1983

When Blood Is Their Argument: Probabilities in Criminal Cases, Genetic Markers, and, Once Again, Bayes' Theorem

Randolph N. Jonakait
New York Law School, farrah.nagrampa@nyls.edu

Recommended Citation
University of Illinois Law Review, Vol. 1983, Issue 2 (1983), pp. 369-422

WHEN BLOOD IS THEIR ARGUMENT:[1] PROBABILITIES IN CRIMINAL CASES, GENETIC MARKERS, AND, ONCE AGAIN, BAYES' THEOREM

*Randolph N. Jonakait**

I. INTRODUCTION

Revolutionary advances in blood typing soon will cause a dramatic increase in the presentation of statistical evidence in criminal trials. Courts have admitted statistics into criminal trials before, and the proper use of this type of evidence has been debated previously. Until now, however, such mathematical evidence has been rare.[2] Recently, however, a number of courts have admitted probability evidence derived from new and complex blood tests. Such evidence may soon be as commonplace as fingerprint testimony. The courts that have admitted this evidence, however, have done so without learning from past discussions about the proper role of statistical evidence. Because our criminal justice system is now at the threshold of an explosion in the presentation of this mathematical testimony, it is important to explore the use and misuse of such statistics in criminal trials.

After surveying the recent and ongoing revolution in forensic serology, this article addresses the proper function of serological probability evidence at trial. The article examines, and rejects, recent commentary suggesting Bayes' Theorem as a possible vehicle for presenting this statistical evidence at trial. The article concludes that at present no workable method exists for the effective and fair introduction of blood marker probability statistics at trial.

---

1. W. Shakespeare, *King Henry V*, act IV, scene i.

* *Associate Professor, Pace University Law School; A.B. 1967, Princeton University; J.D. 1970, University of Chicago Law School; LL.M. 1971, New York University.* *The author gratefully acknowledges the advice and assistance of Dr. G. Miller Jonakait and Linda Tacoma.*

2. One scholar who has advocated a broader use of statistics in criminal trials has concluded, "Although probability theory is used widely as a tool for the production of evidence in some areas of law, its formal use in criminal trials has been severely limited." Braun, *Quantitative Analysis and the Law: Probability Theory as a Tool of Evidence in Criminal Trials*, 1982 UTAH L. REV. 41, 42. *See also* Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 488, 488 n.2 (1970).

II. Blood Groups and Genetic Markers

Until recently, forensic serologists usually could classify blood into only one of four genetically controlled groups.[3] Although it may have been of great value to learn whether a blood sample or stain found at a crime scene came from the victim or the defendant or someone else, the traditional blood typing tests were of little value to a prosecutor since many people fit into each of the four groups.[4]

A decade or two ago, however, serologists began to use a new procedure to type genetic factors in the blood. This procedure, electrophoresis, now allows the serologist to discover many more blood groups than did the traditional tests.[5] Indeed, forensic serologists believe that in the future they will be able to match a blood sample with the individual from whom it was taken.[6]

Although that day has not yet been reached,[7] proponents of ge-

---

3. The four types of the ABO system are types A, B, AB, and O. This is an antigen system. Antigens are grouped by tests relying on agglutination, or the clumping together, of the red cells in the blood. *See* A. Moenssens & F. Inbau, Scientific Evidence in Criminal Cases § 6.11, at 298 (1978).

4. One forensic serologist explains:

> When [ABO] blood grouping was successful, the evidential value of the findings was often limited. For example, a suspect in a homicide may have had a bloodstain on his shirtcuff which investigators believed could have come from the victim. How significant would the following findings be in associating the suspect with the victim? Suppose the suspect was of Group O while the victim was of Group A and the bloodstain on the suspect's cuff was of Group A. We know the stain could not have come from the suspect but on this determination alone it could have come from the victim. On the other hand, it is realized that about forty-two percent of our population is of Group A. Therefore, the finding of Group A blood on the cuff cannot in itself carry much weight.

Baird, *The Individuality of Blood and Bloodstains*, 11 J. Canadian Soc. Forensic Sci. 83, 103 (1978). Such evidence, while it might exclude a suspect from consideration when the relevant blood samples did not match, was of little probative value in establishing a suspect as the guilty party when those samples did match.

5. Many blood components exist in genetically distinct variations. These substances are called polymorphic, which means that although the substances may take different forms, they all serve the same function in the body. Electrophoresis separates some of the polymorphic blood enzymes and other proteins through the use of an electrical current. The blood factors identified by the traditional agglutination tests are often called blood groups or types. The broader term "genetic markers" is applied to the factors identified by the traditional procedure as well as by electrophoresis. For a more complete description of the tests and their history see Jonakait, *Will Blood Tell? Genetic Markers in Criminal Cases*, 31 Emory L.J. 833, 836-43 (1983); Baird, *supra* note 4, at 109-20.

6. Dr. Baird, *supra* note 4, at 286 asserts, "The evolution of forensic blood grouping is likely to reach a point in the next decade where the goal of identifying the individual is achieved." This belief rests on the faith that each person's blood is singular. *See, e.g.*, Diamond, *The Story of Our Blood Groups*, in Blood, Pure and Eloquent 691 (M. Wintrobe ed. 1980):

> From the modest start of four types, the individual characteristics of blood cells have been so greatly expanded that it is now highly improbable that any two people except identical twins would have the same combination of red cell surface markers. In other words, every person on earth is unique in his or her combination of blood groups—a fact recognized only the past thirty years.

7. "Absolute individualization of blood, while theoretically possible, is not a practical goal for any laboratory." Grunbaum, *Potential and Limitations of Forensic Blood Analysis*, in Handbook for Forensic Individualization of Human Blood 2 (B.W. Grunbaum ed. 1981). "It is not yet possible to individualize blood in the same way as one can a fingerprint, but this is because of a lack of knowledge of techniques and not because of the nature of blood." B. Culliford,

netic factor typing believe that they already have a revolutionary evidentiary tool. The tests now nearly can tell if a blood sample came from a specified individual. As one serologist contends, "[W]e are now facing a situation wherein it is almost possible to characterize human bloodstains to the extent that they can be as individualistic as fingerprints."[8] The scientists claim that as they identify each additional genetic marker in a blood sample, they can prove that the sample could have come only from an increasingly smaller percentage of the population. They reason that if genetic markers X, Y, and Z each appear in 50% of the populace, the discovery of one marker in a blood sample merely indicates that half of the world could have been the source of the blood. If two of the markers, however, are found in a sample, the forensic scientist would multiply 50% times 50% and determine that the blood could have originated in only 25% of the population. If the sample contains three markers, only 12.5% of all people could have been the source of the sample. If enough genetic markers are typed or if any of the discovered markers were rarer, the serologist could conclude that the sample could only have come from a small percentage of the population, sometimes so small as to be almost infinitesimal.[9]

This percentage, this number stating how frequently a combination of blood genetic markers appears in the population, is a probability statistic.[10] As evidence in a criminal trial, the value to the prosecution of blood grouping depends precisely upon this probability or frequency figure.[11] Assume, for example, that a person is stabbed to

THE EXAMINATION AND TYPING OF BLOODSTAINS IN THE CRIME LABORATORY 15 (1981). *But see* Selvin, Grunbaum & Myhre, *The Probability of Non-discrimination or Likelihood of Guilt of an Accused: Criminal Identification*, 23 J. FORENSIC SOC. SCI. 27, 30 (1983) [hereinafter cited as *Probability*]: "Not surprisingly, using a large number of [genetic marker] systems makes it possible, at least theoretically, to discriminate between two samples of blood so that innocent individuals are almost certainly excluded."

8. Baird, *supra* note 4, at 121.

9. A genetic marker system is one in which a substance in the blood appears in genetically distinct variations. Such a substance is said to be polymorphic, which means "having different molecular forms but the same biochemical function." Grunbaum, *supra* note 7, at 3. Thus phosphoglucomatase (PGM) is polymorphic and can take different forms such as PGM 1 or PGM 2-1. Each of the different forms of PGM or any other genetic marker system is called a phenotype. Thus, type A blood is a phenotype of the ABO system. *See* Jonakait, *supra* note 5, at 839-40. At least 62 blood genetic marker systems have been identified. *See Joint AMA-ABA Guidelines: Present Status of Serologic Testing Problems of Disputed Parentage*, 10 FAM. L.Q. 247 (1976). Forensic laboratories do not attempt to type all of these systems in a blood sample. Instead, in addition to PGM, other systems commonly classified are adenylate kinase (AK), adenosine deaminase (ADA), esterace D (EsD), erythrocyte acid phosphotase (EAP), and haptoglobin (Hp).

A biostatistician has calculated that only one person out of every 1297 has the most common phenotypes for nine frequently typed genetic marker systems. Selvin, *Statistical Analysis of Blood Genetic Evidence*, in HANDBOOK FOR FORENSIC INDIVIDUALIZATION OF HUMAN BLOOD AND BLOODSTAINS 177 (B.W. Grunbaum ed. 1981).

10. *See* Braun, *supra* note 2, at 45-46.

11. "The usefulness of these [genetic marker] determinations in the criminal justice system is dependent upon a knowledge of the frequency of occurrence of these genetic factors in a general population or in a specified subpopulation." Grunbaum, *supra* note 7, at 3. *See also* Grunbaum, Selvin, Myhre & Pace, *Distribution of Gene Frequencies and Discrimination Probabilities for 22 Human Blood Genetic Systems in Four Racial Groups*, 25 J. FORENSIC SCI. 428, 428 (1980) [herein-

death, and a suspect is placed into custody a short while later. Bloodstains are found on the defendant's clothes. If that blood came from the victim it would be a strong indication of guilt. The blood tests, however, cannot show that the sample actually is the victim's blood. Instead, they might show that the bloodstain and the victim's blood do not match. Even if markers from the two samples do match, however, this match is not very probative because many other people may have the same genetic markers in their blood.[12] Instead, the genetic marker evidence is useful only when the prosecution can prove that the blood samples match and that a limited portion of the population has this particular combination of genetic markers. Thus, the prosecution wants to present evidence not only that the new blood tests showed the same genetic markers in both samples, but also that only one in 1000 people, say, have these particular markers.

These blood tests already are widely used by forensic laboratories throughout the country.[13] Because the tests are so new, however, results are only now beginning to appear in the reported criminal cases. Because the individualization of blood could be important in any criminal trial where blood has been shed, these blood tests will soon be seen in a burgeoning number of cases.[14] Since the end result of the genetic marker evidence is a probability or frequency figure, the criminal courts will face an unprecedented rise in the introduction of statistical testimony. Fair trials demand a careful consideration of the proper use of such statistics. A famous California case, *People v. Collins*,[15] triggered much of the debate surrounding the proper use of statistical evidence in criminal cases.

III. THE PRODUCT RULE

A. People v. Collins

In *Collins*, an assailant knocked down Juanita Brooks in an alley near her house and took her purse. John Bass saw a woman flee from the alley and get into a partly yellow automobile. Bass described the woman as white, with blonde hair worn in a ponytail. The car was driven by a black man with a mustache and beard. The victim was not able to make any identifications, but Bass identified Collins as the

after cited as *Gene Frequencies*]; Baxter, Grouping of *Blood Stains: Present and Future Trends*, 12 CAL. W.L. REV. 284, 285-86 (1976).

12. It might be no more probative than proving that both the killer and the defendant were right-handed or had brown hair.

13. Testimony in one trial indicated that the new blood tests were used by the FBI research laboratory and over one hundred other forensic labs in this country. State v. Washington, 229 Kan. 47, 52, 622 P.2d 986, 990-91 (1981).

14. This assumes that the courts will admit the testimony based on the scientific tests. I have argued elsewhere that since the reliability of the forensic applications of these recently developed procedures has not been proved, such testimony should not now be admitted into criminal trials. *See* Jonakait, *supra* note 5, at 911-12.

15. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

driver. Collins and his white wife, Janet, were charged with the robbery. Collins was black with a beard and a mustache. Janet had a blonde ponytail, and the couple drove a partly yellow automobile.

During the jury trial, the defendant attacked the Bass identification by attempting to show that at a lineup shortly after the incident Bass had been unsure of his identification. Furthermore, Collins contended that on the day of the robbery Janet had been wearing clothes different from those that witnesses had described as being worn by the woman who took the purse. To shore up what may have been perceived as a weak identification, the prosecutor called as a witness a college mathematics instructor. The witness first explained the basic "product rule" of probability, which states that "the probability of the joint occurrence of a number of *mutually independent* events is equal to the product of the individual probabilities that each of the events will occur."[16] Thus, the odds of getting a head on a flip of a coin is one in two or one-half, and the odds of getting two heads on two flips are one in four, or one-half times one-half.[17]

The prosecutor in *Collins* continued by having the mathematician apply the product rule to the case at hand. The mathematician assumed probabilities for the individual characteristics which the accused couple supposedly had in common with a robber couple. Thus, the chances of a partly yellow automobile was assigned a probability of 1/10; a man with a mustache, 1/4; a woman with a ponytail, 1/10; a woman with blonde hair, 1/3; a black man with a beard, 1/3; and an interracial couple in a car, 1/100. Using the product rule, the instructor multiplied these odds together and concluded that only one couple in 12,000,000 would possess all these characteristics. As the California Supreme Court concluded, "under this theory, it was to be inferred that there could be but one chance in 12 million that defendants were innocent and that another equally distinctive couple actually committed the robbery."[18]

The California Supreme Court reversed and held:

> [T]he prosecution's introduction and use of mathematical probability statistics injected two fundamental prejudicial errors into the case: (1) The testimony itself lacked an adequate foundation both in evidence and in statistical theory; and (2) the testimony and the manner in which the prosecution used it distracted the jury from its proper and requisite function of weighing the evidence on the issue of guilt, encouraged the jurors to rely upon an engaging but logically irrelevant expert demonstration, foreclosed the possibility of an effective defense by an attorney apparently unschooled in mathematical refinements, and placed the jurors

16. *Id.* at 325, 438 P.2d at 36, 66 Cal. Rptr. at 500 (emphasis in original).
17. *See* Fairley & Moesteller, *A Conversation About* Collins, 41 U. CHI. L. REV. 242, 245 (1974). *See also* Braun, *supra* note 2, at 46.
18. 68 Cal. 2d at 325, 438 P.2d at 37, 66 Cal. Rptr. at 501.

and defense counsel at a disadvantage in sifting relevant fact from inapplicable theory.[19]

The court concluded that the statistical technique was inherently flawed for two reasons. First, the prosecution presented nothing to establish the probabilities for any of the relevant characteristics; instead, odds were merely assumed without any attempt to prove what the probabilities actually were. Thus, the prosecution failed to lay a proper foundation for the introduction of the probability in figures,[20] or, put another way, "there was no effort by the prosecution to connect the numbers with the real world."[21] Second, even if the probabilities of the characteristics had somehow been accurately computed, the *Collins* court further concluded that:

> [T]here was another glaring defect in the prosecution's technique, namely an inadequate proof of the statistical independence of the six factors. No proof was presented that the characteristics selected were mutually independent, even though the witness himself acknowledged that such condition was essential to the proper application of the "product rule" or "multiplication rule."[22]

Mutual independence is also called statistical independence:

> Statistical independence between X and Y means that it is neither more nor less likely that Y happens whether or not X happens. For example, in drawing a card from an ordinary pack of playing cards, let X mean the card is an ace and Y mean the card is black. The probability that the card drawn is black (rather than red) is 1/2, whether or not the card is an ace.[23]

Lack of proof on this issue makes the *Collins* statistics fundamentally unsound; certainly no reason exists to believe that the characteristics were independent of each other. Even if a study of the population showed that one woman in four was blonde and one woman in ten had a ponytail, we cannot necessarily conclude that only one woman in forty is blonde *and* wears a ponytail. Perhaps, for some reason, blondes have more of a penchant for ponytails than women with other hair colors. And certainly,

---

19. *Id.* at 327, 438 P.2d at 38, 66 Cal. Rptr. at 502.

20. The California Supreme Court stated:

> First, as to the foundation requirement, we find the record devoid of any evidence relating to any of the six individual probability factors used by the prosecutor and ascribed to him by the six characteristics. . . . To put it another way, the prosecution produced no evidence whatsoever showing, or from which it could be in way inferred, that only one of every ten cars which might have been at the scene of the robbery was partly yellow . . . or that any of the other individual probability factors listed were even roughly accurate . . . . A foundation for the admissibility of the witness' testimony was never even attempted to be laid, let alone established.

Id. at 327-28, 438 P.2d at 38-39, 66 Cal. Rptr. at 502-03.

21. Fairley & Moesteller, *supra* note 17, at 247.

22. 68 Cal. 2d at 328, 438 P.2d at 39, 66 Cal. Rptr. at 503.

23. Fairley & Moesteller, *supra* note 17, at 244. Braun, *supra* note 2, at 67 states, "Two or more events are independent if the probability of the occurrence of one is in no way affected by the occurrence or nonoccurrence of any of the others."

> it must be much more likely that a man has a mustache if he has a beard than if he does not. Beards and mustaches tend to run together. And, if you have a "girl with blonde hair" and a "Negro man with a beard," the chance that you have an "interracial couple" must be close to 1000 times greater than the estimate of 1/1000.[24]

Thus, the *Collins* opinion correctly[25] indicates that the product rule cannot be applied to identifying characteristics unless a valid foundation is first laid for the probability assigned to each of the characteristics[26] and unless the mutual independence of each of the characteristics is established.

B. Genetic Marker Evidence: Independence

The *Collins* caveats apply to the probability evidence derived from the analysis of blood genetic markers;[27] the product rule cannot be applied to the marker evidence unless accurate population frequency data exist and then only if the markers' appearances are statistically independent of each other. A trial presentation, for example, that blood found at the crime scene contained type A blood and PGM factor 1-2[28] and that the blood of the person on trial contained those same two factors would tell the jury little. Instead, the prosecution would no doubt also want to give the jury information about the frequency of PGM 1-2 in the population. An example of the prosecution's showing would be that 36% of the white population has type A blood and 36% of the white population has PGM 1-2. The product rule would then be presented and the conclusion reached that only 13% of the white population has both type A and PGM 1-2. As *Collins* teaches, however, this conclusion is only supportable if both the elementary percentages are accurate and the factors are statistically independent.

A 1978 study has made a critical examination of mutual independence.[29] This study typed the ABO system and eleven other markers in

24. Fairly & Moesteller, *supra* note 17, at 245.

25. Almost everyone who has reviewed the case points out two fundamental problems with the probability of 1 in 12 million. The elementary probabilities are of dubious origin since the frequencies of the six characteristics are not accurately tabulated for any relevant population and, furthermore, these characteristics are not statistically independent.
Probability, *supra* note 7, at 28.

26. This probability may also be called a population frequency figure since the probability denominates how often the characteristic occurs in a relevant population. *See supra* text accompanying note 10.

27. "A first look at *Collins* thus reveals two requirements for the introduction of statistical analysis in evidence: the prosecutor must introduce evidence as to the probabilities of the individual factors and of the relations among them." Finkelstein & Fairley, *supra* note 2, at 492.

28. *See supra* note 9.

29. Grunbaum, Selvin, Pace & Black, *Frequency Distribution and Discrimination Probability of Twelve Protein Genetic Variants in Human Blood as Functions of Race, Sex, and Age*, 23 J. FORENSIC SCI. 577 (1978) [hereinafter cited as *Frequency Distribution*]. The authors state that they were seeking to verify that "[t]he phenotypic frequencies are statistically independent . . . ." *Id.* at 579.

blood samples taken from a large population. The samples were then evaluated statistically to see if each identified marker was independent of the others. For practical purposes, the study concluded, these genetic markers were mutually independent.[30]

Although concluding that the markers were independent of each other, however, the report found that blood genetic markers are statistically linked to race or ethnic background.[31] According to this study, although 35.8% of the white population (males and females of all ages combined)[32] have type A blood, only 25.9% of blacks do, and although 35.6% of the white population has PGM 1-2, only 29.5% of blacks have that phenotype.[33] Thus, 12.7% of the white population has both these factors, while only 7.6% of blacks do. Consequently, the authors concluded that the difference in phenotypic frequencies means that discrimination probabilities must be computed separately for each ethnic group.[34]

This study does establish, however, that for each racial or ethnic group the phenotypic frequencies are mutually independent. Consequently, for each race the frequency of each genetic factor can be multiplied by the frequency of each of the other phenotypes. The resulting product indicates the frequency of the joint appearance of those genetic blood markers in the population by race or ethnic group. For instance, if 36% of the white population has type A blood and 36% of that popu-

---

30. [I]t may be concluded that these . . . genetic variants behave statistically independently. The group specific component (Gc) showed significant association in five situations. The Gc system is associated with erythrocyte acid phosphatase (EAP) systems for both whites and Chicano/Amerindians, the ABO system in Chicano/Amerindians, and the PGD [6-Phosphogluconate dehydrogenase] system in both blacks and Asians. Nevertheless, the contingency coeffficients are rather small and no substantial error would be incurred by assuming the Gc system to be statistically independent from the other eleven systems investigated.

Id. at 583-84.

31. The genetic marker determinations are made separately for four ethnic groups denominated white, black, Chicano/Amerindian, and Asian. The report concluded, "Some variants showed sizable differences in phenotypic frequencies among the four ethnic groups . . . and other frequencies show less notable differences . . . . Nevertheless, all differences were highly statistically significant . . . for the twelve genetic variants." *Id.* at 581. The connection between ethnic groups and genetic markers has been noted by many investigators. See the studies collected in A. MOURANT, A. KOPEC & K. DOMANIEWSKA-SOBCZAK, THE DISTRIBUTION OF THE HUMAN BLOOD GROUPS AND OTHER POLYMORPHISMS (1976). *See also* NATIONAL CENTER FOR HEALTH STATISTICS, PUBLIC HEALTH SERVICE, U.S. DEP'T OF HEALTH, EDUCATION, AND WELFARE, SELECTED GENETIC MARKERS OF BLOOD AND SECRETIONS FOR YOUTHS, 12-17 YEARS OF AGE: UNITED STATES (1980) [hereinafter cited as PUBLIC HEALTH SERVICE].

32. This study concluded that there was no correlation between the age of the blood donor and the genetic markers. *See Frequency Distribution*, *supra* note 29, at 581. The study also concluded that no correlation between sex and the phenotypes occurred except for one genetic marker system. "[G]lucose-6-phosphate dehydrogenase (G-6-PD) is a sex-linked variant and is expected to have different phenotypic frequencies between males and females . . . ." *Id.* at 580.

33. *Id.* at 582, table 7.

34. *Id.* at 583. The discrimination probability is a different probability than the frequency figure which states what percentage of the population has a specific set of phenotypes. Instead, the discrimination probability is the probability that two randomly selected people will have the same phenotypes. *Id.* at 577. The discrimination probability requires, as does the product rule, that the phenotypic frequencies be statistically independent. *Id.* at 579.

lation has PGM factor 1-2, the two percentages can be multiplied together to reach the conclusion that 13% of whites will have both those factors. Putting it another way, a 13% chance exists that randomly selected persons from the white population will have type A *and* PGM 1-2.[35]

As *Collins* indicates, however, this final percentage is correct only if the initial percentages are accurate. Thus, to use the product rule in this way, one must first know exactly what percentages of the population have type A blood and PGM 1-2.

C. Genetic Marker Evidence: Frequency Data

Since the discovery of polymorphism in blood, many studies have catalogued the frequencies of genetic markers in various ethnic and racial groups throughout the world.[36] Even so, the data collected for the United States have been of questionable reliability; most of the samples have come from nonrandom sources and therefore might not truly reflect the population as a whole. As recently as 1980, a survey of this literature by the United States Public Health Service reported:

> Since E.B. Ford first defined the term "polymorphism," there has been a growing recognition of the existence of numerous polymorphic genetic markers. . . . The differences in frequencies observed in various racial groups and geographic areas have been published in volumes of tables and individual reports. These frequencies have been based on evaluations of numerous groups from various countries. However, very few samples have been random or systematic, and none have been systematic representative samples of the U.S. population.[37]

This 1980 survey reported that even for the most studied of the genetic markers, the ABO system, no data based on a representative sample from the U.S. as a whole or even any one region had been compiled. Instead, the best frequency tabulations were from "a conglomeration of small samples from heterogeneous sources, some dating as far back as 1907."[38]

The reliability of such statistics is open to question for several reasons. The authors noted:

> The importance of current-based statistics cannot be overemphasized. In addition to the fact that blood typing techniques were not as accurate in some of the early studies as they are today, [two

35. "[T]he probability that a randomly chosen individual of given ethnic group possesses the same blood phenotypes as found in a predetermined sample of blood . . . is simply the product of the phenotypic frequencies of the phenotypes considered in the predetermined sample . . . ." *Id.* at 586.

36. The most complete collection of such studies is in A. MOURANT, A. KOPEC & K. DOMANIEWSKA-SOBCZAK, *supra* note 31.

37. PUBLIC HEALTH SERVICE, *supra* note 31, at 2.

38. *Id.* at 3.

researchers] found secular differences reaching statistical significance in samples taken in the same hospital in Iowa over 17 different years. In another study of blood group frequencies of donors from two different donor organizations in the same city during the same time period (Baltimore, July-August 1963), the two samples yielded significantly different distributions for white donors.[39]

Frequency statistics for other genetic markers present even more problems because often the only samplings have been from limited populations. "The blood frequency data on many other blood components depend on statistics from the British population, Norwegian samples, Milwaukee blacks . . . , or Seattle donors . . . , representing the 'best' statistics available."[40] Moreover, the figures often have been compiled from blood donors. The randomness of such samplings is questionable.[41] As recently as 1980 the Public Health Service concluded:

> In summary, although some genetic markers systems have been known for longer than 75 years . . . no reliable frequency data exist on population samples for the U.S. population as a whole, for various U.S. geographic regions, or for sub-classifications by age or sex. Many frequencies reported are derived from blood donor samples that often are not representative of the total population . . . .[42]

The Public Health Service unsuccessfully tried to rectify these flaws by determining the frequencies of genetic markers from a fair sampling of the United States population. These markers were analyzed in the blood taken from "a proportionately representative sample of non-institutionalized youths aged 12-17 years in the U.S. population."[43] This study, however, has a major limitation for forensic purposes: only a few genetic markers were typed. Such markers as EAP and PGM, which have been used in some criminal cases,[44] were not identified by the Public Health Service. Thus, the questions remain about the reliability of the population frequency figures for many of the important genetic marker systems.[45]

One group of researchers, however, recently has set out to amass the necessary information so that forensic scientists can derive valid

39. *Id.* at 4.

40. *Id.*

41. "Although donor samples have often been used for obtaining 'population frequencies,' especially in Great Britain and other European countries, for the United States the representativeness of such samples is open to question." *Id.*

42. *Id.*

43. *Id.* at 1.

44. For example, EAP was typed in State v. Hampton, 294 N.C. 242, 239 S.E.2d 835 (1978); PGM was typed in People v. Stephens, 81 Cal. App. 3d 744, 146 Cal. Rptr. 748 (2d Dist. 1978).

45. The blood genetic marker systems typed by the Public Health Service were the ABO system, the Rhesus (Rh) system, transferrin haptoglobin, and group specific component. PUBLIC HEALTH SERVICE, *supra* note 31, at 2.

probabilities from the genetic markers.[46] This group, lead by Benjamin W. Grunbaum and operating under a grant from the California Office of Criminal Justice Planning, has undertaken "a large-scale study . . . to establish statistical data of a number of genetic variants in human blood."[47]

Although the two resulting studies are large-scale, they are still subject to some of the criticisms levied by the Public Health Service at earlier works. The samples were collected from blood donors in California only. These researchers realized, however, that such a collection procedure was not enough to ensure a sufficient number of donors from certain ethnic groups. Consequently, the group obtained Asian samples from the Blood Bank of Hawaii and Mexican samples from the Banco Central de Sangre of Mexico City. Because of these additional sources, the researchers concluded, "The data were collected so that adequate numbers of individuals were present for four ethnic groups (white, black, Mexican, and Asian)."[48] The group went on to note that the compiled data could apply strictly to the studied population only, but the study concluded that, in a practical sense, the information could be applied nationally:

> [I]t is likely that these results generalize to most of the United States population with minor variation. The minor variation stems from the fact that samples from other parts of the United States may have phenotypic frequencies that are slightly different for the four major California-Hawaii groups.[49]

If this conclusion is correct, then the Grunbaum group's figures should correspond to the Public Health Service data for the genetic markers that both studies typed. Such a comparison indicates that the two sets of frequencies are in general agreement, suggesting that the Grunbaum group's data need not be limited to the specific populations sampled.

Two caveats, however, are essential. First, the frequencies compiled by each study are seldom, if ever, the same. Thus, the Public Health Service indicates that 51.9% of the white population has Gc Sys-

---

46. These researchers noted:

> Civil and criminal investigations often employ analysis of human blood group data as a tool for identification. When two blood samples compared for identification do not match, the innocence of an accused person is established. When two compared samples match, the accused person is not eliminated from suspicion on the basis of blood analysis. In this case, mathematical probabilities are the best method available to assess the likelihood associated with the failure of genetic evidence to give a conclusive answer. The calculation of the probabilities depends on accurate knowledge of the phenotypic or gene frequencies of the genetic variants being compared . . . . A prerequisite for useful probabilities is reliable estimates of the phenotypic and gene frequencies from specific populations.

Gene Frequencies, *supra* note 11, at 428-29.

47. *Frequency Distribution*, *supra* note 29, at 579.

48. *Gene Frequencies*, *supra* note 11, at 429.

49. *Frequency Distribution*, *supra* note 29, at 585.

tem factor 1-1,[50] while in the two Grunbaum group's studies that same frequency is listed at 50.2%[51] and 51.1%.[52] This indicates that none of the figures can be considered as the precise population frequency; instead, each should be treated as a reasonably accurate approximation. This variance is important to the application of the product rule. If a die is accurately constructed, the chances of rolling a "2" on any toss is precisely one in six, and therefore the product rule states that the chances of tossing two consecutive "2's" is *exactly* one in thirty-six. However, if two genetic markers are identified in a blood sample, and a frequency table indicates that each factor appears in 40% of the population, it would not be correct to conclude that precisely 16% of the population had both factors. Instead that resulting product should be treated only as a good approximation of the combined frequency of the two factors. Such a resulting product must be treated with special caution when only a few markers have been identified. For example, if we think of the 40% figure for each of two identified markers as an approximation which more accurately indicates that the true frequency lies in a range between 35% and 45%, then the frequency for the two markers in combination is between $.35 \times .35$, or 12.2% and $.45 \times .45$, or 20%. That eight percentage point range could, of course, be significant. The problem resulting from the approximate nature of the frequency figures, however, lessens as the number of genetic markers identified in a blood sample increases. For instance, assume six markers are identified, each with a supposed frequency of 40%, which truly indicates a range from 35% to 45%. The resulting products would indicate the combined frequency lay between $.35^6$ and $.45^6$, or from 0.2% to 0.8%, or a range slightly over half a percentage point.

The second caveat concerns rare phenotypes. Although different population surveys may each give a frequency figure within a percentage point of the other, if the frequency is rare enough the magnitude of difference between the frequencies given may be very large. Thus, the Public Health Service reported that in whites, rare phenotypes of the Gc system occur in 0.1% of the population.[53] Grunbaum's group, however, reported 0.4%[54] and 0.9%.[55] Thus, although the given frequencies are all within a percentage point, Grunbaum's group indicates that this rare phenotype occurs four to nine times as often as that reported by the Public Health Service. Of course, any product involving the rare phenotype figures will vary up to nine-fold depending upon which figure is selected. Once again, the problem does not seem acute if a large number of markers are identified. For example, if six markers are

50. PUBLIC HEALTH SERVICE, *supra* note 31, at 25 fig. 9. *See also supra* note 30.
51. *Frequency Distribution*, *supra* note 29, at 583 table 7.
52. *See Gene Frequencies*, *supra* note 11, at 433 table 3.
53. PUBLIC HEALTH SERVICE, *supra* note 31, at 26.
54. *Frequency Distribution*, *supra* note 29, at 583 table 7.
55. *See Gene Frequencies*, *supra* note 11, at 433 table 3.

found, one being the rare one with a frequency between 0.1% to 0.9%, and the other five with frequencies of 40%, the probability of all six appearing in conjunction ranges from 0.001% to 0.009%. Either percentage is so small that each is likely to have the same effect on a trier of fact. However, if only the rare phenotype and one other marker with a 40% frequency are identified, the resulting products indicate that the combination occurs either thirty-six times in 1000 or four times in 1000. Those additional twenty-two people might well make a difference to a juror.

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 criticisms about the use of the product rule as in *Collins* do not apply. Consequently, a valid probability can be computed for the joint occurrence of genetic markers by simply multiplying the individual frequencies together.[56]

D. Genetic Marker Evidence and the Product Rule in the Courts

Although computing probabilities from genetic marker phenotyping now has a solid basis, the courts which have admitted such testimony have shown little concern for checking this foundation. In fact, even though some of the trials admitting such evidence were held prior to the publication of the data compiled by the Public Health Service and the Grunbaum group, no case discussed whether the various markers are mutually independent and only a few decisions indicate the sources of the frequency figures.[57] Indeed, the courts generally are un-

---

56. Thus, one study concerned with the probabilities derived from genetic marker evidence has summarized:

> The same issues debated in the case of the People versus Collins arise when blood found at the scene of a crime matches the blood of a defendant. However, evidence from the analysis of blood group systems provides an opportunity to employ valid probabilistic arguments in criminal identification. Blood group genetic variants often have well defined frequencies and are distributed in a statistically independent manner. These two properties, lacking in the Collins case, make it possible to calculate useful quantitative measures of the strength of genetically derived circumstantial evidence.

Probability, *supra* note 7, at 28.

57. *See, e.g.*, State v. Anderson, 308 N.W.2d 42 (Iowa 1981); State v. Fulton, 299 N.C. 491, 263 S.E.2d 608 (1980); State v. Hampton, 294 N.C. 242, 239 S.E.2d 835 (1978). *See also* People v. Bush, 103 Ill. App. 3d 5, 430 N.E.2d 514 (1981). There the defendant challenged the admission of the genetic marker evidence, but the appellate court was unsure of the precise basis of the attack. The court stated, "Defendant also asserts that this evidence was inadmissible for failure to lay a proper foundation, although the specific nature of this argument is not clear." *Id.* at 13, 43 N.E.2d at 520. The court concluded that the defendant was attacking the scientific tests, not the statistics.

aware of the lessons of *Collins*. In one case, however, the reviewing court at least partially delved into the foundation for the statistics presented at trial.

In *State v. Rolls*,[58] a Maine case, a bloodstain on the defendant's pants was analyzed and the ABO, EAP, and PGM factors were classified. The FBI agent who did the analysis testified:

> that approximately 35% of the Caucasian population of the United States possesses type A^1 blood in the ABO classification, that studies done by Scotland Yard reported that approximately 37% of those studied possessed type 2-1 in the PGM group, and that research done at the FBI laboratory showed that approximately 39% of the population possesses the EAP characteristic BA.[59]

The agent then went on to describe the product rule and concluded that "approximately 5% of the population, or one person in twenty, would possess all three blood characteristics which the victim possessed and which were present in the bloodstain on the defendant's pants."[60]

The defendant attacked the resulting conclusion, contending that there was an "insufficient foundation to support admission of this testimony because the witness based his opinion on hearsay about the percentage of the population possessing each of the three blood grouping characteristics in issue."[61] Holding that Maine Rule of Evidence 703 controlled, the Supreme Judicial Court of Maine stated that the evidentiary ruling was primarily a matter for the trial court's discretion.[62] The trial judge was not bound by the expert's assertions that the underlying data generally are regarded as reliable. The court also held that it was insufficient for an expert merely to show that he relies on such data for the preparation of his testimony. The data must be the sort relied upon by experts for purposes other than testifying.[63]

The Public Health Service study indicates that experts would not have accepted the data as reliable. At the time of the *Rolls* opinion no published study contained a representative sampling of the genetic markers in any section of the United States or for the country as a whole. Clearly, the Scotland Yard figures would not meet the objections listed by the Public Health Service since that data could not be representative of this country. In addition, no information was provided on how the British or the FBI frequency figures were compiled. Because nothing indicates that the FBI expert's statistics contained a

58. 389 A.2d 824 (Me. 1978).
59. *Id.* at 827.
60. *Id.*
61. *Id.* at 829.
62. *Id.* The court stated that the rule provides: "The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to him at or before the hearing. *If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.*" *Id.* (emphasis added by the court).
63. *Id.* at 829-30.

fair sampling of the United States population, and certainly not of Maine, the experts in this field, at least as indicated by the Public Health Service, would have regarded the statistics as unreliable.

Nevertheless, the Supreme Judicial Court of Maine upheld the admission of the statistical evidence:

> Before admitting the witness' testimony, the presiding justice heard evidence from which he could have found as a preliminary fact that the blood grouping surveys from which Agent Spalding drew were of a type which were reasonably relied upon by those in the field. On this record, we are unable to say that there was no proper foundation to support admission of Agent Spalding's testimony.
>
> The Defendant's objection to this testimony was more properly directed to its weight rather than its admissibility.[64]

Other cases have also relied on figures compiled by law enforcement agencies.[65] Of course, if these statistics have not been collected from a random sample of a relevant population, the derived probabilities can be challenged. Reliance on the law enforcement compilations present additional concerns, however. First, nothing indicates that such data have been published. Without publication, the scientific community will not have had the opportunity to analyze the study's methods or results. Thus, the reliability of this information will not have been examined by those who are most qualified to do so.

Second, the ability of police forensic labs to do accurate phenotyping is doubtful. In 1979 several persons raised a concern about the reliability of drug testing in those facilities. Responding to these concerns, the Law Enforcement Assistance Administration sponsored a broad and systematic test of the proficiency of American crime laboratories. Crime laboratories from throughout the United States participated in the program. The results of these tests were disconcerting: "It is an understatement to say that the findings of the Test Program are alarming. 'Shocking' would be more precise."[66] In blood analysis testing the report indicated that of the 132 laboratories which responded, ninety-four gave "unacceptable" responses. This represents an "unac-

---

64. *Id.* at 830 (citations and footnotes omitted).

65. For example, in People v. Young, 106 Mich. App. 323, 308 N.W.2d 194 (1981), the state's expert who worked for the Michigan State Police, testified that the use of the product rule was based upon population frequencies compiled from a research project directed by the witness

> involving 1,000 persons in the Detroit area for the purpose of determining the frequency that certain blood factors would be found in the general population. The results were checked internally for their statistical significance and were also compared to figures that had been established in other parts of the United States and around the world.

Id. at 326, 308 N.W.2d at 195. In State v. Carlson, 267 N.W.2d 170 (Minn. 1978), the probabilities were derived from statistics collected by the Minnesota Bureau of Criminal Apprehension "which were in turn correlated with national and international blood-type records." *Id.* at 172 n.2.

66. Imwinkelried, *A New Era in the Evolution of Scientific Evidence—A Primer on Evaluating the Weight of Scientific Evidence*, 23 WM. & MARY L. REV. 261, 268 (1981).

ceptable" rate of 71.2%.[67]

This discussion about the product rule and genetic marker evidence reveals two important points. First, even though the admission of such statistical evidence in past cases may have been subject to challenge both because the necessary mutual independence had not been demonstrated and because the data underlying the frequency products did not accurately represent a relevant population,[68] the courts admitted this evidence without consideration of these concerns. Courts and litigators appear unaware of even the basic lessons of *Collins* and of elementary statistical requirements. Second, as we have seen, researchers now have established that genetic blood markers do appear independently from each other. Furthermore, phenotypic frequencies adequately grounded in the real world have been compiled. Thus, the two glaring flaws present in *Collins* are absent in this newly developed forensic evidence. The product rule can be applied in the future to the individual marker frequencies with a statistically valid result. Thus, if three markers are identified in a blood sample, and the frequency data indicate that each occurs in 10% of the population, it is correct to conclude that the combination of those markers appears at a rate of one person in every thousand.

This conclusion, however, does not end the inquiry. A question much harder to resolve must now be faced. Granted that a statistically acceptable product can be generated by genetic marker phenotyping, what is the proper use of this product, this probability, this frequency, in a criminal case?

IV. Application and Misapplication of Frequency Evidence—The Subclass

A hypothetical situation will be useful in judging the proper uses of the frequency evidence.[69] Assume a New York City lawyer is found slain in the stairwell of the top floor of her Wall Street office building. The medical examiner concludes that she was killed by a .32 caliber bullet and died almost instantly with death put at 4:00 p.m. on a

---

67. *Id.* at 269.

68. Even if reasons existed to exclude the statistics from some earlier cases, nothing indicates that the defendants were harmed by their admission. Thus, in State v. Rolls, 389 A.2d 824, 827 (Me. 1978), the expert testified that approximately 5% of the population had the combined markers of ABO type A_1, EAP type BA, and PGM 2-1. The 1980 figures compiled in *Gene Distribution*, *supra* note 11, at 432-34 table 3, indicate a joint frequency for those phenotypes of 4.7% of the white population and 2.2% of the black population. (Type A blood consists of two variations, type A_1 and type A_2.) The 1978 study in *Frequency Distribution*, *supra* note 29, at 582-83 table 7, only reported the results as type A without giving separate frequencies for the two major subgroups. *See* Jonakait, *supra* note 5, at 836, n.12. In People v. Young, 106 Mich. App. 323, 308 N.W.2d 194 (1981) and State v. Carlson, 267 N.W.2d 170 (Minn. 1978), the courts did not state what phenotypes were identified in the blood samples so it is not possible to compare the frequencies presented in those cases with a product derived from the Grunbaum group figures.

69. This is an adaptation of Finkelstein and Fairley's hypothetical. *See* Finkelstein & Fairley, *supra* note 2, at 496-97.

Wednesday. Her purse is missing. Some of her clothes are in disarray and other signs of a struggle are evident. Near her body is found a penknife she was known to carry. Dried blood found on the knife is examined for genetic markers. Police determine that the victim has different phenotypes than that found in the dried blood. Using frequency tables and the product rule, investigators conclude that the combination of the identified markers occurs at the rate of one person in every 1000.

The police interview most of the 400 employees of the firm. No one saw the victim enter the stairwell or saw anything suspicious on the four floors occupied by the victim's law firm. The police learn that the victim had a boyfriend and obtained a picture of him. The elevator operator for the building tells the police that he saw the boyfriend leave the building at about 4:00 to 4:30 p.m. on the day of the murder.

The police investigate further and find out the relationship between the boyfriend and the victim was stormy. When the police go to the boyfriend's apartment, the landlord tells them that the boyfriend, on the night of the murder, paid the landlord an extra month's rent in advance, asked him to keep an eye on the apartment, and told the landlord that he was not sure when he would be returning.

A month later, the police find the boyfriend in Chicago and bring him back to New York. His blood is analyzed and found to contain the same genetic markers as the blood on the knife. He is charged with murder and put on trial.

The prosecutor introduces expert testimony about the blood analysis and seeks to present to the jury the frequency data, to have a statistical expert explain the product rule, and to present to the jury the conclusion that only 0.1% of the population has blood containing the identified genetic markers. The defendant objects to this statistical presentation claiming that whatever slight probative value this evidence may have is outweighed by the prejudicial impact that the figures will have on the jury.

Certainly the frequency conclusion could be expected to have impact on the jury. Without the statistical data, the prosecution probably can not produce enough evidence to convince a jury that guilt has been established beyond a reasonable doubt. With the statistics, however, the jury probably would convict.[70]

A. Probative Value of Frequency Evidence

The use of the genetic marker evidence in this hypothetical necessitates two conclusions. First, the dried blood was not that of the vic-

70. *See id.* at 497 (the authors conclude about their hypothetical, "Without the print evidence [which was limited so that the expert could only conclude that such prints appear in no more than one case in a thousand] the case probably does not go to the jury. With it the jury probably convicts.").

tim. Second, the blood could have come from the defendant; the genetic analysis does not exclude the possibility that he was the source of the blood.[71] These conclusions, however, are clear from the genetic analysis without reference to the frequency computation.

A much harder issue is what weight should be given to the similarity between the dried blood and the defendant's blood and the statistics indicating that these characteristics occur in only one out of every thousand people. One possible conclusion from this figure is that there is only one chance in a thousand that the blood came from someone other than the defendant. If the evidence is taken that way, which the average juror might be inclined to do, the statistical presentation will be taken as strong evidence of guilt. The juror would then conclude that science has established that it is 99.9% certain that the defendant's blood was found at the scene.

Such reasoning, however, would be a gross misinterpretation of the statistics. The frequency figure does not give the odds that it was the defendant's blood at the scene. Instead, the figure states that out of every thousand people who might have committed the crime, only one, on average, has blood with the identified characteristics; or stating the same thing, out of every 100,000 people who might have been at the crime scene, only 100 have blood with the same genetic factors found on the knife. Viewed this way, the statistic is not a probability indicating defendant's guilt; instead, it is merely a measure of the number of people who might have left their blood at the murder scene. Conversely, these figures indicate how much of the population can be excluded from consideration. Thus, the statistic allows a way to calculate the number of people in a shortened list of suspects, one of whom is the defendant.

The same statistical problems are presented whenever the evidence produces such a probability. Neutron activation analysis (NAA) gives another example. NAA technology allows identification of the basic chemical elements present in a material sample. In one case, blue specks of paint found on a tire tool used to commit a crime were sub-

---

71. Some commentators have stressed that blood grouping in criminal cases can only prove a negative. For example, "Again, it must be emphasized that these results have solely a *negative* value—the tests can only prove that the sample is *not* the blood of the suspect." Sussman, *Supplementary Applications* in PATERNITY TESTING BY BLOOD GROUPING 153 (L. Sussman ed. 1976).

> Before discussing the various tests and analysis of blood specimens one point must be made perfectly clear. Blood as evidence is what is referred to as a negative-positive. No statement can be made that two blood samples are identical, but the serologist can state at times that two blood specimens are not identical. For example, if the blood on the clothing of a suspect is of group A and the blood group of the victim is also group A, the only positive statement that the serologist can validly make is that both bloods are of the same *blood group*. He cannot state that the two specimens came from the same origin. However, if the blood on the clothing of the suspect is group B and that of the victim is group A, the serologist may then state that the blood on the suspect's clothing could not have come from the victim.

MOENSSENS & INBAU, *supra* note 3, § 6.09, at 293 (emphasis in original). *See also* Annot., 2 A.L.R. 4th 500, 506 (1980).

mitted to neutron activation analysis.[72] The analysis identified five elements in the paint. These elements also were found in the same concentrations in the blue paint on defendant's car. The expert also testified that other tests in his laboratory had shown that:

> [T]wo similar-type paints of similar color but of different brands would contain a given element in concentrations within 10 percent of each other only about one time in five. If one assumes that the chance of finding a given amount of one element is independent of the amounts of any other elements present, then the chance that similar-type paints of similar color but of different brands will match concentrations with 10 percent in each of five elements is $(1/5)^5$—one chance in 3,125 or 0.03 percent.[73]

One commentator analyzed the probative value of this frequency figure:

> The light-blue paint comparison indicates that if 10,000 light blue cars which are painted with a brand of paint different from that of the defendant's car are chosen at random, then only three of them will be found to match the paint on defendant's car within 10 percent in all five elements. If this were the only consideration, and if it could be shown that far fewer than 10,000 light-blue cars, all painted with brands of paint different from that used on defendant's car, could conceivably have been the supplier of the tire tool, then the evidence against the defendant is quite powerful. On the other hand, if more than 10,000 light-blue cars might have been involved, there are three in each 10,000 that would give the same activation analysis results when compared with the specks on the tire tool as the defendant's car did.
>
> There is a much more important consideration, however. Even if the activation analysis could *positively* identify the brand of paint, that is, even if it were 100 percent certain that the speck of paint on the tire tool were of the same brand as the paint on defendant's car, the NAA evidence in itself still does no more than reduce the number of possible culprits to those owning or somehow associated with light-blue cars painted with that brand of paint. Therefore, the NAA evidence only points up the defendant as being a member of a restricted subclass of the general population. That the defendant is singled out only as a member of a restricted subclass is a *typical result* of NAA evidence, although the *size* of the subclass varies greatly with the type of thing being compared. Before the NAA evidence can be meaningfully evaluated, it is necessary in each case to determine the size of the subclass.[74]

The same conclusion holds true for the genetic marker evidence.

72. People v. Woodward, No. 108551 (Cal. Super. Ct., San Mateo County, July 7, 1964).

73. Note, *The Evidentiary Uses of Neutron Activation Analysis*, 59 CALIF. L. REV. 997, 1015 (1971) [hereinafter cited as *Evidentiary Uses*].

74. *Id.* at 1016.

The blood analysis in our hypothetical merely indicates that the defendant is a member of a restricted subclass of the population and other evidence indicates that a member of that restricted subclass was at the crime scene. If the population from which the killer would come were small, say 1,000 persons, then the genetic frequency evidence would indicate overwhelming odds that only a few people were in the restricted subclass of people with the identified genetic markers. Since defendant is one of those people, the blood analysis and the statistical conclusions drawn from it would be powerful evidence of guilt. On the other hand, if the suspect population were as large as 100,000 persons then the restricted subclass could be expected to have about 100 members, of whom the defendant is one. Since the genetic analysis does not indicate that any one person in the subclass is more likely to be guilty than any other person in the subclass, the blood analysis would only lead to the conclusion that there is one chance in a hundred that the defendant is guilty.[75] This approach then

> makes the number of suspects critical. Determining this number, however, will usually involve wholly arbitrary decisions. Shall it include only those in the same neighborhood, the same city, state, or the entire country? The jury might be given a range of choices and the probability associated with each choice, but jurors cannot rationally choose when, as is usual, there is no evidence bearing on this issue.[76]

Who, then, should be included in the suspect population for the hypothetical murder? The answer given by a noted scholar is "the approximate number of people who had the opportunity to commit the offense and lack a satisfactory alibi."[77] Who in our hypothetical fits that definition? Perhaps the answer is anyone in the stairwell with the victim at the time of the killing. Of course, if we knew or could ascertain that information, we would seem to know the killer already. Instead, perhaps the suspect population includes anyone on the victim's

---

75. *See* George, *Use and Misuse of Scientific Evidence* in SCIENTIFIC & EXPERT EVIDENCE IN CRIMINAL ADVOCACY 10 (J. Cederbaums & E. Arnold eds. 1975) ("It is also necessary to inquire into the size of the population which is correctly brought within the class to be identified. If the defendant falls within a very small group, then the scientific evidence increases significantly the weight of the circumstantial evidence against him. However, if the population sample is very large, the scientific evidence has little or no relevancy.").

76. Finkelstein & Fairley, *supra* note 2, at 495. *See also* Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971). In discussing Finkelstein and Fairley's hypothetical Professor Tribe states:

> By itself, of course, the "one-in-a-thousand" statistic is not a very meaningful one. It does not, as the California Supreme Court in *Collins* showed, measure the probability of the defendant's innocence—although many jurors would be hard-pressed to understand why not. As Finkelstein and Fairley recognize, even if there were as few as one hundred thousand potential suspects, one would expect approximately one hundred persons to have such prints; if there were a million potential suspects, one would expect to find a thousand or so similar prints. Thus the palm print would hardly pinpoint the defendant in any unique way.

Id. at 1355.

77. Williams, *The Mathematics of Proof*, 1979 CRIM. L. REV. 297, 351 n.7.

floor near the time of death since all of them had the opportunity to kill her. The suspect population might then be 100 people. However, the killer may have gone up or down the stairway and never appeared on the floor of the killing; then perhaps the list of suspects should include anyone in the building near the time of death, say 5,000 people. Then again, it would have taken only a few minutes for anyone working in the Wall Street area to leave their employment and go to the crime scene. All of these people also had the opportunity. Of course, no reason exists for limiting the possibilities of people within a few minutes of the crime scene. If instead, all those within an hour's travel time from the crime scene are considered to have had the opportunity, there might be as many as ten or fifteen million in this group. And, with jet travel, there is really no logical reason to limit the notion of opportunity to the thirty miles or so that surround Wall Street. The only way to place any upper limit on the suspect population is by making assumptions or arbitrary choices. Of course, this gets us back to one of the problems of *Collins*; statistics should not be introduced into a criminal trial when based on unproven assumptions. The limitation on the suspect population requires exactly that.

Furthermore, even if it is reasonably concluded that the killer must have lived or worked at the time of the killing within an hour's travel of Wall Street, and census figures tell the trier of fact how many people are within this location, at least some of those people would be infants or invalids or have completely verifiable alibis. Therefore, none of these people could be the killer. The calculation of the number to be weeded out from the suspect populations is completely indeterminable; it could only be made by arbitrary choice which also once again leads to the problem identified in *Collins*.

Finally, even if a reasonable upper limit on potential suspects could be ascertained and if that number could be winnowed down to eliminate people who could not have done the killing, the number left would still be so enormous that the statistic would have little probative value. Assume only 5,000,000 people are on the potential suspect list, and that eighty percent could establish that they could not have done the crime. Then 1,000,000 people are still in the suspect population. Now apply the frequency evidence to this population. The number of this suspect population must be multiplied by the frequency of the combined genetic markers, one in 1,000 or .001. This calculation indicates that the blood of any one of 1,000 people could have been left at the scene.

Thus, even if assumptions allow us to limit the suspect population as much as has been done here, the genetic evidence that indicated that defendant has blood like that found at the scene and that only one person in 1,000 has such blood only means that defendant is one of 1000 people who could have committed the crime. Put another way, the blood evidence only proves that there is a 1/1,000 chance defendant's blood was at the scene.

B. Prejudicial Impact

If the jury truly understood that the frequency figure merely helps to define a subclass of suspects, most always quite large, then a defendant would not be harmed by the introduction of the frequency figure. Looked at this way, the statistics only lead a very small way towards the prosecutor's goal of proof beyond a reasonable doubt. If, on the other hand, the jury misunderstands and believes that the frequency data is an indicator of the probability of guilt or the probability that the defendant was at the scene then the jury would view this testimony as strong evidence against the defendant when in fact it is not. In such a circumstance, the probative value of the evidence would in fact be greatly outweighed by its prejudicial impact.[78]

The potential for misuse of frequency or probability evidence is demonstrated by one of the neutron activation analysis cases. As discussed above, paint chips were submitted to NAA and probabilities were derived as to how frequently that composition of paint would appear in other brands. Even though only a frequency was reported by the expert,

> *Time* magazine reported the expert witness to have testified that it was "99.98% [sic] certain" that the tire iron came from Woodward's car and "99.999% certain" that it was used to jimmy the door. If the witness actually said this, it should be clear that his testimony was totally unfounded and highly prejudicial, because the NAA measurement was not capable of singling out the defendant so unambiguously. In any event, that the reporter for *Time* thought the expert said this lends credence to a belief that the jury might have thought so also. Therefore, it is doubtful that the correct interpretation of the evidence was made clear to the jury.[79]

C. Judicial Response

Because the potential is great that juries will misuse the genetic marker frequency evidence, an instruction about the limited probative value of that probability would seem appropriate. No such case has been reported.

Likewise, no case has been reported in which the frequencies generated by the new blood tests have been refused admission because they are too prejudicial. A number of defendants have presented this argument by citing a New York case, *People v. Robinson*.[80] An Illinois court summarized such a challenge to the blood grouping statistics:

78. 6 J. WIGMORE, EVIDENCE § 1864, at 643 (J. Chadbourn rev. ed. 1976).
79. *Evidentiary Uses*, *supra* note 73, at 1018.
80. 27 N.Y.2d 864, 265 N.E.2d 543, 317 N.Y.S.2d 19 (1970).

"Defendant relies on the so-called New York rule which states that such evidence is inadmissible because it lacks probative value in that a large proportion of the population will have the same blood type."[81]

Robinson did not involve the recently developed genetic marker tests. There the defendant was convicted of murder. The prosecution proved that the blood type of the semen found in and on the body of the victim was the same as the defendant's blood type. The New York Court of Appeals indicated that this evidence should not have been admitted:

> Proof that defendant had type "A" blood and that the semen found in and on the body of the decedent was derived from a man with type "A" blood was of no probative value in the case against defendant in view of the large proportion of the general population having blood of this type, and therefore, should not have been admitted.[82]

The New York case is distinguishable from other situations involving the more sophisticated procedures. First, the development of the genetic marker tests allows the blood sample to be linked with a much smaller proportion of the population than was done merely with ABO typing. As that percentage gets smaller, the probative worth of the evidence would appear to increase. Perhaps it therefore can be argued that while the evidence in *Robinson* had little value, that situation has now changed. Second, nothing in the New York opinion indicates that the jury was presented the frequency figure for the appearance of this sort of blood in the population. If that number is presented it can be argued that the jurors can correctly assess for themselves the proper probative value of the evidence.

In any event, this New York approach has been widely rejected.[83] As a Michigan court said, "[t]he overwhelming majority of courts allow the use of blood grouping evidence in criminal trials, both to connect victims with defendants and, of interest here, to show that defendants could have left blood stains on crime scenes."[84] Uniformly, the courts have concluded that the statistics derived from the recently developed genetic marker tests were validly admitted and that arguments about

81. People v. Bush, 103 Ill. App. 3d 5, 13, 430 N.E.2d 514, 520 (1982).

82. *Robinson*, 27 N.Y.2d at 865, 265 N.E.2d at 543, 317 N.Y.S.2d at 19-20. *See also* People v. Macedonio, 42 N.Y.2d 944, 366 N.E.2d 1355, 397 N.Y.S.2d 1002 (1977).

83. Indeed, New York itself seems ready to abandon this rule. Recently the New York Court of Appeals has held that the District Attorney could get an order requiring a blood sample from a potential murder suspect who had not yet been charged. Blood was found at the murder scene which was not the victim's and such blood was found at a rate of less than one in a hundred. In dictum, the court went on to state, "[m]oreover, the relative rarity of the assailant's type of blood relegates arguments as to remoteness to the realm of weight rather than admissibility . . . ." Matter of Abe A., 56 N.Y.2d 288, 299, 437 N.E.2d 265, 271, 452 N.Y.S.2d 6, 12 (1982). In a footnote, the court continued, "Suffice it here to say that *People v. Macedonio* . . . and *People v. Robinson* . . . cited for the contrary proposition . . . each dealt with type A blood, which is found in 40% of the population." *Id.* at 299 n.4, 437 N.E.2d at 271 n.4, 452 N.Y.S.2d at 12 n.4.

84. People v. Horton, 99 Mich. App. 40, 50, 297 N.W.2d 857, 862 (1980).

the weak probative value go merely to an assessment of the weight of the evidence.[85] Indeed, one case has indicated that the limited worth of this evidence is so blatantly apparent that a jury could not be prejudiced by it.

In *State v. Fulton*,[86] a robbery occurred in Winston-Salem, North Carolina. The victim was cut. Dried blood was found on the defendant's clothing. The genetic markers discovered in the dried blood matched the victim's blood. Frequency tables and the product rule led to the conclusion that 11% of the population had this particular combination of genetic factors. On appeal, the defendant argued that the frequency data should not have been admitted because it "could only mislead the jury into believing that the particular blood type allegedly found on defendant's shoes came from an extremely limited source when in fact the source actually encompassed a very large number of people."[87] The North Carolina Supreme Court rejected that contention:

> Had the blood grouping test shown that the blood on defendant's tennis shoes did not belong to the same group as [the victim's] blood, this would have been highly significant in defendant's favor and would tend to exonerate him. On the other hand, since the blood test showed that the victim's blood group was the same as the blood on defendant's shoes, the test was relevant but weakly probative in character because 11% of the population has the same blood type as [the victim]. In a city the size of Winston-Salem, this 11% would encompass several thousand people whose blood could have been on defendant's shoes. The challenged statement is therefore mildly unfavorable to the defendant but essentially harmless because its probative value is so minute. Certainly no prejudice resulted.[88]

The court's analysis is disingenuous. The court's conclusion totally ignored the possibility that the jury misused or misweighed the evidence. Nothing indicates that the jury was informed that the frequency data merely defined a subclass of people of whom the victim was merely one of many. The jury was never told that the evidence only indicated a chance of one in thousands that the dried blood was indeed the victim's. If this limited import of the evidence was not clearly and authoritatively explained, the jury may well have concluded that the probability had a much stronger probative effect. In that event, the jurors would have decided that the blood evidence alone indicated an 89% chance that the dried blood was the victim's. If the

85. *See, e.g.*, People v Young, 106 Mich. App. 3d 323, 328, 308 N.W.2d 194, 197 (1981); State v. Washington, 229 Kan. 47, 622 P.2d 486 (1981). *See also* People v. Gillespie, 24 Ill. App. 3d 567, 321 N.E.2d 398 (2d Dist. 1974); Annot., 2 A.L.R. 4th 500 (1980).

86. 299 N.C. 491, 263 S.E.2d 608 (1980).

87. 299 N.C. at 496, 263 S.E.2d at 611.

88. *Id.* at 496-97, 263 S.E.2d at 611.

jury did misuse it in this way, the evidence, instead of being of little value to the prosecution, would have been taken as strong evidence of defendant's guilt. Of course, we do not know how the jury used the probability evidence. Apparently the court felt that the proper use of the statistic was self-evident. If so, presumably no harm to the defendant resulted. Only speculation and assumptions can tell if the court's conclusion was correct.

The court's analysis, however, indicates to trial judges that such frequency data should not be admitted in the future. The court conceded that probability statistics are weak evidence at best. As we have seen, properly interpreted, the evidence is almost devoid of value. On the other hand, the evidence carries with it the potential for extreme prejudice to the defendant. Surely, then, the prejudicial impact of the evidence greatly outweighs its probative value, and the frequency should not be admitted.[89]

Not surprisingly, scholars familiar with statistics have concluded that this approach of using frequency figures in criminal trials "ought to be abandoned because it is appropriate to extremely few situations, and those can be handled without statistical analysis."[90] Others would allow use of these statistics only when accompanied with a clear explanation of what the frequency means.[91]

---

89. In State v. Carlson, 267 N.W.2d 170 (Minn. 1978), the state produced an expert in a murder case who:

> stated that only .85 percent of the population would have blood with the same combination of ABO, PGM, and EAP characteristics as the victim's blood and the matching bloodstains found on [the defendant] Carlson's jacket. [Another expert] testified that based on his studies there was a 1-in-800 chance that the foreign pubic hairs found on the victim were not Carlson's and a 1-in-4,500 chance that the hairs found clutched in the victim's hand were not Carlson's.

Id. at 175.

On appeal, the defendant raised the issue of whether an expert witness may express his findings in terms of mathematical probabilities. *Id.* at 173. The court first analyzed other cases which had held that statistical evidence should not have been admitted. The court concluded that in those cases an adequate foundation had not been laid for those numbers; here, "[i]n contrast, the foundation for the experts' testimony in the present case was properly laid, based upon empirical scientific data of unquestioned validity." *Id.* at 176. The court, however, did not end its analysis there and went on to state:

> Our concern over this evidence is not with the adequacy of its foundation, but rather with its potentially exaggerated impact on the trier of fact. Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven . . . . Diligent cross-examination may in some cases minimize statistical manipulation and confine the scope of probability testimony. We are not convinced, however, that such rebuttal would dispel the psychological impact of the suggestion of mathematical precision . . . . For these reasons, we believe [the expert's] testimony that there was only a 1-in-800 chance that the foreign pubic hairs found on the victim did not come from the accused and an even more remote 1-in-4,500 chance that the head hairs did not belong to the accused was improperly received.

Id. at 176. The court did not specifically rule on the frequency derived from the blood grouping.

90. Finkelstein & Fairley, *supra* note 2, at 490.

91. In discussing the hypothetical case in which a fingerprint was said to occur once in every thousand people, Professor Tribe concludes:

> To be sure, the finding of so relatively rare a print which matches the defendant's is an event of significant probative value, an event of which the jury should almost certainly be

An alternative method of presenting the evidence without misleading the jury is available. Instead of presenting the frequency data, only the size of the subclass derived from the frequency should be presented to the jury. Thus, in the *Fulton* case the jury would not be told that 11% of the population had blood like the victim's and the blood found on defendant's clothing; instead they would be told that the dried blood could have come from any of thousands of people in the Winston-Salem area and that the victim was merely one possible source from these thousands. This represents the proper use of such evidence. The defendant, therefore, could not be prejudiced by it. Of course, presented in this way the weakness of the evidence is clear, and the essential task of ascertaining the probably indeterminable size of the suspect population becomes apparent.

D. *The Extremely Infrequent Frequency*

Some authorities who have considered the proper use of frequency data in criminal trials, while agreeing that probabilities of one in 1,000 are by themselves misleading indicators of guilt, have indicated that if a small enough probability is validly generated by the product rule, it should be accepted as proof of guilt. Thus in discussing the *Collins* case, Michael Finkelstein and Professor William Fairley, after pointing out that the product presented there represented "the frequency of couples meeting the description of the one placed at the crime," went on to state, "[i]f a sufficiently precise estimate could be made that the frequency of such couples in the Los Angeles area was one in twelve million, it would be possible to state within reasonable margins for error that there was only one such couple in the Los Angeles area."[92]

> informed. Yet the *numerical index* of the print's rarity, as measured by the frequency of its random occurrence, may be more misleading than enlightening, and the jury should be informed of that frequency—if at all—only if it is also given a careful explanation that there might well be many other individuals with similar prints. The jury should thus be made to understand that the frequency figures do not in any sense measure the probability of the defendant's innocence.

Tribe, *supra* note 76, at 1355 (emphasis in original).

Glanville Williams has stated:

> It is very important that when statistical information is presented to an inexpert tribunal like a jury, explanations should be given as to its bearing on the issues. There is always the possibility of statistical evidence of great complexity being placed before the court, which even the judge may be incompetent to assess without assistance, and which may gain unmerited credence by reason of its apparently scientific character.

Williams, *supra* note 77, at 299. Braun, however, states that such a frequency should be admitted when such "evidence is coupled with other evidence linking the accused to the scene of the crime or the victim, the concern over the admissibility of the partial transfer evidence need not be addressed because that evidence would be just one piece among others." Braun, *supra* note 2, at 59-60, 77-80. Braun apparently gives little credence to the notion that such a statistic may be too prejudicial to be admitted because its probative effect may be misweighed by the jury. Instead, he merely concludes, without further analysis, "[t]here are adequate safeguards to protect a factfinder from being overwhelmed by mathematical analysis where the state is attempting to identify the accused and the criminal as the same party . . . ." *Id.* at 62.

92. Finkelstein & Fairley, *supra* note 2, at 494. *See also Probability*, *supra* note 7, at 28 ("It should also be noted that a probability such as 1 in 12 million presented in the *Collins* case, when

On the other hand, even if a statistically valid product of one in 12,000,000 were obtained from genetic markers, the introduction of this frequency still could be misleading. An appendix in *People v. Collins* indicates the reason. The court contended that even with a frequency of one in 12,000,000, if the population which contained the suspect couple were large enough, a good chance exists that more than one couple would possess the suspect's characteristics. The court derived a formula which indicated that if the population were 12,000,000, then the chances were about 41% that another couple besides the accused one would have the fatal traits. The opinion concluded,

> Thus the prosecution's computations, far from establishing beyond a reasonable doubt that the Collinses were the couple described by the prosecution's witnesses, imply a very substantial likelihood that the area contained more *than one* such couple, and that a couple *other* than the Collinses was the one observed at the scene of the robbery.[93]

At first glance the court's conclusion may not be intuitively clear, but an example should help clarify it. Assume that we have 120,000,000 balls; 119,999,990 are black and ten are white. The frequency of the white balls in the combined population is one in 12,000,000. Assume further that all the balls are thoroughly mixed and fed in equal numbers into ten giant urns. Each urn now contains a population of 12,000,000 balls. One of our urns now becomes a representative of the population of the area where the *Collins* robbery oc-

---

derived from well-defined and independent frequencies, could reflect the guilt or innocence of a defendant."). In the context of genetic markers, a statistician has stated that when the chance is less than one in a thousand that randomly selected persons will have the analyzed phenotypes, but the defendant's blood matches, "the inference is often made that the accused is unlikely to be identical for the analyzed phenotypes by chance and, therefore, it is likely that the blood found at the scene of the crime belongs to he defendant." Selvin, *supra* note 9, at 195.

In the context of the *Collins* case, however, Professor Tribe pointed out that

> even if the product rule could properly be applied to conclude that there was but one chance in twelve million that a randomly chosen couple would possess the six features in question, there would remain a substantial possibility that the guilty couple did not in fact possess all of the characteristics—either because the prosecution's witnesses were mistaken or lying, or because the guilty couple was somehow disguised.

Tribe, *supra* note 76, at 1336.

In addition, the calculation of a frequency as small as the one in *Collins* can only confidently be done if extensive empirical data first has been gathered:

> But as a practical matter the court was right to doubt that the prosecutor could show uniqueness. A derivation of such extraordinarily small probabilities with any useful degree of precision would be extremely difficult. In most cases, the estimate of the population frequency of evidentiary traces (of hair or incomplete fingerprints, for example) will have to be made on the basis of samples numbering at most a few thousand. As a result, probabilities of the magnitude involved in *Collins* would require an inference, based on a few thousand trials, that an event would occur once rather than more than once in millions of trials. Such an inference inevitably involves powerful assumptions which cannot be adequately supported without extensive data. Except in cases where the number of suspects is sharply limited, it will almost never be practically possible to gather enough data to sustain a conclusion of uniqueness with any confidence.

Finkelstein & Fairley, *supra* note 2, at 494-95.

93. 68 Cal. 2d 319, 335, 438 P.2d 33, 43, 66 Cal. Rptr. 497, 507 (1968).

curred. We start sifting through this urn and find a white ball. Does that mean it is the only white one present in that urn? Not necessarily—if the balls were randomly deposited in the ten urns, then we can expect some of the urns to contain no white balls, some to contain one, and some to contain more than one. Indeed, it should be clear that even though one white ball has been found, the chances are good that another white one still is in the urn. The formula presented by the *Collins* court attempted to give the probability for this possibility; that is, if one white ball is found in the urn, the odds of finding another one in the same urn. The court's conclusion was 41%.

The mathematical analysis of the *Collins* appendix has been criticized and the criticism illustrates two important points. First, the proper use of statistics is apparently so elusive that even those attempting to teach others about the errors of their ways cannot agree among themselves what those errors are. Thus, one of the reasons the court's formula was not correct, according to Finkelstein and Fairley was that "it assumes a sampling of the population with replacement of the sampled couples, instead of sampling without replacement."[94] The critics stated that "the difference between replacement and nonreplacement is critical."[95] Consequently, the court's conclusion that there is a 41% chance that another couple with the accused's characteristics exists in a population of 12,000,000 couples overstates the odds.[96]

In a later article, Professor Fairley and Frederick Mosteller seem to conclude that the *Collins* court was correct as long as the suspect population was 12,000,000. No mention is made that the court's analysis overstated the odds of finding another couple because the court's approach posited sampling with replacement.[97]

Professor Robert P. Charrow and Robert L. Smith have discussed both the *Collins* appendix and the Fairley and Mosteller article. They agree that the *Collins* court was in error, but conclude that the court may have *under*stated the odds of finding another couple with the robbers' characteristics, and those chances, instead of being 41% may have in fact been as high as 65%.[98]

These articles reveal why lawyers and judges should be wary of the use of statistics in a criminal case. Statistics is such a minefield that

94. Finkelstein & Fairley, *supra* note 2, at 493 n.12.

95. *Id.*

96. As Finkelstein and Fairley stated,

> [The *Collins* court's figure] obviously has nothing to do with the likelihood that a couple answering the description of the accused was correctly charged. For if there was only a single ball in the urn representing a couple with the characteristics of the accused, the court's formula would still yield a substantial probability of duplication (the same ball being picked twice) although by hypothesis the accusation was correctly made.

Id.

97. *See* Fairley & Moesteller, *supra* note 17, at 250 n.13.

98. Charrow & Smith, *A Conversation About "A Conversation About* Collins", 64 GEO. L.J. 669, 676 (1976).

not even those who study the field agree where the traps lay. Since the scholars cannot always reach a consensus, can we really expect judges and juries not trained in mathematics to evaluate properly the admissibility, relevancy, or weight of the statistics? As Charrow and Smith conclude:

> Unfortunately, probability theory is a field unusually rich in paradoxes—paradoxes that are based on "misguided intuition." Consequently, if the courts ever hope to employ mathematical techniques for the evaluation of circumstantial evidence, they must exercise extreme care.[99]

This advice to the courts is meaningless, however, if those professing knowledge about probabilities cannot even agree as to what the errant insights are.[100]

These criticisms highlight a second important point. No matter what the proper formula is for determining the odds of finding a second couple in *Collins*, the size of the population is the key concern. Thus, Fairley and Mosteller calculate that if the frequency of the robbers' characteristics is one in 12,000,000, but the population were only 250,000, then the chances of finding a second couple in addition to the accused pair would be about 1%. On the other hand, if the frequency were one in 25,000 and the population were still 250,000, then the odds of finding the second couple are 2,200 to 1.[101] As the authors conclude, "[T]hat shows how important it is to get a fair estimate of the proportion of robber-like couples . . . ."[102]

This, of course, returns the analysis to the fundamental problem of determining the size of the suspect population which, as discussed above, cannot be established by any statistically valid method: it merely must rest on unverifiable assertions. An unverifiable assertion, as the *Collins* court recognized, cannot be the basis for the admission of statistics into criminal cases.

The genetic marker evidence in the hypothetical case produced a statistically valid frequency of one in 1,000. That frequency, however, furnishes very weak proof because it only helps to define a subclass of suspects and that subclass, even if calculable, an extremely doubtful proposition, would invariably be large. On the other hand, if the true nature of the evidence is not understood by the trier of fact, the frequency evidence may be taken as strong proof against the defendant. If so, the prejudicial impact of the evidence greatly outweighs its pro-

---

99. *Id.* at 677-78.

100. Perhaps this illustrates the point made by Tribe: "The very mystery that surrounds mathematical arguments—the relative obscurity that makes them at once impenetrable by the layman and impressive to him—creates a continuing risk that he will give such arguments a credence they may not deserve and a weight they cannot logically claim." Tribe, *supra* note 76, at 1334.

101. Finkelstein & Moesteller, *supra* note 17, at 251-52.

102. *Id.* at 253.

bative value. Thus, this analysis indicates that the frequency figure should not be introduced into the criminal trial.

V. PROBABILITIES, CRIMINAL TRIALS, AND BAYES' THEOREM

In the hypothetical case above, we concluded that without the blood genetic marker frequency the defendant would probably be acquitted; with it, he would be convicted. This intuition is open to some doubt.

The error, according to some commentators, is to approach the genetic marker frequency isolated from the rest of the evidence in the case. Taken separately, the evidence does have little probative worth, but normally one does not weigh each piece of evidence alone. Instead, jurors view evidence at trial as a unit. Individual facts are melded in the jury's mind, to be taken in context of all the evidence as interpreted by each juror's life experience. Juries find guilt in a "culmination of probabilities."[103] Juries often give considerable weight to fairly insignificant evidence if that evidence supports an event about which the jurors already have persuasive proof.[104] Conversely, juries often discount evidence which would otherwise be highly persuasive because the evidence does violence to prior beliefs held by the jury. This "cumulative perspective" controls the probative significance of evidence when statistics are not involved.[105] Finkelstein and Fairley,[106] and other commentators,[107] argue that the same perspective should apply

103. Finkelstein & Fairley, *supra* note 2, at 497.

104. *Id.*

105. *Id.*

106. *Id.*

107. As Glanville Williams explains:

> Take a criminal trial where a number of facts are established, with different degrees of probability but all pointing to the guilt of the defendant. The jury or magistrates must consider whether the only reasonable explanation of the facts is that the defendant is guilty, or whether there is a fair possibility that the accumulation of apparently suspicious facts is accidental. For example, suppose that evidence is given that a palm print similar to the defendant's was found on the murder weapon, and that a similar print would on average be made only by one person in a million. It cannot logically be inferred from this evidence, standing alone, that the odds on the defendant's guilt are a million to one. Assuming a relevant population of 10 million people, there will be about 10 people who will make the same print. If the police, happening to have the defendant's palm print on file, arrested him and charged him with the murder merely on that evidence, the probability of his being the culprit (ignoring, as we must, any previous convictions of his) would be no more than 1 in 10, and even that figure leaves out the possibility that the murderer planted the defendant's print on the weapon, or that the defendant did the deed but lacked mens rea, and so on. However, if there is other evidence against the defendant the palm-print evidence takes on a different colour. The jury then have to ask themselves: what is the chance that a person who is already involved in deep suspicion, as the defendant is, would turn out to be one person in a million whose palm print corresponded with that on the weapon? That this should happen by coincidence is not a chance in 10 but one in a million, and it would make no difference that the police first suspected the defendant because he made the same palm print as that on the weapon.

Williams, *supra* note 77, at 351. *See also* Tribe, *supra* note 76, at 1350. Professor Tribe states:

> But the fact that mathematical evidence *taken alone* can rarely, if ever, establish the crucial proposition with sufficient certitude to meet the applicable standard of proof does not imply that such evidence—*when properly combined with other, more conventional, evidence in*

when statistics are involved.

Bayes' Theorem is a statistical formula which describes this process of considering the totality of evidence.[108] It is best described as a "quantitative description of the ordinary process of weighing evidence."[109] To employ the theory one hypothesizes the probability of event *A* occurring. Statisticians have labelled this occurrence the "prior possibility."[110] Next, one conducts observations and experiments, "not to supercede but to modify the earlier probability, resulting in a 'posterior probability' by means of the Bayes formula."[111] Commentators have stressed that the uses of this mathematical formula are limited. They caution that Bayes' Theorem cannot by itself give the probabilities that a given statement is true or false, or that a given occurrence will or will not occur.[112] Instead, Bayes' Theorem is simply a method of revising a prediction of probability in light of subsequently obtained observations. The theorem instructs one how the previous estimates of probability should be interpreted given the receipt of further data.[113]

Glanville Williams illustrates the use of the theorem in a criminal case:

> Suppose that I am a juror in a criminal case, and am pretty satisfied of the defendant's guilt on the basis of a signed confession which the defendant has rather unconvincingly disputed. Since I am an unusually numerate juror, I attach the figure of 0.95 to the degree that I am convinced of guilt. I am not prepared to convict with this degree of doubt, so I look around for further evidence. My mind goes to the fact that the defendant fled when he heard that the police were after him. It is perfectly proper for the jury to take this into account, for which purpose it is necessary to try to estimate the degree of probability of guilt to attach to it. Normally

---

> *the same case*—cannot supply a useful link in the process of proof. Few categories of evidence indeed could ever be ruled admissible if each category had to stand on its own, unaided by the process of cumulating information that characterizes the way any rational person uses evidence to reach conclusions. The real issue is whether there is any acceptable way of combining mathematical with non-mathematical evidence. If there is, mathematical evidence can indeed assume the role traditionally played by other forms of proof.

Id. (emphasis in original).

108. Reverend Thomas Bayes, in *An Essay Toward Solving a Problem in the Doctrine of Chance*, PHILOSOPHICAL TRANS. OF THE ROYAL SOCIETY (1763), suggested probability judgments based on intuitive guesses should be combined with probabilities based on frequencies by the use of what has come to be known as Bayes' Theorem, a fairly simple formula . . . .

Tribe, *supra* note 76, at 1351 n.69 (citing I. GOOD, PROBABILITY AND THE WEIGHING OF EVIDENCE 62 (1950)). "Bayes' Theorem is a relatively simple rule discovered by the Reverend Thomas Bayes in 1763, yet first applied to this type of inquiry only a few years ago." Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065, 1083 (1968). Another writer states, "Bayes' Theorem is no more than a plausible common sense description of jury behavior." Kornstein, *A Bayesian Model of Harmless Error*, 5 J. LEGAL STUD. 121, 126 (1976).

109. Tribe, *supra* note 76, at 1351 n.69.

110. *See* Williams, *supra* note 77, at 346, upon which this description leans.

111. *Id.* at 347.

112. Kaplan, *supra* note 108, at 1083.

113. *Id.*

the jury would not be expected or encouraged to quantify the degree, but I am going to do it, by trying to estimate how many innocent persons would flee from the police. It would depend to some extent on the gravity of the charge; here the charge is, shall we say, murder. It would also depend on what evidence the defendant thought the police had against him. Even an innocent person may be frightened into a flight from justice if he knows that the police have evidence that happens to incriminate him, and if he believes that he cannot convincingly explain it away. Here the defendant knew that the police had found in his possession the remains of the poison used by the murderer. Even so, I cannot imagine that many innocent people would be so foolish as to take off, and I estimate the number as, at most, 20 in 100 of those who flee. I am content to take this as a firm figure, and accordingly, 80 in 100 of those who flee are guilty.

My estimate of the probability of guilt on the basis of the confession was made without reference to the question of running away. It covered both those hypothetical defendants who did not run away and those who did. However, I am now directing my attention to the fact that this defendant did run away, and accordingly I wish to raise my estimate of the probability of guilt above 0.95. The question is: by how much?

Let us call the probability of guilt on the basis of the confession (the prior probability, meaning the one that one starts with) pG. This is 0.95, and the probability of innocence (p not−G) is 0.05 (1 − 0.95). Let us call the second event, running away, E. Hidden within pG are two separate figures: the probability of G on the assumption that E has taken place (pG/E), and the probability of G on the assumption that it has not (pG/not−E). The former figure is greater than pG, and the latter is less than pG. Our task is to find pG/E.

We need not go into the algebra. The procedure can be reduced to simple terms, which are vouched for by the statisticians. Write down, in one column, the prior probabilities, and in the next column the two variants of pE: the probability of the defendant running away and being guilty (pE/G), and the probability of his running away and being innocent (pE/not−G). They are in the proportion of 0.8 to 0.2, and for the purpose of the calculation may be taken as these two figures. Multiply the two sets of figures together.

Priors		Event		Multiply
pG	0.95	pE/G	0.8	0.760
p not−G	0.05	pE/not−G	0.2	0.010

The posterior probabilities will be in the proportion 0.76 : 0.01. But these do not add up to 1. Unity is restored to the total by an arithmetical trick: add them together, and divide the sum into each.

Add	Posteriors
0.76	$0.76 \div 0.77 = 0.99$
0.01	$0.01 \div 0.77 = 0.01$
0.77	

Behold! The two fractions (rounded off to two points of decimals) now add up to 1. The posterior probability of guilt on the assumption of E (pG/E) is 0.99 (actually a trifle more than 0.987, but still a significant increase from 0.95), and the posterior probability of innocence on the assumption of E (p not-G/E is 0.01).

Bayes' Theorem can be applied to each piece of evidence. The hypothetical example assumed as one of its facts that the police found the poison used for the murder in the defendant's possession. This is an independently incriminating fact: in the previous calculation it counted for innocence, not guilt, so it is proper to do another calculation weighing it for guilt. Let me assume that nine out of ten persons against whom such evidence is given are guilty of murder; the actual figure is probably higher. Doing another Bayes' calculation, the probability of the defendant's guilt now rises from 0.987 to 0.9985.[114]

To use Bayes' Theorem in the hypothetical blood marker case, the jury would first estimate the probability of guilt from all the non-statistical evidence. Of course, no precise measure would exist for this number; instead, the jurors would have to make a subjective determination. Then, using Bayes' Theorem, the jurors would factor in the genetic marker frequency, that is, that only one in a thousand people have blood like that found at the crime scene and possessed by the

---

114. Williams, *supra* note 77, at 347-48. Professor Kaplan illustrates the theorem in another setting:

> If we are attempting to discriminate between two mutually exclusive hypotheses, Bayes' Theorem can be written in the following form: $Q_1 = LQ_0$. In this formula, Q_0 refers to our previous idea of the odds on one hypothesis and L is the ratio of the probability that the observed evidence would have occurred under that hypothesis to the probability that the same evidence would have occurred under the other. Q_1, of course, refers to the odds on the hypothesis calculated after receipt of the additional data. Bayes' Theorem thus quantifies the commonsense conclusion that if a given piece of new evidence more probably would have occurred under hypothesis A than under hypothesis B, the receipt of that piece of evidence should raise somewhat our previous estimate of the odds that hypothesis A is correct. L, the probability ratio, is generally a more familiar item for us to compute than is the likelihood of the correctness of the hypothesis. Thus, let us assume that an oil geologist may be able to say that if a field of unknown productivity is in fact commercially productive, the probability is two-thirds that oil will be discovered (and hence one-third that it will not be) in a well drilled to a depth of 10,000 feet at a random spot in the field. The geologist may also be able to say that if the field is not commercially productive the probability is four-fifths that oil will not be discovered in the well. If a well is then drilled and proves unproductive, Bayes' Theorem will tell the geologist that he should now change his estimate of the odds that the field is a commercially productive field by multiplying his prior idea (Q_0) by the probability of 1/3 / 4/5 or 5/12. The odds that the field is productive are now less than half as great as before the well was drilled.

Kaplan, *supra* note 108, at 1083-84. Mathematical derivations as well as verbal descriptions of the formula can be found in Finkelstein & Fairley, *supra* note 2, at 498-501; Tribe, *supra* note 76, at 1350-54.

defendant. Here Finkelstein and Fairley have made the task easier since they have compiled a table of calculations of Bayes' Theorem for various values.[115] They have concluded that if the prior assessment of guilt based on the non-statistical evidence is .01, .1, .25, .50, .75, and the frequency evidence is one in 1,000 or .001, then Bayes' Theorem leads to the subsequent probabilities .909, .991, .997, .9990, or .9996, respectively. Therefore, if the prior odds of guilt were calculated to be as low as one in ten, the subsequent probability will be over 99%.

A. The Bayes' Theorem Debate

Finkelstein and Fairley, strong proponents of the use of Bayes' Theorem in criminal trials,[116] indicate that one of the Theorem's prime advantages is that the probative force of the frequency evidence is no longer dependent upon the size of the suspect population.[117] Bayes' Theorem, according to Finkelstein and Fairley, has additional advantages. They stress that the use of quantitative measures of guilt will force conscientious jurors to consider the evidence more circumspectly, and exclude impermissible evidence such as appearance or prejudice.[118] Further, Bayesian analysis, according to the authors, would demonstrate to the jury that an apparently damning probability figure like one in 1,000 could be relatively insignificant if prior suspicion were relatively weak.[119]

---

115. Finkelstein & Fairley, *supra* note 2, at 500.

116. Finkelstein and Fairley give the following outline as to how the jury could be instructed to use Bayes' Theorem:

> An expert witness could explain to jurors that their view of the statistical evidence should depend on their view of the other evidence. He might then suggest a range of hypothetical prior probabilities, specifying the posterior probability associated with each prior. Each juror could then pick the prior estimate that most closely matched his own view of the evidence . . . . To minimize the possibility that a prosecutor would prejudice the defendant's case by choosing only highly incriminating "hypothetical" prior probabilities, an expert so testifying should be required to show the posterior probabilities associated with a broad range of prior estimates. Such a procedure would also foreclose the chance that jurors would consider the expert as interjecting his own opinion as to the appropriate prior.

Id. at 502.

117. The statement that prints with particular characteristics occur with a frequency of one in a thousand persons means only that a defendant who has such a print is a thousand times more likely to have left it than someone selected at random from the population. By itself, this is not a meaningful statistic for measuring probability of guilt. As we have seen, a defendant could be a thousand times more likely to be guilty than someone selected at random and still more likely to be innocent than guilty. The comparison with a random selection is irrelevant. The jury's function is not to compare a defendant with a person selected randomly but to weigh the probability of defendant's guilt against the probability that *anyone* else was responsible. Bayes' Theorem translates the one-in-a-thousand statistic into a probability statement which describes the probative force of that statistic.

Id. at 502 (emphasis in original).

118. Glanville Williams similarly concludes, "[W]orking out Bayes' theorem is good discipline because every assumption must be clearly specified . . . . [T]he ordinary juror who follows this ordinary procedure does not really know what his thought-processes are, the Bayesian juror does so to a much greater extent." Williams, *supra* note 77, at 349.

119. Probably the greatest danger to a defendant from Bayesian methods is that jurors may be surprised at the strength of the inference of guilt flowing from the combination of their

The conclusion that frequency evidence, such as that generated by genetic markers, can best be integrated into criminal trials though the use of Bayes' Theorem has been vigorously challenged, most notably by Professor Laurence Tribe.[120] Tribe first concludes that the Bayesian approach should not be permitted because it will not lead to the precision that its proponents claim and because it will distort the fact-finding process by overwhelming the non-quantifiable aspects of a criminal case.

As we have seen, Bayes' Theorem can only be used when a probability of guilt exists apart from the frequency evidence. Tribe argues, however, that jurors are not equipped to assess this starting point since few have ever done such probability estimations. Consequently a number picked by a juror will be "spuriously exact."[121] Because the application of Bayes' Theorem thus "compels the jury to begin with a number of the most dubious value,"[122] Tribe argues that the use of that technique at trial would yield inaccurate and misleadingly precise conclusions.

Inaccuracies also will result because the evidence of guilt will be counted more than once in assessing the various probabilities. The jury, if it is to operate in the traditional way, will only assess the probabilities after the receipt of all the evidence. Certainly, knowledge of the genetic marker evidence will color the juror's assessment of the

---

> prior suspicions and the statistical evidence. But this, if the suspicions are correctly estimated, is no more than the evidence deserves.

Finkelstein & Fairley, *supra* note 2, at 517.

Finkelstein and Fairley state that they made an informal, nonrandom survey of intuitions of guilt based upon their hypothetical case. They found that the addition of the frequency evidence

> was thought to raise the probability of guilt, but assessments of the weight of this evidence varied widely, and the subjects were uncertain how to treat the new information. In most cases the assessments were not as great as they would have been if the probabilities had been computed in accordance with Bayes' theorem.

Finkelstein & Fairley, *supra* note 2, at 502 n.33.

120. *See* Tribe, *supra* note 76. Tribe wrote his article as a response to Finkelstein and Fairley. Finkelstein and Fairley responded to Tribe in Finkelstein & Fairley, *A Comment on "Trial By Mathematics"*, 84 HARV. L. REV. 1801 (1971) [hereinafter cited as Finkelstein & Fairley II]. Tribe answered this rebuttal in Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810 (1971) [hereinafter cited as Tribe II]. *See also* Braun, *supra* note 2; Williams, *supra* note 77. For an application of Bayes' Theorem to the criminal appellate process, see Kornstein, *supra* note 108, at 140-44.

121. Tribe, *supra* note 76, at 1359.

122. *Id.* Kornstein indicates that the initial probability may be reached by averaging and therefore

> it may be totally inaccurate because it ignores group dynamics. Averaging jurors' estimates simply does not duplicate the process of give-and-take that occurs in the jury room. An average fails to account for conformity effect, cognitive dissonance, and other psychological interactions. Consequently, a jury might well arrive at a real verdict quite different from the verdict dictated by the formula.

Kornstein, *supra* note 108, at 140. Braun contends, "There is . . . no guarantee that a jury would have an adequate evidentiary basis to quantify a degree of belief in the accused's guilt, and even if such a basis existed, jurors may lack the ability to translate that degree of belief into a number." Braun, *supra* note 2, at 53. Glanville Williams states, "The most obvious objection to attempting to translate the probability of ordinary life and of the law courts into mathematical terms is that our data are too vague to justify precise statement." Williams, *supra* note 77, at 305.

probabilities arising from the rest of the case. This means that the marker evidence will be counted twice—once to set the prior probability and then again in Bayes' Theorem. This, of course, overstates the value of the genetic evidence.[123]

Bayes' Theorem, according to Tribe, also presents additional problems of the "overbearing impressiveness of numbers" and "the dwarfing of soft variables."[124] Tribe is concerned that the Bayesian approach will lead the jury into concentrating on the statistics at the expense of more impressionistic evidence because this latter evidence cannot be easily reduced to numbers.[125] As a result, the jury might be led into believing that the formula tends to be conclusive on the question of guilt when, in fact, it cannot be. In the hypothetical case, Bayes' Theorem could not generate a probability as to the defendant's guilt.

---

123. *See* Tribe, *supra* note 76, at 1366-68. *See also* Tribe II, *supra* note 120, at 1816 n.31:

> [E]ven if the prior could be estimated without any consideration of the statistical evidence, the proposed application of Bayes' Theorem—at any point in the trial—would entail a distorted outcome whenever some or all of the evidence that underlay the prior is conditionally dependent upon the statistical evidence, a circumstance whose presence or absence cannot be feasibly determined in any given case.

Williams also gives an illustration of this problem:

> At what point in the charge does one start calculating the prior probability? The police arrest a man who is behaving suspiciously at dead of night, and afterwards obtain confirmatory evidence that he has committed theft. Is the prior probability based on the man's suspicious behavior, or is it based on the fact that he has been charged, or just on the fact that he is a man? It cannot reasonably be based on the fact that the man has been charged, because the charge was based on evidence known to the police, and to use the charge to create a prior probability, and then to use the evidence to create a posterior probability, would be to use the evidence twice over.

Williams, *supra* note 77, at 350. Dr. Braun argues:

> An additional problem with using subjective probability to weigh the evidence is that jury determinations, however tentative, that are made before all the evidence is in must be prejudicial to the defendant. . . . It is doubtful, however, that a juror could arbitrarily ignore a piece of evidence already heard to make an initial determination of culpability and thus use that determination in conjunction with the evidence initially ignored to arrive at a quantifier expressing the probability that the accused committed the relevant criminal act.

Braun, *supra* note 2, at 54.

124. Tribe, *supra* note 76, at 1361. Glanville Williams summarizes:

> There is a danger in the combination of subjective judgment and mathematics, as Professor Tribe points out. If it were all presented to a jury, they might be hypnotised by the end-figure of 0.9985, and attend insufficiently to its fragile evidential base. If firm statistics took the place of some of the conjectured figures, while other conjectured figures remained (as they very likely would), it might be even worse. Professor Tribe calls this danger "the dwarfing of soft variables." Further danger would occur if the jury were asked to estimate pG, the probability of guilt on the basis of only part of the evidence. The traditional view is that the jury must consider the evidence as a whole before reaching a decision. A jury invited to apply Bayes would very probably make a mistake, or have their attention diverted from important issues.

Williams, *supra* note 77, at 348-49.

125. 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.

Tribe, *supra* note 76, at 1365-66.

Even if it could prove with certitude that the blood at the scene came from the defendant, it could not prove that he either killed the victim or, if he did, that he was guilty of a crime. For instance, perhaps he arrived at the crime scene, saw the victim attempting to commit suicide, and was cut in a struggle. Perhaps he fought with someone besides the victim at the scene and that other person was the killer. Perhaps he got in a fight with the victim. After she cut him, he may have fled and someone came along later who killed her. The defendant may have killed her in self-defense, or even accidentally. Perhaps he killed under an emotional state which mitigates murder to manslaughter, or perhaps he killed her, but was insane.

Bayes' Theorem cannot speak to those possibilities. At best, the formula only gives a probability that defendant was present at the crime scene, nothing more. It does not speak to guilt. That limitation, however, might be lost as the theorem is applied. "One consequence of mathematical proof, then, may be to shift the focus away from such elements as volition, knowledge, and intent, and towards such elements as identity and occurrence . . . ."[126]

Tribe goes on to argue, however, that even if the application of Bayes' Theorem in a criminal trial were workable and accurate, the Bayesian methods would still "undermine the presumption of innocence, erode the values served by the reasonable doubt standard, and exacerbate the dehumanization of justice."[127] Tribe concludes that the supports for the presumption of innocence will be knocked away because

> the trier is forced by the Finkelstein-Fairley technique to arrive at an explicit quantitative estimate of the likely truth at or near the trial's start, or at least before some of the most significant evidence has been put before him . . . .

---

126. *Id.* at 1366. Tribe also argues:

> But then we have come full circle. At the outset some way of integrating the mathematical evidence with the non-mathematical was sought, so that the jury would not be confronted with an impressive number that it could not intelligently combine with the rest of the evidence, and to which it would therefore be tempted to assign disproportionate weight. At first glance, the use Finkelstein and Fairley made of Bayes' Theorem appeared to provide the needed amalgam. Yet, on closer inspection, their method too left a number—the exaggerated and much *more* impressive $P(X/E) = .997$—which the jury must again be asked to balance against such fuzzy imponderables as the risk of frameup or misobservation, if indeed it is not induced to ignore those imponderables altogether.
>
> What is least clear in all of this is whether the proponents of mathematical proof have made any headway at all. Even assuming with Finkelstein and Fairley that the accuracy of trial outcomes could be somewhat enhanced if *all* crucial variables could be quantified precisely and analyzed with the aid of Bayes' Theorem, it simply does not follow that trial accuracy will be enhanced if *some* of the important variables are quantified and subjected to Bayesian analysis, leaving the softer ones—those to which meaningful numbers are hardest to attach—in an impressionistic limbo. On the contrary, the excesssive weight that will thereby be given to those factors that can most easily be treated mathematically indicates that, on balance, more mistakes may well be made with partial quantification than with no quantification at all.

Id. at 1365 (emphasis in original).

127. Tribe II, *supra* note 120, at 1815.

> . . . A juror compelled to derive a quantitative measure . . . of the defendant's likely guilt after having heard no evidence at all . . . cannot escape the task of deciding just how much weight to give the undeniable fact that the defendant is, after all, not a person chosen at random but one strongly suspected and hence accused by officials of the state after investigation.[128]

Finkelstein and Fairley's proposal, however, does not just undercut the presumption; instead, their suggestion totally disregards it. Bayes' Theorem, if it works correctly for frequency evidence such as that derived from genetic markers, should also work for the receipt of every piece of evidence as long as a probability can be assigned the evidence. How, then, should that first piece of evidence be treated? Suppose a robbery in which the victim convincingly identifies the defendant at trial. The jurors believe that the probability of the identification being correct is 99.9%. This would seem to indicate that the juror is convinced that the odds are one in 1,000 that the wrong person is on trial. Can Bayes' Theorem be applied to this situation? With just this one probability, the formula is useless since the theorem only allows modification of an already existing probability. Two possibilities exist in this situation. First, the jurors may use their 99.9% figure as the initial probability which will be modified by the use of Bayes' Theorem provided further evidence is received. This approach acts as if the presumption of innocence did not exist.

On the other hand, if a reasonable doubt can be quantified, a concept at least implicit in the use of Bayes' Theorem,[129] the presumption of innocence also should be quantifiable. One approach to quantifying the presumption of innocence would conclude that, before any evidence is admitted, the presumption indicates no chance that the defendant is guilty. In Bayesian terminology that translates to mean an initial probability of zero. If so, Bayes' Theorem cannot be applied since the formula requires the multiplication of this initial probability by the probabilities assessed for the received evidence. This multiplication would always result with a probability of zero, and a defendant could never be found guilty. Instead of treating the presumption of innocence as zero, Professor Kaplan has argued that the defendant's chances of guilt, measured before the receipt of evidence, are the same as the chances of guilt for anyone else in the country. Using this logic, he has assigned the probability of guilt for the purposes of assessing the presumption of innocence.[130]

---

128. Tribe, *supra* note 76, at 1368-69.

129. *See infra* text accompanying notes 130-36.

130. Courts, however, make no effort to instruct the jury on the magnitude of [the initial odds of guilt] or on how they should compute it. But courts do emphasize the presumption of innocence. The instruction on this presumption is more than a direction to the jury not to consider as evidence against the defendant the fact that he has been indicted and is being tried. It is a direction, at the very least, that the jury not assume at the outset of their consideration of the case that the defendant is more likely guilty than not. Thus the jury is being told

This numerical assessment of the initial probability for purposes of applying Bayes' Theorem still leads to ludicrous results. In the the hypothetical robbery, the jury agrees that there is a 99.9% chance that the identification is correct. Presumably, the normal jury will convict based upon this evidence alone. If the triers apply Bayes' Theorem, however, with the presumption of innocence treated as the prior probability and set at 1/200,000,000, the will reach the conclusion that the probability of guilt is only about five in 1,000,000.[131]

Assume further that the prosecutor not only introduces this identification but also a confession by the defendant. The jurors are so impressed with the reliability of the statements that the jurors conclude that there is only one chance in 1000 that the confession is not truthful. With these two pieces of evidence surely a conviction should result. However, if the jury starts with the presumption of innocence as 1/200,000,000 and then applies Bayes' Theorem successively to the probabilities derived from the identification and the confession, the

---

> that [the initial odds of guilt are] certainly not more than one-half. We can, indeed, go further than this and regard the presumption-of-innocence instruction as a direction to the jury to begin their consideration of the case by assuming that . . . the initial odds on guilt . . . are to be considered "quite small." One could even speculate that a reasonable value of [the initial odds of guilt] is 1:200 million—that this defendant is no more likely a priori to be the guilty party (assuming that there is a guilty party) than is anyone else in the United States.

Kaplan, *supra* note 108, at 1085-86.

131. In applying Bayes' Theorem to a hypothetical criminal case, Finkelstein and Fairley made some assumptions. *See* Finkelstein & Fairley, *supra* note 2, at 498-500. Tribe challenges the validity of those assumptions and goes on to state that, without the assumptions, Bayes' Theorem will have such a "messy form" that its application will have "strained the system beyond its breaking point." Tribe, *supra* note 76, at 1362-65. However, if we accept the proponents' assumptions and apply them here, Bayes' Theorem takes the following form:

Posterior probability =

$$\frac{\text{prior probability}}{\text{prior probability} + [\text{probability that defendant not guilty derived from presented evidence} \times (1-\text{prior probability}).]}$$

The prior probability here is the presumption of innocence or 1/200,000,000 or .000000005. Since the identification indicates a .999 chance of guilt, the identification would indicate the defendant was not guilty only one in 1000 times or .001. Thus the posterior probability equals

$$\frac{.000000005}{.000000005 + (.001 \times .999999995)} =$$

$$\frac{.000000005}{.000000005 + .000999999995} = .000005$$

Selvin, *supra* note 9, at 31-32, presents a different derivation of Bayes' Theorem. Selvin indicates that

$$\text{posterior probability} = \frac{1.0}{1.0 = [\text{probability of not guilty derived from presented evidence X} \ (1\text{-prior probability/prior probability}).]}$$

The denominator would equal

$$1.0 + (.001 \times 199999999) =$$
$$1.0 + [.001 \times (.999999995/.000000005)]=$$
$$1.0 + 199999.999 + 200001.999.$$

Thus, the posterior probability equals 1.0/200002 or five in 1,000,000.

probability of guilt merely becomes five in 1,000.[132] Once again acquittal is required. Surely something is wrong.

Thus, if the presumption of innocence is factored into Bayes' Theorem, ludicrous results occur. Finkelstein and Fairley ignore this. Their proposal can only work if the jury calculates the initial probability of guilt after the receipt of at least the first piece of evidence and without reference to the presumption of innocence. In other words, Bayes' Theorem can only work if the presumption of innocence disappears from consideration.

Tribe also believes that the application of Bayes' Theorem to criminal trials will destroy our present doctrine of reasonable doubt since the theorem presupposes that reasonable doubt can be quantified. Thus, if a jury utilizing the theorem concludes that the probability of guilt is 99% and convicts, the jury has, at least implicitly, reached the conclusion that a 1% chance that the defendant is not guilty does not constitute a reasonable doubt.[133] Although it may be natural to assume that reasonable doubt can be quantified,[134] Tribe contends that "the system does *not* in fact authorize the imposition of criminal punishment when the trier recognizes a quantifiable doubt as to the defendant's guilt. Instead, the system dramatically—if imprecisely—insists upon as close an approximation to certainty as seems humanly attainable in the circumstances."[135]

---

132. The posterior probability obtained by modifying the presumption of innocence by the receipt of the identification is .000005. This now becomes the prior probability which is modified by the probability assigned to the confession. Thus, the prior probability, in Finkelstein and Fairley's formulation equals

$$\frac{.000005}{.000005 + (.001 \times .999995).}$$

133. Tribe contends,

> An inescapable corollary of the proposed method, and indeed of any method that aims to assimilate mathematical proof by quantifying the probative force of evidence generally, is that it leaves the trier of fact, when all is said and done, with a number that purports to represent his assessment of the probability that the defendant is guilty as charged.

Tribe, *supra* note 76, at 1372. "Even when the number measures only one element of the offense and omits an element like intent . . . it sets an upper bound on the probability of guilt . . . ." *Id.* at 1372 n.138.

> For example, if the figure of 99.7 per cent represents only the probability that the print was the defendant's, leaving a margin of doubt of .3 per cent on the issue of the print's identification, it seems clear that there will be an even *larger* margin of doubt on the ultimate issue of the defendant's guilt.

Tribe II, *supra* note 120, at 1818 (emphasis in original).

134. Glanville Williams states that

> it is natural to assume that "reasonable doubt" in a criminal case can in principle be quantified. Everyone realises that legal proof cannot in fact at present be translated into a precise numerical figure, but we have never decisively dismissed the notion that in principle, or in the ideal world, it could be.

Williams, *supra* note 77, at 297. Another writer notes, "While attempts at quantification show that 'beyond a reasonable doubt' denotes the probability of guilt somewhere around .90, the standard is indeed vague, and thus likely to be a source of disagreement." Kornstein, *supra* note 108, at 143.

135. Tribe, *supra* note 76, at 1374 (emphasis in original). Finkelstein and Fairley describe "beyond a reasonable doubt" differently: "When we say that defendant is guilty beyond a reason-

If we allow the quantification of proof beyond a reasonable doubt, we are stating that our criminal system, as part of its normal functioning, tolerates the conviction of innocent people. If proof beyond a reasonable doubt were set at 99.5%, then the criminal system frankly condones the punishment of one innocent person out of every two hundred, a result from which most people "recoil."[136] Our courts would convict admittedly innocent defendants. If proof beyond a reasonable doubt is seen as a normative goal of certainty, however, the conviction of the innocent is not considered to be an acceptable result of the criminal system. Instead, it is an abberation—a breakdown in that system. The quantification of a reasonable doubt, as Bayes' Theorem requires, is a statement that our society is willing to condone the conviction of numbers of innocent people. Tribe contends that a just society should never be willing to tolerate such a result.[137]

Tribe's arguments, however, have recently come under attack by Professors Saks and Kidd.[138] They strongly contest Tribe's assertion that the jury will overvalue a statistical presentation at the expense of the other evidence. Relying upon various empirical studies, they con-

---

able doubt, we mean that the evidence has brought us to a state of belief such that if everyone were convicted when we had such a belief the decisions would rarely be wrong." Finkelstein & Fairley, *supra* note 2, at 504.

136. Kaplan, *supra* note 108, at 1073. Dr. Braun states:

> To develop a quantifier after hearing all the evidence and to use it to decide the ultimate question of guilt or innocence of the accused would, at the very least, add a superfluous step to the decisionmaking process and would, at worst, dehumanize the judicial process at the expense of the accused. . . . Such a quantifier could be used only in conjunction with the establishment of a threshold probability value, which would, if exceeded, indicate a belief beyond a reasonable doubt. Thus, the quantifier would serve only to shift the philosophical emphasis from minimizing the chance of convicting an innocent defendant to establishing a tolerance level for how many innocent people the juror was willing to convict to ensure conviction of the guilty.

Braun, *supra* note 2, at 51-52. *See also* Kornstein, *supra* note 108, at 139. *But see* Williams, *supra* note 77, at 305-06.

> Using numbers has the advantage of making one face the unpleasant fact that convicting in a criminal case on a probability of 0.95, high though that may seem at first sight, involves the probability of convicting one innocent person in 20. This is more unwelcome when the conviction is of murder than when it is of a parking offense. Still, the prospect of having to acquit 19 murderers because you are afraid of convicting one innocent person is also uninviting.

Id.

137. *See* Tribe, *supra* note 76, at 1372-77. Glanville Williams summarizes:

> The tribunal in a criminal case must feel sure of guilt. It must have no doubts—apart from unreasonable doubts and fanciful doubts, as is sometimes said. If a juror is able to calculate that he is 99 per cent. sure, then he already has a doubt, in legal theory, and ought not to convict. . . . Of course, magistrates and reflective jurors know that mistakes are inherent in any human institution. In that sense, they know that there is always the possibility of making a mistake when convicting. But it is a general and theoretical doubt. What the law demands is that on this occasion, in respect of this defendant, you must feel sure. To depart in any way from this requirement would erode the protection given by the criminal process to defendants who may be falsely accused. It would be a public proclamation of callousness towards the possibility of convicting the innocent, and it would dilute confidence in the justice of the legal system.

Williams, *supra* note 77, at 306 (emphasis in original).

138. Saks & Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 L. & SOC'Y REV. 123 (1981).

clude that people who have to make decisions based upon complex information use simplifying mental strategies in order to make their choices.[139] Contrary to what Tribe assumes,[140] Saks and Kidd contend that the empirical results indicate that in using the simplifying processes, people do not overvalue the quantitative or probabilistic evidence. Rather, people give too little weight to the probabilities.[141] Saks and Kidd conclude that jury verdicts will be more accurate if jurors are allowed and encouraged to use mathematical tools:[142]

---

139. *Id.* at 127-31. Saks and Kidd label the simplifying stategies as "heuristics."

140. Saks and Kidd state that Tribe's conclusions are based on unproven assumptions: "Influential as Tribe's paper has been, like much legal scholarship, it is 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 merely plausible, not because any evidence is adduced on their behalf." *Id.* at 125.

141. *Id.* at 134. The information presented by Saks and Kidd support one of Tribe's contentions. Tribe maintained that jurors will not be able to set proper prior probabilities once they learn the frequency evidence, that knowledge of the genetic marker evidence will color the assessment of the probability of guilt arising from the rest of the case. *See supra* text accompanying note 123. Saks and Kidd state:

> One fact, however, can be unambiguously derived from the extensive literature on the psychology of decision making. People tend to be overconfident in their judgments. Not only do individuals tend to overestimate how much they already know, but they also tend to underestimate how much they have just learned from facts presented in a particular context. Once they do know an outcome, people fail to appreciate how uncertain they were before learning of it. . . .
>
> Essentially, people find it difficult to disregard information that they already possess. Telling people that an event has occurred causes them to report that the event was more likely to have happened. Furthermore, hearing such information does not also cause them to report that the information affected their perceptions or decisions. People do not appreciate the extent to which hearing new information has an effect on their judgments.
>
> Why do people tend to be overconfident in their judgments? One possibility is that individuals reinterpret previous information in light of new information, so that the two sets of information are integrated into a coherent whole. The "old" view of these events is assimilated into the "new" correct view in such a natural and immediate fashion that the assimilator is unaware that his or her perspective has been altered. The outcome psychologically is that the person reports that he or she really knew the answer or held the same opinion previously, and that a discrepancy never existed between initial reactions and the apparent conclusions.

Saks & Kidd, *supra* note 138, at 143-44. Therefore, knowledge of the genetic marker evidence will become assimilated into the nonstatistical evidence without the jurors being aware of the assimilation. As a result, jurors will inevitably set the prior probability higher than is warranted by just the nonstatistical evidence.

Professors Saks and Kidd go on to suggest that this "hindsight effect" is an inherent problem in criminal trials.

> In a criminal trial, people are first given the "answer"—that is, the defendant. Then, the evidence is provided, and the fact finder is asked whether the evidence does in fact prove the conclusion. This arrangement seems especially prone to hindsight. Each of the bits of evidence will appear more likely to lead to the defendant than they would have if the defendant were not already known. Analogizing from the hindsight experiments to the "fact finders" at trial, the evidence will seem to point more surely to the answer than it did when the investigators were developing the evidence. It may be that the high standard of proof required for a finding of guilt makes up for the peculiarity (and consequent distorting effects) of the way the question is posed: answer first. It is noteworthy that only criminal proceedings are framed this way and only criminal proceedings require the highest standard of proof. An interesting alternative procedure might be to experiment with trials in which the evidence is presented first and fact finders are asked which of several defandants, if any, is the guilty party. Under such conditions, fact finders, lacking the judgmental bias produced by hindsight, would probably be less sure of their judgments than is true with the existing criminal trial structure.

Id. at 145.

142. Professors Saks and Kidd contend "that while certain errors and harm may be inherent

Our suggestion is modest, and most lawyers should find it comfortingly traditional. Namely, experts ought to be permitted to offer their data, their algorithms, and their Bayesian theorems. The errors that may be introduced will be subjected to adversarial cross-examination. Various formal mathematical models do have room for errors—variables omitted, poor measurements, and others that Tribe has cogently presented. But so do intuitive techniques. Properly employed and developed, the former can have fewer. It is up to opposing counsel to unmask the errors. Moreover, as a matter of developing and introducing new tools from what might be called decision-making technology, the identification of flaws does not imply that the tools ought not to be used. The proper question is whether the tool, however imperfect, still aids the decision maker more than no tool at all.[143]

B. Once Again, Bayes' Theorem

In the debate over the use of Bayes' Theorem, time has made Professor Tribe the *de facto* winner. Although these issues have been debated for some time, nothing in the reported cases indicates that Bayes' Theorem has been presented to juries or applied by the triers of fact. This, therefore, might seem an effective end of the discussion.

The debate needs to be started anew, however, for Bayes' Theorem has again been proposed for use in criminal trials—this time explicitly in conjunction with blood genetic marker evidence.[144] Because the use of genetic marker phenotyping recently has become widespread in forensic laboratories and because criminal cases now are subject to a flood of such evidence,[145] the use of Bayes' Theorem in criminal litigation may become extensive. Before Bayes' Theorem becomes widely employed by trial courts, the theorem should be scrutinized closely.

Three recent commentators, Selvin, Grunbaum, and Myhre contend that Bayes' Theorem can be used to calculate the probability that blood found at the crime scene was the defendant's blood. Their approach follows that of Finkelstein and Fairley. Further, these com-

---

even in the proper use of probabilistic tools, even more harm may be inherent in not using them." Saks & Kidd, *supra* note 138, at 125. After surveying the empirical studies, they conclude:

> Unaided individuals tend to have great difficulty incorporating quantified variables, give excessive weight to bits and pieces that happen for whatever reason to be salient, base their decisions on less information (often the less useful information) than do mathematical models, and apply their decision policies inconsistently. . . . This presents an interesting set of concerns about human decision making that contrasts with Tribe's concerns about mathematical decision making. The problems associated with drawing inferences from probability evidence, problems Tribe would like to see the courts avoid, are not avoided by dumping the data, quantitative as well as nonquantitative, into the mental laps of human decision makers, armed only with their intuitition.

Id. at 147.

143. *Id.* at 148.

144. *See Probability*, *supra* note 7. The same proposal, in a somewhat shorter form, is also presented in Selvin, *supra* note 9, at 195-96.

145. *See supra* text accompanying notes 13-14.

mentators claim that Bayes' Theorem allows the trier of fact to calculate the probability that a person whose blood matches the blood found at a crime scene was present (denoted: P (present/coincidence)), by using Bayes' Theorem to combine the prior probability of the defendant's presence at the scene (P(present)) derived from all the nongenetic evidence and the frequency derived from the genetic data.[146] These commentators maintain that where C equals the genetic frequency figure, Bayes' Theorem leads to the following equation:

$$P(\text{present/coincidence}) = \frac{1.0}{1.0 + C[P(\text{not present})/P(\text{present})].}$$[147]

The authors give a simple illustration.[148] If the prior probability is assumed to be .50 (and consequently P(not present) = .50), and the genetic analysis of blood found at the scene matches the defendant's with that phenotype combination occurring in the relevant population with a frequency of .0097, then applying Bayes' Theorem, the subsequent probability of presence equals 1.0/((1.0 + .0097(.5/134)) = 1.0/(1.0 + .0097(1)) = 1.0/(1.0 + .0097) = .990. In other words, the probability that it was defendant's blood at the scene is 99%.

This proposed application of Bayes' Theorem is basically the same as that advocated by Finkelstein and Fairley and is therefore not surprising. What is surprising is that the proposal is made without any reference to the criticisms of the Finkelstein and Fairly approach. Nowhere are the concerns raised by Tribe and others even mentioned. Instead, the article actually furnishes an example of one of the dangers listed by Tribe. He contended that the Bayesian approach would result in the hard numbers overwhelming the soft variables. Thus, the trier of fact would lose sight of the limitations of the statistical approach. Almost as if to further this contention, the authors write about their Bayesian formula:

> This expression shows that when the value C decreases, the P(present/coincidence) approaches 1.0 for realistic values of P(present). That is, when the probability is small that a person who was assumed not present at the crime scene matches the given blood phenotypes, then it can be inferred that the defendant was, in fact, at the crime scene *(i.e., not falsely accused)*.[149]

As discussed above,[150] presence at the scene does not equal guilt. If the authors of this proposal, sophisticated in the uses of statistics, cannot keep this distinction clear, can we expect a lay jury to understand the limitations of Bayes' Theorem?

This proposal does recognize, however, the difficulty, if not the im-

146. *Probability*, *supra* note 7, at 31-32.
147. *Id.* at 32.
148. *Id.*
149. *Id.* (emphasis added).
150. *See supra* text accompanying notes 124-26.

possibility, of accurately ascribing a probability to all the non-genetic evidence.[151] The proposal goes on to conclude that, "[t]ypically, all that can be said is that if C is small then P(present/coincidence) will be close to 1.0 for all but very small values of P(present)."[152] In other words, if the frequency derived from the genetic evidence is low, Bayes' Theorem will establish that a defendant was at the crime scene, except in those situations in which the prosecution presents little against the defendant. Presumably the authors had this in mind when earlier in their article they stated:

> It should also be noted that a probability such as 1 in 12 million presented in the Collins case, when derived from well-defined and independent frequencies, could reflect the guilt or innocence of a defendant. However, this value is not the probability a person who matches the characteristics in question is falsely accused.[153]

The authors believe such a frequency indicates guilt. They apparently reason that, when plugged into Bayes' Theorem, the result will be an almost conclusive probability that defendant was at the scene of the crime, except in those rare cases where the assessment of non-genetic evidence leads to an absurdly small prior probability. A look at a slightly different application of the theorem will show that the authors' conclusion is not always correct; the very rare frequency may not truly prove guilt.

C. *Applying Bayes' Theorem to Others Than the Defendant*

A person can defend against a criminal charge by proving that someone else committed the crime. Of course, the defendant does not actually have to establish the identity of the true criminal; all he need show is a reasonable possibility that another person is the culprit.[154] If that is done, then the prosecution has not proved the case against the defendant beyond a reasonable doubt and the defendant must be acquitted.

If Bayes' Theorem truly is valuable in assessing the probability of a defendant's guilt, it should be just as useful in weighing the odds that someone else is guilty. The Bayesian proponents, however, have ignored this application. Perhaps another look at *People v. Collins*[155] indicates why.

As we have just seen, Selvin, Grunbaum, and Myhre contend that

---

151. "In most cases a realistic estimate of the P(present) does not exist and, therefore, the P(present/coincidence) cannot be accurately established." *Probability*, *supra* note 7, at 32.

152. *Id.*

153. *Id.* at 28.

154. Finkelstein and Fairley seem to recognize this when they state, "The jury's function is not to compare a defendant with a person selected randomly but to weigh the probability of defendant's guilt against the probability that *anyone* else was responsible." Finkelstein & Fairley, *supra* note 2, at 502 (emphasis added).

155. *See supra* text accompanying notes 15-26.

when the frequency evidence gives a small probability, Bayes' Theorem supports the inference that the defendant was not falsely accused. Apparently a validly derived probability like that in *Collins* leads to the conclusion of guilt.[156]

As seen above, even with the *Collins* frequency, if the suspect population were large enough the odds would be good that another couple besides the Collinses would have the characteristics of the robbers.[157] Assume that another such couple is discovered. The startling fact is that if the slightest piece of evidence can be found from which the trier of fact could calculate some miniscule prior probability of guilt, then Bayes' Theorem would indicate that this other couple is guilty, too. For instance, assume that a second couple with the described characteristics of the Collinses is interviewed, and they state that at the time of the robbery they were home alone together. Although the story is not implausible, it is more likely that the guilty couple would be unable to produce an independently verifiable alibi than an innocent couple. Of course, the jurors might reason that many innocent people could not give an ironclad alibi. Thus, the jurors might conclude that this second couple's failure to give a verifiable alibi is an indicator of guilt, but only a very slight one. Assume the jury assesses the prior probability based on this evidence as one in 10,000, or almost no chance that this other couple is guilty based on this one piece of weak evidence alone. Now, however, assume that the court instructs the jury to use Bayes' Theorem. The jury would calculate that with the previously set frequency for the described characteristics of one in twelve million, the probability that this other couple was at the crime scene becomes more than 99.9%, even though the other evidence of guilt was negligible.[158]

Thus, if the Collinses had the resources to search the entire population, locate another couple with similar characteristics, and then find the slightest evidence which could be interpreted as some indicator of guilt, the Collins couple would have to be acquitted. The application of Bayes' Theorem to this second couple would indicate that the other couple is guilty. Since only one couple could have committed the crime, the Collinses would have to be found not guilty. Of course, if the other couple were put on trial, they could produce the same analysis about the Collinses, and another acquittal would result. Bayes' The-

156. *See supra* text accompanying note 91.

157. *See supra* text accompanying notes 93-102.

158. The frequency of one in 12,000,000 equals .0000000833. Using the formula found in *Probability*, *supra* note 7, at 32, and used in the text *supra* note 147, we have

$$\frac{1.0}{1.0 + .0000000833(.9999/0001)} = \frac{1.0}{1.0 + .0000000833 \times 9999} =$$

$$\frac{1.0}{1.0 + .000833} = 99.92\%$$

orem, applied this way, does not simply ensure an acquittal for the Collinses, but ensures an acquittal for everyone.

D. The Hypothetical Case and Bayes' Theorem

If Bayes' Theorem works as its advocates contend, then it should be proper to apply the theorem not only to the defendant, but also to other potentially guilty parties. As just seen, however, one can imagine a situation where such application leads to ludicrous results. Outlandish outcomes are not limited merely to cases where the frequency is as rare as it was in *Collins*. A look at our hypothetical murder[159] with an imaginary dialogue between jurors will illustrate why Bayes' Theorem cannot be used in criminal cases.

Assume the jury has heard all the evidence against the defendant along with a lucid explanation of Bayes' Theorem. The jury concludes that the defendant's relationship as the victim's boyfriend is some evidence of guilt. The jurors reason that murderers often know their victims, and therefore one who knew the victim is more likely to be guilty than a total stranger.

The jurors also believe that the defendant's presence in the building at about the time of the murder is further evidence of guilt. Therefore, if the state proves that the defendant was in or could have been in the building at the time of the fatal attack, the defendant is more likely to be guilty than a person who does not have such evidence presented against him. Moreover, the jurors believe that because the defendant left town shortly after the death, his conduct shows his awareness of his guilt and tends to prove that the defendant is the murderer. Finally, the jury considers the stormy relationship between the defendant and the victim. They conclude that a person in a calm relationship is less likely to harm his partner than is a person in the midst of a tempestuous affair. This, too, is evidence of guilt.

The jurors concentrate on this non-genetic evidence and calculate from it the probability that the defendant is guilty.[160] They all agree that this evidence by itself does not constitute proof beyond a reasonable doubt. Instead, they finally agree that the non-genetic evidence indicates a 25% chance of guilt.[161]

The jurors then apply Bayes' Theorem by factoring in the genetic data which indicate that the frequency of the combined phenotypes is .001. After the calculations, the jury now concludes that the probability

159. *See supra* text accompanying note 69.

160. To see how workable Bayes' Theorem is, it might be useful for the reader to see what subjective probability he derives from such evidence.

161. Of course, the genetic marker evidence only tends to prove that the defendant was present at the crime scene and not that he was guilty. For simplicity's sake, we are making the logical leap that if he had been present, then he is guilty.

of guilt is .997.[162]

Although the jurors realize that proof beyond a reasonable doubt has not been defined in numerical terms, all agree that a 99.7% chance of guilt is above that figure. The jurors start to leave their room to deliver the guilty verdict, when one of them says, "Hold it. Bayes' Theorem requires that we find the defendant not guilty." This lone juror continues:

"I see two sides to proof beyond a reasonable doubt. Looking at the evidence against the defendant, we must be convinced beyond a reasonable doubt that he committed the crime. We have done that and seem to have concluded that he is guilty. If proof beyond a reasonable doubt does exist against the defendant, however, then we are also saying that no reasonable possibility exists that someone else did the killing. But if such a reasonable possibility does in fact exist, then I think that we must acquit. I guess a reasonable possibility must be the difference between one and whatever probability constitutes proof beyond a reasonable doubt. If we had decided that proof beyond a reasonable doubt required the prosecutor reach a 99% probability of guilt, then, of course, we would have acquitted if the proof had only reached 98%. I think, therefore, that if a 2% chance exists that someone else did the crime, we have a reasonable doubt as to the defendant's guilt and must set him free."

The foreman replies, "But there can't be this 2% chance. You should have listened more carefully to the statistician who testified. He said that certainty in probability equals one. For example, it is certain that a coin when flipped will come up either a head or a tail. It follows then that the probability of either a head or a tail on a coin toss must equal one. These two possibilities on a coin toss are mutually exclusive of one another, that is, 'the occurrence of one necessarily precludes the occurrence of the [other].'[163] The statistician explained that the probability of one mutually exclusive event or another occurring is the sum of the probabilities that each will occur separately.[164] When it is certain that one of the mutually exclusive events has to occur, the sum of probabilities equals exactly one. Thus, there are six mutually exclusive events to the roll of a die. The sum of the probabilities of each occurring separately equals precisely one. Thus, whenever mutually exclusive events are identified and they exhaust the possibilities, the sum of the separate probabilities will equal one.

"We have decided that the murderer's blood is on the knife," continued the foreman. "The proof is clear that only one person's blood is

---

162. This can be determined from the table presented by Finkelstein and Fairley. *See* Finkelstein & Fairley, *supra* note 2, at 500.

163. Braun, *supra* note 2, at 70-71.

164. *See id.* at 71. "Another way to view mutual exclusivity is to note that the probability of two such events occurring simultaneously is zero because, by definition, such mutual occurrence is impossible." *Id.* at 71 n.139.

on that knife. We have concluded that there is only one killer. Therefore, if we know whose blood it is, we know who the killer is. Now, either it is the defendant's blood on that knife and he is guilty or it is someone else's blood. These two events are mutually exclusive and they take in all the possibilities. That means that the sums of the probabilities of these two events must equal exactly one.

The lone juror says, "We agree."

The foreman forges on. "Then, since we have determined that the odds of the defendant's guilt is 99.7%, that must mean that the chances of everyone besides the defendant being guilty are 0.3%. It just cannot be that someone else has a probability of guilt of 2%."

The lone juror counters, "From what you say, you must agree with me that if I can convince you that a person exists who has a probability of guilt of, say 15%, then the defendant's probability could only be 85%."

"If you can convince me of that impossibility, I will switch my vote since I agree that an 85% chance of guilt is not proof beyond a reasonable doubt."

The Henry Fonda-like juror continues, "We know that the blood on the knife occurs at a rate of one in 1000 people. I think that we can all agree that there are about fifteen million people in the metropolitan New York City area. That means that there must be 150,000 people around who have blood which matches that found at the scene. So far we have only concentrated on analyzing the evidence against one of those people, the defendant. I am positive, however, that if all those people were investigated, we would find evidence of guilt against some of them. Don't you think that the odds are overwhelming that at least one of those 150,000 took a sudden trip shortly after the murder? We considered the defendant's journey as evidence of guilt which went into our calculation of the prior probability. Obviously the trip of any of those 150,000 should also be considered as evidence of guilt. We would now have a prior probability of guilt for this other person. We then should apply Bayes' Theorem concerning this probability and the genetic frequency evidence to see what probability of guilt exists for this other person.

"Another example," he continued, "Perhaps my view of human nature is too cynical, but I am willing to bet a bundle that if the police could ask all 150,000 where they had been at the time of the murder, some, either because they were guilty, just contrary, didn't like the police, had forgotten, or didn't want their spouse, boss, or parents to know where they had actually been, would not answer truthfully. If the defendant had given a false alibi we would have considered that some evidence of guilt, and therefore it should be considered evidence of guilt for those others.[165]

"Certainly the odds are high that one of those 150,000 recently spent money which he or she apparently did not have a short time before. Since the victim's purse was taken, if the defendant had recently spent a sum of unaccounted for money, we would have seen that as evidence of guilt. Consequently, sudden unexplained spending by any of the others of our 150,000 should be treated as evidence of guilt.

"And so on. The odds seem overwhelming that if all 150,000 were investigated as fully as the defendant, we would learn details about some of them that would tend, if only slightly, to show that person to be guilty. What are the probabilities of guilt from a sudden trip after the crime or from a false alibi? If we assessed the prior probability at only one in 10,000 from such evidence, then the genetic data and Bayes' Theorem would indicate a 9.1% chance of guilt for this other person. That's more than the 2% that you said couldn't exist."[166]

The foreman answers, "I see your point. Perhaps you are right that if we knew all the facts, there might be somebody who would appear to be at least slightly guilty and who has blood like the defendant's. In fact, I might agree with you and say if all 150,000 were investigated, we certainly could expect to find somebody with enough of a probability of guilt from the non-genetic evidence that Bayes' Theorem coupled with the marker frequency would prove a reasonable chance that some person other than the defendant was guilty. But we can't reach that conclusion. The judge told us we were to use our common sense and reach a verdict based on the evidence or the lack of evidence presented in court. We were expressly forbidden to speculate on matters not in evidence; we were told that we could only reach deductions that were supported by the evidence. While I think that it is a fair conclusion from the evidence that people other than the defendant have the same combination of genetic markers, it is the purest speculation to assume that any evidence of guilt would be introduced against them. Since we are forbidden to speculate about matters not supported by the evidence, we cannot acquit on the grounds you suggest. I guess the defendant should have introduced such evidence if he wanted us to acquit for that reason."

"Well, I thought that the defendant didn't have to prove anything; that the burden of proof was on the prosecutor. I doubt that there is any way that the defendant could identify those with blood like his. I wonder how the police do. Now that I think about it, I bet they don't.

165. *See, e.g.*, 2 J. WIGMORE, EVIDENCE § 279, at 141-42 (J. Chadbourn rev. ed. 1979). People v. Deitsch, 237 N.Y. 300, 303, 142 N.E. 670, 671 (1923).

166. Using the formula in the text accompanying *supra* note 147, the subsequent probability of guilt equals

$$\frac{1.0}{1.0 + .001(.9999/.0001)} = 9.1\%$$

I bet they first investigate and when they find someone whom, for whatever reason, they suspect, they try to get his blood. If it doesn't match, they cross that person off the suspect list. If the blood does match, they assume that person is guilty and concentrate their investigative resources to find all possible incriminating evidence against that person. I bet they never continue to seek anyone else with the matching blood much less try to get evidence of guilt against someone else. Frankly, I can't imagine how the defendant could make up for this police deficiency. Thus, even though my logic and common sense tell me that if we could apply Bayes' Theorem to the other people with the matching blood, we would find some with a reasonable chance of guilt, those other people have not been identified and brought before the jury for us to make an assessment of the probabilities of their guilt."

"So we are agreed that we must convict the defendant."

"Oh, no. I think that deductions based upon evidence presented in court clearly show that a reasonable doubt exists. We know that she was killed in her office building. We were told she worked for one of those legal factories; somebody said that her firm, including lawyers and support staff, employed four hundred people. We know that the building had fifty floors. Doesn't the evidence support the conclusion that roughly 5,000 people were in the building at the fatal time?"

"I guess so."

"It's obvious that the killer was in the building when she died. Indeed we said that it was evidence of guilt that the defendant was seen in the building near the time of her death. Therefore presence in the building is also an indicator of guilt for everyone else there when she died. Since the murderer had to be in the building and we have concluded that about 5,000 people were there then, we must mean that anyone then present, based on that fact alone, has a one chance in 5,000 of being the murderer."

"That seems to make sense."

"Furthermore, since the genetic combination occurs once in every thousand people and since there were 5,000 people in the building, five people must have the blood markers we are looking for."

"Hold it. Just because the frequency is one in 1,000 and the population is 5,000 that does not mean that there will be five people with markers."

"Okay. I agree. However, even though the number could be greater or fewer than five, I think that we can conclude the chances are overwhelming that at least one other person in the building has blood like the defendant's.[167] The prosecutor want us to use probabilities through the use of Bayes' Theorem to find the defendant guilty. Surely, therefore, it is all right to use statistics to reach the conclusion

167. *See supra* text accompanying notes 93-102.

that someone else in the building besides the defendant has blood like that found at the scene."

"That only seems fair."

"We said that each person in the building had a one in 5,000 chance of being guilty. The probability of guilt of the other person assessed from the non-genetic evidence is therefore one five-thousandth. If we use this prior probability in Bayes' Theorem along with the frequency probability of one in 1,000, my calculation indicates that this other person's probability of guilt is 16.7%.[168] If we analyze the evidence this way, the most defendant's probability of guilt would be is 83.3%. We agreed that that was not proof beyond a reasonable doubt and we must acquit."

"Wait a minute. We don't know anything about this guy. Maybe he was tied up in a meeting with twenty people from three to five on the day of the killing. We wouldn't say there was a 16.7% chance of his being guilty then."

"True. But, then again, maybe he was seen shortly after the murder with a cut. In that case, our initial assessment of guilt would have been higher than one in 5,000. And remember, we have just considered the possibility that only one other person in the building has the same genetic markers as the defendant when, of course, there are probably more than one. In any event, your suggestion of an alibi for this other person is pure speculation—something you have convincingly reminded us we cannot do. Instead, we have just used the evidence, drawn reasonable conclusions from it, and then applied Bayes' Theorem as the prosecution witness taught us to do. From this, we have concluded that it is reasonable to believe that a 16.7% chance exists that someone besides the defendant was the killer. We were told the burden of proof was on the prosecutor. If this other person has an alibi or some other evidence to show that he was not the killer, it was up to the prosecutor to do this in order to remove what we agreed would be reasonable doubt. He hasn't done that. The evidence indicates a reasonable doubt. We must acquit."[169]

168. Once again, using the formula in the text at *supra* note 147, the subsequent probability of guilt equals

$$\frac{1.0}{1.0 + .001(.9998/.0002)} = 16.67\%$$

169. The flaw in the hypothetical jurors' reasoning might seem obvious; the jury has not properly set and adjusted the prior probabilities. Something more basic than that is wrong, however. This flaw is inherent in the use of Bayes' Theorem in criminal trials—the sum of the probabilities calculated by the theorem for all possible suspects need not equal one.

A briefer example should illustrate. Assume that the nonstatistical evidence proved conclusively that the murderer was one of two people and each had an equal chance of being guilty. The prior probability of guilt for each is .50. Suppose the investigation revealed that each suspect has blood which matches the blood found at the scene of the crime. Such an occurrence is possible. *See* Terasaki, *Resolution by HLA Testing of 1000 Paternity Cases Not Excluded by ABO Testing*, 16 J. FAM. L. 543, 549 (1978) (Identical twins have the same genetic markers in their blood). If the

VI. Conclusion

Because genetic markers occur independently from each other and because studies are now available showing how frequently specified markers appear in the population, a valid probability for a combination of genetic markers can be calculated. The introduction of this frequency figure into a trial, however, is of little probative value by itself. On the other hand, if the evidentiary effect of that probability is misjudged, as it easily may be, the prejudicial impact of the probability will be great.

Most commentators who have studied the introduction of probabilities into criminal trials agree that the dangers of introducing such a frequency figure are great. Instead, some have argued that a probability such as the one derived from genetic marker evidence should be admitted into criminal trials through the use of Bayes' Theorem. Others have argued that the theorem should be barred because it would give a spuriously exact probability, tend to distort trials by a concentration on the quantifiable evidence, and undermine our notions of the presumption of innocence and proof beyond a reasonable doubt.

Most certainly, however, Bayes' Theorem should not be used simply because it often does not work.[170] Since, as one scholar has noted, Bayes' Theorem is "the only plausible mode of integration [of mathematics into the trial process] yet proposed"[171] and since this approach fails, we are still looking for the proper way to present the probabilities derived from the blood genetic marker evidence.

---

blood occurred at a frequency of one in 1,000, Bayes' Theorem indicates a subsequent probability of .9990 for each suspect. *See supra* text accompanying note 115.

170. Glanville Williams describes a hypothetical situation in which it is clear that a defendant is guilty, but Bayes' Theorem indicates only a 10% chance of guilt. Williams, *supra* note 77, at 347-49.

171. Tribe, *supra* note 76, at 1350.