Population Genetic Models

3

JOHN BUCKLETON

Contents

3.1 Introduction

This chapter will discuss those population genetic models used for assigning a profile or preferably a match probability. Three models — the product rule, the subpopulation correction, and an admixture model — will be discussed.

The interpretation of DNA evidence often requires the assignment of a probability to the chance of observing a second copy of a particular genotype in a certain population. Implicit in this apparently simple statement, there are many questions about what this probability should be, how it should be assessed, and upon what other information, if any, should it be conditioned.

In common usage, the word "frequency" is often substituted for "probability." Hence a genotype probability will become a genotype frequency. This is a slight loss in accuracy in the use of nomenclature, but it allows us to slip into common usage. A frequency really should have a numerator and a denominator, e.g. 3 in 25, where we have counted 3 particular outcomes out of the 25 possible. Since most genotype probabilities are very small, they are not estimated by direct counting. Hence, strictly, they are not frequencies.

The frequentist approach to interpreting evidence will report this genotype frequency, f.

Under the logical approach for these hypotheses:

H_p: The DNA came from the suspect, and
H_d: The DNA came from a male not related to the suspect,
the likelihood ratio

$$LR = \frac{1}{\Pr(G_c|G_s,H_d)} = \frac{1}{f} \tag{2.3}$$

The standard response to our inability to directly assess these frequencies has been to attempt to model them using a population genetic model. However, certain cautions should be considered with the concept of a true genotype probability. First among these cautions is to consider what would represent a "fair and reasonable" assignment of probability. It would be tempting to suggest that a fair and reasonable assignment would be one that was near the true value. If we consider the values of probabilities that will be generated by 13-locus CODIS or 10-locus SGM$^+$ multiplexes, we realize that they are very small. It would be very difficult, if not impossible, for us to determine their true values. In fact, this would typically require the genetic typing of the whole population of the world, and the values would change constantly as individuals were born or died.

Is there a requirement for the fair and reasonable probability assignment to equal the true value? Interestingly, the answer is no. If we consider the profile probabilities of 13-locus CODIS profiles (or 10-locus SGM$^+$ profiles), it is certain that most genotypes do not exist. There are more possible genotypes (about 10^{23}) at these 13 loci than there are people. Therefore, only about 1 profile in 10^{14} can exist.[842] True profile frequencies will take the values

$$\frac{n}{6{,}000{,}000{,}000}$$

where $n=0, 1, 2\ldots$ and the population of the world at a given instant is taken for illustration as 6 billion.

Most genotypes will thus have true frequencies of 0. These are of no interest to us because they do not exist and will not occur in casework. It is the remaining ones that are of interest. For those that do exist we know that the suspect has this genotype, but we must remember that we are interested in the probability of obtaining this genotype from someone other than the suspect.

All our probability assignments will differ from the true frequencies. Even if we move to the superior conditional probabilities, these will typically be small numbers whereas the actual frequencies are 0, 1, 2, or more in 6 billion. We distinguish between the actual frequency of a genotype and its probability. The frequency of a genotype will be a probability only if we could conceive of carrying out an experiment of randomly sampling, with replacement, individuals chosen from the population of the world at a given instant.

The assignment of a probability to a multilocus genotype is an unusual activity. Few other fields of science require such a probability assignment. The field of genetics is well established, but largely concerns itself with things such as allele probabilities or genotype probabilities at one or a very few loci. Therefore, the attempt by forensic scientists to assign probabilities to multilocus genotypes is a relatively novel experiment peculiar to forensic science. It may be based on genetics and statistics, but it is a new extension of previous methods, broadly speaking attempting to go where no science has gone before.

These probabilities cannot be directly measured by any mechanism that we can envisage. Ian Evett has discussed his view of whether these probabilities can be considered estimates at all:

> Probability is a personal statement of uncertainty. In the DNA context, I take some numbers (that are estimates of things like allele proportions and F_{ST}'s) and stick them into a formula. Out comes a number and on the basis of that I assign… a probability.

> That is a personal, subjective probability, which incorporates a set of beliefs with regard to the reliability/robustness of the underlying model. So, whenever you talk about estimating a probability, I would talk about assigning a probability.
>
> Thus I would not say, as you do… that the probabilities are "untestable estimates." I would ask — "is it rational for me to assign such a small match probability?"

We cannot directly compare our probability assignments to true values. We may be able to test the process by which these probabilities are assigned, but in casework we will be unable to test the final probability assignment. This makes it most important that these inherently untestable probabilities are assigned by the most robust methods.

In this chapter, the options currently in use to assign these genotype probabilities are discussed. In addition, we consider a third option that has been suggested by Bonnie Law. This model was designed to cope with the phenomenon of admixture.

3.2 Product Rule

This is the simplest of the available population genetic models. It is deterministic as opposed to stochastic.[211] This means that it assumes that the populations are large enough that random effects can be ignored. It was the first model implemented in forensic DNA analysis, having previously been used for a number of years in blood group analysis. It is based on the Hardy–Weinberg law and the concept of linkage equilibrium.[805,806] Both these concepts have been extensively discussed. However, it is worthwhile making a few comments that are specifically relevant to forensic science.

3.2.1 Hardy–Weinberg Law

This concept was first published in 1908,[392,826] although simplified versions had been published previously.[151,611,878] This thinking developed naturally following the rediscovery of Mendel's work.[546] It concerns the relationship between allele probabilities and genotype probabilities at one locus. In essence, the Hardy–Weinberg law is a statement of independence between alleles at one locus.

The Hardy–Weinberg law states that the single-locus genotype frequency may be assigned as the product of, allele probabilities

$$P_i = \begin{cases} p_{i1}^2, & A_{i1} = A_{i2} \\ 2p_{i1}p_{i2}, & A_{i1} \neq A_{i2} \end{cases} \tag{3.1}$$

for alleles A_{i1}, A_{i2} at locus *i*. This will be familiar to most in the form

$$\begin{cases} p^2 & \text{homozygotes} \\ 2pq & \text{heterozygotes} \end{cases}$$

This law will be exactly true in all generations after the first if a number of assumptions are met. It may also be true or approximately true under some circumstances if these assumptions are not met. The fact that the equilibrium genotype frequencies are obtained after one generation of random mating means that we do not need to enquire into the deep history of a population to describe the genotype frequencies at one locus[211] if these requirements are met. It also means that any perturbation from equilibrium is likely to be rectified rapidly. This is not exactly true for populations with overlapping generations, such as humans, where equilibrium is achieved asymptotically as the parental population dies. A few other exceptions to the rule that equilibrium is achieved in one generation are given in standard population genetic texts such as Crow and Kimura.[211]

The assumptions that make the Hardy–Weinberg law true are that the population is infinite, randomly mating, and there are no disturbing forces. Inherent in this law is the assumption of independence between genotypes: specifically, that the knowledge of the genotype of one member of a mating pair gives no information about the genotype of the other. Consider what would happen if the population was finite, as indeed all populations must be. The knowledge of the genotype of one member of a mating pair slightly reduces the probabilities for these alleles in the other member, since one or two copies of these alleles have been "used up." This effect is very minor indeed unless the population is quite small or the locus extremely polymorphic. Most human populations may be numbered in tens of thousands or more individuals.

The assumption of random mating assumes that the method of selection of mates does not induce dependence between genotypes. This is often translated comically and falsely along the lines "I did not ask my spouse his/her genotype before I proposed." When the assumption of random mating is questioned, no one is suggesting that people who are genotype *ab* deliberately go and seek partners who are type *cd*. What is suggested is that geography, religion, or some other socioeconomic factors induce dependence. This will be discussed later, but the most obvious potential factor is that the population is, or more importantly has been in the past, divided into groups that breed more within themselves than with other groups.

A consequence of the assumption of an infinite population and random mating is that the allele proportions are expected to remain constant from one generation to the next. If the population is infinite, randomly mating, and the allele proportions do not change, then the Hardy–Weinberg law will

hold in all generations after the first. This is true whether or not the Hardy–Weinberg law holds in the first generation, the parental one. It therefore describes an equilibrium situation that is maintained indefinitely after the first generation. Note that it does take one generation of random mating to achieve this state. Such a stable state would describe an equilibrium situation and hence this state is often called Hardy–Weinberg equilibrium (HWE).

There are, however, a number of factors that can change allele proportions. These are referred to as disturbing forces. The term is derived from the fact that they change genotype proportions from those postulated by HWE. These factors include selection, migration, and mutation. There are comprehensive texts available describing the effect of these forces on both allele proportions and on HWE, and they will not be discussed at length here. In this chapter we will simply consider how close the Hardy–Weinberg assumptions are to being fulfilled, and what the probable consequences of any failure of these assumptions may be. Remember a model may be useful even though it is not an exact description of the real world.

3.2.2 Linkage and Linkage Equilibrium

HWE describes a state of independence between alleles at one locus. Linkage equilibrium describes a state of independence between alleles at different loci.

The same set of assumptions that gives rise to HWE plus an additional requirement that an infinite number of generations has elapsed also lead to linkage equilibrium. This result was generalized to three loci by Geiringer,[331] and more generally to any number of loci by Bennett.[54]

However, recall that HWE is achieved in one generation of random mating. Linkage equilibrium is not achieved as quickly. Strictly the state of equilibrium is approached asymptotically, but is not achieved until an infinite number of generations have elapsed. However, the distance from equilibrium is halved with every generation of random mating for unlinked loci or by a factor of $1-r$, where r is the recombination fraction, for linked loci. Population subdivision slows this process.[421]

It is worthwhile discussing the difference between linkage equilibrium and linkage, as there is an element of confusion about this subject among forensic scientists. Linkage is a genetic phenomenon and describes the situation where one of Mendel's laws breaks down. It was discovered in 1911 by Morgan[555,556] working on *Drosophila*. The discovery was a by-product of his team's studies of inheritance that had largely led to the confirmation of the chromosomal theory of inheritance. The first paper on gene mapping appeared in 1913.[740]

Specifically, the phenomenon of linkage describes when alleles are not passed independently to the next generation. The physical reason for this phenomenon had been identified by 1911 and related to the nonindependent segregation of alleles that are sufficiently close on the same chromosome.[597]

The state of linkage can be described by the recombination fraction or by the distance between two loci. Typical data for distance may be expressed in centiMorgans (cM) or in physical distance in bases. In humans, 1 cM is assumed to equal approximately 1000 kb.

The physical distance may be converted to a recombination fraction by standard formulae.[a] Recombination fractions tend to be different for each sex. Distances may be given separately or sex-averaged.

Linkage disequilibrium is a state describing the relationship between alleles at different loci. It is worthwhile pointing out that linkage disequilibrium can be caused by linkage or by other population genetic effects such as population subdivision. This will be demonstrated later.

Therefore, it is incorrect to advance the following line of logic.

A: The loci are on different chromosomes or well separated on the same chromosome.

Which implies that

B: There is no linkage.

Which implies that

C: There is no linkage disequilibrium.

Modern genetic understanding would state that the progression from statement A to statement B is logical and grounded on experimental observation. However, the progression from statement B to statement C is not supportable without additional data.

Linkage disequilibrium has been noted for very closely linked loci. For example, Gordon et al.[366] investigated 91 unrelated Afrikaners and observed linkage disequilibrium between pairs of loci separated by 0.00, 0.00, 0.54, 2.16, 2.71, 3.68, 5.28, and 5.51 cM on chromosomes 1, 2, 5, 11, 20, and 21. Such linkage disequilibria have been used to estimate the time since the last bottleneck for various populations[522] and may give interesting anthropological information. Deka et al.[227] investigated linkage disequilibrium and identified Samoans as an interesting study group plausibly because of a recent bottleneck. Szibor et al.[750] investigated linkage disequilibrium between alleles at loci on the X chromosome for a sample of 210 males. The loci investigated contained three linkage groups from a total of 16 loci. They observed disequilibrium only for alleles at the loci DXS101 and DXS7424. This is an example of the well-known phenomenon that linkage does not necessarily imply linkage disequilibrium.

[a] See Chapter 1, footnote c.

The CODIS loci HUMCSF1PO and HUMD5S818 are both located on chromosome 5 and are reported to be separated by 25 cM.[30] This translates to a recombination fraction (Haldane) of 0.39. This would be expected to have no effect at the population level, but in restricted circumstances may have a moderate effect in paternity testing or disaster victim identification.

The most likely causes of linkage disequilibrium for unlinked or loosely linked loci are population genetic effects such as population subdivision or admixture.[154,421] These will be discussed in some detail later.

If the population is in linkage equilibrium, then a multilocus genotype probability (P) may be assigned by the product of single-locus genotype probabilities (P_i):

$$P = \prod_i P_i \tag{3.2}$$

3.2.3 Consideration of the Hardy–Weinberg and Linkage Equilibrium Assumptions

There are five assumptions for the Hardy–Weinberg law to hold and one additional assumption for linkage equilibrium to hold. In this section each of these assumptions will be considered with regard to whether or not they are true, and in particular to how far from true they may be.

3.2.3.1 Infinite Population

This assumption is clearly violated to greater or lesser extents, depending on the size of the population. In addition, there is ample evidence for the existence of population bottlenecks in the past. The effect on disturbing the equilibrium in the present is likely to be very limited for most realistic populations unless a relatively recent bottleneck is suspected. Recall that one generation of random mating is sufficient to restore HWE. Any effect is most likely to occur for rare alleles.

Crow and Kimura[211] give

$$\Pr(A_i A_i) = p_i^2 - p_i(1 - p_i) f$$
$$\Pr(A_i A_j) = 2p_i p_j(1 + f)$$

where N is the number of individuals and $f = 1/(2N - 1)$ We see that any departure from equilibrium is expected to be very small for most realistic values of N.

3.2.3.2 No Mutation

One of the assumptions for Hardy–Weinberg and linkage equilibrium is that there is no mutation at the loci in question. With regard to the commonly used STR loci, this assumption is clearly violated. In fact, we believe

that the STR loci are mutational "hot spots," with mutation rates above much of the coding DNA but probably less than the VNTR loci or mitochondrial DNA.

Various treatments have been offered that deal with change in allele frequencies due to mutation or to the effects of mutation and selection.[267] If, however, we accept that these loci are selectively neutral, then the most realistic situation that we need to consider is the situation of mutation and genetic drift. The effect of mutation, of the type observed at STR loci, on a divided population is that it tends to oppose the effect of drift. If drift is tending to remove genetic variation from separated subpopulations, mutation tends to reintroduce it. When a mutation occurs at an STR locus, it tends to add or subtract a single repeat, with mutational losses or gains of multiple repeats being much more rare (see Chapter 10 for a summary of mutation references). This mode of mutation fits well with a theoretical model, the stepwise mutation model, that was first proposed by Kimura and Ohta.[462]

If we consider two populations that have become separated or isolated, then they begin to evolve separately and their respective allelic frequencies tend to drift apart. This process will be associated with an increase in relatedness within the separated subpopulations and can be quantified by an increase in the inbreeding coefficient θ. The effect of stepwise mutation to alleles already present is to lower relatedness and hence θ.[285,671,672] This may seem odd. The people are still related, but their alleles can no longer be identical by descent as they are no longer identical. The equilibrium situation that may result is given by Evett and Weir.[267] Whether drift or mutation is the dominant factor depends on the product $N\mu$, where N is the population size and μ the mutation rate. If $N\mu \ll 1$, the population will typically be moving toward fixation for one allele, which means that genetic drift forces are dominant. If $N\mu \gg 1$, then mutation is the dominant force and multiple alleles will be present.[577]

This effect can be elegantly demonstrated using simulation software. Two programs have been offered by forensic programmers — Gendrift (Steve Knight and Richard Pinchin, FSS) or Popgen (James Curran, University of Waikato[b]) — and there are others in the population genetics community.

It would be unwise, however, to assume that mutation is a completely benign phenomenon from the perspective of decreasing associations between individuals. The exact nature of the mutational process does have a serious effect on the departures that may be observed and the validity of models to correct for them. This is discussed briefly later.

[b] The latter program is available free from James Curran's website: http://www.stats.waikato.ac.nz/Staff/curran/Popgen95.zip.

3.2.3.3 No Migration Into or Away from the Population

Allele probabilities will change if migration occurs into or away from the population. Emigration from a moderately sized population has very little effect since the subtraction of a few alleles from the gene pool alters the allele probabilities very little. Immigration of alleles into the population from a different population can have a much more marked effect. Such gene migration is often accompanied by physical migration of people, but this is not necessarily a requirement.

To consider this issue, it is critical to consider the interaction of migration and our definition of population. Most of our current definitions of population have both an ethnic and a geographical basis. Consider the New Zealand population. We currently subdivide this arbitrarily into Caucasian, Eastern Polynesian (Maori and Cook Island Maori), Western Polynesians (Samoans and Tongans), and Asians. The physical migration of a British person to New Zealand would represent migration of alleles into the New Zealand Caucasian gene pool. The intermarriage of Caucasians and Maori would represent migration of Caucasian genes into the Eastern Polynesian gene pool without necessarily involving any physical migration of people. The fact that this is treated as a migration of genes INTO the Eastern Polynesian gene pool is dependent on how we intend to (arbitrarily) define the ethnicity of the resulting progeny.

The effect of migration on equilibrium is dependent on the difference in allele frequencies between the donor and recipient populations.[267] Hence the physical migration of British people to New Zealand is likely to have a very small effect on the equilibrium situation of New Zealand Caucasians since the allele frequencies in the two populations are similar. However, the migration of Caucasian genes into the Eastern Polynesian gene pool is much more likely to disturb the equilibrium since the populations have more differing allele probabilities.

3.2.3.4 No Selection

It is difficult to find experimental data that bear directly on the issue of whether or not there is selection at the STR loci used for forensic work. This is clearly an area that warrants further scrutiny. The general field is very active in human population genetics. At this stage, most of the argument in favor of there being little or no selection at STR loci relates to the fact that these loci are noncoding and hence do not produce any gene products. Theoretically then, any mechanism for selection would have to operate by an indirect route, say by hitchhiking on other advantageous or disadvantageous genes, or by affecting DNA packing, replication, or repair.

The STR loci are intronic. Introns are thought to have "invaded eukaryotic genes late in evolution, after the separation of transcription and translation."[538,539] When first studied, these DNA sections were thought to be

nonfunctional and were termed "junk DNA." Mattick[538,539] argues convincingly for a role for at least some intronic products in gene expression and postulates that they were a critical step in the development of multicellular organisms.

Makalowski[532] discusses the origin of the phrase "junk DNA" and reinforces the modern conception that this DNA may have important functions. If this theory is eventually accepted, as would seem inevitable, then the question would arise as to whether there is a function for the specific intronic segments used in forensic work.[149]

The observation of greater microsatellite diversity among Africans[448] is consistent with the out of Africa event and a selectively neutral model. However, greater diversity among Africans is certainly not proof of selective neutrality.

Mitochondrial DNA shows a deviation from selective neutrality; however, this is postulated to be the result of a selective sweep in modern humans outside Africa.

Selection is a proven phenomenon in some blood group systems such as ABO and Rhesus.[152] A mechanism has been proposed for the selective interaction between ABO and Haptoglobin.[567] However, these genes are clearly coding and produce important gene products. Hence direct selective mechanisms are expected.

Selection by association with disease loci is a mechanism that may possibly affect STR loci. Such associations at other loci are known.[596] The effect of a selective sweep caused by the appearance of an allele favored by selection at a nonforensic locus has not been considered in detail in the forensic literature. However, unless such a sweep is recent, this is unlikely to have much effect on the modern state of equilibrium (although it may have had an effect on modern allele probabilities).

Neuhauser[577] compares random drift and selection and notes that if $Ns \ll 1$, where N is the population size and s is the selective advantage of one allele over another, for a two-allele locus, then selection does not have much effect, and the locus acts almost as if it were neutral.

A theoretical model for estimating mutation rates at di-, tri-, and tetranucleotides from the distributions of their allele sizes was given by Chakraborty et al.,[166] who note the departure of the predictions of the model from directly observed values. This led Chakraborty et al. to an interesting discussion of whether there is any evidence of constraints in the number of DNA repeats at a locus, which may be evidence for the existence of selection. They conclude that the shape of modern allele distributions is inconsistent with the existence of constraints.

In summary, there are reasonable theoretical reasons to believe that these loci are selectively neutral or nearly so. No direct evidence for strong selection at forensic loci has been reported, but how hard have we looked for it? Equally, there is little direct experimental evidence for selective neutrality.

3.2.3.5 *Random Mating*

Of the various assumptions given, this is the one that has deservedly attracted the most attention. It is clear that we do not select our mates on the basis of their DNA genotypes at the STR loci. Most of us do not even know our own genotype at these loci. We also believe that these genotypes have no physical manifestation, which is to say that they do not affect the phenotype of an individual. Hence we should be unable to detect these genotypes by looking at a person. This should preclude some inadvertent selection of genotypes. However, it would be wrong to assume from this that random mating is a fair assumption.

Crow and Kimura[211] discuss the two main types of nonrandom mating: inbreeding and assortative mating. Assortative mating is not discussed here. There is considerable evidence that it does occur in humans. For instance, an intelligent person is more likely to marry another intelligent person. Jared Diamond[231] discusses this in some detail in his popular science book *The Rise and Fall of the Third Chimpanzee*. In the STR context, I believe that the issue of importance is inbreeding.

What is alleged is that the population is made up of subpopulations[510,511] whose members preferentially mate within their subpopulation, possibly for religious, language, or other reasons, but more probably just because of geographical proximity (for an excellent review, see Excoffier[283]). This is termed inbreeding. In the past, people traveled a lot less than they do now. The notion of marrying the "girl or boy next door" is not universal nor is it totally unknown. It is important to note that there is no suggestion that subpopulations are completely isolated from each other. All that is required is any departure from a completely random choice of mates. The more isolated the subpopulations, the larger the effect, but partial isolation will also lead to some subpopulation effects.

In lectures on DNA around the world, I have performed a trial with the various classes. Unfortunately I have not kept the results, which would make an interesting section. However, the general flavor of them can be reported. What was asked was for people to give the "ethnicity" of their four grandparents. Table 3.1 gives the results for the area around my desk at the laboratory at the FSS at Trident Court in Birmingham, U.K. Each cell represents one individual's self-declared ethnicity for their four grandparents.

This experiment would not meet minimum survey standards; however, let us treat them as a demonstration rather than as evidence. First let us note that this arrangement does not look random. Too many ethnicities occur together. For instance, there are four Chinese entries and four Indian entries together. Let us assume that we separated these two individuals out as being of a different "race." What we are left with still does not look like a random arrangement. For instance, there are four Greek Cypriots and two Iraqis together. Let us assume further that we take these out and put them into different categories.

Table 3.1 Self-Declared Ethnicity of Some Staff at the FSS Laboratory, Trident Court in 2002

Irish, Irish, Irish, Irish	Swiss, Swiss, Swiss, Swiss
English, English, English Irish	English, English, English English
English, English, English, English	Chinese, Chinese, Chinese, Chinese
Welsh, English, English, Scottish	English, English, English, English
Scottish, Scottish, English, English	English, English, Irish, Scottish
English, English, English, English	English, English, English, Scottish
Hungarian, Scottish, Scottish, English	English, English, English, Scottish
English, English, English, English	Greek Cypriot, Greek Cypriot, Greek Cypriot, Greek Cypriot
English, English, English, English	Irish, Irish, Iraqi, Iraqi
English, English, English, Scottish	Indian, Indian, Indian, Indian

Still, what we are left with does not look random. There are too many Irish and Swiss together. If we could peer deeper into the past, we might find that the people reporting "English" have differing amounts of Celtic, Scandinavian, or Saxon heritage.

This experiment has worked wherever I have tried it: in New Zealand, Australia, the United States of America, and the United Kingdoms of Great Britain and Northern Ireland. I, personally, do not believe that the modern human population is the result of random mating. I do believe that we are the result of an evolutionary process whereby our ancestors mated in groups to a greater or lesser extent. This is breaking down in modern times, but the process is far from complete.

This leads us to the classical consideration of the Wahlund principle.[801] Assume that a certain area is made up of two or more subgroups that breed within each group but not to any large extent between the two groups. Further assume that there are some allele probability differences between these groups. Then even if the subpopulations themselves are in HWE, the full population will not be. An example is given in Table 3.2.

First we note that the mixed population is not in HWE even though each subpopulation is. Next we note the classical Wahlund effect in which all the probabilities for homozygotes are increased above Hardy–Weinberg expectation. The total heterozygote probabilities are generally decreased, although individual heterozygotes may be above or below expectation. Note that in this example two of the heterozygotes are below expectation, whereas one is above. The total for all the heterozygotes will always be down (which is really the same as saying the total of the homozygotes is always up).[267,836]

The same subpopulation phenomenon will induce between locus dependence, that is, it will induce linkage disequilibrium. This is more complex but not harder to demonstrate. In Table 3.3 we give a numerical demonstration.

Table 3.2 An Example of the Wahlund Effect

Allele	*a*	*b*	*c*
Subpopulation 1	0.7	0.2	0.1
Subpopulation 2	0.2	0.1	0.7

Genotype	Subpopulation 1	Subpopulation 2	1:1 Mix	Hardy–Weinberg expectation
aa	0.49	0.04	0.2650	0.2025
bb	0.04	0.01	0.0250	0.0225
cc	0.01	0.49	0.2500	0.1600
ab	0.28	0.04	0.1600	0.1350
ac	0.14	0.28	0.2100	0.3600
bc	0.04	0.14	0.0900	0.1200

This table shows the "correct" genotype proportions and two incorrect calculations. The first incorrect calculation proceeds by combining the two subpopulations and then using the population allele probabilities — this incorrectly assumes Hardy–Weinberg and linkage equilibrium in the population. This is the type of error (although greatly exaggerated) that would occur if we assumed that a structured population was homogeneous. The second incorrect calculation (again carried out on the combined population) proceeds as if we had performed some sort of testing and had abandoned the assumption of HWE, but instead had used observed genotype proportions and then multiplied across loci. This approach is a better method to assign probabilities as it corrects for Hardy–Weinberg disequilibrium; however, it fails to account for linkage disequilibrium.

The third approach was adopted, incorrectly, by Buckleton and Weir in some of their early recommendations, but is now abandoned. It appears later in this chapter as the "Cellmark wrinkle" in the descriptions of the O.J. Simpson case. It persists in recommendations by other authors but should be superseded.

Inspection of these numbers shows that the "correct" probabilities for two loci cannot be determined if the population structure is ignored. Proceeding from either the population allele probabilities or the population genotype probabilities will give incorrect answers.

The demonstration that the multiplication of population genotype probabilities gives an incorrect answer shows that linkage disequilibrium can be induced by population substructure whether or not the loci are physically linked. Loci that are on different chromosomes may, therefore, be in disequilibrium[514,576,590,591] and expressions have been derived to estimate the magnitude of the disequilibrium.[267,836] In fact, almost any instance of disequilibrium in the forensic literature involves loci that are on different chromosomes.

Table 3.3 Two-Locus Genotype Probabilities for a Population Consisting of Two Subpopulations in Equal Proportions

Allele	Subpopulation 1	Subpopulation 2
Locus 1		
a	0.7	0.2
b	0.2	0.1
c	0.1	0.7
Locus 2		
d	0.5	0.2
e	0.2	0.4
f	0.3	0.4

	dd	*ee*	*ff*	*de*	*df*	*ef*
	1:1 Mix Correct					
aa	0.062	0.013	0.025	0.052	0.077	0.036
bb	0.005	0.002	0.003	0.005	0.007	0.004
cc	0.011	0.039	0.040	0.040	0.041	0.079
ab	0.036	0.009	0.016	0.031	0.045	0.023
ac	0.023	0.025	0.029	0.036	0.043	0.053
bc	0.008	0.012	0.013	0.015	0.017	0.025
	1:1 Mix from Alleles					
aa	0.025	0.018	0.025	0.043	0.050	0.043
bb	0.003	0.002	0.003	0.005	0.006	0.005
cc	0.020	0.014	0.020	0.034	0.039	0.034
ab	0.017	0.012	0.017	0.028	0.033	0.028
ac	0.044	0.032	0.044	0.076	0.088	0.076
bc	0.015	0.011	0.015	0.025	0.029	0.025
	1:1 Mix from Genotypes					
aa	0.038	0.027	0.033	0.048	0.061	0.058
bb	0.004	0.003	0.003	0.005	0.006	0.006
cc	0.036	0.025	0.031	0.045	0.058	0.055
ab	0.023	0.016	0.020	0.029	0.037	0.035
ac	0.030	0.021	0.026	0.038	0.048	0.046
bc	0.013	0.009	0.011	0.016	0.021	0.020

Some of the most common causes of disequilibrium are population genetic effects, such as the existence of subpopulations, and such disequilibria occur for the same reasons as the Wahlund effect.[484,485]

This disequilibrium phenomenon is sufficiently understood that decay rates for linkage disequilibrium for nonlinked loci have been calculated and appear in standard texts.[267 (pp. 127–129),421,836] The dependency effects are not expected to be large for loci with low mutation rates. There is a slight tendency for the dependencies to rise with the number of loci.[488,843]

We give examples later using the ESR data for Eastern Polynesians. Analysis of these data suggests disequilibrium regardless of the chromosomal position of the loci. In this particular case, the most likely explanation is not population subdivision but the effects of admixture with Caucasians. The population in the U.S. described as Hispanics may be showing the same admixture effects or this may be the result of subpopulations, or both.[c] The Hispanic population is often subdivided into South-Eastern and South-Western Hispanic.

Conversely, loci that are closely linked on the same chromosome may be in equilibrium (or near it). In fact, there is no absolute relationship between the position on a chromosome and the state of independence between loci. However, as a generalization, Hudson[421] notes "loosely linked loci are typically observed to be near linkage equilibrium in natural populations…. In contrast…very tightly linked loci often show some signs of linkage disequilibrium."

There is growing evidence of a block-like structure to linkage disequilibrium. This implies that some regions of the genome are closely linked and others are unlinked. This structure can, obviously, be produced by recombination hot spots, but interestingly can also be produced without such hot spots.[882]

In summary, a lack of random mating, in particular the existence of subpopulations with different allele probabilities, will cause Hardy–Weinberg and linkage disequilibrium. The proportions of the different subpopulations and the differences in their allele probabilities will affect the magnitude of this disequilibrium. The larger the differences in the allele probabilities between the differing subpopulations, the larger the resulting disequilibria. Excoffier[283] notes that population subdivision will also produce a larger number of observed alleles, with an excess of rare alleles.

The first human populations that came under intense scrutiny by the forensic community were the Caucasian populations of the U.K. and the U.S. These populations comprise subpopulations arising from different areas of the U.K. and Europe. Studies have suggested that there are only minor differences between these Caucasian subpopulations in Europe or the U.K. *per se*. Although these differences are real,[79,152,566] they are small and hence they give rise to very small disequilibrium effects. The effect of these disequilibria is a very mild bias in the product rule toward the assignment of a genotype probability that is too low.

3.2.3.6 An Infinite Number of Generations

Loci that are on different chromosomes or well separated on the same chromosome will assort in a Mendelian manner. The linkage disequilibrium associated with such loci is expected to halve with every generation,[267] and hence

[c] This possibility appears to have received recent acceptance from Budowle and Chakraborty at least in the published literature, both previously strong supporters of the use of the product rule.[60]

will approach equilibrium asymptotically, but never quite get there, if the disturbing force is removed. Linked loci will also approach equilibrium but more slowly, depending on the rate of recombination between the loci. An example of very tightly linked loci that are near equilibrium is given by Mourant, when he discusses the Rhesus blood group (a set of three linked loci) in Australian Aborigines.[566]

3.2.3.7 Summary

It was a pity that the first population extensively studied by the forensic community was the Caucasian population. This is because this population is probably one of those nearest to Hardy–Weinberg and linkage equilibrium of the large modern human populations. Hence it was the least likely to educate us on departures from equilibrium and how to manage these. At that time we did not understand the weakness of our independence tests, and this contributed to our misunderstandings. We return to this subject in Chapter 5.

This section is closed with a quote from Wild and Seber: "What often happens is that, in the absence of knowledge of the appropriate conditional probabilities, people assume independence. … this can lead to answers that are grossly too small or grossly too large — and we won't know!" [865] The situation in DNA is probably not this bad, but the warning is real nonetheless.

3.2.4 How Big Is the Potential Departure If We Use the Product Rule?

It has become accepted wisdom that the error induced by ignoring subpopulation effects may be of the order of a factor of 10. This was based on the comparison of the product rule estimator using various databases as the source of the allele probability estimates. Budowle et al.[127,128] and Hartmann et al.[396] compared the product rule assessment calculated from different subpopulation databases and demonstrated that over 80% of assignments were within a factor of 10 of each other. This approach compares an estimate with an estimate. There has been considerable discussion about the bias inherent in this analysis due to sampling effects,[691] but we have difficulty deciding how much can be read into the results of these discussions.

The conclusions arising from these studies require further validation. It is not totally different to the situation where two students give the same answer in a test. It would be unwise to assume that because they gave the same answer they are both correct.

In addition, we must expect an effect from the number of loci and the populations under consideration. The more the loci, the larger the potential effect of population subdivision. Certain populations are expected to show larger departures than others.

Much later, Gill et al.[355] investigated the magnitude of this bias and refined Budowle's method. Using this modified approach, Gill et al. calculated the product rule assignment for a ten-locus genotype using allele probabilities from the relevant subpopulation and this probability when estimated from an averaged European database (see Table 3 of Gill et al.). They found that the difference between these two estimates may be of the order of two, three, or even four orders of magnitude. Further, they show that almost any of the available adjustment methods, such as a subpopulation correction or even the use of minimum allele probabilities, if applied sensibly, will compensate in part or in full for this effect.

The comparison of an estimate with an estimate is interesting, and would give us some confidence that the effect of changing the database is minor. However, it does not show that either estimate is within a factor of 10 of the true value. It is the latter question that is of forensic interest: How far is our estimate from the true value? The suggestion that the difference between the product rule estimate and a hypothetical true value is a factor of 10 must be taken as a hypothesis with some empirical support. It cannot be taken as proved as we cannot know the true value. Even the simulations by Curran et al.[d] described later in this chapter do not truly compare this estimate to a true value. They simply compare the difference between the product rule assignment and that which would occur under certain population genetic events. It is a simple fact that we cannot measure the difference between the product rule estimate and a true value. Nor can we measure this difference for any other population genetic model. The simulations seek to bring evidence to bear on this matter, but they are, in my opinion, a long way short of scientific proof.

It is often assumed that cosmopolitan populations do not exhibit subdivision. While this may be true, there are also instances where it may not. If the population is old and well mixed, there should be very little, if any, population subdivision. However, a cosmopolitan population may be something like that of London or New York, which consist of people with very different genetic backgrounds who live in the same area. This is exactly the situation where we expect subpopulation effects.

3.2.5 Populations Separating By Genetic Drift

If we accept that the loci that we consider in forensic applications are selectively neutral, then we expect the main evolutionary force producing differences between separated populations to be the random drift of allele

[d] This follows a set of concepts discussed between Mulligan J. and myself during *R v Karger*.[639] I am indebted to His Honour for sharing his insight in this matter, which is often hard to convey in a court situation.

probabilities. This is an extensively researched subject and is only covered very superficially here.

Even if all other evolutionary forces were absent, the allele probabilities in one generation would still differ slightly from the previous one. This difference is caused by the random transmission of alleles to the new generation. For large populations, this effect is very small and takes a long time to be observable. However, for smaller populations the effect may be quite rapid.

The difference between populations that are diverging by drift is often characterized by a parameter θ or F_{ST}, which may be treated as synonyms for the purposes of this text. This parameter is often termed the between-person coancestry coefficient. It is a very useful parameter for characterizing the subpopulation effect; however, it is both difficult to visualize and to measure. For the purposes of this section, it will be adequate to consider it as a measure of the genetic distance between subpopulations. The larger the distance between subpopulations, the longer we assume that they have been separated and the higher θ will be.

It turns out that θ may also be considered as a measure of the relatedness between people in the subpopulation. If this subpopulation has been separate from others for some time, then people in this subpopulation will be more related to each other than they would be to a person taken from a different subpopulation. To help give a feel for the size of θ values, consider that first cousins would have $\theta = 0.0625$.

A formula relating θ to the time since separation is given in many standard texts:[836]

$$\theta_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

where t is the time since separation in generations and N is the effective size of the population (strictly a monoecious population in which selfing is allowed). Evett and Weir[267] discuss the avoidance of selfing and show that the above model is a close approximation. Crow and Kimura[211] give

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f}$$

for the effective size of the population (N_e) when separate sexes of number N_m and N_f are present. When the sexes are present in equal numbers, $N_m = N_f = N/2$ and hence $N_e = N$. Crow and Kimura discuss the effect of differing numbers of progeny on N_e.

If mutation of the infinite alleles type is added to the model, then the opposing forces of drift and mutation may form an equilibrium state, given in several texts:[267,836]

$$\hat{F} \approx \frac{1}{1 + 4N\mu}$$

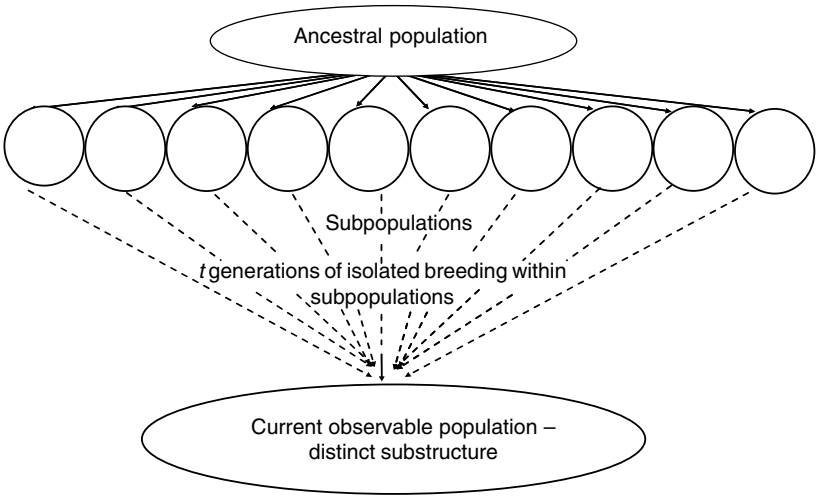

Figure 3.1 Simplified population model. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

where $\hat{F}$ is the equilibrium value of the between-person inbreeding coefficient and μ is the mutation rate.

3.3 Simulation Testing

3.3.1 Product Rule

Curran et al.[215] consider the question: How wrong could the product rule estimator be if the population was subdivided into ten subpopulations and the θ value was approximately 0.03? A computer simulation that allowed the liberty of using the true match probability referred to as the "Gold Standard" examined this question. Populations with known amounts of substructure were produced by dividing a population and allowing it to breed by random mating only within the subpopulations for a suitable number of generations to create the required amount of structure (see Figure 3.1). The ratio of the product rule estimator to the true match probability was then compared. This simulation demonstrated the subpopulation effect but it does not include the effect of mutation. Nor can we truly claim that this is the true match probability. It is certainly the probability if the populations satisfy certain genetic assumptions, but how accurately these assumptions apply to the human condition is the real question.

The Curran et al. results are reproduced in Figure 3.2. Data points above the line given by ratio = 1 indicate that the assignment is conservative with

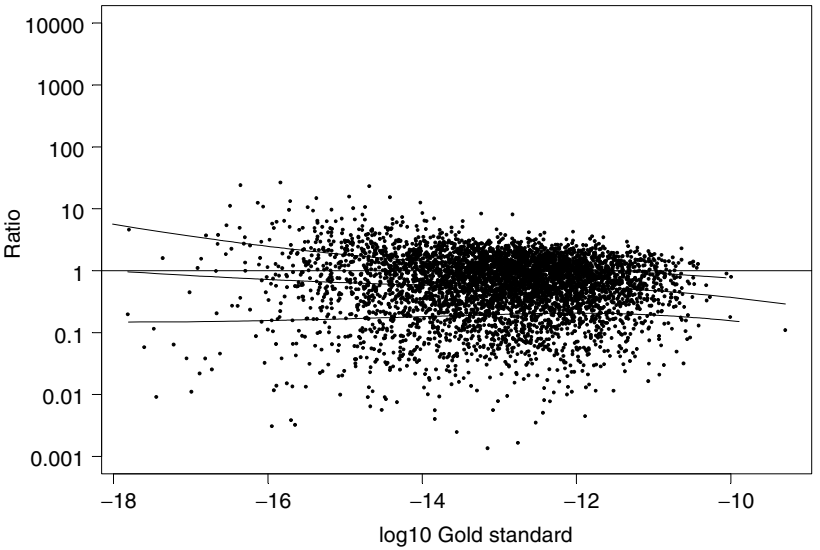

Figure 3.2 Ratio of the naïve product rule profile frequency to the true profile frequency for a population with true inbreeding coefficient $\theta = 0.03$. The median and quartile trend lines are fitted. 64.7% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

respect to the true value. Data points below this line indicate that the estimate is nonconservative. The product rule assignment is seen to be nonconservative for 64.7% of the 50,000 simulated profiles (given the above conditions). The first thing that we note is that this number is greater than 50%. In other words, the product rule estimator has a mild bias in favor of the prosecution if the population is subdivided. This effect is most pronounced when the profile is common. The simulation is for ten loci. The effect would be greater for more loci and less for fewer loci.

In 14.7% of simulated profiles, the estimate was less than one tenth of the true value. By this we are saying that in 14.7% of cases the product rule estimator is incorrect and favors the prosecution by more than a factor of 10. Indeed, a number of estimates differ by more than a factor of 100. This effect is not a result of sampling error because the simulation has been set up to remove all effects of sampling error. Sampling error would add additional uncertainty to these estimates and would spread the results up and down on the graph. We emphasize that usually the subpopulation effect is mild and we do not wish to overemphasize it. The result could be viewed as not substantially different from the conclusion of Budowle et al.: that 80% of estimates were within a factor of 10 of each other.[122,127,128]

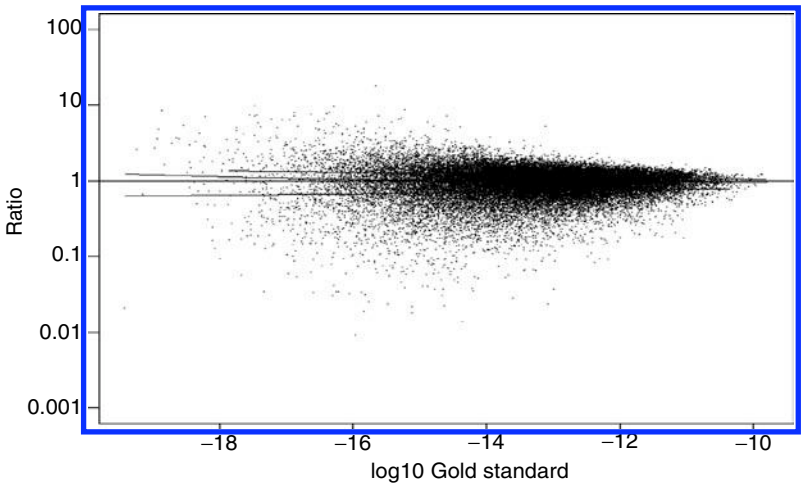

Figure 3.3 Ratio of the naïve product rule profile frequency to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.01$. The median and quartile trend lines are fitted. 51.8% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

The choice of 3% as a value for θ is somewhat arbitrary and would be excessive for Caucasian populations in the U.S. However, it may be more appropriate for Hispanic populations and may, indeed, be an underestimate for Amerinds. The subpopulation effect would be smaller for smaller θ.

In Figure 3.3, we reproduce the equivalent graph with the subpopulations bred to $\theta = 0.01$. In this case, 51.8% of samples returned values less than 1, compared with 64.7% for $\theta = 0.03$. The bias is seen to be very small in this instance. (Do not be deceived by the mean trendline being above 1 at the left. This is expected and is more than compensated for by it being slightly below 1 at the right hand end.) Only a few values lie outside a factor of 10 of the true answer.

It can be seen from these experiments that the product rule estimator has a very small bias in favor of the prosecution in most cases where the population is subdivided. The magnitude of this bias is not large, and it is important not to overemphasize it. However, it is real and is not the result of sampling uncertainty. It will be larger for strongly subdivided populations and smaller for less subdivided populations. The effect may be more than a factor of 10. This finding adds an important verification relative to a true match probability.[e] It does put into perspective comments such as "implementation of the product rule is a reasonable best estimate,"[395,486,509] which must be qualified

[e] Of course this is not a "true match probability" either, but it is the true match probability under THIS model.

with our current understanding that the product rule is unlikely to be an unbiased estimator.

The Curran et al. simulations do not include a specific consideration of mutation. Consideration of an infinite allele mutational process has suggested that this may have a significant effect on the estimation process:

> The product rule probability always underestimates the two-locus match probability. For highly mutable minisatellite loci, these probabilities can differ by an order of magnitude or more… the degree of underestimation worsens for more loci.[488]

This statement is for an infinite allele mutation model and may not be appropriate for a stepwise mutation model. However, it does suggest that further research is warranted if the product rule is to be used.

3.3.2 NRC II Recommendation 4.1

NRC II recommendation 4.1 offered a correction for Hardy–Weinberg disequilibrium caused by the Wahlund effect. It was suggested that a correction upward in frequency be applied to correct for the expected upward bias produced by population subdivision, and further that this correction should be applied only to homozygotes. No correction was recommended for heterozygotes since, on average, these should have a downward bias (recall that individual heterozygotes may be displaced from expectation in either direction). This comment is generally true for the event of population subdivision, but would be untrue for populations undergoing admixture. In admixing populations, the number of heterozygotes is likely to be elevated.

The recommendation suggests that

$$P_i = \begin{cases} p_{i1}^2 + p_{i1}(1 - p_{i1})F, & A_{i1} = A_{i2} \\ 2p_{i1}p_{i2}, & A_{i1} \neq A_{i2} \end{cases} \tag{3.3}$$

where F is the within-person inbreeding coefficient and not the between-person inbreeding coefficient, θ, as written in NRC II.

This recommendation is a logical way of correcting for Hardy–Weinberg disequilibrium, but makes no attempt to correct for linkage disequilibrium. It will suffer from the same approximations that are revealed in Table 3.2 for the 1:1 mix from genotypes. Hence it will still have a very mild tendency to underestimate multilocus genotype probabilities.

Curran et al. tested recommendation 4.1 by comparing this assignment with the "Gold Standard Profile Frequency" for a population with a true inbreeding coefficient $\theta = 0.03$ created by simulation. This is reproduced in Figure 3.4. In this simulation, 54.4% of values are less than 1 (reduced from

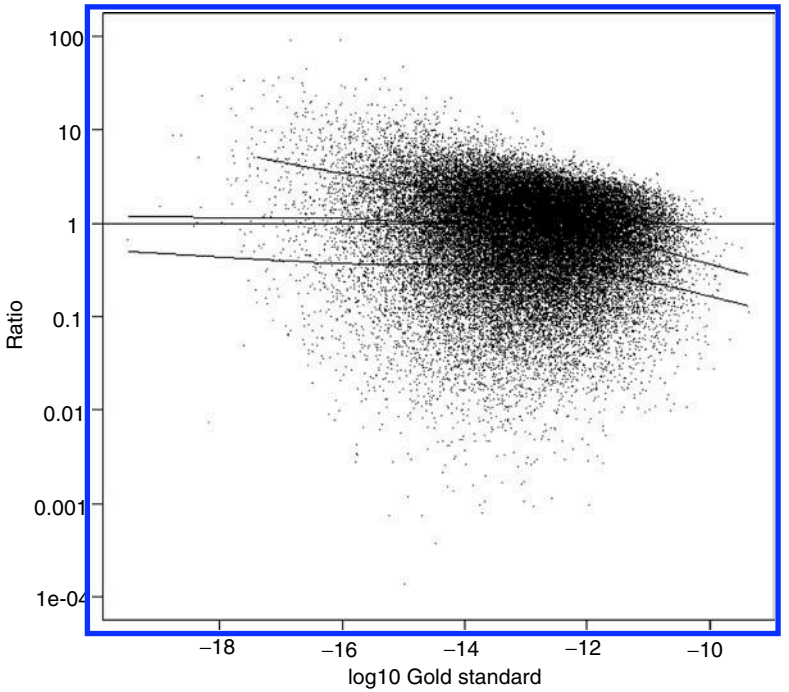

Figure 3.4 Ratio of the Recommendation 4.1 profile frequency ($\theta = 0.03$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.03$. The median and quartile trend lines are fitted. 54.4% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

64.7% for no correction). We see that this estimator still has a small prosecution bias and some undesirable variance properties.

3.3.3 The Subpopulation Formulae

If it is difficult to calculate the genotype probability in the population due to the effects of population subdivision, can we calculate it in the subpopulation of the suspect? We note that the subpopulation of the suspect may not be known, may not be easily defined, and almost certainly has not been sampled.

A potential solution has been offered by Balding and Nichols and has found widespread acceptance both in the forensic and the legal communities. The formulae[29,36,41,267,585] calculate the conditional probability of a second profile matching the stain from the subpopulation of the suspect given the profile of the suspect.

These formulae follow from a formal logic given initially by Balding and Nichols and appearing as Equations (4.10) in NRC II and (4.20) in Evett and

Weir, but they date back to the work of Sewall Wright[873] in the 1940s. A reasonably gentle derivation appears in Balding and Nichols.[39]

$$P_i = \begin{cases} \dfrac{[3\theta+(1-\theta)p_{i1}][2\theta+(1-\theta)p_{i1}]}{(1+\theta)(1+2\theta)}, & A_{i1}=A_{i2} \\ \dfrac{2[\theta+(1-\theta)p_{i1}][\theta+(1-\theta)p_{i2}]}{(1+\theta)(1+2\theta)}, & A_{i1}\neq A_{i2} \end{cases}$$

$$P = \prod_i P_i \tag{3.4}$$

Let us call the profile found at the scene of a crime profile C with genotype G_c. We will write the probability that the offender has this profile as $\Pr(G_c)$. Such a probability is called a profile probability, as the probability is not conditioned on any other information. Recommendation 4.1 is an attempt to calculate this probability.

However, let us consider whether the probability of a second copy of a certain genotype is raised slightly if one other person is known to have this genotype. There are many reasons why this may be true. But initially we will merely assume that it is true. If we had no knowledge as to whether or not this genotype had ever been found previously in an individual, then, indeed, we would be required to resort to a profile probability and Recommendation 4.1 may be an appropriate method. The "true" value of most of these profile probabilities would be 0 as discussed in Chapter 2.

However, we invariably have the information that at least one copy of the profile exists. We have seen it in the suspect. In other words, we are not talking about the vast majority of profiles that do not exist, we are talking about one of the few that do, indeed, exist in the real world.[842] Let us call the genotype of the suspect G_s, and we note that G_s and G_c are the same. In other words, the suspect could be the source of the stain at the scene. We are interested, however, in calculating the probability that a second person has this profile given that the suspect has it. This is written as $\Pr(G_c|G_s)$ and is called a match probability. It will be the same as the profile probability $\Pr(G_c)$ only if the knowledge that one person has the profile has no impact on our assessment that a second person has the profile. This is the assumption of independence discussed at the start of this chapter.

For the various population genetic reasons given above, we expect the assumption of independence to nearly hold, but to be violated in a minor way, in real populations. The main reason for this is population subdivision and relatedness. The fact that one person has the profile slightly increases the probability that his/her relatives or other members of his/her subpopulation have the profile. We are therefore led to the consideration of match probabilities.

It has been assumed that application of these formulae requires an assumption of independence between loci.[280, 311] This follows from the way that the single locus probability assignments are assembled into a multilocus probability assignment. Indeed these are multiplied and this gives the impression of an assumption of independence.

However, this is not true and was explicitly stated in Balding and Nichols' original paper:[36]

> Further, we have restricted attention to the suspect's sub-population and hence concerns about the Wahlund effect and correlations among loci can be ignored. Therefore the whole profile match probability is, to a close approximation, the product of the single-locus probabilities.

For those who prefer to investigate this statement in an algebraic way, some formative thoughts are given in Box 3.1. The subpopulation formulae of Balding and Nichols were designed to give an estimate of the match probability in the same subpopulation as the suspect. Most implementations of this approach apply this correction (in an overly conservative manner) to the whole racial group to which the suspect belongs rather than simply applying it to the subpopulation of the suspect. This is an understandable response to the difficulties in defining the subpopulation of the suspect, which most often is unknown, and not definable even if known. Equally the proportion of this subpopulation in the population is likely to be unknown. However, the approach of applying the correction to the whole "race" usually results in the correction becoming an "overcorrection" and hence gives rise to considerable conservativeness (or even performs in an overly conservative manner[f]) in the probability assignments.

Over the years I have received a lot of adverse criticism to the use of this correction regarding the difficulties in defining the subpopulation of the suspect. The difficulties can be demonstrated by taking almost any person and considering the question: "To what subpopulation does he belong?" Consider a Caucasian resident of New Zealand, born in London to New Zealand parents. He has Irish, Scottish, Norwegian, and English ancestors. It is almost impossible to define a subpopulation for him. This would be true of most people. This is termed a "population-centered approach" and it can be depicted graphically (see Figure 3.5). In this arbitrary graphic are placed circles depicting the Irish, Scottish, and English subpopulations. These all overlap in differing ways. Where should we now place Norwegian? Nor have we

[f] Clearly the term "overly conservative" used here has no objective definition. Rather it is a subjective term used to imply a very strong bias in favor of the defendant.

Box 3.1 Linkage Equilibrium and Conditional Probabilities (J.S. Buckleton and C.M. Triggs)

Consider two loci (locus 1 and 2). The crime stain has genotype G_c^i at locus i. The suspect matches and hence has genotype G_s^i at this locus. We note that $G_c^i = G_s^i$ for each of the loci, i, examined. We require $\Pr(G_c^1, G_c^2 | G_s^1, G_s^2)$. Using the third law of probability,

$$\Pr(G_c^1, G_c^2 | G_s^1, G_s^2) = \Pr(G_c^1 | G_c^2 | G_s^1, G_s^2) \Pr(G_c^2 | G_s^2, G_s^2)$$

Balding and Nichols' equation (Equation (3.4)) approximates this as

$$\cong \Pr(G_c^1 | G_s^1) \Pr(G_c^2 | G_s^2)$$

This is not an assumption of independence between G_c^1 and G_c^2.
One condition that will make this true is if

$$\Pr(G_c^1 | G_c^2, G_s^1, G_s^2) = \Pr(G_c^1 | G_s^1) \quad \text{and} \quad \Pr(G_c^2 | G_s^1, G_s^2) = \Pr(G_c^2 | G_s^2)$$

Looking at the first equality, we note that this does not imply independence between G_c^1 and G_c^2 unconditionally but rather implies that G_c^1 is independent of G_c^2 and G_s^2 in the presence of G_s^1. In other words, G_c^2 and G_s^2 provide no further information about G_c^1 given G_s^1. The truth of this assumption depends on our belief in the population genetic model.

The second equality requires that G_c^2 is independent of G_s^1 in the presence of G_s^2. The Balding and Nichols' equations are not a simple assumption of independence between loci.

The model upon which Balding and Nichols' equations (Equations (3.4)) are based assumes Hardy–Weinberg and linkage equilibrium at the subpopulation level (as well as some other assumptions). This is an explicit assumption of disequilibrium both within a locus and between loci at the population level. It is therefore seen that Balding and Nichols' formulae correct for that component of linkage disequilibrium that is caused by population subdivision.

really been specific enough. Should we have said "Graham" rather than Scottish? Hence the argument goes: subpopulations are indefinable.

However, the problem is illusionary. This can be shown by a similar graphic. Consider the same population but from a suspect-centered approach. The suspect has a number of close relatives: siblings, parents, and children. He also has more distant relatives: uncles, cousins. Further out he

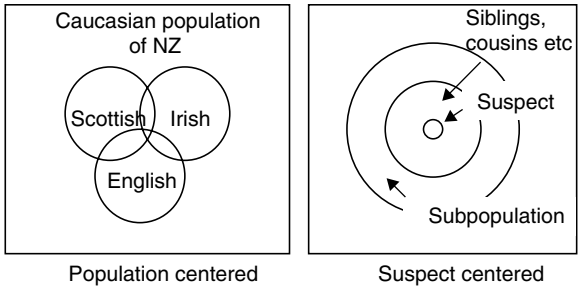

Figure 3.5 Diagrams depicting the population centered and suspected centered views of defining a subpopulation.

has second cousins and so forth. Beyond this there are a number of people to whom he is related more remotely. He may not know these people and there is probably no collective name for them. These are his subpopulation.[99,g]

Curran et al. use this same simulation approach to test how the "correction" advocated by Balding and Nichols[36] would perform.

Figures 3.6 and 3.7 reproduce the ratio of the "Balding and Nichols' θ corrected probability" to the true match probability for populations with true inbreeding parameters $\theta = 0.01$ and $\theta = 0.03$, respectively. In this experiment, Curran et al. have used the correct θ value created by the simulation when they applied Balding and Nichols' formula and have applied it to the whole population. In other words, there is no inherent conservativeness in the θ value *per se*, but there is a conservancy in that the correction is applied to the whole population rather than the subpopulation of the suspect alone. We can see that "θ corrected probability" has a strong bias in favor of the defendant, as expected. Few values lie below the ratio = 1 line and most are strongly conservative especially at the "rare" end on the graph.

This approach should remove any tendency of the product rule or Recommendation 4.1 to underestimate the genotype probability from population subdivision, but could potentially leave unaccounted subdivision of the subpopulation, possibly called sub-subpopulation division. The above simulations suggest that there is a substantial bias in the subpopulation formulae toward the direction of overestimation of the genotype probability. Since it is likely that sub-subpopulation effects will be markedly less than

[g] Subpopulations do not end, they fade out. We could envisage persons who are progressively more and more remotely related to the suspect. This could be approximated, if necessary, by bands of persons with differing θ values or better by the use of the general formulation whereby each pair of persons has a θ appropriate for their relationship. For this diagram, we take an arbitrary boundary to the subpopulation. The further out we push the boundary, the more people who are included in the subpopulation but the smaller the average value of θ.

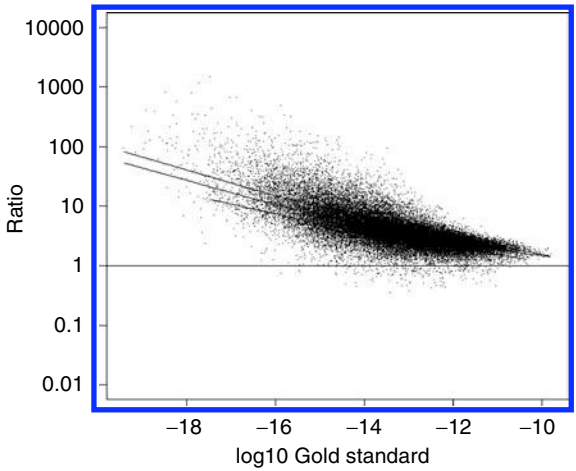

Figure 3.6 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.01$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.01$. 0.5% of samples have values less than 1. The median and quartile trend lines are fitted. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

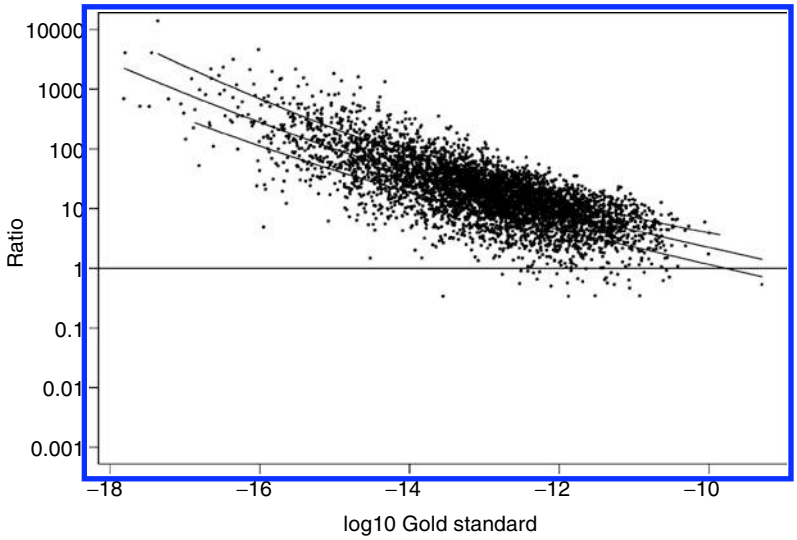

Figure 3.7 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.03$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.03$. 0.8% of samples have a ratio of less than 1. The median and quartile trend lines are fitted. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

subpopulation effects, it seems very unlikely that there will be any remaining bias toward underestimation.

Most laboratories actually exceed this level of conservativeness in that they tend to use a conservative value for θ. For example, the U.K. Forensic Science Service use a value of 0.02, whereas 0.005 could probably be justified for the Caucasian population of the U.K. Curran et al., using the simulation approach, also tested this. Figures 3.8 and 3.9 give the results from these simulations where the true population inbreeding coefficient $\theta = 0.005$, but 0.01 or 0.02 was used in the Balding and Nichols' correction.

This added level of conservativeness, that is, using a conservative value of θ, simply introduces increased conservativeness in the performance of the Balding and Nichols' estimator.

A criticism of this approach points out that this conditional probability is the probability assignment for a certain genotype in the same subpopulation as the defendant, not in the population as a whole.[121,129] This is indeed correct. It is sometimes suggested that these formulae are, therefore, only applicable if it is known that the true offender, if not the suspect, must be from the same

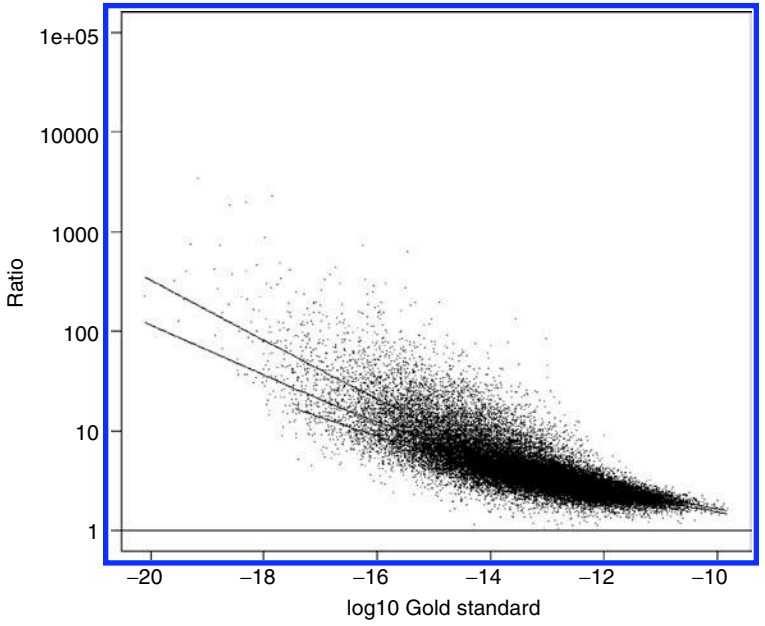

Figure 3.8 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.01$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.005$. The median and quartile trend lines are fitted. 0% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

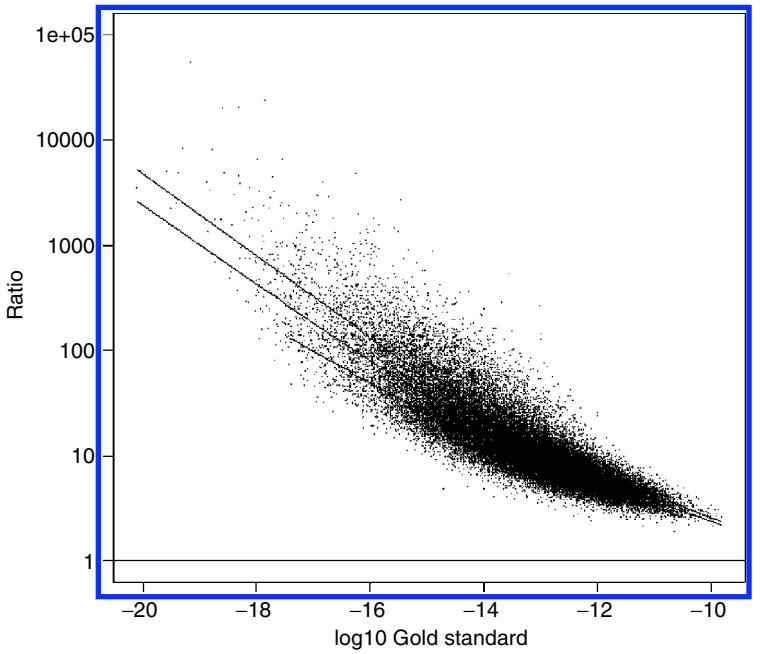

Figure 3.9 Ratio of the Balding profile frequency ($\theta = 0.02$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.005$. The median and quartile trend lines are fitted. 0% of values are less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

subpopulation as the suspect. This argument can be easily examined by simple mathematical exploration. But before we do that we ask: can the product rule be used only if it is known that all possible offenders are *not* from the same subpopulation as the suspect or are *not* related to the suspect? This is the logical corollary of the argument of Budowle et al.[121,129] If we pursue this line, we will eliminate all possible estimators.

We will assume arbitrarily that each person is as likely as any other to be the true offender if the suspect is innocent. This assumption is very unlikely to be realistic in practice for many reasons, not the least of which is that those people close to the crime scene have a higher chance of being the offender, and persons in remote locations have a lesser chance. Assume further, for example, a population of which 10% are in the same subpopulation as the suspect.

To demonstrate these effects, we generated simulated allele proportions randomly between 0.02 and 0.20 (Table 3.4) and examined the relative contribution to the estimated match probability. In this simulation, 11 loci were

Table 3.4 Simulation of Allele Proportions Randomly between 0.02 and 0.20, and Relative Contribution to the Estimated Match Probabilty

Locus	Pr(Allele 1)	Pr(Allele 2)	Product Rule	Subpopulation	Ratio
1	0.15	0.19	0.0564	0.0724	1.3
2	0.03	0.05	0.0027	0.0085	3.1
3	0.08	0.16	0.0254	0.0384	1.5
4	0.06	0.15	0.0184	0.0305	1.7
5	0.16	0.08	0.0267	0.0398	1.5
6	0.20	0.04	0.0159	0.0297	1.9
7	0.12	0.11	0.0256	0.0380	1.5
8	0.03	0.07	0.0040	0.0110	2.8
9	0.15	0.03	0.0101	0.0212	2.1
10	0.19	0.03	0.0097	0.0227	2.3
11	0.08	0.10	0.0173	0.0281	1.6
12	0.09		0.0082	0.0240	2.9
13	0.18		0.0310	0.0551	1.8
Assigned probability			1.32E−24	6.31E−21	4780
Weighted probability assignment				6.33E−22	

set as heterozygotes and two as homozygotes. The product rule and the subpopulation corrected probability assignments were calculated. For the subpopulation correction, we used $\theta = 0.03$. If we assume that the product rule relates to the 90% of the population who are not members of the subpopulation, and the subpopulation correction relates to the 10% who are members of this subpopulation, we arrive at a weighted probability assignment given.

We see that the weighted probability assignment is different to both the product rule and the subpopulation corrected estimate. But it is almost totally dominated by the contribution of the 10% of the population who are in the same subpopulation as the suspect. The contribution from the product rule is almost irrelevant. In fact, a reasonable approximation could be obtained by simply multiplying the subpopulation probability estimate by its fraction in the population, completely ignoring the product rule contribution. However, if the correction is applied to the whole population rather than simply the subpopulation, as is customary, this is likely to result in an "overcorrection," as previously discussed and demonstrated by simulation. Hopefully this simple example can settle the discussion on the subject of product rule or subpopulation correction. We have a choice: Do we want to be slightly under or more substantially over with our estimate?[h]

[h] Bear in mind that we do not know the true answer. Hence the words "over" and "under" are relative to the "gold standard" which, in itself, is the result of a model.

3.4 Discussion of the Product Rule and the Subpopulation Model

If we are able to show by population genetic studies that the effects of population subdivision are so minor that we are prepared to ignore them, then it is permissible to use the product rule as a first-order approximation provided that it is understood that it is probably slightly biased in favor of the prosecution. A useful review of various approaches is made by Gill et al.[355]

The belief on which the use of the product rule is based can arise only from well-constructed population genetic examinations[600] that assess the population genetic subdivision at the genetic level. This is vital rather than assessment at the geographical level, which may be peripheral, especially in countries settled largely by recent colonization. This is because geographic samples in, say, the U.S., taken from Caucasians from different states or cities, are unlikely to express the underlying genetic diversity. Suppose that we took two samples each of, say, 33% Scottish, 33% English, and 33% Italian. The allele frequencies demonstrated by these two samples will probably be very similar. However, if we compare comparable samples drawn separately from the Scottish, English, and Italian populations, we will find small but real differences between them.

A common and reasonable response is that the difference between the product rule estimate and a fair and reasonable assignment of the evidential value is not forensically significant.[127,128] This is probably true in many instances; however, there is divergent evidence. For instance, in the identification of war victims from the 1991–1995 war in Croatia, Birus et al.[69] found an unexpectedly high number of false matches between skeletal remains and the relatives of missing persons. They attribute this to substructure in Croatia and warn:

> Although genetically and statistically sound and widely accepted, calculations that we perform today produce numbers that might not be fully applicable in all situations. One of the factors not included in these calculations (the product rule) is the effect of local inbreeding.

It remains important to understand that the commonly applied approach of independence testing in no way measures the extent of departure from equilibrium, and cannot be used to estimate the difference between the product rule assignment and a fair and reasonable assignment.[230, 503, 504, 511, 584, 665]

Therefore, the statement that the potential error is not forensically significant, if true at all, cannot be based on independence testing. Again it can only be investigated at all, and certainly not proved, by a population genetic

model or perhaps by experiments of the type pioneered by Tippett[315] in the case of less vanishingly small probabilities.

It may be interesting to note the expected behavior of these two approaches, if indeed the requirement of independence is not fulfilled. If we pick a genotype at random, irrespective of whether it is known to exist or not, then recommendation 4.1 is likely to provide a fair and reasonable probability assignment (note that although it is fair and reasonable, it is not necessarily the true value). However, if we now add the additional information that one person, the suspect, has this profile, then we have two options.

First, we could ignore this additional information and still proceed with Recommendation 4.1. This is no longer an unbiased approach. In fact, using Recommendation 4.1 the probability assignment is likely to have a small bias in favor of the prosecution because the knowledge that we have ignored increases the probability that a second copy of this genotype exists. The extent of this bias is dependent on how large or small are the dependence effects.

Second, we could follow the logical Bayesian approach, which does, in fact, lead to consideration of the conditional probabilities such as $\Pr(G_c|G_s)$ discussed above. These have a remarkable robustness to deviations both from Hardy–Weinberg and linkage equilibrium and as such, we believe, represent a more fair and reasonable probability assignment. However, we accept that, as implemented, they appear to represent an overcorrection. For a discussion on implementation in the U.K., see Foreman et al.[313] (unfortunately not generally available).

This difference between the two approaches is as fundamental as the difference between unconditional probabilities and conditional ones.[267,840] An approach based on mathematical logic leads us to the conditional probabilities. In fact, it would appear that some former major proponents of the validity of the product rule have now modified their position in the face of increasing data.[60,120,121,134,154,743]

There is no possibility of experimentally verifying probability assignments this small. They represent, in multilocus cases, extrapolation way beyond anything that can be experimentally examined.

It must be accepted that, like the product rule, the subpopulation formulae rely on a population genetic model, albeit one that is more robust and concedes doubt correctly to the defendant. Whereas it is possible to say that the product rule is mildly biased towards the prosecution, it is not possible to state whether or not the subpopulation formulae are also biased. It is at least theoretically possible that they are conservative, and the experimental evidence given here suggests that this is so.

A discussion of the ethics of this debate is given by Beyleveld,[62] who also discusses some of the pressures that have been brought to bear on independent bodies, when considering these issues.

3.4.1 Effect of Mutation

The effect of mutation on the assessment of multilocus genotype probabilities has recently been considered. Laurie and Weir[488] warn of the consequences of mutation of the infinite allele type on the estimation process. This model may be a reasonable model for minisatellites, although a consensus has not yet been developed.

Laurie and Weir suggest that the assumption of independence understates the two-locus match probabilities for such loci. The effect increases with increasing mutation rate. For loci with high mutation rates, the two-locus probabilities may differ substantially from the product of single-locus probabilities. They show that these dependency effects accumulate across loci: "These results indicate a potential concern with using the product rule to compute genotypic match probabilities for highly mutable loci."[488]

In loci with high mutation rates, alleles stand an increased chance of being recent and rare. "Hence, if two individuals share alleles at one locus, they are more likely to be related through recent pedigree, and hence more likely to share alleles at a second locus."[488]

This conclusion may hold for the infinite alleles model. This model is unlikely to be applicable to STRs and the effect of mutation on between-locus dependencies at these loci has yet to be settled.

If we restrict ourselves to the question — Do the Balding and Nichols' formulae give an adequate assignment of the match probability in the subpopulation of the suspect? — we again must accept the impossibility of experimentally testing such multilocus estimates.

We are left with examining the validity of the assumptions of the model and simulation results. This matter is elegantly considered by Graham et al.,[370] who point out that the assumptions of the Balding and Nichols' model include a steady-state population and a mutation model in which the allelic state after mutation is independent of the state prior to mutation. Both of these assumptions are untenable. Graham et al.[370] investigate the consequences of a generalized stepwise model and conclude: "[the Balding and Nichols] theory can still overstate the evidence against a suspect with a common minisatellite genotype. However, Dirichlet-based estimators [the Balding and Nichols' formulae] were less biased than the product rule estimator, which ignores coancestry."

Laurie and Weir finish with the conclusion:

> The method of adjusting single-locus match probabilities for population structure [the Balding and Nichols' equations] when multiplied across loci has been shown empirically to accommodate the dependencies we have found for multiple loci.

3.4.2 Admixture

Previously we have described a population genetic model designed to cope with population subdivision. This may describe the evolutionary event where one population continually splits into two or more populations that subsequently evolve separately. Human history is more complex than this and no pretence was ever made by the authors of these approaches that they were exact descriptions of the evolution of actual human populations.

What happens when the rate of gene flow into a population becomes very large?

This may describe the modern evolutionary events in many populations. Populations such as the New Zealand Maori were once much more isolated than they are now. However, they were never completely isolated as the Polynesians were great navigators and there is considerable evidence of extensive trading networks across large distances in the Pacific. With the large-scale settlement of Aotearoa (New Zealand) by Pakeha (Caucasians), gene flow of Caucasian genes into the Maori population was initiated and seems to have been sudden and considerable. The modern New Zealand Maori population is thought to contain no full-blood Maori.[491]

This is a different evolutionary event to the small-scale migration treated in modifications of the subpopulation model. It warrants separate treatment with a different population genetic model. We will refer to this model as the "admixture model."

Admixture in the Americas is common, with individuals having ancestors who may be Caucasians, Native Americans, Asians, or Africans.[688] It has been estimated that 15–25% of the African-American gene pool is derived from the Caucasian population.[606]

Chakraborty and Kidd[161] suggested that estimation of profile frequencies using average allele frequencies and the product rule may be recommended as the number of individuals in the population with mixed ancestry increased. This is partially because random mating in the admixed population restores the within-locus disequilibrium in the population and the between-locus disequilibrium is halved after each generation.[160] However, this thinking applies more to a future equilibrium situation and not to the transitional state that most admixing human populations demonstrate. In the transitional state, there is pronounced correlation between loci, whether the admixed population is defined to exclude pure blood individuals or not. This can be demonstrated by extreme examples such as Table 3.5. Note that in the crossed offspring, every individual is genotype *abcd* and hence this population is in Hardy–Weinberg and linkage disequilibrium. Real examples will show much milder effects.

Law[491] describes an alternative and preferable model for this situation. This model is based on the concept that alleles are independent within and between loci conditional on the pedigree (essentially an assumption of Mendelian

Table 3.5 Hypothetical Admixture Between Two Populations

	Allele probabilities	Pop 1	Pop 2
	Locus 1		
	Allele *a*	1	0
	Allele *b*	0	1
	Locus 2		
	Allele *c*	1	0
	Allele *d*	0	1
Genotype probabilities	*aa*	*ab*	*bb*
	Pop 1 × Pop 1		
cc	1	0	0
cd	0	0	0
dd	0	0	0
	Pop 1 × Pop 2		
cc	0	0	0
cd	0	1	0
dd	0	0	0
	Pop 2 × Pop 2		
cc	0	0	0
cd	0	0	0
dd	0	0	1

segregation). The model allowed for differing mating patterns, number of parental populations, and genetic distance between populations. Comparisons were made with the product rule estimate using average allele frequencies.

Law concludes that "as the genetic distance and the number of parental populations increases, the difference between the match probability calculated using (the Law admixture model and) the product rule increases. The maximum difference can be larger than (a) factor of more than 10,000 for a six loci genotype."

The Law model can also be compared with the estimate that would be produced if the substructure model of Balding and Nichols were used for a population undergoing admixture. This analysis suggests that a conservative estimate of θ could be used in Balding and Nichols' equation along with the allele frequencies from the whole admixed population. Since we are using a model where the inbreeding coefficient θ does not have its usual interpretation, it is better to rename it as the "equivalent inbreeding parameter q" and to understand that we are simply seeking that value for q which gives us approximately equal estimates when compared with the admixture model.

Law concludes that:

> …there are genotypes which require an equivalent inbreeding coefficient that is greater than the genetic distance between the

parental populations especially when there are three or more parental populations. However, the spread of the estimated equivalent inbreeding coefficients is reasonably large as different genotypes may be affected by admixture to differing degrees depending on the difference in allele frequencies. Using the maximum estimated equivalent inbreeding coefficients is likely to overestimate the match probability since such an extreme estimate is most likely (to be) due to rare alleles in one of the parental populations. The 95th percentile of the equivalent inbreeding coefficient may provide a more appropriate value of q.

This analysis suggests that the use of a value for q that is the same as the genetic distance between the parental populations may be an adequate compensation for admixture effects (see Table 3.6). If a more accurate estimation is required, the Law algorithm is preferred.

3.4.3 Allelic Dropout

Occasionally the situation occurs when one allele can be reliably scored but it is ambiguous whether or not there is a second allele. This situation is handled using the "F" designation in the U.K. and the "N" designation in New Zealand. Using the product rule the, say, 16, F genotype is assigned a frequency $2p_{16}$ (strictly this should be $p_{16}(2 - p_{16})$[102]). This approximation has been referred to extensively as the "2p rule." Using the subpopulation

Table 3.6 Median, Upper Quartile, 90th, 95th Percentiles, and the Maximum for q

Number of Parental Populations	Admixture Proportions	Genetic Distance	50%	75%	90%	95%	max
2		0.03	0.01	0.02	0.02	0.02	0.04
	Equal	0.05	0.01	0.02	0.02	0.03	0.06
		0.10	0.03	0.04	0.05	0.06	0.10
		0.03	0.01	0.01	0.02	0.02	0.05
	Unequal	0.05	0.01	0.02	0.03	0.03	0.06
		0.10	0.04	0.05	0.07	0.08	0.13
3		0.03	0.02	0.02	0.03	0.03	0.05
	Equal	0.05	0.03	0.04	0.05	0.05	0.08
		0.10	0.08	0.10	0.12	0.14	0.23
		0.03	0.01	0.02	0.02	0.03	0.05
	Unequal	0.05	0.02	0.03	0.04	0.05	0.08
		0.10	0.05	0.07	0.09	0.10	0.14

Reproduced with the kind permission of Dr. Law.

Table 3.7 Conditional Probabilities for the "θ" Designation Assuming Two Different Conditioning Situations

Genotype of stain	Suspect	"2p Equivalent"	"$p_{16}(2-p_{16})$ Equivalent"
16θ	16,16	$2\dfrac{2\theta+(1-\theta)p_{16}}{1+\theta}$	$\dfrac{2\theta+(1-\theta)p_{16}}{1+\theta}\left(2-\dfrac{3\theta+(1-\theta)p_{16}}{1+2\theta}\right)$ that is always less than the 2p equivalent
		$2\dfrac{\theta+(1-\theta)p_{16}}{1+\theta}$	$\dfrac{\theta+(1-\theta)p_{16}}{1+\theta}\left(2-\dfrac{2\theta+(1-\theta)p_{16}}{1+2\theta}\right)$
	16, x		that is always less than the 2p equivalent

correction, the probability assignment depends on the genotype of the suspect and any other conditioning genotypes. To demonstrate this, we condition only on the suspect's genotype below, and the extension to conditioning on additional genotypes follows by the same method (see Table 3.7). However this approach does not adequately model drop out. It is preferable to use the models discussed in Chapter 8.

3.4.4 Arbitrary Limits

Foreman and Evett[311] have suggested that "case specific match probabilities should not be calculated as a matter of principle." Instead they suggest the use of "general figures." Below we give the calculated figures for the most common profiles for an SGM$^+$ 10-locus match and the suggested reported value:

- 1 in 8300 for siblings which they suggest reporting as 1 in 10,000.
- 1 in 1.3 million for parent/child reported as 1 in a million.
- 1 in 27 million for half siblings or uncle/nephew reported as 1 in 10 million.
- 1 in 190 million for first cousins reported as 1 in 100 million.
- 1 in 2.4 billion for members of the same subpopulation reported as 1 in a billion.
- 1 in 5 billion for unrelated persons also reported as 1 in a billion.

This is an extension of an older Metropolitan Police Forensic Science Laboratory policy of truncating match probabilities at 1 in 10 million.[382]

This approach is or has been accepted practice in the FSS and at Forensic Alliance in the U.K. Foreman and Evett motivate their approach by stating that "the independence assumptions are sufficiently reliable to infer probabilities that are of the order of 1 in tens of millions" but that SGM$^+$ case specific match

probabilities would "invoke independence assumptions to a scale of robustness which we could not begin to investigate by statistical experiment. ..."

I admit the pragmatism and intuitive appeal of this approach; however, it really is a long way away from my own philosophy. My objections would range from the practical to the philosophical and will be mentioned briefly here.

- The relative reliance upon independence assumptions and Mendel's laws differs markedly between the calculations for siblings to the use of the product rule. For siblings, most of the procedure leading to a probability assignment is based on the assumption that alleles assort in a Mendelian fashion and only to a very small extent on independence assumptions within a population. Hence these calculations are much less affected by uncertainties about independence.
- If we can support probability assignments of 1 in tens of millions using Tippett testing (see Chapter 5) but not lower, how are we to support assignments of 1 in a billion?
- The probability assignments that are advocated in this chapter are really based on belief in a model. They are not based on independence testing or Tippett tests at all.
- A limit of 1 in a billion is not likely to induce further refinements of the model, or simulate further sampling and study.
- What would we do if we added more loci?

In general, I would vastly prefer to assign a probability, whatever it may be, without a limit but to accept and make explicit that very low probabilities cannot be verified experimentally.

3.4.5 Same Source?

The reasonable question has arisen: when can a DNA profile match be considered proof that two DNA samples have come from the same source? The FBI announced a policy on this in November 1997.[410] The term "same source" is used in this discussion to describe this situation as it best approximates the underlying forensic question. Other terms such as "uniqueness," "source attribution," and "individualization" have been used elsewhere. This has led to considerable discussion of the use of these terms, which has also produced useful philosophical debates about their meaning. I cannot do justice to these arguments and simply direct the reader to the well-written work by Champod and Evett[173] on the equivalent subject in the area of fingerprints (see also the response by McKasson[540] and the more balanced commentary by Crispino[206] or the excellent writing of Inman and Rudin[429]).

The question of whether we can ever base a conclusion of common source on a probabilistic argument has also been examined, most notably by Stoney,[734,735] Champod,[168] and Evett et al.[280] In the DNA context we can see that, using the current population genetic models, the more loci we add, the smaller are the match probabilities produced by our model. There are three important points with regard to this. First, that the estimated match probability derived from the model can approach zero but never actually equal zero. Second, that estimates of very small match probabilities arising from models cannot be directly tested. They are as reliable or unreliable as the models themselves. Third, we recognize that we are considering an extreme extrapolation using these models. We are not operating near the center of their prediction range where they are more testable and tested. The models have been extensively tested in this central range and there is some considerable reason to believe that they are robust there, but they are still models and the probabilities produced by them are still untestable.[i]

To conclude the same source from a probabilistic model, someone has to decide that the probability estimate produced by that model at this extreme end of extrapolation is sufficiently reliable that it can be trusted and the probability is sufficiently small that it can be ignored. Stoney[735] terms this the "leap of faith."

Inman and Rudin[429] describe this situation, "at some subjective point, most qualified authorities would agree that, for practical applications, the likelihood … is so small that it can be ignored." In the text following this quote, they very clearly set out the subjective nature of this decision.

There has been considerable argument about whether a scientist should do this or leave the matter to the court. Certainly in England and Wales, the court direction appears to be that the scientist should not be the person who decides whether the probability is small enough to ignore.[201]

Inman and Rudin[429] agree:

> It is the purview of the fact finder to draw inferences from circumstantial evidence, and, of course, potentially individualizing physical evidence is circumstantial evidence. However, there are pieces of information that only science can legitimately provide to the fact finder, such as population frequencies, transfer and persistence data, and limitations of the evidence and the test.

It is unclear whether the scientists should even be the persons who decide on the reliability of the model. It is regrettable to me that, as we add more loci, we extrapolate the model further and further, but little new experimental

[i] To quote Ian Evett: "is it rational for me to assign such a small match probability?"

data into the reliability of the model at this extreme are being produced. Robertson and Vignaux[659] complained about a similar lack of fundamental research in the area of fingerprints:

> In these cases it seems that the expert is giving evidence of identity when, and only when, in his judgement the probability of getting the evidence assuming the alternate hypothesis is so small that it does not matter what the numerator or even the prior odds are. At what point this is reached seems to be a matter of judgement and experience and there most writers on expert evidence are content to let the matter rest. This may have had the unfortunate effect of removing the incentive to carry out the basic research to build appropriate models. Intellectually, this is unsatisfactory and further work is required to understand the processes involved in making these decisions. In the meantime the proposal that all forms of scientific evidence be given in the form of a likelihood ratio is a counsel of perfection.

Returning to DNA profiling, Budowle et al.[129] make the reasonable distinction between the judgement in one particular case and the judgement in all potential cases. We could imagine a criterion that was considered reasonable in an individual case and Budowle et al. suggest "99% confidence."[j] They go on to suggest that this may correspond with the term a "reasonable degree of scientific certainty." This term has been selected because of its legal implications.

> From the medical model has come the phrase "to a reasonable scientific certainty." Both the judicial system and some experts have latched onto this phrase as a convenient way to render an opinion as fact. As convenient as it might be, it is a non sequitur. As we have repeatedly discussed throughout this book, the notion of scientific certainty does not exist. In our opinion, scientific experts should refrain from resorting to that phraseology in expressing their opinions.[429]

Budowle et al.'s method stems from a suggestion by NRC II who discussed the use of the formula $p_x \leq 1 - (1 - \alpha)^{1/N}$, where p_x is the match probability, N is the size of the suspect population, and $1 - \alpha$ is the confidence interval. They give an example using a 99% confidence interval $(1 - \alpha) = 0.99$ implying $\alpha = 0.01$ and $N = 260{,}000{,}000$, the approximate population of the U.S. This suggests

[j] This term needs thought. There is a distinction between the use of the words confidence and probability.

a match probability of $p_x = 3.9 \times 10^{-11}$. It is suggested that the estimated p_x be decreased by a factor of 10 to provide additional conservativeness. Weir[840] correctly points out the flaws in this approach which unreasonably assumes independence of trials.

Also included in the original publication is a brief mention of relatedness. In particular, they recommend typing of relatives. The typing approach to dealing with relatedness is admirable, but is applied only rarely in the U.S., the U.K. or New Zealand. In the absence of typing, they suggest that the match probability for brothers be calculated or that calculations should be performed (when required) for three classes of people: unrelated, subpopulation members, and relatives. They do not give a specific formulation of how to amalgamate the contribution from relatives and unrelated people, directing the reader, correctly, to Balding.[34]

This division of the population into unrelated, subpopulation, and related persons is akin to the coarse division undertaken by Balding. The unifying formula suggests that it is the weighted sum of all three contributions that should be considered and not simply one or the other of these probabilities.

The unifying formula will assign a posterior probability to the hypothesis that the suspect is the donor of the stain material. This appears to be the probability that is desired in "source attribution." However, the unifying formula will require an assignment of prior probabilities and this cannot be avoided. This may appear as a fatal flaw and indeed it is worrying. It is central to the concerns about the concept of "source attribution" and "a reasonable degree of scientific certainty." We see therefore that any approach to assigning a posterior probability involves a prior. This is, of course, not an original insight and was reported as long ago as 1983[257] in forensic science and much earlier in other sciences.

There is an interesting interplay between the prior for the suspect and the probability that someone else possesses this profile. Balding and Donnelly[37] explained this:

> Finally, we remark that the magnitude of the size biasing effect… is related to the prior distribution. Intuitively, the effect occurs because, under the hypothesis of innocence, two distinct τ-bearers[k] have been observed. Such an observation stochastically increases the number of τ-bearers, thus decreasing the strength of the evidence against the suspect and decreasing the probability of guilt. Decreasing the prior probability of guilt increases the chance that the suspect and criminal are distinct, hence increasing the

[k] This is the term used to describe the people carrying the matching profiles: in this case, the defendant and the true perpetrator.

effect of size biasing. (David Balding and Peter Donnelly quoted with the kind permission of CRC Press)

This effect can easily be illustrated. Suppose that we have a certain profile at a crime scene and that this matches a suspect. But the suspect, for whatever reason, cannot have been the donor (his prior is 0). Then the probability that someone else possesses this profile goes from whatever value it was before to 1.

> Consider a crime scene DNA profile which is thought to be so rare that an expert might be prepared to assert that it is unique. Suppose that, for reasons unrelated to the crime, it is subsequently noticed that the crime scene profile matches that of the Archbishop of Canterbury. On further investigation, it is found to be a matter of public record that the Archbishop was taking tea with the Queen of England at the time of the offense in another part of the country. (You may consider your preferred religious leader, beverage, and head of state in place of those named here.) A reasonable expert would, in light of these facts, revise downwards any previous assessment of the probability that the crime scene profile was unique. However, this is just an extreme case of the more general phenomenon that any evidence in favour of a defendant's claim that he is not the source of the crime stain is evidence against the uniqueness of his DNA profile.[34] (David Balding, quoted with the kind permission of Science and Justice)

The supposition that the Budowle et al. approach is necessarily conservative is of concern. An appeal is often made at this point to the increase in the frequency assignment by a factor of 10 and the relatively large value chosen for N (260 million). The factor of 10 was intended to compensate for potential sampling error or subpopulation effects or both. Examination of the unifying formula suggests that it may be inadequate especially when many loci are considered. It is also likely to be inadequate to compensate for both subpopulation effects and sampling error, and certainly cannot compensate for the effect of uneliminated brothers.

Budowle et al. make it clear that this approach is designed for a case-by-case application. If we misapply this method to the question of "are such profiles unique in the U.S.," we will soon be embarrassed. There are 3.38×10^{16} pairs of people in the U.S. If we use the estimated match probability suggested for the 99% confidence interval $p_x = 3.9 \times 10^{-11}$ and assume that the factor of 10 recommended as additional conservativeness was included, then $p_x = 3.9 \times 10^{-12}$. If this match probability is exactly correct (recall that it is only an estimate), then there will be an expectation of about 132,000 matching pairs

Table 3.8 Size of Databases That Give the Expectation of One Match

	$\theta = 0.00$	$\theta = 0.03$
U.S. African-Americans	43,000,000	11,000,000
U.S. Caucasians	34,000,000	9,300,000
U.S. South-Western Hispanics	21,000,000	5,900,000

of unrelated people in the U.S. In fact, a database of about 716,000 profiles all with a match probability of $p_x = 3.9 \times 10^{-12}$ would have an expectation of about 1 match. In reality, full CODIS profiles produce match probability estimates less than this. Bruce Weir[844] estimates that we would expect a full CODIS match among unrelated people if the databases were of the size shown in Table 3.8.

Despite the careful words in the paper of Budowle et al., my suspicion is that it will be read as providing a declaration of uniqueness among all people and hence such an adventitious match will cause public embarrassment. Certainly the view is developing among the public that DNA profiles are unique.

The situation is probably slightly worse when we consider relatives. The expected number of matches when relatives are included in the population or database will be larger. It is likely that there are a number of pairs of persons matching at the 13 CODIS loci in the whole U.S. population of 260 million. Many of these matching sets will be brothers. The chance that two of these are involved in the same crime is small, but the matches will eventually be revealed as the sizes of databases increase and will embarrass forensic science if we have declared such profiles unique.

Findlay and Grix[299] have studied juries and report a strong preexisting prejudice that is pro-DNA. It is likely that many jury members wrongly believe that all DNA findings represent certain identification. It would be worrying to foster this belief.

My feeling is that we would be unwise to conclude the same source because it is not our place to do so. If we do so, I would prefer the standard to be much higher than previously suggested AND I would like us to make transparent that we have subjectively decided to round a probability ESTIMATE off to zero. On balance I cannot see much positive coming from a policy of declaring a common source.

3.4.6 Animal and Plant DNA

We are starting to see the use of animal DNA in criminal proceedings (for an excellent review, see Halverson[386]) when, say, blood from a shot dog may have been transferred to an offender. Animal and plant DNA is extensively used in

wildlife and conservation science to investigate illegal hunting and other risks to protected species. The population genetic arguments given above apply to all species, except that in many cases subpopulation effects and inbreeding are more severe outside humans.[52]

3.5 A Complex Case Example — DNA Evidence and Orenthal James Simpson[1]

In June 1994, O.J. Simpson was 47 years old. He was one of the most respected sportsmen in the U.S. and he had just been charged with the double murder of his estranged wife Nicole Brown Simpson and her friend Ronald Goldman. This precipitated a trial with media coverage unprecedented in U.S. history. DNA evidence was about to be center stage.

In his early sporting career, O.J. had been the star running back for the University of Southern California, winning the Heisman Trophy in 1968. His professional career was with the Buffalo Bills until his retirement in 1979. That same year his first marriage to Marguerite Whitley, his teenage sweetheart, ended. The couple had three children, a son Jason, daughter Arnelle, and a second daughter, Aaren, who accidentally drowned at the age of two.

O.J. had met Nicole Brown in 1977. She was aged 18, he 30 at the time. Nicole had been born in Frankfurt, Germany to a German mother and a U.S. military serviceman father.

O.J. was inducted into the football hall of fame in 1985, his first year of eligibility.[498] He had married Nicole the same year and the couple later had two children: Sydney born in 1986, and Justin in 1988. However by 1992, Nicole had left O.J. after what was presented at the trial as a history of abuse and violence. In 1993 police were summoned to Nicole's residence after the now estranged O.J. had kicked in the door, screamed obscenities, and had beaten her Mercedes-Benz car with a baseball bat. Official records listed 62 separate incidents of physical and mental abuse by Simpson toward his wife. One of these incidents occurred in 1985 and involved Detective Mark Furhman, who was to feature prominently later in the investigation and trial. Furhman later recalled that this incident was "indelibly pressed" into his memory.[507]

At 10:20 PM on Sunday, June 12, 1994, there was the sound of a dog barking at 875 South Bundy Drive in the Brentwood district of LA. Shortly before midnight, Akita, Nicole's dog, paws splashed with blood, had led neighbors to the scene of the murders. Nicole, aged 35, was face down with her throat slashed almost through. To her right lay the body of a male later identified as Ronald Goldman, aged 25, a waiter at the fashionable Mezzaluna restaurant.

[1] This section was written by John Buckleton and Christopher Triggs.

Nicole and Ronald had known each other for six months, but there was no suggestion of a romance between them. On the night of the murders, he had been delivering Nicole's mother's reading glasses, which had been left in the restaurant. Next to the bodies were keys, a blue knit cap, a beeper, a blood-spattered white envelope, and, nearer to Nicole's body, a bloodstained left-hand leather glove. Bloody shoeprints and spots led from the bodies toward the back of the property.

This book deals largely with the interpretation of DNA after it has been analyzed in the laboratory. It neglects the huge and vitally important fields of evidence collection, recording, and handling. This section seeks in a small way to redress this imbalance. There have been understandable complaints that we forensic scientists have not learnt the lessons necessary from this and other similar cases.

The autopsy was performed on June 14 by Dr. Irwin Golden. It showed injuries to both of Nicole's hands, which suggests that she had defended herself. From the cut to the throat, the pathologist concluded that the attacker was right handed.

Mr. Goldman had been clubbed from behind and stabbed 19 (or 28[498]) times.

White towels had been used by the detectives to soak up blood[498] to allow easier approach to the bodies. *This is an unwise practice*.[m]

Detective Mark Furhman was the 17th officer to sign in at the scene. After initial inspections, instructions had been issued that O.J. should be told personally of the tragedy. Furhman had volunteered. He knew Mr. Simpson's house was two miles from Nicole's from the previous visit. At the trial, the defense claimed that Mr. Simpson was "targeted" by the police. However, it would be normal for an ex-husband to be a suspect early in an investigation and this would not be an issue as long as an open mind was maintained. Detective Vannatter, the head of the team of detectives at the scene, has subsequently insisted that O.J. was not being treated as a suspect at this time. However, events suggest that he was. For instance, the Goldmans were not informed personally of their son's death although O.J. had been. Forensic staff were initially called to O.J.'s Rockingham house. *But a valid complaint would relate to the sending of any personnel from one crime scene to another potential scene. The issue of cross contamination would immediately arise and should have been stringently guarded against. If the same staff must go to both scenes then strict precautions must be taken, such as overalls (clean or disposable), overshoes, and fresh gloves.*

[m] We use the italics here to signify personal commentary as opposed to the historical narrative.

Two cars containing detectives went to O.J.'s residence at the junction of Rockingham Avenue and Ashford Street. Outside, parked badly, was Mr. Simpson's 1994 white Ford Bronco. Detective Furhman pointed out to a fellow officer what appeared to be blood inside the Bronco near the driver's door handle. This supposedly represented reasonable cause and so Detective Furhman climbed over the wall and unlocked the gate. *If this evidence represented probable cause, then this was now a crime scene and different personnel should have been summoned.* Dr. Henry Lee, a criminalist employed by the defense, presented arguments that Furhman must have opened the vehicle[498] since the blood was not visible with the door closed.

There had been no response from the front door. At the back of the house were three guest bungalows. Brian "Kato" Kaelin, a friend of Nicole's, was staying in one. Arnelle Simpson, O.J.'s daughter, in another. She let the officers into the main house. No one was present. O.J. Simpson had taken the 11:45 flight to Chicago to attend a convention of the Hertz Rental Company scheduled for the next day. He appeared in advertisements for this company and his presence at this conference had been expected. The flight had been booked well in advance.

Mr. Kaelin was then interviewed. He and O.J. had been together for dinner at a McDonalds in Santa Monica and had returned to the house at 9:40 PM. At 10:45 Kaelin had heard three banging noises from the rear of the building near an air-conditioning unit. He went outside to inspect the source of these noises and had seen a limousine parked outside the gate. This was the vehicle previously ordered by Simpson to take him to LA airport to catch the prebooked flight to Chicago. A few minutes later O.J. had appeared and Kaelin had helped Allan Park, the chauffeur, to load some bags into the vehicle. O.J. had insisted on holding onto a small black bag.

Allan Park later testified that he had been instructed to arrive at Rockingham no later than 10:45. He had arrived early and first called on the buzzer at 10:40. He received no answer. At 10:50 he spotted a tall, well-built, black man who had hurried up to the house from the Rockingham gate. He had tried the buzzer again and had spoken with Simpson who came down 10 minutes later carrying a bag. Park testified that Simpson was sweating and that he requested that the air-conditioning in the limousine be turned on. Park also testified that it was a cool night.

Furhman returned to the house with the news that he had found a blood-stained right glove in a dark narrow walkway between the bungalows.[498] He had already started the search of this secondary scene. There were blood spots leading out of the west gate into Rockingham. Other red marks were present inside the Bronco on the driver's door and the console near the passenger's side. Another trail of blood spots led up to the front door of the house (see Figure 3.10).

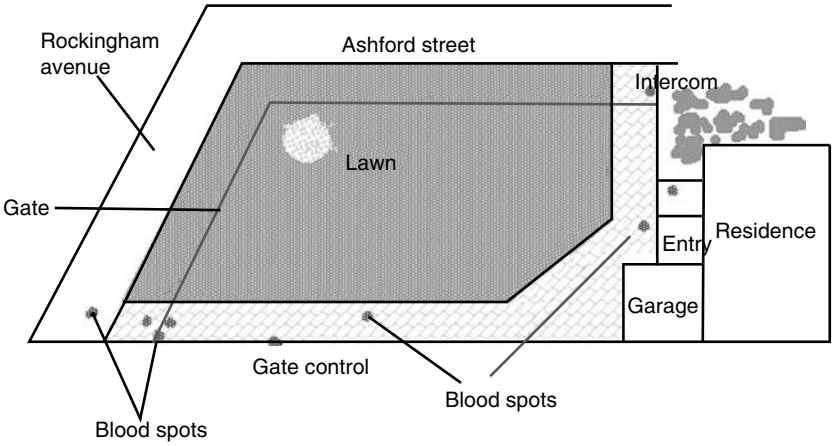

Figure 3.10 The layout of part of Mr. Simpson's Rockingham residence. Reproduced with kind permission from Professor Douglas Linder, University of Missouri Kansas City School of Law.

Detective Vannatter had instructed Furhman to drive to Bundy to check whether the glove at Rockingham matched the one beside the bodies. Furhman did this and then returned to Rockingham. Later the defense would raise the suggestion of the deliberate planting of evidence. *These two episodes certainly create the potential for cross contamination and as such these actions were inviting criticism.*

At this time the detectives had called O.J. at his hotel in Chicago. They reported that his reaction to the news was puzzling, in that he did not ask for details of the deaths. *Highly subjective comments like this are of debatable value. They are unlikely to be admissible in court, nor should they be admissible.*

At 07:10 on June 13, Dennis Fung, an LAPD criminalist, and his assistant Andrea Mazzola, a trainee, arrived at Rockingham. Of subsequent interest in the trial was that they were called to the secondary scene first and not the primary scene at Bundy. *Not every laboratory has the resources to send different teams to different scenes. Many forensic scientists have examined multiple scenes from the same case.*[n] *However, a policy of different personnel for different scenes is clearly advisable especially if one is the crime scene and the other a suspect's domicile.*

At the murder scene a blanket from elsewhere at the scene had been thrown over Nicole's body, presumably by detectives, to protect her from photographers. The motivation was to allow Nicole dignity in death, but the

[n] John Buckleton admits that on more than one occasion he has examined multiple scenes of the same case.

evidential implications had not been well thought through. *Many scene of crime investigators now carry sterile plastic coverings for purposes such as this.*

Later that day O.J. returned from Chicago. He was detained and led to Bundy in handcuffs. This was recognized as improper treatment, and he was released. Detective Vannatter, while unlocking the handcuffs, noted that the middle finger of Simpson's left hand was bandaged. Simpson — reported by detectives as confused — stated that he had cut himself in LA while retrieving his cellphone from his Bronco vehicle. He had reopened the wound in Chicago on a broken glass in the sink. When his hotel room at O'Hare Plaza Hotel was checked, a broken glass was present in the bathroom sink.

At this time, a blood sample had been taken from O.J. and passed to Vannatter and then to Fung at Rockingham. *It has been questioned why the reference sample was taken to the scene and not directly to the laboratory.* Fung and Mazzola had by now bagged and tagged a pair of navy blue bloodstained socks found in the master bedroom at Rockingham.

The 90-minute chase of O.J.'s Ford Bronco on the 17th of June was viewed by an estimated 95 million people. The vehicle was televised driving slowly down LA freeway 405 followed by numerous police vehicles. Al Cowlings, O.J.'s friend and former teammate, was at the wheel. Simpson had a .357 Magnum pressed to his own head in the back seat. In a bag were $US8000, his passport, a fake moustache, and beard. Earlier that day he had failed to appear for arraignment on charges of double murder. Eventually the vehicle had driven sedately back to Rockingham. The following day Mr. Simpson was charged with double murder.

3.5.1 The Evidence

Hairs had been found on Mr. Goldman's shirt and inside the knit cap. These were described in evidence as consistent with having come from O.J. Hairs on the glove found at Rockingham (the Rockingham glove) were consistent with having come from Nicole and Ronald.

Fibers in the Ford Bronco matched fibers on the Rockingham glove and the Bundy knit cap. Blue/black cotton fibers on Ronald's shirt matched the socks that had been found in O.J.'s bedroom. Cashmere fibers from the knit hat matched the glove lining. One glove with this type of lining was at the scene anyway, so the finding of the fibers was not *per se* a connection between Bundy and Rockingham.

The dark brown leather, cashmere-lined, size extra-large gloves had been manufactured by Aris Gloves. This style of glove had only been sold by Bloomingdale's in New York City. Between 1989 and 1992, 240 pairs had been sold, two of these, on December 20, 1990, to Nicole. Photographs were produced of Mr. Simpson wearing gloves of this type in 1993 and 1994. Richard

Rubin, the vice president and general manager of Aris Gloves, testified that he had measured Mr. Simpson's hands as size extra-large.

The police had surmised that the single set of shoeprints at Bundy implied a single killer. The shoeprints showed a waffle-type pattern and were later identified as Italian-made Bruno Magli shoes selling for US$160. They had been sold in 40 stores across the U.S. and 300 pairs of size 12 shoes had been sold in total. Simpson wore size 12 shoes, but our literature search has been unable to ascertain if the prints at the scene were definitely identified as size 12. Simpson later denied ever owning a pair of Magli shoes. However, a photograph was eventually produced of him wearing this type of shoe at a stadium in New York in September 1993.

Henry Lee visited the Bundy scene 13 days after the murder. He found extra shoeprints on a piece of paper, an envelope, and in photographs of Goldman's blood-soaked jeans. This undermined the prosecution's single-killer premise. William Bodziak demonstrated, using photographs taken on the 13th of June, that the extra shoeprints were not there on the day after the murder. Presumably they had occurred after the scene was released. Lee's version of these events does not appear in his book (but he did present photographs of the shoeprints and marks). He has offered to provide his view by correspondence, but it was not available at the time of writing. Without having heard his response, it would be unwise to draw a conclusion.

The DNA profiles of 45 bloodstains were typed and subsequently presented in court. In many cases, these stains were divided and analyzed by two or three separate laboratories. Only the most superficial summary of the results of the typing is presented here. The most important results are considered below.

DNA on the Rockingham glove was consistent with being a mixture of DNA from Mr. Simpson and the two victims. In total, 11 subsamples from this glove were typed. The most extensively typed subsample, item 9:G3, taken from the inside back of the ring finger is discussed in detail. This subsample was typed at eight RFLP and two PCR loci and was found to match Ronald Goldman. Other subsamples on this glove were found to match O.J. Given the subsequent "planting" defense, the presence of blood matching O.J. on the glove is of interest.

Samples of blood, items 47–50 and 52, had been taken from what became known as the Bundy walk. These matched O.J. The samples were taken by Fung and Mazzola on the 13th before O.J.'s blood had been sampled. The first PCR result became available on the 14th. Tampering, if it occurred, had to occur in this window. The five control samples for this batch of items had been unaffected by contamination. Item 52 was the most fully typed, portions of the analysis having been done at one or more of the Los Angeles Police Department laboratory, California Department of Justice, and

Cellmark. Eventually this item was typed at a total of five RFLP and seven PCR loci. The match probability for the PCR loci was estimated as 1 in 240,000 and 1 in 170 million for the VNTR loci. To many this was the strongest evidence.

The defense argued that the items had been incorrectly stored in plastic bags in a hot truck. The parent DNA had completely degraded and the result matching O.J. had come from contamination in the laboratory, allegedly from Mr. Yamauchi's gloves as he had prepared the samples. The defense further argued that the control samples could not be relied on in such a laboratory. A defense explanation of the RFLP results was required, but never given. The RFLP technique requires much more DNA to obtain a result, typically from a stain about the size of a quarter; hence, to explain O.J.'s profile being present, it is necessary to posit gross contamination. Spot 49 (but not 52) in the sequence of five spots had also been tested by conventional serological methods. These would also require the grossest of contamination to register a false result. The match probability for these serological tests was approximately 1 in 200.

The blood on the Rockingham socks was consistent with having come from Nicole. This blood was typed at 14 RFLP and 7 PCR loci. The RFLP match produced a match probability[o] estimate of 1 in 4.4 billion for the Cellmark RFLP set and 1 in 4×10^{10} for the California DOJ set. The PCR result was 1 in 45,000. The two RFLP numbers cannot be simply multiplied as they share two loci but, as Weir points out, numbers are barely necessary.

The defense presented considerable evidence to support their "planting" suggestion. The blood on the socks had not been noticed by Fung when he collected them on June 13, by the defense when they examined them on June 22nd, nor by an LAPD criminalist doing an evidence inventory on June 29. The defense presented evidence that EDTA, a substance used as a preservative for the blood sample tubes drawn from people, was present in the sample of blood recovered from the sock, suggesting that it may have come from Nicole's reference sample. The FBI disputed this finding.

Stains were collected from the rear gate at Bundy on July 3rd and matched O.J. Fung had presumably overlooked these. These were typed by the LAPD using RFLP producing a 1 in 57 billion match probability and by the California DOJ using PCR producing a 1 in 520 match probability. Due to the late collection of this sample, it came under attack as potentially planted. The question of whether these stains contained EDTA and why they were in better condition than the samples taken much earlier was hotly debated.

[o] Quoting the most common result calculated from the five databases used.

Stains were collected from the Bronco on June 14 and later on August 26, over two months after the murders. These stains were consistent with having come from O.J. Simpson, Nicole, or mixtures of blood from O.J. and Nicole, from O.J. and Ronald, or from all three. Various defense arguments weakened much of the probitive value of these findings. Had the stain recovered from the partial shoeprint on the carpet in the Bronco which matched Nicole been transferred by Furhman who had travelled to Rockingham from the Bundy scene? The controls had failed for item 31, which was consistent with having come from O.J. and Ronald. According to the defense, the stains collected on August 26 had been planted and this supposition was bolstered by the fact that a theft had occurred from the vehicle while it was in police custody, reinforcing the view that it was not securely stored.

Match probability statistics were produced for each of these bloodstains and many others.[835,838] One of the authors, John Buckleton, was working with Bruce Weir at this time.

> My part in the saga involved the statistics and the statistics themselves were barely central. Weir and I had advised Cellmark on their data and processes and they presented the statistics we recommended. Weir had presented evidence himself. I was his assistant and my part was repeating his calculations, in the U.S. initially and later in the U.K., after I relocated due to contract obligations. I am the colleague he refers to later, along with Richard Pinchin, Steve Knight, and Ian Evett. I reproach myself for not being in LA and being of more use in the checking.

At this time, match probabilities were still calculated using the product rule except for the "Cellmark wrinkle." This was used at one locus that had failed independence testing. At this locus, the observed genotype probabilities were used. A 99% confidence interval for the match probability was estimated by bootstrapping. All the match probabilities were very small.

> Should we just say that it was O.J.'s blood at Bundy or Nicole's blood on the sock? Weir and I debated it. In the end we didn't. This moral high ground led to a complex report. I still find it hard, today, to amalgamate the information from all the different items and different laboratories. At the time we were unsure whether or not we should multiply the results for different loci from different laboratories where the databases and protocols were different and where independence testing of the various loci between the different laboratories had not been done. We were expecting a severe challenge.[242,415,457,583,767]

I would do things differently now. I routinely use the subpopulation correction and appropriate values for θ. I would still apply sampling error estimation but use the Bayesian posterior rather than the bootstrap. I have no qualms about multiplying results from different laboratories where independence testing had not been done. This latter is largely because I have abandoned any faith that independence testing informs at all about the population genetic model.

3.5.2 The Trial

The trial lasted 133 days, produced 50,000 pages of transcript, called 126 witnesses, and produced 857 pieces of evidence. The defense team, eventually dubbed the "dream team," included Robert Shapiro, Barry Scheck, Johnny Cochran, Peter Neufeld, and William Thompson. Appearing for the prosecution were Marcia Clark, Christopher Darden, Rockne Harmon, and George "Woody" Clark.

The defense hired a jury consultant who found that black, middle-aged women were Mr. Simpson's strongest supporters. Of the 200 African-Americans polled, 44% stated that they had been treated unfairly by the LAPD at least once.[759] The jury included eight blacks, most of them middle-aged women.

Forensic scientists should not allow themselves a view on guilt or innocence. But some of the evidence looked strong. The defense soon undermined much of that.

Detective Furhman was questioned about racism:

Bailey: "You say under oath that you have not addressed any black person as a nigger or spoken about black people as niggers in the past 10 years, Detective Furhman?"

Furhman: "That's what I'm saying."

Later, on September the 5th, a 10-year-old set of tapes made by a North Carolina writer researching racism in the LAPD was played. Furhman could be heard using the word "nigger." Worse, the tapes were littered with gloating admissions that he and other officers had often planted evidence on suspects to secure convictions. There were 42 instances of "nigger" and 18 instances admitting participation in police misconduct in order to incarcerate criminals, including planting evidence. Furhman bragged about stopping interracial couples for no reason, he spoke of his desire to put black people in a pile and burn them, and that he was against having women in the police force because they would not engage in cover-ups.[507] On September 6, Furhman invoked his 5th Amendment rights.

U.S.A. Today and Gannett News Service had previously published a survey from legal and media databases itemizing 85 instances since 1974 of prosecutors knowingly or unknowingly using tainted evidence.[455]

Furhman was not the only detective to come under scrutiny. Detective Vannatter's statement that O.J. was not a suspect on the night of the 12th or the morning of the 13th stretches credibility. He also stated that he had entered Simpson's home without a warrant because of the risk that there was another victim. Vannatter had access to the blood sample from O.J. taken on the 13th that had been handed to him. He had carried it around rather than logging it as police procedures required.[507] The suggestion that O.J.'s blood was planted was strengthened by the "missing" 1.5 ml. Thano Peratis had testified that he had drawn 7.9–8.1 ml. In all, 1 ml was used for DNA testing and the toxicology department measured the remainder on receipt in their section as 5.5 ml. Peratis, by this time too ill to come to court, altered his testimony in a video. This process denied the defense the right of cross examination.

Mr. Fung's testimony lasted three weeks. He had 11 years of forensic experience and had examined 500 scenes. He was questioned about the blanket used to cover Nicole (there was never a suggestion that he had personally placed it over Nicole). Could hairs from the blanket have transferred onto Nicole? He was shown a crime scene photo with his hand ungloved when it should have been gloved. He was questioned about taking only representative samples from the Bronco and the incorrect placing of blood samples into plastic bags where they could deteriorate.

Mazzola, Fung's trainee assistant, was cross examined by Neufeld. She had collected most of the blood samples without supervision. Videotape showed her resting a hand on a dirty footpath, wiping tweezers with a dirty hand, and dropping several blood swabs.

Evidence was produced that the blood on the socks had occurred by "compression transfer," implying that the blood had not got there while O.J.'s foot was inside the sock. There was also the disputed finding of EDTA in the blood from the sock.

Finally, Darden asked Simpson to put on the gloves. To guard against contamination and hazard to Simpson, he donned latex gloves, and then the leather glove. Simpson stated: "They're too tight."

The RFLP technology was not seriously questioned by the defense.

Mullis, the Nobel Prize winning inventor of PCR, stated that he felt the technology was not ready for forensic application. Listing his interests as drug taking, womanizing, and surfing, he was eventually not called by the defense but other witnesses more than adequately spoke of the contamination risks.

Alan Dershowitz, who advised the defense, stated on TV that the probability of a known wife beater actually killing his wife was very small (1/10,000). This statement is somewhat misleading as pointed out by Good.[360,361] Let

B: the event that a man beats his wife.
M: the wife was murdered.

The statistic quoted on TV was close to $\Pr(M \mid B) = 0.0001$.

But the actual evidence is that Mr. Simpson beat his wife AND his wife was murdered. We are interested in the probability that Mr. Simpson is the murderer GIVEN that Nicole was murdered AND Mr. Simpson had beaten her. Let the event that Mr. Simpson is the murderer be *G*. We require $\Pr(G \mid M, B)$, which is quite different to $\Pr(M \mid B)$. Hibbert[405] and Good[360,361] give this as approximately 0.33 revised later to approximately 0.9.

Professor Weir had either advised others or produced much of the statistics himself. In the end, working in his hotel room he produced three and four person calculations for the mixtures in the Bronco.

> The "hard times" referred to in the title of this column apply to what happened next. The time for my scheduled testimony was moved forward two weeks, and I was called to Los Angeles before completing my mixture calculations. I was able to extend my computer program there to handle three unknown contributors instead of the two that had ever been considered before, but was unable to fax my results to colleagues in England for checking because the hotel's computer would not recognize a change in area codes in the United Kingdom. On the afternoon of Thursday June 22, Judge Ito ordered me to perform additional calculations for four unknown contributors before I could testify the following morning! Another late night session with my lap-top computer in a hotel room, and no opportunity for careful checking. In my written report to both prosecution and defense it was obvious that I had left out a term in the calculations — a term that I had correctly included in the calculations I did in my office during normal waking hours. Reviewers of a scientific paper would have noted such an inconsistency and simply called for a correction, but opposing lawyers in a trial are free to use such errors to discredit an expert. Never mind that the errors concerned only a very small number of the calculations, and did not alter the overwhelming evidentiary strength of the matching DNA profiles in all those bloodstains which came from only one person. Subsequently I have developed the algebraic treatment that circumvents the need for those hurried computer calculations…. I do not believe that statisticians should agree to perform detailed analyses in hotel rooms, especially if they are going to be on national TV the next day. (Weir[837] reprinted with permission

> from Stats. © 1996 by the American Statistical Association. All rights reserved)

Neufeld cross examining Weir: "The numbers on the board are biased against Mr. Simpson, isn't that correct?"

Weir: "As it turns out, it looks that way."

The number in question was for a mixed stain on the Bronco steering wheel (item 29). The relevant number went from 1 in 3900 to 1 in 1600. Weir had sighted the error himself, made the correction himself, and put the matter before the court. The complexity of four person calculations was substantial.

> I could not repeat the calculations by hand and Steve Knight and Richard Pinchin had to write software to enumerate the large number of possibilities. It took us a long time to repeat Weir's calculation and the relevant exchanges in court were over before we had done this. In the context of the trial the observation that we were fallible counted more than the number itself. To me, of course, this is not news but it does emphasise the value of independent checking. [John Buckleton]

The profiles from item 29, the Bronco steering wheel and the three reference samples, are given in Table 3.9. The 1.3 allele was not observed in item 29; hence, putatively, Ronald was excluded. This implied the presence of an unknown DNA source. However, the spot from the 4 allele was weak and the issue of whether the 1.3 allele had "dropped out" arose. If we accept that the 4 allele is not from Ronald Goldman, then this is the only allele out of 400 from 45 stains not included in one of the principals' profiles.

The press statements were not flattering:

LA Times[578]

> Dry as sand and just as digestible. (Peter Arenella, UCLA law professor)

Table 3.9 Profiles Considered from Item 29

Locus	Item 29	OS	NB	RG
DQα	1.1,1.2,4	1.1,1.2	1.1,1.1	1.3,4
LDLR	AB	AB	AB	AB
GYPA	AB	BB	AB	AA
HBGG	ABC	BC	AB	AA
D7S8	AB	AB	AB	BB
Gc	ABC	BC	AC	AA

Reproduced from Weir[835] with the kind permission of Nature and Professor Weir.

> More mind-numbing statistics of all sizes with little real meaning to the case, even assuming jurors had any clue about their significance. (Myrna Raeder, Professor of Law, Southwestern University)

In the end, the matter was settled largely on other considerations. The prosecution summation included emotive sections such as

> … it is because he hit her in the past. And because he slapped her and threw her out of the house and kicked her and punched her and grabbed her around the neck … and it's because he used a baseball bat to break the windshield of her Mercedes back in 1985. And it's because he kicked her door down in 1993 … It's because of a letter he wrote her … June the 6th, talking about the IRS. It's because he stalked her … and the fuse is burning. … the fuse is getting shorter, the fuse is getting shorter, and there is about to be an explosion, he is about to lose control, and he is about to lose control like he did on those earlier occasions. And sure he didn't kill her on those earlier occasions in October '93 or in 1989. But that was then and back then the fuse was a lot longer. But now the fuse is way short and it is awfully short … . how do we evaluate this, when a man takes a baseball bat to his wife's car and beats the "F" out of it? If nothing else, it sends a message to her. It instills fear, wouldn't you agree? And would you agree it suggests to her that this can happen to you, that maybe you'll be next? That fuse is burning. It's burning in 1985 … the fuse is lit. It's burning, but it's a slow burn. (Darden, closing argument, reprinted with kind permission from Cotterill[197])

Perhaps the best metaphor from the defense alluded to the glove in particular and the evidence in general: "If it doesn't fit, you must acquit."[197] Mr. Simpson was acquitted on October 3, 1995.[835]

Thagard[759] has studied possible lines of reasoning by which the jury may have reached this verdict. He mentions the inference from Nicole's history of cocaine use that drug dealers may have been involved.

Bayesian inference in the hands of Thagard[759] and JavaBayes gives a posterior of 0.72 that Mr. Simpson was guilty and 0.29 to the alternative that drug dealers were the killers. It also assigns a posterior of 0.99 to the proposition that the LAPD framed Mr. Simpson. 0.72 is well below our subjective level for "beyond reasonable doubt," and in our opinion is entirely consistent with acquittal. Three of the jurors, Cooley, Bess, and Rubin-Jackson, described their conclusions as based on reasonable doubt. "I'm sorry, O.J. would have had to go if the prosecution had presented the case differently, without a doubt. As a black woman it would have hurt me. But as a human being, I would have to do what I had to do." (Juror Carrie Bess)[759]