CHAPTER 4

INTERPRETING THE PROBATIVE VALUE OF DNA TYPING: POPULATION GENETICS AND STATISTICAL ANALYSIS

Once a DNA match has been declared, the other difficult task of the scientist is to interpret it. Like traditional fingerprinting, DNA evidence is not based on certainty but on some probability analysis. Occasionally a DNA fingerprinter would say that there is one chance in 50000 or 500000 that a randomly selected Hispanic male have the same DNA profile as that found in both crime scene sample and the sample collected from the accused. In this connection the scientist must answer two fundamental questions in order to accept his result. They are (1) what is the probability of the crime scene profile if it came from the accused? (2) What is the probability of the crime sample profile if it came from some unknown person? Therefore, a scientist who had performed the DNA typing must establish the uniqueness of a match. A paper published by D.J. Balding provides a clear understanding of calculating the probability that the accused is the person who possesses a given DNA profile.

1 In a DNA typing it is easy for the scientist to exclude a person by saying that there is no match but in the case of declaring a match it is a difficult task for him to establish that the match is a true match. The reliability of a match depends on the reliability of the assessment made by a scientist in interpreting it.

2 It is the duty of the DNA typing scientist not only to say that two samples exhibit the same pattern but also to suggest that the pattern is extremely rare. See, Peter J. Neufeld and Neville Colman, "When Science Takes the Witness Stand", 262 Scientific American 18 (1990).

3 Normally the reliability of a match probability always depends on the first question. However, in a forensic testing due to the lack of the identity of the actual culprit, it becomes fundamental to answer the second question. See, Richard Lempert, "Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes", 13 Cardozo L. Rev. 303 (1991).

4 See, D.J. Balding, "When can a DNA Profile be Regarded as Unique?", 39 Science and Justice 257-250 (1999) as quoted in Ian W. Evett, Lindsey A. Foreman, Graham
According to him the uniqueness of a DNA match depends on the population of all the other people who can be considered as possible suspects for the crime at issue. This hints that all other persons in a population must be eliminated from the probability that any one of them may have the same match as the accused have. Thus one can come to the conclusion that the smaller the match probability, the stronger the evidence in support of the proposition that the crime sample came from the accused. Therefore, in order to get reliable DNA evidence, commentators recommended that a DNA match must be confirmed only after taking an estimate of "Match Binning Frequencies".

---


5 If the relative frequency of the incriminating fragment size is large, that means, if many people would have the same match, then the said match would have a less probative value. See, David H. Kaye, "DNA Evidence: Probability, Population Genetics and the Courts", 7 Harv. J. L. & Tech. 101 (1993).

6 However, some commentators has stated that there is no need to make a probability estimate from the DNA match made by a scientist who had followed a proper laboratory protocol. For example, Andre A. Moenssens in his article published in Jurimetrics journal has commented as follows:

Evidence of DNA match made by a scientist who followed the proper laboratory protocols is admissible without drawing statistical inferences there from. A witness can simply testify to having performed the necessary steps and having determined that the two samples examined match in all their characteristics. The expert, can further, testify that in her experience she has never encountered two different individuals who exhibited the identical DNA characteristics and for that reason the witness's conviction of identity can be stated with a reasonable degree of scientific certainty. ... The next step of extrapolating calculations as to the probability of random matches is not an essential step to DNA identification testimony. But inertia being what it is, and serologists having testified for years to mathematical odds against random matches in traditional blood groupings, the practice was inevitably adopted by the DNA profiling community. See, Andre A. Moenssens, "DNA Evidence and its Critics – How Valid are the Challenges", 31 Jurimet. J. 87(1990).

7 "Match Binning Frequencies" a're the estimate made by the DNA scientist after comparing the frequencies of the alleles found in the crime sample gene with the random samples collected from the relevant population. For more details see, Richard C. Lewontin and Daniel L. Hartl, "Population Genetics in Forensic DNA Typing", 254 Science 1745 (1991).
2. **Population Genetics in DNA Matching**

Determination of DNA matching in accordance with the population genetic setup is one of the most difficult stages in the DNA interpreting process. In this stage scientists will use random chance of finding another individual with the same DNA profile within the population of all possible suspects. Before going to the process it is important to know what is population genetics. Population genetics means and contains a group of population defined by the anthropological and ancestral identity of individuals. Thus for forensic purposes the reference population is divided into certain major racial or sub groups like Caucasians, Blacks, Hispanics etc. Some groups were again sub divided considering the geographical aspects. Thus once the ethnic group of the suspect has been identified from his physical appearance or by other means, DNA frequencies can be determined using population databases.

3. **Method of Calculating Random Match Probabilities**

In forensic DNA typing if a DNA collected from the crime scene matches with the sample collected from the suspect, the scientist must prove that the suspect is the only possible source of the specimen. The inference is based on

---


10 The probative value of the DNA evidence always depends on the “random match probability” estimate given by the scientist. The smaller the match probability, the stronger the evidence in support of the proposition that the crime sample came from the suspect. Therefore, to catch this uniqueness, scientific studies recommends increasing the number of loci’s to get smaller match probabilities. In England, with the help of a new technique known as “SGM plus” they are analyzing more than 10 loci, producing the match probability in the range of 1 in 10 billion to 1 in 100 trillion. See, Ian W. Evett and Lindsey A. Foreman, “DNA Profiling: A Discussion of Issues Relating to the Reporting of very Small Match Probabilities”, [2000] *Crim. L. R.* 341 at 347.
two assumptions generally accepted in the forensic DNA community. They are
(1) Hardy-Weinberg Equilibrium and (2) Linkage-Equilibrium.

4. The Hardy-Weinberg Equilibrium in Forensic DNA Typing

The underlying principle of Hardy-Weinberg equilibrium is that allele and
genotype frequencies do not change from one generation to next. An English
mathematician Godfrey Hardy and a German physician Wilhelm Weinberg
discovered the equation. In forensic DNA typing, Hardy-Weinberg law justifies
the assumption of statistical independence implicit in formulas used to calculate
the probability that the DNA patterns of a specimen and of a suspect would
match by chance alone. The basics of Hardy-Weinberg law is as follows:

If two heterozygous individual's mates (a gene having 'B' as a dominant
allele and 'b' as the recessive allele) their offspring will inherit the alleles as
follows: (1) 25% of their offspring's will be homozygous for the dominant alleles
"BB" (2) 50% are heterozygous like their parents "Bb" and (3) 25% are
homozygous for recessive allele "bb" and were different from their parents.

Results of random union of the two gametes produced by two individuals,
each heterozygous for a given trait. As a result of meiosis, half the gametes
produced by each parent which carry allele B; the other half allele carry allele b.
Here 0.25 alleles will be BB, 0.25 alleles will be Bb, 0.25 alleles will be Bb and
the balance 0.25 alleles will be bb.

Results of random union of the gametes produced by an entire population
with a gene pool containing 80% B and 20% b may result in 0.64% allele as BB,

---

11 M. Krawczak and J. Schmidtke, DNA Fingerprinting (2nd ed. Bios Scientific
12 Technical Comments, "Forensic DNA Tests and Hardy-Weinberg Equilibrium", 252
Science, 1038 (1991)
0.16 allele as Bb, 0.16 allele as Bb and the balance 0.04 alleles as bb. The respective 0.64%, 0.32% and 0.04% were got from the calculation as follows:

64% homozygous for BB i.e. 0.8 x 0.8 = 0.64.

32% Bb will be heterozygotes i.e. 0.8 x 0.2 x 0.2 = 0.32.

4% bb will be homozygous i.e. 0.2 x 0.2 = 0.04.

Thus 96% of the generation will be homozygous and the balance 4% will be heterozygous.

Keeping this calculation in mind one has to consider whether the 0.04% recessive alleles may disappear in future generations. This does not happen because after each mating the recessive alleles will be duplicated. The proportion of the allele b will remain as 0.04% after every generation.

Now one can find out the equilibrium by an algebraic analysis of the problem using the expansion of the binomial (a + b)^2

(a + b)^2 = a^2 + 2ab + b^2.

Here, if 'a' represents the frequency of one gene in the total number of genes in a population and 'b' represents the frequency of its single allele. So, a + b = 1.

a^2 = the fraction of the population homozygous for 'a'.

b^2 = the fraction homozygous for 'b' and

2ab = the fraction of heterozygotes.

Here, a = 0.8 and b = 0.2 thus (0.8 + 0.2)^2 = (0.8)^2 + 2 (0.8) (0.2) + (0.2)^2.

= 0.64 + 0.32 + 0.04

Here b^2 = 0.04, so b = 0.2, the frequency of the b allele in the total population. Since a + b = 1, a = 0.8 and allele B makes up 80% of the gene pool.
Because B is completely dominant over b, one cannot distinguish the Bb alleles from the BB ones by their phenotype. But substituting in the middle term (2ab) of the expansion gives the percentage of heterozygous alleles. 2ab = (2)(0.8)(0.2) = 0.32. So, recessive genes do not tend to be lost from a population no matter how small their representation.

Thus Hardy-Weinberg law shows that gene frequencies and genotype ratios in a randomly breeding population remain constant from generation to generation\(^{13}\). However, Hardy-Weinberg law had certain limitations. Hardy-Weinberg law can be applied only if following conditions are satisfied: (1) The population must be large; (2) The mating must be a random mating; (3) No mutation must be there. The frequency of gene 'B' and its allele 'b' will not be in Hardy-Weinberg equilibrium if the gene 'B' and its allele change due to mutation. However, scientists opined that mutation only play a minute role in evolution; (4) No gene migration. All human beings are made up of genes of the population in their own locality. If the members of one population mated with the members of another, the local population of that particular area may develop a distinct gene pool; (5) No genetic drift. If the population is so small then there is a chance that certain essential members to constitute a gene pool may be eliminated. If so the frequency of an allele may begin to drift toward higher or lower values.

5. Linkage Equilibrium

Linkage equilibrium is an assumption, which means that all the genotypes at each locus are randomly assorted within the population. This allows the calculation of the probability of each independent locus by the product rule. For example, one can multiply the match of five independent loci to get the total probative value. This will be useful only in multi locus fingerprinting and not in single locus.

6. How Frequency of Alleles Calculated

The application of population genetics in forensic DNA fingerprinting is to determine the frequency with which a specific allele occurs within a given ethnic group. If the frequency of the allele in question is quite common in the concerned population, the probative value of the evidence will be very low and it gains no value in forensic purposes. If the likelihood of matches in the defined population is reduced, that gives DNA identification technology more value in forensic examination. For example, the Rh-positive blood types are common; therefore a scientist will give evidence that its frequency in the relevant population is large. At the same time the Rh-negative allele was found in the general population in somewhat uncommon and hence its frequency is very high. However, in the case of DNA fingerprints, scientists are using the alleles that occur in the polymorphic section of the genome, hence the likelihood that the samples will match is much smaller. Therefore, it is necessary to determine the frequency with which each individual allele occurs in a particular population. The determination starts with the finding of the frequency of the individual bands (alleles) in the autorad. This is by a method adopted by the United States
Federal Bureau of Investigation called “binning”\textsuperscript{14}. The FBI in a scientific paper\textsuperscript{15} published this method and it has been elaborately discussed in a leading United States decision, \textit{United States v. Yee}\textsuperscript{16}. The method was summarized as follows:

The FBI in their laboratory has developed a table of frequencies. The frequencies of the alleles corresponding to the DNA sample that is being tested are then determined by reference to this table. For constructing the reference table they have developed a method called “fixed bin method”. As a first step FBI has randomly collected the DNA samples of 200 or 300 persons from the relevant population (normally Caucasian for Americans). Then these collected DNA’s were subjected to DNA fingerprinting using five or six probes, the laboratory routinely uses or intended to use. Then the loci identified by these probes were assigned to the predetermined bin. DNA forensic laboratories are regularly using “size markers”\textsuperscript{17} in the R.F.L.P: typing process. The bins were established with reference to the size markers that were applied in each test. With the help of these size markers an allele (band) was determined to have a particular fragment size and it would be placed in the bin, which included that size. Thus each random sample collected from the population data base were tested by the probes and each of the alleles (bands) detected are put in the bin

\textsuperscript{14} Binning is a method generally used for determining the individual band frequencies.


\textsuperscript{17} The testing laboratories use the size markers during gel electrophoresis. The size markers are solutions composed of DNA fragments of known, predetermined fragment lengths. During gel electrophoresis they will run adjacent with test DNA’s. After the completion of the process they will finally appear in the autorad as an array of bands. The length of the sample DNA can easily trace out with the help of this size markers because the length of the fragments of size markers are known lengths. By comparing the size of the test DNA fragments with the size marker fragments, which is nearest in location to the test DNA fragments, their correct length can be determined. See, \textit{U.S. v. Yee}, \textit{Ibid} at 172.
corresponding to its base pair size as determined by the size markers on the gel. In fixed bin method, a bin would consist of all alleles falling in between two adjacent DNA markers. The frequency of any allele in this bin would be defined as frequency of all alleles in the bin. Any allele identified will fall within one of the fixed bins and will be assigned a frequency based upon that bin. The FBI was applying a standard safety measure known as “collapsing” the bins that contained less than five bands (alleles). The frequency of each bin is determined by dividing the total number of bands located in a bin by the total number of bands generated from all the data base samples tested for that probe. If the frequency for various bands were determined the overall frequency for the DNA profile is calculated. This calculation is made by probe by probe basis i.e. the frequency of the bands under a single probe will be calculated at one time. Thus frequencies of the bands under all the probes will be calculated one by one.

7. Calculation of the Frequency of Bands in Homozygote and Heterozygote Bands

After the completion of the DNA analysis, the DNA profile at a single profile may shows two bands or one band. If the DNA profile shows two bands they are called heterozygote and if the DNA profile shows only one band it is homozygote. The frequency calculation in homozygote and heterozygote is different. In the case of heterozygote the FBI calculates the frequency by using a formula taken from the Mendelian genetics. The formula is $2pq$. Here ‘p’ is the bin frequency associated with one band and ‘q’ is the frequency associated with

\[ 2 \times \binom{p}{1} \times \binom{q}{1} \]

---

20 Ibid.
21 Ibid.
other band. In the case of homozygous, band frequency is calculated using a formula in traditional genetics i.e. 'p' square.\textsuperscript{22}

If the frequency of the bands (alleles) from different probes were calculated, the aggregate frequency of the total DNA profile is determined by using a rule in mathematics called the "product rule". With the help of product rule one can multiply the frequencies of different probing. Thus one can get the total probative value of the frequencies associated with different loci.\textsuperscript{23}

8. Lewontin - Hartl's Hurdle on Population Genetic Application of Frequency Calculation

The population genetic application of DNA fingerprinting is based on two assumptions; the Hardy-Weinberg equilibrium (allelic independence within the loci) and Linkage equilibrium (allelic independence among loci). The application of the product rule is also based on these assumptions. However, the later studies made by two American geneticists, R.C. Lewontin and Daniel L. Hartl shows that there are some theoretical misunderstandings regarding the application of the population genetics in frequency calculation.\textsuperscript{24} The major criticism leveled against the calculation of match probability is using multiplication (product rule). They challenged the major assumption in population genetics that racial groups are genetically homogenous. Their observation is as follows:

\textsuperscript{22} Ibid.

\textsuperscript{23} Ibid. The scientific community admitted the validity of using this product rule in frequency calculation. The product rule can validly apply only if the sums to be multiplied are statistically independent. In the case of frequency calculation, the frequencies associated with different loci can be multiplied together because the occurrence of the genetic events at one loci are considered as independent of the occurrence of the genetic events of the other loci. Here the law of linkage equilibrium will be applied. This allows the calculation of the total random match probability i.e. \(G_1 \times G_2 \times G_3 \times G_4 \times G_5\) (here the frequency of five locus's were multiplied using product rule).

Both these claims are based on misinterpretations of population genetics theory. More importantly, they ignore a considerable body of evidence indicating genetic substructure within what are called the “Caucasian”, “black, and “Hispanic” populations. The census populations designated “Caucasian”, “black”, and “Hispanic” are actually each made up of multiple sub populations that are genetically diverse. Consequently, with current available data, the current method of estimating the probability of a match by multiplying together the frequencies with which of individual VNTR pattern occurs in a reference database is unjustified. Furthermore, the magnitude and direction of the error depends on the particular VNTR locus, the bands observed, and the reference database. Hence, it cannot be ascertained whether the estimates as currently calculated are biased for or against any particular defendant. On the other hand, although the current method is flawed, it is not irretrievable, and not suitable data could be gathered that would allow acceptable estimates to be made. Perhaps a better solution would be to abandon the current method altogether and replace it with one or more of the alternative approaches.\textsuperscript{25}

Thus the study conducted by Lewontin and Hartl shows that there are varying subgroups in the major ethnic groups like Caucasian, Blacks and Hispanics. The presence of these subgroups with varying DNA patterns may contradicts the assumptions that guarantee independence of alleles at a specific locus and makes doubt on the validity of multiplying “genotype” frequencies across loci.\textsuperscript{26} According to them, these ethnic subgroups within each database

\textsuperscript{25} Ibid. at 1746. Dr. Richard Lewontin had testified these aspects in Yee and in U.S. v. Jakobetz, 747 F. Supp. 250 (D. Vt., 1990).

tend to mate endogamously (i.e. within a specific subgroup) with persons of like religion or ethnicity or who lives within close geographical distance. Such endogamous mating tends to maintain genetic difference between subgroups. This may result in the differences in the DNA frequencies of a given DNA fragment. A DNA fragment, which is common in some subgroups, may not be included in the major ethnic group like Hispanic or Caucasian. Thus there is no reference “Caucasian” database that would be meaningful for all these different subgroups. Moreover, they argue that a VNTR combination that is very rare in reference database might, for example, be much more common in the suspect’s particular group, thus raises the chances that he will be incorrectly identified as the criminal. After the evaluation they came to the conclusion that the probability of the random match of a given VNTR phenotype should not be estimated for Caucasians, Blacks and Hispanics with the present method of multiplication and the databases presently available. Instead they have propounded certain alternatives. Among them the relevant thing to be considered is to look at the allele frequencies within each subgroup and then multiply them along with the common alleles in the racial group.

However, later studies conducted by certain biostatistician’s shows that the fear about population substructure is baseless. Studies conducted by two biostatistician’s, Neil J. Risch and Bernard Delvin on the frequency of matching alleles for large numbers of pairs of different people in laboratory databases shows no false matches across four or five loci and rates of matches on subsets
of loci that do not depart markedly from the expected values given independence of alleles across loci.\textsuperscript{27} Regarding this they observed:

The observed independence of matching among loci, both in the FBI and Lifecodes data sets, provides no support for claims of linkage disequilibrium within ethnic groups. Indeed, if linkage disequilibrium among loci does exist, it has little effect on the probability of two random individuals having matching genotypes.\textsuperscript{25} Although we find the probability of a matching DNA profile between unrelated individuals to be vanishingly small, especially at five loci, related individuals, in particular identical twins and siblings, have a far greater probability of matching genotypes. For identical twins, the probability is 1.0, while for siblings it is (0.25) or 0.001. Therefore, in the forensic setting, we conclude that an innocent suspect has little to fear from DNA evidence, unless he or she has an evil twin.\textsuperscript{28}

In order to settle the conflict on population sub structure between the two major groups in the scientific community, National Research Council of United


\textsuperscript{28} Ibid, at 19-20. Similar criticism has also been given by other two scientists in population genetics. According to them Lewontin and Hartl’s claim against population genetics was wrong. They pointed out that Lewontin and Hartl’s use old and inappropriate blood group data to bolster their contention that allele frequency differences between human groups might affect the calculated probability of match between two samples. Moreover they argued as follows:

Lewontin’s work is often cited as showing that substantial genetic variation exists within the major racial groups. However, when Lewontin’s approach is applied to smaller levels of population structure, the majority of genetic differences are still found to be within villages or parishes. These results demonstrate the truism of biological diversity of individuals, even in extremely subdivided groups. The reality of human evolution shows that even though marital preference is nonrandom at every level at which one can define populations, its effect on deviation from HWE of genotype frequencies of linkage equilibrium is minimal. No new population genetic principles are needed to apply this thesis to forensic DNA typing. See, Ranajit Chakraborty and Kenneth K. Kidd, "The Utility of DNA Typing in Forensic Work", 254 Science, 1735 (1991).
States implemented a new principle known as "ceiling".²⁹ The principle provided in the report of the committee is as follows. The rule says that the traditional multiplication method will give better results even if there is a sub stratum in the ethnic group. The ceiling method involves two important steps: (1) For each allele at each locus, determine a ceiling frequency that is an upper bound for the allele frequency that is independent of the ethnic background of a subject; and (2) To calculate a genotype frequency, apply the multiplication rule, using the ceiling frequencies for the allele frequencies. Committee suggested that random samples of 100 individuals must be drawn from each 15-20 populations, each representing a group relatively homogeneous genetically; the largest frequency in any of these populations or 5%,³⁰ whichever is larger, should be taken as the ceiling frequency. For example, if two locus has been studied in two populations, allele 'A' of the locus 1 in the Indian and English population shows the frequency of 7% and 10% and allele 'B' shows the frequency of 12% and 9%. Allele 'C' of the locus 2 shows the frequency of 3% and 6% and allele 'D' shows the frequency of 9% and 13% respectively, the ceiling rule will assign the values as 10% for allele 'A', 12% for allele 'B', 6% for allele 'C' and 13% for allele 'D'.³¹

The ceiling principle was seriously criticized by the scientists as having no statistical or genetic foundation.³² They argue that the principle was too


³⁰ A minimum of 5% is considered as mandatory for each allele and should be substituted if the frequency estimate is small in all groups.


conservative and erroneous. Therefore, the National Research Council in its 2nd report abolished the ceiling principle and recommended to follow the normal method of calculating match probability.\(^{33}\)

The other important limitation while interpreting the match of a DNA test result is the chance of having similar match to any of the near relatives of the suspect. The probability that the forensic sample would match a relative of the person who left it is considerably greater than the person who is selected randomly from the general population. An individual will receive half of the genes from his father and another half from his mother, therefore it is very difficult to distinguish between siblings i.e. between brother and brother or sister and sister. A far simpler and straightforward approach suggested is to obtain profiles from near relatives to establish unequivocally whether a relative has an identical profile.\(^{34}\)

9. Statistical Interpretation of DNA Profiles

It is an accepted fact among scientists in the scientific community of forensic DNA Fingerprinting that DNA test results are meaningless unless it is conveyed to the court in a statistical form.\(^{35}\) There are mainly two reasons for this (1) like traditional fingerprinting, the value of the DNA evidence is not certain. It depends on the probabilistic calculation. (2) Because of the involvement of

---


\(^{35}\) According to Walls, “Among forensic scientists there is a growing recognition that in many cases the results obtained yield their maximum information only if statistical methods and calculations of probability are used”. See, Walls, “Ten Years of Forensic Science 1964-73”, [1974] *Crim. L. R.* 505; see also William C. Thompson, "Are Juries Competent to Evaluate Statistical Evidence?", 52 L. & Contemp. Probs. 9 (1989).
population genetics, the probability of a match report could be given only in numerical terms. However, there is a difference of opinion among persons in this field that whether judges are capable in evaluating the probabilistic statistical testimony provided by the DNA scientists. The general concern regarding this is that judges may admit technical evidence at its face value without any evaluation either because of their ignorance in the subject or due to a special status given to the technical witnesses or their subject.

In the previous topic it is clearly dealt with the importance of the application of the procedures regarding the estimation of the frequency of the DNA match profile in the population concerned. However, this estimation in itself is not enough to provide a sufficient basis for the judiciary in determining the guilt or innocence of an accused. What court needs from an expert is the likelihood ratio, which provides the probability of the evidence given guilt and the probability of the evidence given innocence. Nevertheless, an expert is not expected to state the guilt or innocence of the accused, which is exclusively within the province of the

36 However, in some laboratories they are using a different method to convey the evidentiary value of DNA test results. For example, in England, the Forensic Science Service (FSS) Laboratory at Birmingham uses a set of descriptive terms to explain the strength of matching probabilities. They use certain English terms in an ascending order to convey the strength of evidence: inconclusive, weak support, moderate support, moderately strong support, strong support, very strong support and conclusive.


38 See, Michael Freeman and Heler Reece, Science in Court (Dartmouth Publishing Co., Ltd. 1998).
court of law. In the leading English decision, R v. Adams, when the Criminal Appeal Court was considering the evidentiary value of the statistical evidence given by a forensic DNA expert, Philips LJ, Jowitt and Keene J J., observed:

It was important that a scientist, in giving evidence, should not go in to matters, which are for the jury. He should explain the nature of the DNA match and give the random occurrence ratio; he may be able to say how many people with matching characteristics are likely to be founding the United Kingdom or in a more limited sub group. This will often be the limit of the evidence which he can properly and usefully give. He should not be asked his opinion of the likelihood that it was the defendant who left the crime stain.

Thus in this case court has given the maximum extent of a scientific expert's testimony. Moreover, court gave a suggestion regarding the role of the expert in explaining the significance of the DNA evidence in the following words:

Provided that he has the necessary data, and the statistical experience, it may be appropriate for him then to say how many people with the matching characteristics are likely to be found in the United Kingdom or perhaps in a more limited relevant sub group, such as, for instance, the Caucasian sexually active males in the Manchester area.

Historically, it has been declared through precedents that an expert, whatever his efficiency in the field he has, is not allowed to express an opinion on the issue which the court had to decide (the so called Ultimate issue rule) see, R v. Anderson and Others, [1971] 3 All ER 1152 and State v. Pali Ram, AIR 1979 S.C. 14.


ibid. at 671. In this case the DNA expert gave evidence that he was sure that the accused was the offender.

ibid.
Thus from the above discussion it is submitted that the calculation of a scientific DNA evidence always come before a court of law in a probabilistic manner. The great task of a judicial officer starts here. He is not only concerned with the evaluation of the scientific evidence on record but also other non-scientific evidence, which may have an equal or more probative value. Here he has to combine the evidentiary value of both scientific and non-scientific evidence in an equal proportion. In this situation an important difficulty, which has been regularly faced by courts, are that in almost all criminal trials in which scientific evidence is adduced, prosecutors may present the evidence favorable to them. For example, assume that the frequency of a DNA profile is 1 in a million and that there is a match of the defendants profile and the DNA sample collected from the forensic sample, the prosecutor may express that there is only 1 in a million chance that the defendant is innocent. This is called as “The Prosecutors Fallacy”. The correct statement prosecutor ought to have given regarding this is “the chance of obtaining this DNA profile if the DNA in the

---

43 Thompson and Schumann coined the phrase “The Prosecutors Fallacy”. They have also coined another phrase known as the “Defense Attorney's Fallacy”. Under this phrase the defense attorney when faced with the frequency of a DNA profile of 1 in a million may say that this implies that in a country, which has a population of 300 million, 300 individuals possess the same DNA profile, all of which could have committed the crime. He may further argue that any one of those 300 individuals could have committed this crime and how court can reliably come to the conclusion that the accused in this case has committed the crime. See, Thompson, WC and Schumann, EL, “Interpretation of Statistical Evidence in Criminal Trials: Prosecutors Fallacy and Defence Attorneys Fallacy”, 11 Law and Human Behaviour 167 (1987) as quoted in Bernard Robertson and G.A. Vignaas. Interpreting Evidence: Evaluating Forensic Science in The Court Room (John Wiley and Sons, 1995), p.91. Bernard Robertson has stated that these fallacies can eliminate through a balanced approach by comparing the probabilities of the evidence given both the prosecution and the defence stories. This gives one a likelihood ratio, which tells the value of the evidence and the impact that it should have on the case as a whole. See also, Balding and Donnelly, “The Prosecutors Fallacy and DNA Evidence”, [1994] Crim. L. R. 711; Mathews, “Improving the Odds on Justice”, New Scientist 12 (1994).
forensic sample came from an individual other than the defendant is 1 in a million.\footnote{44}

10. Match Probability and Likelihood Ratio in Interpreting DNA Profile

Rationality in decision-making is the soul of the fact-finding process. If accurate fact-finding is the motto of an adversarial trial process, more importance must be given in achieving it. It is also a well-known fact that the evidentiary value of a particular event could be correctly measured only by way of probability,\footnote{45} which has its root in logic. However, mere logic and commonsense cannot give accurate results in every situation. Therefore, the new trend among legal luminaries is to adopt a quantifying way of fact determination. Here subjective elements of commonsense are converted into numerical values to get a mathematical precision. In this section it is discussed, how far mathematics can be adopted and adapted in the legal field for attaining a maximum rationality in decision-making process especially in criminal adjudication involving scientific evidence.

Probabilistic evaluation of scientific evidence is not a novel thing in the history of fact-finding process. U.S. courts started depending on mathematical tools from the end of 1960’s. It was in People v. Collins,\footnote{46} the Californian Court

\footnote{46} 66 Cal. 2d 319 (1968). During this period “hypothetical testing” has been widely used for cases involving racial discrimination. But no similar attempt has been made by

\footnote{44} 44 In the first sight one may think that both statements expresses a similar meaning, but in fact they are different and gives different consequences. By the first one, the prosecutor is actually making a statement about the guilt or innocence of the defendant conditional on the frequency of the DNA profile of the forensic sample. On the other hand, the second statement establishes only the frequency of the DNA profile in the population.

\footnote{45} 45 The probabilistic evaluation of evidence is an accepted thing in the law of evidence. The important phrase “beyond a reasonable doubt” which is the cornerstone of the criminal trial itself justifies the application of the probabilistic evidence in criminal cases. By the very old phrase what law expects is not certainty but a maximum satisfaction that the accused is not prejudiced by the lack of evidence, before he is going to be convicted.
of Appeal considered the legality of the evidence adduced by the prosecution completely depending on the statistical inference made by an instructor of mathematics. The brief facts of the case reveal that the accused Collin and his companion was arrested and tried by the State for the offence of robbery. When the case came up for trial, the prosecution has produced a mathematical instructor as an attempt to establish the guilt of the accused with the help of certain formulas in mathematics. As a first step, prosecutor has formulated certain characteristics of the accused basing on victims descriptions, such as the accused was a Negro man, had a mustache and beard, drove an yellow car with a girl having a blond pony tail hair. After the narration he assigned certain values (probabilities) to these characteristics and asked the mathematician to find out the value of the evidence presented. Applying the "product rule" he multiplied these probabilities and concluded that the figure of 1/12,000,000 represented the probability of any couple possessing all the characteristics of the accused.

The jury convicted both accused. They preferred appeal to the Supreme Court of California, reversing the conviction of the trial court on several grounds. One of the main grounds was that the court as unwarranted in this case found the

---

Ibid. at 326
application of the product rule because the characteristics for which the
probabilities were assigned are not statistically independent.\(^{48}\)

The Collins case later became the reason for heated discussions among
the members in the legal community regarding the application of the statistical
methods in fact-finding process. It was the joint work of Michael O. Finkelstein
and William B. Fairley in *Harvard Law Review*\(^{49}\) for the first time, which got the
public attention in this area. They proposed a new approach based on the well-
known statistical theorem, the "Bayes theorem" devised in the 18\(^{th}\) century by
Rev. Thomas Bayes. "Bayes theorem" will help the fact finder in two ways (1) it
will help to update the belief in a hypothesis with the help of new evidence. (2) It
will help to combine different types of evidences in a case (for example to
combine scientific and non-scientific evidence).

11. Some Basic Concepts in Statistical Evidence

(A) Hypothesis

Before the application of the Bayes theorem, a hypothesis or assertion must
be formulated. It is simply a belief of the fact-finder relating to a particular thing.
The hypothesis of a thing may be either true or false, but the person who
formulated it is not sure about it. Therefore, the person who formulated the
hypothesis will always try to prove its truth. For example, in a criminal trial a
judge will always try to prove his hypothesis 'accused is guilty in murdering the
victim'. The Bayes theorem helps the fact finder in proving the truth of a hypothesis
with the help of additional evidence.

\(^{48}\) For the correct application of the product rule the events to which probabilities were
assigned must be independent. If it is not independent the events cannot be
multiplied to get a total sum.

\(^{49}\) Michael O. Finkelstein and William B. Fairley, "A Bayesian Approach to Identification
(B) Probability or Odds (Prior and Posterior)

The probability is the scale by which the truth of a hypothesis is measured. Probabilities were formed by assumptions and attributions. It is supported with evidence as a condition. Therefore one can say that the existence of a probability depends on the quality of the evidence adduced. For forensic purpose, the value of probabilities is considered to be in between 0 and 1. If the probability is valued as 0 then one can come to the conclusion that the premise is false. On the other hand if the value of the probability is 1 then the premise is considered as true. The probability can be demonstrated as percentage if it is multiplied by 100%.\(^5\)

In a hypothesis testing there will be prior and posterior probabilities. In order to apply Bayesian theorem for updating the belief of a hypothesis, a judge must form certain prior probabilities. These prior probabilities are formed before getting any evidence. For example, to determine the truth of a hypothetical statement “the accused is guilty in murdering the victim”, a judge may take two prior probabilities (1) the probability that the accused killed the victim, and (2) the probability that the accused did not killed the victim. The values of these probabilities were then updated by additional evidence. Regarding this Finkelstein has said:

A judge might improve his performance further by using the statistically determined prior probability solely as a benchmark for what he believed to be an “average case”. Where in his opinion the facts showed that the case

\(^5\) For example, if a probability of 0.07 is multiplied by 100% one will get 70%. For more details see, Bernard Robertson and G.A. Vignaux, *Interpreting Evidence: Evaluating Forensic Science in The Court Room* (John Wiley and Sons, 1995), pp.11-16; Mike Redmayne, “Science, Evidence and Logic”, 59 *Mod. L. Rev. 747* (1996); David Hodgson, “A Lawyer Looks at Bayes Theorem”, 76 *Austl. L. J. 109* (2002).
was either stronger or weaker than usual, he could subjectively adjust the prior accordingly.\textsuperscript{51}

Posterior probability is the value of the prior probability after it is updated by additional evidence. This posterior probability will again come as a prior probability if any new evidence is there.

Odds are another form of mathematical method to establish the probability. Different from probability it will be always in the form of ratios. It will give a clear idea of how much more likely is an assertion is true than it is false. For example, whether the blood stained cloths found in the possession of the accused is that of the victim, can be given in odds as 0.50:0.50 i.e. half to half.

(C) Likelihood Ratio

Likelihood ratio is the ratio between the probability of the further evidence if the hypothesis is true and the probability of the further evidence if the hypothesis is false. It was in U.K. for the first time they proved that if the scientific evidence were presented in likelihood ratios, it would give better results. The National Research Council of U.S. in its second report has stated that computations of the likelihood ratio for the hypothesis that the defendant is the source of the sample as a proper method.\textsuperscript{52} In an Australian case \textit{R v. Karger}\textsuperscript{53} an expert explained likelihood ratio as:

\begin{itemize}
\item \textsuperscript{51}Michael O. Finkelstein and William B. Fairley, "A Bayesian Approach To Identification Evidence", 83 \textit{Harv. L. Rev.} 489 (1970). However, the formulation of prior probabilities was seriously criticized, because they were formulated by the judge according to his subjective satisfaction. Ignorance about certain necessary facts connected with the predictions may affect the output of the prior probabilities.
\item \textsuperscript{52}See, The Evaluation of Forensic DNA Evidence, Committee on DNA Forensic Science: an Update, National Research Council (National Academy Press, Washington, D.C. 1996).
\item \textsuperscript{53}[2002] S.A.S.C. 294.
\end{itemize}
The likelihood ratio is becoming a more commonly used form of presenting not only DNA evidence but other forms of forensic evidence. It is based on the fact that, in a court situation, generally, the court is faced with two competing scenarios, or points of view; the prosecution point of view that the accused has left the crime stain, and the defence point of view, or hypothesis, which is that some other unrelated person has left the stain. The calculation we do doesn't estimate the probability that the person left the stain, but what we can do is estimate the probability that the crime stain and the reference sample match, if the accused left it, or if someone else left it. You can then weigh up those two competing probabilities in what is called the likelihood ratio. 

12. DNA Fingerprinting Database for Identifying Criminals

DNA fingerprinting can be used for identifying a real culprit as well as to exonerate an innocent. Now the policy of the law enforcement machinery in advanced countries is to catch the real culprits without making much disturbance to others. Formation of a DNA database system was evolved from this forethought. DNA fingerprinting database is simply a database consists of DNA samples collected from the crime scene or from individual donors. It is a biological profile of a person, which will be kept by the legal machinery as a permanent record. The normal procedure of every DNA database system is that if a DNA sample has been obtained in connection with a crime, it is sent to a DNA laboratory for analysis and the results are stored in a central database. Once it is stored in the database system, then it can be used for comparing other.

54 Ibid. at Para 14.
55 The list of donor groups i.e. who is included in the donor group varies substantially from jurisdiction to jurisdiction.
samples in the database as well as against the new samples obtained from the individual donors and the samples collected from the crime scene.\textsuperscript{56}

Initially the database system was formed as an effective tool for crime solving. It was in United States, the world's first DNA information system was introduced. The FBI under the United States Department of Justice (DOJ) formed this in 1989. FBI started this project as a combined DNA index system (CODIS). CODIS is a software-based system, which contains two indexes. One index is the convicted offender index contains DNA profiles of felons convicted of violent crimes and the other index is the forensic index contains DNA profiles from crime scenes.\textsuperscript{57} In 1998, the FBI created another system known as the National DNA index System (NDIS). This is an electronic system in which the Federal, State and local law enforcement agencies can contribute DNA samples and thereby activate the DNA database system in finding culprits.\textsuperscript{58}

In 1995, the world's first DNA database of persons convicted of an offence punishable by imprisonment was established in the UK.\textsuperscript{59} By March 1998, the database contained the DNA profiles of more than 2,55000 suspects and convicted persons and 30,000 profiles developed from materials found at crime scenes.\textsuperscript{60} In 1998, three other European countries, Austria, Germany and Netherlands introduced their DNA databases.\textsuperscript{61}

\textsuperscript{56} See, Paul E. Tracy and Vincent Morgan, "Big Brother and his Science Kit: DNA Databases For 21\textsuperscript{st} Century Crime Control", 90 J. Crim. L. & Criminol. 635 at 642 (2000).
\textsuperscript{57} Ibid.
\textsuperscript{58} Ibid.
\textsuperscript{59} Adrian Linacre, "The UK National DNA Database", 351 The Lancet 1841 (2003).
\textsuperscript{60} Mike Redmayne, "The DNA Database: Civil Liberty and Evidentiary Issues", [1998] Crim. L. R. 437.
\textsuperscript{61} See, P.M. Schncider, DNA Databases for Offender Identification in Europe-the need for Technical, Legal and Political Harmonization, in Proceedings of the 2\textsuperscript{nd} European
In Australia, at federal level, a national DNA database system known as the National Criminal Investigation DNA Database (NCIDD) has been established. It is operated by the Federal government agency known as CrimTrac. The database will contain DNA profiles from each of the participating States and Territories. Profiles will be shared between jurisdictions. All DNA profiles held in New South Wales databases will be available on the national database system. Indexes in New South Wales include a missing persons index; offenders index; suspects index; unknown deceased persons index; volunteers (limited purposes) index; and a volunteers (unlimited purposes) index. Australian database system is under the strict control of, Crimes (Forensic Procedures) Act 2000.62

The unresolved issue in connection with this topic is regarding the determination of the list of persons who may be included in the DNA database system. The type of person to be included in the DNA database differs from jurisdiction to jurisdiction which includes sex offenders, persons convicted of felony offences, all persons convicted, persons arrested for felony offences, all persons arrested, and total population. In UK in March 2003 it was announced that the UK police will have the right to retain indefinitely the DNA profiles from all people from whom DNA is collected, whether the person is charged with a crime or not.63 Robert William and Rony Dunkan in their scientific commentary published in Nature have

opined that it is better to take DNA profile of all individuals at birth.64 According to them this tendency would not only act as a deterrent from crime for all members of the community, but would make the task of catching criminals easier for police.65

The use of DNA database to identify suspects, in effect creates some personal liberty issues in the mind of persons arguing for human rights.66 They argued that DNA profile contains data of person's genetic characteristics, and if it is disclosed to some others, his privacy will be infringed. But it is clear from the observation made by some scientists that there is no room for such argument, because the forensic scientists are using 'non-coded' loci for identification purpose. In a non-coded region of DNA, genes are not available; therefore the storing of DNA database will no: raise any issue against genetic privacy. However, authorities should take extreme care in disposing of forensic samples after conducting DNA analysis.

65 Ibid.