumann, *supra* note 179, at 171. For the anti-probabilists, here is Prof. Wright: "The fact that judges, jurors, and lay persons ignore the base rates and instead focus on the particularistic evidence is, contrary to the assertions of the Bayesians, highly rational." Wright, *supra* note 36, at 1062.

218. Thompson & Schumann, *supra* note 179, at 173.

219. Goodman, *supra* note 179, studied the impact on mock jurors of different phrasings of the mathematical evidence in another manner. Working with a pool of prospective jurors on call at the King County Superior Court in Seattle, she presented them with a transcript of an arson case where an expert testified that gas found at the scene matched gas found in the defendant's alleged accomplice's car. *Id.* at 148. In one variation, some of the subjects received the expert testimony in terms of the *odds* of a random match, *i.e.*, the odds that the gas would match by chance was 1/1000. In the other variation, the rest of the subjects received the testimony in terms of the *percentage* likelihood of a random match, i.e., the chance of a random match was one-tenth of 1%. *Id.* at 151. In other words, whereas Thompson and Schumann tested the difference between inclusive/exclusive hard statistics and the probability of a random match, Goodman tested the difference between two phrasings of the probability of a random match. Goodman's conclusions were correspondingly less dramatic: the two groups came to virtually identical estimates of guilt, with the only difference being that the odds group was slightly less confident of its judgment, presumably because odds are a less-well-known and understood concept than percentages. *Id.* at 183-84.

subjects were not given both arguments at once, however; instead, half of the subjects first received the prosecution argument and half first received the defense argument. After reading the first argument the subjects were asked three questions: first, whether they believed the logic and reasoning of the argument was correct; second, whether they thought the detective should revise his estimate of the suspect's probable guilt in light of the blood type evidence; and third, what they thought the detective's estimate of the probability of guilt should be in light of the blood type evidence. The subjects were then given the other argument and asked to answer the same three questions. As to whether each argument was correct, fifty subjects (68.5 percent) labeled the argument consistent with the Defense Attorney's Fallacy as correct, while twenty-one (28.8 percent) found the argument consistent with the Prosecutor's Fallacy to be correct. Only sixteen subjects (22.2 percent) found both arguments to be incorrect. The order in which the arguments were presented did not significantly affect the ratings regarding correctness.[220] With respect to the responses to questions regarding to what extent, if any, the detective should revise his estimate of guilt in view of the blood type evidence, only four responses (three percent) were consistent with the respondents' having bought into the Prosecutor's Fallacy, while eighty-two of the responses (56 percent) were consistent with the Defense Attorney's Fallacy. Also of interest is the fact that slightly over one-fifth of the respondents gave inconsistent answers, *i.e.*, one response after reading the first argument consistent with one fallacy and another response after reading the second argument consistent with the other fallacy.[221]

Three conclusions can be drawn from the Thompson and Schumann experiments, although they must remain quite tentative in view of the relatively small number of subjects tested and the fact that the setting was not a real trial context. First, many people do not understand mathematics with any level of sophistication, or perhaps at all. As Thompson and Schumann concluded:

> These experiments indicate that people are not very good at drawing correct inferences from associative evidence and incidence rate statistics. They are strongly influenced by subtle and logically inconsequential differences in how these statistics are presented. They are unable to see the error in crude arguments for fallacious interpretations of the evidence, and their judgments of probable guilt are strongly influenced by such arguments.[222]

Second, presenting evidence in the form of probability of a random match is more likely to cause subjects to confuse that probability with the probability

220. Thompson & Schumann, *supra* note 179, at 178.

221. *Id.* at 179.

222. *Id.* at 181. *See also* Goodman, *supra* note 179. "Most mock-jurors have little facility with probabilities, and were unable to convert odds to percentages and vice versa. . . ." *Id.* at 203.

of the defendant's guilt than when evidence is presented in the form of empirical statistics. Third, while Thompson and Schumann found it to be unclear why the Prosecutor's Fallacy was less prevalent in the second experiment than in the first, the most likely conclusion is that people who are initially amenable to the Prosecutor's Fallacy when it is unrebutted, as in experiment one, can fairly easily see the error of their ways when a contrary defense argument is presented, as in experiment two. Thus, there is good reason to hope that a minimal percentage of jurors would buy into the Prosecutor's Fallacy if a competent defense attorney explained to them why they should not do so.

In a similar study in 1986,[223] Goodman reached conclusions that are consistent with the findings of Thompson and Schumann, except in one respect. Goodman used as subjects 233 undergraduate psychology majors. She presented them with information regarding a murder case where one of the items of evidence against the defendant was that the defendant's blood type matched a blood trace at the crime scene, and it was established that the trace could not have come from the victim. Some subjects were presented with testimony that the blood type found at the scene was possessed by ten percent of the population; some that the blood type was possessed by five percent of the population, some that it was possessed by one percent of the population, and some that it was possessed by one-tenth of one percent of the population. All subjects were informed that the suspect population consisted of 100,000 persons.[224] The subjects were then required to answer questions concerning their evaluations of the case. Goodman reached three conclusions from this data that corroborate the Thompson and Schumann results: (1) the subjects were very bad at manipulating the numbers—many were unable to calculate how many persons in the suspect population would have the blood type found at the scene;[225] (2) the subjects "underutilized" the evidence compared with what Bayes' Theorem would suggest;[226] and (3) the incidence of the Prosecutor's Fallacy was rare (1.6 percent).[227] She reached one conclusion somewhat inconsistent with Thompson and Schumann: she found that the incidence of the Defense Attorney's Fallacy was not as widespread as Thompson and Schumann found.[228] Goodman's study also allowed her to reach one conclusion about which Thompson and Schumann's study was silent—she found that the subjects treated the mathematical evidence as being of virtually equal weight no matter which of the four statistics the subject received. Thus, the group that received the ten percent figure (which would leave 10,000 possible suspects), gave the evidence the same weight as the group that received the one-tenths of one percent (which

223. *See* Goodman, *supra* note 179.
224. *Id.* at 101-104.
225. *Id.* at 131-32.
226. *Id.* at 134-35.
227. *Id.* at 132-33.
228. *Id.* at 133-34.

would leave only 100 possible suspects).[229] Goodman concluded, "[T]he student mock-jurors were insensitive to the variations . . . or, alternatively, . . . they found it difficult to accord appropriate weight to probabilistic evidence."[230]

The teaching of the social science studies on the issue of unfair prejudice is clear (albeit tentative): valid empirical statistics are relevant and not unfairly prejudicial, assuming competent defense counsel. Probabilities of a random match, though, while relevant, are unfairly prejudicial because they do not add any information not already provided by the empirical statistics, but do have a significant potential for causing the jury to engage in an inappropriate line of reasoning.

D. Probabilities of Guilt

1. Category Three: Non-Empirical Probabilities of Guilt Incorporating Empirical Statistics Without Bayes' Theorem

Evidence in this category is characterized by the utilization of empirical statistics being phrased in terms of probability of a random match, with the probability being so miniscule that it virtually leaves no other suspect in the world besides the defendant. Prosecutors always argue that such evidence is not really being used to convince the jury to subjectively probabilize the guilt calculation, but rather is simply being used nonprobabilistically to exclude all possible alternatives inconsistent with guilt. This nonprobabilistic argument rings true with an item of evidence like a fingerprint, which is based on the truly empirical statistic that no two individuals have identical fingerprints. Fingerprint evidence is virtually never used probabilistically.[231] Rather, it is used as the ultimate empirical statistic. But the cases that are included in this category are not based on statistics that are nearly as empirical as those underlying fingerprints. Indeed, it was often a close call whether the cases in this category should instead be placed in category four, as involving non-empirical statistics. The statistics underlying probability calculations in category three cases obviously are built on shaky foundations and have high probabilities of error. Further, the expert testimony in each case could well have been presented nonprobabilistically, yet the prosecution reached to create probabilities. This indicates that the prosecution in these cases probably was seeking to induce the jury to reason probabilistically to guilt, rather than to simply eliminate in a nonprobabilistic fashion alternatives inconsistent with guilt.

---

229. *Id.* at 133-35.

230. *Id.* at 135.

231. *But see* Hicks v. Scurr, 671 F.2d 255 (8th Cir. 1982) (fingerprint expert unable to find sufficient points for comparison to make a clear match, but testified there was a one in seven million chance of a random match. The court held the testimony to have been properly admitted.); Commonwealth v. Drayton, 386 Mass. 39, 434 N.E.2d 997 (1982) (fingerprint expert testified that there was a one in 387 trillion chance of a random match. The court held it was error to state the fingerprint testimony probabilistically, but held that the error was harmless.).

There are only six cases in this category, but each is factually interesting, and taken together they are legally significant because they stand in stark contrast to *Collins*. Unlike *Collins*, courts in five of the six of these cases seem not to be much concerned with the validity of the underlying statistics, or, where the product rule was applied, with the mutual independence of the statistics. None of the courts in these six cases seem particularly worried about the probabilistic evidence having the effect of probabilizing the "beyond a reasonable doubt" burden of persuasion, nor that jurors will be unfairly prejudiced to equate the probability of a random match with the probability of guilt. In short, these cases show a distinct tendency toward adopting the probabilist position. The cases will be discussed in chronological order.

The 1969 New Hampshire case of *State v. Coolidge*,[232] is the only significant probabilistic evidence case to be decided in the decade subsequent to *Collins*. There, to prove that the defendant had been at the crime scene, an expert testified concerning particle matches obtained by vacuuming the victim's clothing and vacuuming the defendant's clothes and automobile. Out of a visual comparison of the sweepings, the expert selected for comparison forty particles from the victim's clothing with forty particles obtained from the defendant's clothes and automobile. The expert found twenty-seven matching pairs among the forty pairs of particles. He then testified that based upon previous studies made by him, the probability of finding similar particles in sweepings from different places was one in ten, that each matching set was independent from any other, and that thus the probability of finding twenty-seven similar particles and sweepings from independent sources was only one in ten to the twenty-seventh power.[233] In one sentence, the New Hampshire Supreme Court held that any objections to this testimony went to weight rather than its admissibility.[234] The court thus demonstrated a near complete lack of concern with the validity of the database upon which the expert based his testimony, as well as with the mutual independence of matches. Further, as two critiques pointed out, the court seemed oblivious to the fact that the expert's math was drastically wrong:

> Specifically, he failed to consider in his probability statement the number of samples that he examined in order to find twenty-seven matching particles. That number is an essential ingredient in any probability statement of this nature. If, for example, three hundred independent pairs of samples had been examined, with a probability of similarity of 1/10 for each pair, we would expect about thirty matching pairs. A finding of only twenty-seven pairs would argue against physical contact between defendant and the victim.[235]

---

232. 109 N.H. 403, 260 A.2d 547 (1969), *rev'd on other grounds*, 403 U.S. 443 (1971).

233. State v. Coolidge, 109 N.H. 403, 418-19, 260 A.2d 547, 558-59 (1969), *rev'd on other grounds*, 403 U.S. 443 (1971).

234. *Id.* at 419, 260 A.2d at 559.

235. Broun & Kelly, *supra* note 7, at 47. *See also* Tribe, *supra* note 31, at 1342:

Despite these defects, Coolidge's conviction was upheld.

In the 1978 Arizona case of *State v. Garrison*,[236] the defendant was charged with first degree murder. The prosecution called a dentist who testified that there was an eight in one million probability that bite marks found on the deceased's body were not made by the defendant. The doctor based the figure on unspecified articles and books in the medical field.[237] The Arizona Supreme Court cavalierly accepted the validity of the foundation undergirding the probability figure and upheld the conviction.[238] A blistering and convincing dissent showed the probability to be without foundation and thus unfairly prejudicial.[239]

The third case in this category, the 1983 Georgia decision in *Williams v. State*,[240] involved prosecution of the defendant for a notorious series of slayings of black children in Atlanta. Some of the key evidence against Williams came from alleged matches between carpet fibers found on several of the victims and the carpeting in Williams' car and home. As to the carpeting in Williams' car, two forensic experts testified that they had received information from General Motors that only 620 out of over two million cars in the Atlanta area had the kind of carpeting found on the floorboard of Williams' 1970 Chevrolet station wagon.[241] It was unexplained how General Motors, which presumably could only know how many such cars it had *shipped* to the Atlanta area, could know how many such cars there *actually* were in that area given that cars are objects which are frequently moved, traded, sold or junked. As to the carpeting in Williams' home, the manufacturer was not so obliging as to provide an exact figure concerning how many homes in the Atlanta area had that kind of carpeting. Undeterred by this lack of a reliable database, the creative prosecutor, through an investi-

> Most significantly, the court was evidently unaware that the relevant probability of finding 27 or more matches *out of 40 attempts*, was very much larger than 1/10 to the 27th—larger in fact, by a factor of approximately 10^{10}. Indeed, even the forty particles chosen for comparison were visually selected for similarity from a still larger set of particle candidates—so large a set, conceivably, that the probability of finding 27 more matches in sweeping over such a large sample even from 2 entirely different sources, could well have been as high as 1/2 or more.

Id. n.40.

236. 120 Ariz. 255, 585 P.2d 563 (1987).

237. State v. Garrison, 120 Ariz. 255, 258, 585 P.2d 563, 566 (1987).

238. *Id.* at 258-59, 585 P.2d at 566-67.

239. *Id.* at 260-63, 585 P.2d 563, 568-71 (Gordon, J., dissenting).

Bite mark evidence is not usually phrased in probabilistic form but rather is offered under the theory that each person's dentition is unique. While courts almost always uphold the admissibility of bite mark evidence, it has been convincingly argued that there is no empirical data to support any probability of a random match or that each person's dentition is unique. *See* Wilkinson & Gerughty, *Bite Mark Evidence: Its Admissibility Is Hard to Swallow*, 12 W. St. U.L. Rev. 519 (1985).

240. 251 Ga. 749, 312 S.E.2d 40 (1983). The probabilistic aspects of this case are set forth much more clearly in a law review note than they are in the appellate opinion itself. *See* Note, *The Odds of Criminal Justice in Georgia: Mathematically Expressed Probabilities in Georgia Criminal Trials*, 1 Ga. St. L. Rev. 131 (1984).

241. Williams v. State, 251 Ga. 749, 824, 312 S.E.2d 40, 98 (1983) (Smith, J., dissenting).

gator, proceeded to create a statistic using a bare minimum of data and an abundance of assumptions. Briefly, the methodology followed by the prosecutor and investigator was as follows. They obtained from the carpeting manufacturer information that during 1971 and 1972 the combined sales of the brand of carpeting found in Williams' home and another brand was slightly over 16,000 square yards in a ten-state region including Georgia. Then assuming that twenty square yards was a reasonable amount of carpet for an average room, they concluded that approximately 820 rooms in the ten-state region contained such carpeting. Assuming that these rooms were equally distributed throughout the ten-state area, the expert concluded that there were likely to be eighty-two such rooms in Georgia. The prosecution then indulged in two assumptions that were "very, very beneficial to the defense," namely that each house containing such carpeting had only one room carpeted with it, and that all of the rooms to be found in Georgia were in the Atlanta area. Then, using a figure for the number of homes in the Atlanta area obtained from the Atlanta Regional Commission, the investigator calculated that the odds of randomly selecting a home containing carpeting similar to that in Williams' were one in 7792.[242] In closing argument, then, the prosecutor applied the coup de grace. First, he argued that "there would be only one chance in eight thousand that there would be another house in Atlanta that would have the same kind of carpeting as the Williams home."[243] This, of course, was a drastic misstatement concerning the expert testimony since it had "established" that there were eighty-one other homes in the Atlanta area containing such carpeting, and thus the probability of another such home was in fact one, not one in eight thousand.[244] The prosecutor then reminded the jury that only about 600 cars of the two and one-half million in the Atlanta area had carpeting similar to Williams' station wagon, and then inexplicably figured the odds of choosing such a car at random to be about one in 5000.[245] Then, assuming that the occurrence of the carpeting in the 620 cars and the carpeting in the eighty-two homes were mutually independent, he used the product rule to multiply the one in 5000 probability of a random matching car with a one in 8000 probability of a matching random house to arrive at the conclusion that there was a one in forty million chance that a random selection would turn up a person who owned both such a car and such a house.[246] Not satisfied with this figure, the prosecutor then stated that it would be more realistic to assume that a homeowner using that particular carpeting would probably carpet at least four rooms with it, and thus the probability of a random match was more like one in 150 million.[247] The Georgia Supreme Court in one sentence

242. *See* Note, *supra* note 240, at 141.

243. *Id.* at 150.

244. *Id.* at 150-51.

245. *Id.* at 151-52. Actually using the prosecution's figures of 620 such cars and 2.5 million cars in the Atlanta area, the probability of picking such a car at random would be one in 4032.

246. *Id.* at 152.

247. *Id.*

approved of this prosecution approach, stating that inferences suggested to the jury "may include those based on probabilities."[248] The Georgia Supreme Court thus appears oblivious to the problems concerning the validity of the underlying statistics, their mutual independence, and the tendency of this probabilistic evidence to probabilize the "beyond a reasonable doubt" standard.

In the 1984 Alaska case of *Huf v. State*,[249] the prosecution sought to prove that hairs found in a cap left at the scene of a sexual assault belonged to the defendant. It called an FBI agent who testified that during the last seven years he had performed about 10,000 examinations of the hair of roughly 10,000 different people, and on only two occasions had he found hair from two people so similar that it could not be distinguished. He also testified that head hair and pubic hair are "two independent events."[250] The prosecutor argued in closing that there was a one in 5000 chance that a random person would have scalp hair like Huf's, and a one in 5000 chance that the random person would have pubic hair like Huf's, and thus the likelihood of anyone having both of the same kinds of hair was one in 5000 times one in 5000 via the product rule.[251] The court agreed that the prosecutor's argument was "misleading"[252] without specifying how. The court then held, however, that any error was harmless because of the other strong evidence in the case against the defendant and because the expert's testimony was strong in its inclusive/eliminative effect, even without the prosecutorial overstatement.[253] Although it is difficult to figure out why the court believed the prosecutor's argument to be "misleading," probably the best explanation is that the court was not satisfied that the head and pubic hair statistics were mutually independent. While the court was concerned with mutual independence, it did not seem concerned with the underlying validity of the expert's statistics or with the possibility of unfair prejudice.

An intriguing 1986 Iowa case is *State v. Klindt*.[254] A husband was charged with murdering his wife by killing her, dismembering her, and then throwing her into the Mississippi River. A torso later turned up downstream which could not be positively identified as the defendant's wife. At trial, the state undertook to show the probability that the body was the wife of the defendant by asserting that the torso contained a combination of genetic markers that would be found in only twenty-seven out of 10,000 people. The state apparently could not, however, find any known samples of the alleged victim's blood with which to compare the genetic markers found in the torso, so the state took samples of the alleged victim's parents' blood, which allowed an expert to calculate that the parents of the defendant's wife

248. *Williams*, 251 Ga. at 786, 312 S.E.2d at 73.
249. 675 P.2d 268 (Alaska Crim. App. 1984).
250. Huf v. State, 675 P.2d 268, 269 (Alaska Crim. App. 1984).
251. *Id.*
252. *Id.*
253. *Id.* at 269-70.
254. 389 N.W.2d 670 (Iowa 1986).

were 107.8 times more likely than a random couple to have produced an offspring with the genetic markers found in the torso. The really creative portion of the prosecution's presentation, however, was in calling a statistician to testify that the torso was very likely to be that of the defendant's wife. Investigating officers had developed a list of all white females who had been reported missing in a four-state area bordering on the upper Mississippi River as of April 16, 1983, the date the torso was discovered.[255] The list originally contained the names of seventeen persons, but was narrowed by eliminating those who had obviously identifying characteristics such as scars. Four missing women remained on the list, including the alleged victim. The statistician was provided with data concerning the race, sex, age range, and blood type of the torso, that the torso had borne a child, had had an episiotomy, and had not been surgically sterilized. The alleged victim was not excluded by any of these factors. The statistician obtained some information concerning these characteristics regarding the other three missing women, and, using data on the frequency of some of these conditions among the general population from statistics provided by the United States Department of Health and Human Services, plus the genetic marker evidence, testified that the probability was over ninety-nine percent that the torso was the alleged victim's rather than any of the other three missing women. The appellate opinion is ambiguous regarding exactly what information was obtained by the statistician and how he used it to reach the final probability.[256] The Iowa Supreme Court found the testimony to have been properly admitted because the underlying data was "adequately established" and "this background evidence was of a type which could reasonably be relied upon by experts in the field."[257] The court distinguished *Collins* because the testimony in the case before it "did not purport to identify the perpetrator of a crime, as in *Collins*. It did not even purport to show the torso was actually Joyce Klindt's. It only showed that the chances were much stronger that the torso was Joyce's than of any of the other three missing women."[258] The court was certainly being disingenuous in this last statement—if the evidence did not purport to show that the torso was actually the alleged victim's, then it was irrelevant. Further, while it is true that the testimony did not purport to identify the culprit, but rather the victim, the issues of who were the perpetrator and the victim were virtually synonymous in the case because if the torso was found to be that of the defendant's wife, there were no other suspects for the murder other than the defendant. As to the court's holding that the foundational statistics and the expert's manipulation of them were proper, it is difficult to critique that holding given that the specifics of what the expert did are not apparent from the opinion. It is apparent, however,

---

255. It cannot be ascertained from the opinion how far back in time the officers went to develop this list.

256. State v. Klindt, 389 N.W.2d 670, 673 (Iowa 1986).

257. *Id.*

258. *Id.* at 674.

that the court ignored two factors regarding the foundation that seem to be quite significant. First, the database was limited to a four-state area even though it is certainly possible that the torso belonged to a person who would have been reported missing in some other state. Second, the database was limited to those persons who had been reported missing, while it was certainly possible that the torso belonged to someone who had not been reported missing. Thus, there was a significant likelihood that the final pool of four candidates was incomplete. Even the addition of one more candidate to the pool could have radically changed the calculation.

The final case in this category is the 1988 Wisconsin case *State v. Pankow*.[259] The defendant, who babysat for children in her home, had three infants die in her care during a five-year period. She was charged with murder. Her defense was that the children had all died of Sudden Infant Death Syndrome ("SIDS"). At trial the state undertook to demonstrate how improbable it would be for three such deaths to have occurred in one household. To do so it called a statistician who was asked to determine the probability of three infants dying of SIDS in the same household during a five-year period given that twenty children were cared for in the home, the deaths occurred after nine o'clock in the morning, and the children were over six months of age. The statistician was also supplied with "certain generally accepted data: two SIDS deaths occur per 1000 live births; 90% of SIDS deaths occur under six months of age; and 90% of SIDS deaths occur between midnight and nine o'clock in the morning."[260] The statistician calculated that the probability of the natural occurrence of three such deaths would be 1000 times smaller than 9.1 in one trillion, or stated differently, that such an event would occur at random once every 600,000 years.[261] The Wisconsin Court of Appeals found that the underlying statistics were derived from generally accepted data, and that the case was different from most of its probabilistic predecessors because the evidence was not introduced to prove identity or to prove cause, but merely to meet the defense theory that the deaths were attributable to SIDS.[262] The court here, like the Iowa Supreme Court in *Klindt*, is being disingenuous regarding the purpose for which the evidence was offered. Since the evidence was offered to rebut the defense theory concerning the cause of death, it is hard to see how the court could conclude that the evidence was not being offered to support the prosecution's claim that the defendant was the cause of death. From a statistical standpoint, the major factor ignored by the court was that there was apparently no showing that the three statistical figures used by the statistician were mutually independent.

Cases in this category show that prosecutors are still beguiled by the power of numbers. The cases also show that not all courts have been as

---

259. 144 Wis. 2d 23, 422 N.W.2d 913 (Ct. App. 1988).
260. State v. Pankow, 144 Wis. 2d 23, 37-38, 422 N.W.2d 913, 918 (Ct. App. 1988).
261. *Id.* at 38, 422 N.W.2d at 918.
262. *Id.*

skeptical as the *Collins* court regarding foundational matters and prejudicial effect. Further, the cases tend to show that courts do not have a strong grasp of mathematics. The relative scarcity of cases in this category, however, suggests that *Collins* has significantly deterred prosecutors from even attempting to offer creative statistical evidence. Had *Collins* been decided differently, one would expect a whole raft of creative uses of mathematical evidence appearing in appellate opinions in criminal cases.

2. Category Four: Non-Empirical Probabilities of Guilt Developed Without Empirical Statistics and Without Bayes' Theorem

The characteristics of evidence in this category are that the statistics have no plausible basis in empirical research, that the resulting probability of a random match is so miniscule as to realistically leave the defendant as the only possible culprit, and that Bayes' Theorem is not used. It is more clear here than in category three that a prosecution argument that it is merely using the evidence in nonprobabilistic fashion to eliminate all possible alternatives inconsistent with guilt should not be accepted. By probabilizing evidence as to which no empirical statistics exist, it seems clear that prosecutors seek to induce juries to reason toward a subjective probability of guilt. *Collins* and its less illustrious predecessors (*Risley*, *Miller* and *Sneed*) stand squarely against admission of evidence in this category. *Collins* is so generally approved and respected that one would expect its deterrent effect to render the number of cases in this category minimal. This deterrent effect is borne out by the fact that only five cases subsequent to *Collins* fall into this category. One would also expect that these cases would not pass appellate muster, but surprisingly in two of the cases prosecution techniques that seem virtually identical to *Collins* have avoided reversal.

Let us begin, however, with two cases where the result was consistent with *Collins*. The 1983 California case of *People v. Cella*[263] is singular in that it is the only case where the *defendant* was offering mathematical evidence against the prosecution. Cella was being investigated for white collar crimes. A state agent made an illegal search from which he made up a list of ten nonexistent companies that were then turned over to the state. Months later the state served several search warrants against the defendant, one of which contained a list of ten nonexistent companies in the same order as had been provided by the agent as a result of the earlier illegal search. The defendant argued that the later search warrant must have been a result of the earlier illegal search and, thus, evidence obtained by the later search should have been suppressed. In support of this argument, he offered the testimony of a mathematician that the chances were two or three in 100,000 that a list of ten items would randomly appear in the same order twice.[264] The California Court of Appeals found that the trial judge had properly rejected the evidence:

263. 139 Cal. App. 3d 391, 188 Cal. Rptr. 675 (1983).

264. People v. Cella, 139 Cal. App. 3d 391, 404, 188 Cal. Rptr. 675, 683 (1983).

> Such technique blindly overlooks such profound problems of integrating mathematic purity with the countless nonmathematic variables that exist in reality. It quantifies with deceptive exactitude "fuzzy imponderables." If mathematical probabilities are to be of any use in the courtroom setting, *all crucial variables must be quantified exactly*. . . . There could be any number of reasons, not quantified by the mathematician, why 10 corporate names would appear twice in the same order. Our duty does not include speculation as to the quality of proof to be gained from misapplied statistics.[265]

Accordingly, the court sustained the conviction.

In another California white collar crime case in 1984, *People v. Louie*,[266] the defendant, a physician, was charged with making false statements to obtain or affect unemployment benefits. The prosecution sought to prove that the defendant had submitted diagnoses of vague, impossible-to-check-on sorts of illnesses at a much higher rate than other physicians. His diagnoses on 196 disability claims were categorized and then compared with the number of diagnoses filed by other health practitioners in the vicinity. The defendant had filed many more claims in each category than had been filed in total by the other practitioners with whom he was compared. Not content to rest with this, the prosecution called a mathematics professor who testified that the likelihood that random chance accounted for the differential in diagnoses in the first three categories was three in ten million, and that the likelihood as to the first four categories was "off the tables."[267] The California Court of Appeals held that the statistics lacked foundation because there were a host of real world factors that could account for the defendant's patient population having different characteristics than those of other physicians.[268] The court thus found the testimony "utterly useless as proof of criminal intent."[269]

Of the two cases that seem non-*Collins*-like in their approach, a 1987 Illinois case, *People v. Prewitt*,[270] can probably be explained by virtue of the fact that the defendant had not objected to the probabilistic testimony at trial and thus the question before the court was whether the admission of the testimony was plain error. Exactly what the prosecution had argued is unclear from the appellate opinion, which simply states that, "Starting from the proposition that 20% of the population are non-secretors and by assigning the defendant's individual characteristics with certain probabilities, the prosecutor deduced that defendant, being a non-secretor, was certainly the offender. Aside from the initial proposition, the prosecutor based his theory entirely on conjecture."[271] The court found the prosecutor's approach

265. *Id.* at 405, 188 Cal. Rptr. at 684.
266. 158 Cal. App. 3d Supp. 28, 205 Cal. Rptr. 247 (1984).
267. People v. Louie, 158 Cal. App. 3d Supp. 28, 41, 205 Cal. Rptr. 247, 257 (1984).
268. *Id.* at 47-48, 205 Cal. Rptr. at 261-62.
269. *Id.* at 48, 205 Cal. Rptr. at 262.
270. 160 Ill. App. 3d 942, 513 N.E.2d 977 (1987).
271. People v. Prewitt, 160 Ill. App. 3d 942, 948, 513 N.E.2d 977, 981 (1987).

not to be prejudicial enough to constitute plain error because the trial judge had instructed the jury that closing arguments are not evidence.[272] In reaching this result, the court distinguished *Collins* by stating, "This is not a situation wherein the prosecutor produced a mathematician or other expert witnesses to testify to the veracity of the statistics. Such reference to expert testimony may well unduly influence a jury."[273] This is a misreading of *Collins* because the prosecutor in that case made clear that the expert was not testifying to the "veracity of the statistics" but only to the veracity of the product rule. Nonetheless, given that the procedural context of the case on appeal was that the defendant had not objected to the mathematical approach at the trial level, this case probably does not make great inroads against *Collins.*

A 1983 Indiana case, *Roach v. State,*[274] though, appears to be a virtual replay of *Collins* with a completely opposite result. There the defendant was charged with burglary. Investigators at the scene found broken glass and muddy shoe prints. Muddy shoes containing a sliver of glass were later seized from the defendant. At trial the testimony was elicited from a lab technician that the wear pattern on the seized shoes matched the shoe prints found at the scene. Further, the lab technician testified without any apparent foundation that the chances were one in a 1000 that the glass from the bottom of the seized shoes came from a source other than the glass at the burglary scene. The prosecutor in closing argument used this testimony in a way that is so strikingly *Collins*-like that it is worthy of a substantial quotation:

> The testimony clearly was by Bruce Boaz, the lab technician, that mathematically the chances are one in one thousand that the glass from the bottom of [Roach's] shoes came from other than a common source. A thousand to one that, this glass goes with the exhibit from the scene. Next we've got testimony about footprints, footprints again found at the scene out in the muddy field. The testimony was that, by Officer Stump, that these shoes found on the feet of Tim Roach matched the size, the wear and the general tread of the prints that he found in that field. What's the possibility first that the individual who was at this scene, if it wasn't Tim Roach, had the same size shoes? I don't know what the chances are. One in a hundred, one in fifty, one in twenty-five? How about one in ten? Next what is the possibility that an individual out there had the same wear pattern on the bottom of his shoes that Mr. Roach did? The shoes were worn and showed to be worn out in the same way that the shoes that Mr. Roach had. Again, one in a hundred, one in fifty, one in twenty-five? Let's be fair. How about one in ten? I don't know. Put any value that you think. What's the chance that some random individual out there at the scene had the same shoes

272. *Id.* at 948, 513 N.E.2d at 982.
273. *Id.* at 948, 513 N.E.2d at 981.
274. 451 N.E.2d 388 (Ind. Ct. App. 1983).

> with the same tread pattern? Again, I don't know. One in a hundred? Let's go with one in ten.
>
> * * *
>
> These are assumptions. I'll grant you that. But you put the amount in there that you think. Okay. What's the chance, then, that these footprints were the same size and the same wear pattern and the same tread pattern and belonged to somebody else? Multiply those out. One, two, three zeroes. One in a thousand? Maybe that's a little heavy. Let's try one in a hundred, to be reasonable about it. Next what's the possibility that this individual with glass in his shoe had the same footprints and ultimately ended up in the Defendant's car? The Defendant's car found at this particular scene. Again, I don't know. What is the possibility? One in a thousand? One in a hundred? I don't know. Be fair. How about one in a hundred? One in ten? I don't know. Put any value that you feel. Taking these and using the fair figures, one in a thousand here on the glass, what's the possibility that all these things happened, all these circumstances happened all at the same time? One in a thousand here. One in a hundred here. One In [sic] a hundred here. I don't know. Maybe it's a little bit more. Maybe it's a little bit less. But using these amounts, there's three, four, five, six, seven zeroes. One in ten million, chances of all these things happening.

In a strikingly un-*Collins*-like opinion, the Indiana Court of Appeals upheld the conviction:

> Although prosecutors may not argue facts not in evidence, they may properly argue their own analysis of the evidence. In assigning probabilities to the various pieces of evidence, the prosecutor did not state as a fact that any specific probability was correct. The only exception was the probability that glass in Roach's shoe would match glass found at the scene—estimated by the state's expert as one in a thousand. Thus, the prosecutor merely supplied a method of analyzing the evidence in the record, leaving the jurors free to assign any statistical probability to the various facts. The prosecutor did not improperly state facts not in evidence here; thus, we find no prosecutorial misconduct.[275]

The court sought to distinguish *Collins* because in the case before it the jurors had not been "required to accept unproven assumptions as facts."[276] In so stating the court ignored the fact that neither were the jurors in *Collins* required to accept the calculations as fact; indeed, they were urged to develop their own calculations. While the court indicated that it would be unwilling to allow the equation of mathematical probability with proof beyond a

---

275. Roach v. State, 451 N.E.2d 388, 392 (Ind. Ct. App. 1983).
276. *Id.*

reasonable doubt, it rather implausibly found that the prosecutor had not attempted to equate the two in his argument.[277] Either the court in this case was not thinking carefully and clearly, or it was buying into most of the premises of the probabilist position.

The other case in this category does not involve the use of mathematical evidence at trial; rather, it involves an appellate judge analyzing a case in a manner strikingly like that argued for by the prosecutor in *Collins*. The defendant had been convicted of murder and sexual assault. Several pieces of evidence, including a fingerprint, a comb, some hair, and car keys placed him at the victim's home. On appeal, the defendant argued that the evidence was insufficient to support the conviction. In rejecting this argument, Judge Posner reasoned as follows:

> Suppose that the probability that the fingerprint was not his (or, as he argues, was put on the can months earlier when he was shopping in the store where it was bought) is .01 (a generous estimate); the probability that the comb was not his is .50; the probability that the hair was not his is .30; and the probability that someone else discarded Miss Ayer's car keys near his mother's house is .05. Then, assuming these probabilities are independent of each other, the probability that Rowan was not in Miss Ayer's house at a time near when she died is only .000075 (.01 x .50 x .30 x .05), which is less than 100th of one percent. . . . True, it would not follow that he had killed her; someone else might have entered the house before or after him, and done the deed. But that is exceedingly unlikely . . . and does not cast substantial doubt on his guilt.[278]

Thus, probabilist hearts beat not only within the chests of prosecutors, but some judges as well.

3. Category Five: Non-Empirical Probabilities of Guilt Incorporating Empirical Statistics Via Bayes' Theorem

The scholarly debate regarding Bayes' Theorem dates back to Finkelstein and Fairley's article in 1970.[279] Finkelstein and Fairley posed the hypothetical where a woman has been found murdered, there is some non-mathematical evidence pointing to her boyfriend, and the police have found a palm print at the scene which would match only one out of every 1000 people, including the defendant.[280] Finkelstein and Fairley advocated using Bayes' Theorem to portray to the jury the effect that this evidence should have on their assessment of the defendant's guilt.[281] They proposed developing a table illustrating what the posterior probability would be after application of

277. *Id.* at 393.
278. Rowan v. Owens, 752 F.2d 1186, 1188 (7th Cir. 1984).
279. Finkelstein & Fairley, *supra* note 23.
280. *Id.* at 496.
281. *Id.* at 498.

Bayes' Theorem based upon differing prior probabilities. In the case of the palm print, if the jurors had a prior probability of guilt based on the nonmathematical evidence of ten percent, the palm print evidence should increase that probability of guilt to 99.1 percent. If the prior probability were fifty percent, the posterior probability should be 99.9 percent. If the prior probability were seventy-five percent, the posterior probability should be 99.96 percent.[282] Finkelstein and Fairley argued that this was a perfectly legitimate way of assessing the impact of new evidence.

Professor Lawrence Tribe took issue with Finkelstein and Fairley.[283] He argued that there were two insurmountable legal barriers to Finkelstein and Fairley's proposal. First, he argued that the application of Bayes' Theorem would require the jurors to formulate prior probabilities of guilt before all the evidence was in, whereas the presumption of innocence requires the jurors to believe completely in the defendant's innocence until they retire to the jury room to deliberate.[284] Second, Tribe argued in conformance with the anti-probabilists' position that such use of Bayes' Theorem would tend to quantify the "beyond a reasonable doubt" standard of proof, which is legally impermissible:

> In short, to say that society recognizes the necessity of tolerating the erroneous "conviction of some innocent suspects in order to assure the confinement of a vastly larger number of guilty criminals" is not at all to say that society does, or should, embrace a policy that juries, *conscious of the magnitude of their doubts in a particular case*, ought to convict in the face of this acknowledged and quantified uncertainty. It is to the complex difference between these two propositions that the concept of "guilt beyond a reasonable doubt" inevitably speaks. The concept signifies not any mathematical measure of the precise degree of certitude we require of juries in criminal cases, but a subtle compromise between the knowledge, on the one hand, that we cannot realistically insist on acquittal whenever guilt is less than absolutely certain, and the realization, on the other hand, that the cost of spelling that out explicitly and with calculated precision in the trial itself would be too high.[285] (emphasis added.)

Finkelstein and Fairley responded to these two arguments.[286] As to the argument that formation of a prior probability was inconsistent with the presumption of innocence, they argued that jurors are constantly evaluating and reevaluating the probability of the defendant's guilt throughout the

282. *Id.* at 500.

283. Tribe, *supra* note 31.

284. *Id.* at 1368-72. This objection has been echoed by other commentators. *See, e.g.*, Cohen, *supra* note 37, at 100; Jaffee, *supra* note 26, at 968, 971, 987; Jaffee, *supra* note 7, at 983; Tyree, *supra* note 176, at 234.

285. Tribe, *supra* note 31, at 1375.

286. Finkelstein & Fairley, *A Comment On "Trial By Mathematics"*, 84 HARV. L. REV. 1801 (1971).

trial.[287] As to the argument that Bayes' Theorem would impermissibly quantify the burden of persuasion, Finkelstein and Fairley argued that the burden was in fact probabilistic, and that in any event the probability in their hypothetical concerned the likelihood of a trace having been made by the defendant, not the likelihood of the defendant's guilt.[288]

Professor Tribe also put forth three reasons to doubt that jurors were competent to deal with evidence in the way advocated by Finkelstein and Fairley. First, he argued that jurors would find it impossible to assign prior probabilities.[289] Second, Tribe argued that jurors would be unable to ignore the mathematical evidence in forming their prior probability and thus would tend to "double count" the significance of the mathematical evidence.[290] Third, Tribe argued that the mathematical evidence would tend to "dwarf the soft variables":[291]

> The syndrome is a familiar one: if you can't count it, it doesn't exist. Equipped with a mathematically powerful intellectual machine, even the most sophisticated user is subject to an overwhelming temptation to feed his pet the food it can most comfortably digest. Readily quantifiable factors are easier to process—and hence more likely to be recognized and then reflected in the outcome—than are factors that resist ready quantification. The result, despite what turns out to be a spurious appearance of accuracy and completeness, is likely to be significantly warped and hence highly suspect.[292]

Finkelstein and Fairley responded by arguing that jurors could easily enough form prior probabilities.[293] As to the "dwarfing of soft variables" argument, Finkelstein and Fairley contended that the mathematical evidence concerning the trace evidence is certainly admissible in the case without Bayes' Theorem, and thus it only makes sense to use Bayes' Theorem to help the jury give proper weight to the evidence that they might otherwise misevaluate.[294]

One reason why Tribe's article has stood the test of time is that it formulated many of the anti-probabilist arguments in a manner that has never been improved upon. Tribe, however, was only partially anti-probabilistic. He did not contest the validity of probabilistic calculations, but instead argued that the costs of using them in the trial process would be

---

287. *Id.* at 1808. *See also* Friedman, *supra* note 56, at 734 n.6 (not inconsistent with presumption of innocence for jurors to assign a positive, albeit tiny, prior probability to defendant's guilt); Lempert, *supra* note 49, at 464 (appropriate for jurors to start with a prior probability of one over one less than the number of persons in the world).

288. Finkelstein & Fairley, *supra* note 23, at 1808-09.

289. Tribe, *supra* note 31, at 1358-59. *See also* Braun, *supra* note 7, at 50-51.

290. Tribe, *supra* note 31, at 1366-68.

291. *Id.* at 1361.

292. *Id.* at 1361-62.

293. Finkelstein & Fairley, *supra* note 23, at 1802-03. *See also* Kaye, *supra* note 31, at 42, 52-53.

294. Finkelstein & Fairley, *supra* note 23, at 1806.

too great.[295] Thus, he did not mention one important argument that anti-probabilists who completely reject the validity of probabilists' tenets in the proof process have later proffered. That argument, which has already been mentioned,[296] is that multiplication with respect to items of evidence is inappropriate because each item of evidence is an all-or-nothing proposition: if it is believed, then other items of evidence based thereon can be considered, whereas if it is rejected, the line of reasoning ends.

These debates concerning the applicability of Bayes' Theorem to the trial process have consumed countless law review pages. With respect to the case law, however, Bayes' Theorem rarely turns up. The only context in which Bayes' Theorem has been utilized is in sex offense cases where the alleged victim became pregnant and delivered a child, allegedly as a result of illegal sexual activity, and the prosecution attempts to prove that the defendant was the father by use of mathematical evidence. In such cases, in addition to a probability of a random match (known in the paternity area as the "probability of exclusion") and a "paternity index," the ratio that compares the alleged father's likelihood of producing the child's phenotypes with the likelihood of a random man selected by police doing so,[297] a third probability known as the "probability of paternity" can be calculated. First, a prior probability of paternity is developed. Then the "paternity index" can be used as the other input into Bayes' Theorem to arrive at a posterior "probability of paternity." Three courts in criminal cases have addressed the admissibility of such a "probability of paternity."[298] In a 1983 case, the Minnesota Supreme Court, the most anti-probabilist court in the country, cited Professor Tribe in holding a probability of paternity to be inadmissible in the criminal case: "[T]here is a real danger that the jury will use the evidence as a measure of probability of the defendant's guilt or innocence, and that the evidence will thereby undermine the presumption of innocence, erode the values served by the reasonable doubt standard, and dehumanize our system of justice."[299] In a 1988 case, the Wisconsin Supreme Court

295. *See supra* note 41.

296. *See supra* text accompanying note 37.

297. For a discussion of these two probabilities, see *supra* note 198 and accompanying text.

298. Two other cases used the term "probability of paternity" but upon examination of the evidence that was offered, it appears that they actually involve either a "probability of exclusion" or a "paternity index." People v. Alzoubi, 133 Ill. App. 3d 806, 479 N.E.2d 1208 (1985); State v. Thompson, 503 A.2d 689 (Me. 1986).

299. State v. Boyd, 331 N.W.2d 480 (Minn. 1983). Although it is unclear from the opinion what prior probability the expert witness used in coming to the "probability of paternity", it seems likely that he used the normal assumption of such experts that there was a 50% prior probability of the defendant's paternity. Even probabilists reject this arbitrary 50% figure and argue that the prior probability should depend on the jurors' assessment of the nonmathematical evidence in the particular case. *See, e.g.*, Ellman & Kaye, *supra* note 172, at 1149-52. Nonetheless, one court in a civil case has mandated the use of the 50% prior probability because permitting the jury to form its own prior probability would be too confusing. Commonwealth v. Beausoleil, 397 Mass. 206, 490 N.E.2d 788, 797 n.19 (1986).

found that while the "probability of exclusion" and the "paternity index" could be admitted at trial, the "probability of paternity," "although clearly relevant," could not be admitted because, "[I]t is antithetical to our system of justice to allow the state, through the use of statistical evidence which assumes that the defendant committed the crime, to prove that the defendant committed the crime. Because the probability of paternity assumes the fact that it is used to prove, it is inadmissible."[300] It is not clear what the court meant by this statement. One possibility is that the application of Bayes' Theorem requires a prior probability of guilt, which is legally impermissible. Another possibility is that the calculation assumes that the defendant had intercourse with the child's mother.

A 1987 decision by the North Carolina Supreme Court is apparently the only decision having permitted Bayes' Theorem to be used in a criminal case.[301] There, an expert used Bayes' Theorem to show how the genetic evidence would affect the jurors' prior probabilities at the "weak end" level, at the medial level, and at the high range.[302] The only objection raised by the defendant was that the expert had been allowed to go further and give her subjective opinion that he was the father, and thus the court's apparent approval of the prosecution's approach is merely dicta.[303] Nonetheless, the court did state, "In the present case, Dr. McMahan's testimony on the use of the [probability of paternity] was unquestionably of assistance to the trier of fact."[304] Thus, the North Carolina Supreme Court became the first and so far only court to approve of the use of Bayes' Theorem in a criminal case.

Some of the social science research speaks to the possible impact of Bayesian calculations upon jurors. Finkelstein and Fairley suggested that the question whether jurors will be helped, misled, or confused by the application of Bayes' Theorem was a topic that should be subjected to empirical research.[305] One recent study has been specifically directed to that question.[306] This study by Faigman and Balgioni was conducted on 180 volunteers enrolled in continuing adult education courses at several community colleges in Virginia. The volunteers were asked to read a transcript regarding evidence in a breaking-and-entering case where the culprit had cut himself on broken glass from the window used to enter the store. The transcript contained direct and cross-examination of three witnesses who did not give mathematical evidence: the arresting officer; an eyewitness who saw a car similar to the

---

300. State v. Hartman, 145 Wis. 2d 1, 15-16, 426 N.W.2d 320, 326 (Wis. App. 1988).

301. State v. Jackson, 320 N.C. 452, 358 S.E.2d 679 (1987).

302. *Id.* at 459, 358 S.E.2d at 682. The odd thing about the report of the expert's testimony is that Bayes' Theorem was used to calculate the "paternity index" when in fact the "paternity index" is one of the inputs into Bayes' Theorem to arrive at the "probability of paternity." Whether the expert, the court, or both were confused is impossible to ascertain.

303. *Id.* at 459-60, 358 S.E.2d at 682-83. The court agreed with this argument.

304. *Id.* at 460, 358 S.E.2d at 683.

305. Finkelstein & Fairley, *supra* note 286, at 1806.

306. Faigman & Baglioni, *supra* note 179.

one driven by the defendant outside of the scene of the crime; and the defendant, who gave vague and inconsistent testimony regarding merchandise that might have been stolen in the break-in that was found in his apartment, his whereabouts on the night of the burglary, and where he had received cuts on his arm (allegedly from his construction job). Then a physician who had taken a blood sample from the defendant testified that the defendant's blood type under the ABO system matched that found at the scene, and told the jurors what percentage of the population had that particular blood type. Finally, a statistician testified regarding how Bayes' Theorem would evaluate the blood grouping evidence. He presented a chart to the jurors that displayed four prior probabilities ranging from one percent to eighty percent and their accompanying posterior probabilities.[307] There were two variations in the experiment. One related to the type of blood to which the physician testified: one group of subjects received testimony that the defendant's blood type of A was possessed by forty percent of the population; the second group that the defendant's blood type of O was possessed by twenty percent of the population; and a third group that the defendant's blood type of AB was possessed by five percent of the population. The second variation related to the "probes" inserted in the materials by the researchers. On each probe the subjects were asked to state the likelihood that the blood found in the stereo shop was the defendant's blood; were asked two questions specific to the testimony given just prior to the probe; and were asked to make a determination of guilt based on the evidence already heard. For some of the subjects these probes were placed in three places—before the physician's statistical evidence, after the physician's statistical evidence, and after the statistician's testimony. For a second group the probes were placed only after the physician's statistical evidence and after the statistician's testimony. As to a third group of subjects, the probe was placed only after the statistician's testimony.[308]

The numerical results of the Faigman and Baglioni study are not easy to synopsize (unlike those in the Thompson and Schumann studies)[309] because Faigman and Baglioni used more sophisticated mathematical tools, such as standard deviations, to evaluate the responses. The conclusions reached by Faigman and Baglioni, then, are easier to state in prose than in numerical terms. First, they found that virtually all of the respondents "significantly underutilized" the statistical evidence.[310]

> Indeed, except in the AB blood-group condition (*i.e.*, presented with the 5% figure) and where the subjects explicitly stated a prior

---

307. *Id.* at 5-6.

308. *Id.* at 6-7.

309. *See supra* notes 216-22 and accompanying text.

310. Faigman & Baglioni, *supra* note 179, at 13. This, of course, marks Faigman and Baglioni as probabilists since their conclusion that the respondents "underutilized" the evidence was based upon their findings that the respondents did not make as much of the evidence as Bayes' Theorem would indicate.

> probability (*i.e.*, the three-probe condition), respondents virtually ignored the statistical evidence. Further, the subjects did not conform to the expectations of either critics [citing Tribe] or proponents [citing Finkelstein & Fairley] of the courtroom use of Bayes' theorem; they were not overwhelmed by this statistical theorem, nor did they accept the statistician's conclusions. Overall, subjects who estimated a prior probability on the first probe revised their estimates after the statistician testified as did subjects who estimated a prior probability at the end of the transcript. Yet, for both of these groups, the revisions remain significantly below the probabilities about which the statistician had testified (*i.e.*, a Bayesian model).[311]

Second, Faigman and Balgioni concluded that requiring the subjects to state a prior probability before the physician testified sensitized subjects to the blood group evidence and made them more likely to give greater weight to that evidence. "This finding suggests that explicit quantification of the nonstatistical evidence may increase the utilization of the statistical evidence."[312] Third, the researchers concluded that there was no evidence to support Professor Tribe's assertion that jurors were likely to "double count" the mathematical evidence by letting it contaminate their prior probability.[313] Finally, Faigman and Baglioni concluded that while the respondents on average viewed the statistician's testimony as accurate, they did not give it very much weight. "The statistician was only given as much weight as the eyewitness, who admitted to drinking the night of the burglary, and the defendant, whom the subject saw, on average, as only 36% likely to be telling the truth. . . . Apparently, subjects felt as the following subjects succinctly put it: 'I personally don't put much weight on statistical deductions as proof of anything.'"[314] Faigman and Baglioni, as good probabilists, concluded their report as follows: "The results . . . suggest, contrary to Tribe's assertion, that an expert's Bayesian formulation will not overwhelm the average trier of fact. Courts, it seems, should be less concerned with jurors being overwhelmed by the complexity of statistical techniques and more concerned with impressing upon jurors the relevance of those techniques."[315] Anti-probabilists, on the other hand, contend that it shows the innate wisdom of juries that they do not reason in Bayesian or any other probabilistic fashion.[316]

CONCLUSION

It is the author's hope that this article has demonstrated three things about mathematics in the criminal proof process. First, that the significance

311. *Id.* at 13-14.
312. *Id.* at 15.
313. *Id.*
314. *Id.*
315. *Id.* at 16.
316. Wright, *supra* note 36.

of these issues is going to wax, not wane, as statistical data becomes increasingly available. Thus, it behooves those involved in the legal system to make the effort necessary to gain a basic understanding of the issues, however unappetizing the prospect to those who are not mathematically inclined. Second, while the topic is initially imposing to the non-mathematically inclined, it *is* understandable if one is willing to devote a modicum of time and effort to master the basics. These include the mathematical tools—the distinctions among data, statistics, and probabilities, the three theories of probability and the two mathematical rules, the fundamentals of the probabilist and anti-probabilist positions, and the *Collins* case. These basics enable one to recognize and categorize mathematical evidence into one of the five categories, and then to intelligently analyze the evidence's strengths and weaknesses. Third and finally, an understanding of the basics empowers the non-mathematically inclined to enter into the debate about the proper role of mathematics in the proof process. This is too important a topic to be left solely in the province of the mathematically sophisticated.