A Framework for Interpreting Evidence

2

JOHN BUCKLETON[a]

Contents

This book is intended as a discussion of the interpretation of DNA evidence. However, there is nothing inherently different about DNA evidence that sets it qualitatively aside from all other forensic evidence or even all evidence.[167,663] Hence, it is important that DNA evidence is considered as one form of evidence and not as something completely separate. We come immediately to the issue of setting evidence into a framework that is appropriate for court. This has been the subject of entire books by more informed authors,[659] but it is by no means settled. The issue revolves around a basic contrast: the tools best fitted to interpret evidence coherently are also those that appear to be most problematic to explain to a jury or layperson. Does the

[a] I acknowledge many valuable discussions over the years with Drs Christopher Triggs and Christophe Champod, which have contributed to material present in this chapter.

system used by the scientist also have to be the one presented in court? This is a question that is only just beginning to be asked, let alone answered. Parts of this section follow Triggs and Buckleton[783] reproduced with the kind permission of Oxford University Press.

The interpretation of the DNA results has been a key area for debate in the DNA field ever since the inception of this form of evidence.

> The statistical interpretation of DNA typing results, specifically in the context of population genetics, has been the least understood and, by definition, the most hotly debated issue of many admissibility hearings. The perceived incomprehensibility of the subject, has led to a recalcitrance of the judicial system to accept DNA typing.[678,679]

This statement by Rudin and Inman is not only true but is also very interesting. DNA evidence is actually much simpler and more extensively studied than most other evidence types. Many evidence types, such as toolmarks and handwriting comparison, are so complex that at present they defy presentation in a numerical form. Numerical assessment is attempted in glass and fiber evidence in New Zealand and the U.K., but the issues in both these fields are far more complex than in the DNA field. It is the very simplicity of DNA evidence that allows it to be presented numerically at all. And yet, as Rudin and Inman point out, there is still much debate about how to present this evidence.

It could be argued that the presentation of scientific evidence should bend to conform to the courts' requirements. Indeed a court can almost compel this. There have been several rulings[b] on this subject by courts, which have been used to argue for or against particular approaches to the presentation of evidence. An instance of this could be the Doheney and Adams ruling.[201] More specifically the Doheney and Adams ruling has been, I believe erroneously, read as arguing against a Bayesian approach and for a frequentist approach (discussed further later).[c] However, a fairer and more impartial appraisal of the various methods offered for interpretation should proceed from a starting point of discussing the underlying logic of interpretation. Only as a second stage should it be considered how this logic may be presented in court or whether the court or jury have the tools to deal with this type of evidence. There is little advantage to the situation "wrong but understood."[658,660]

[b] For reviews of some court cases in Australia, New Zealand, the U.K. and the U.S., see References 319, 329, 401, 649, 660, 661, 662, and 663.

[c] Robertson and Vignaux[663] give a more eloquently worded argument in support of this belief.

> To be effective in the courtroom, a statistician must be able to think like a lawyer and present complex statistical concepts in terms a judge can understand. Thus, we present the principles of statistics and probability, not as a series of symbols, but in the words of jurists.[362]

What is proposed in this chapter is to consider the underlying claims of three alternative approaches to the presentation of evidence. These will be termed the frequentist approach, the logical approach, and the full Bayesian approach.[429] The first of these terms has been in common usage and may be familiar.[585] I have adopted a slightly different phrasing to that in common usage for the second and third approaches and this will require some explanation. This will be attempted in the following sections. The intention is to present the merits and shortcomings of each method in an impartial way, which hopefully leads the reader to a position where they can make an informed choice. Juries may misunderstand any of the methods described, and care should be taken over the exact wording. In fact, it is clear that care must be taken with all probabilistic work and presentation.[35,101,287,472,473,474,475,476,477,493,503,554,664,764,766,769] One famous statistician's evidence has been described as "like hearing evidence in Swahili, without the advantage of an interpreter."[604]

Comparisons of the potential impact on juries of the different methods have been published.[299,365,754,755] It is necessary to countenance a situation in the future where the desirable methods for interpretation of, say, a mixture by simulation are so complex that they cannot realistically be explained completely in court.

It is important that the following discussion is read without fear of the more mathematical approaches as this fear wrongly pressures some commentators to advocate simpler approaches. It is probably fair for a jury to prefer a method for the reason of mathematical simplicity, but it would be a mistake for a scientist to do so. Would you like your aircraft designer to use the best engineering models available or one that you can understand without effort?

2.1 The Frequentist Approach

At the outset it is necessary to make clear that the use of the frequentist approach in forensic science is related to, but not identical to the frequentist approach in probability theory.[519,652] The frequentist approach in forensic science has never been formalized and hence is quite hard to discuss. It appears to have grown as a logical framework by a set of intuitive steps. There are also a number of potential misconceptions regarding this approach,

which require discussion and will be attempted. To begin, the approach will be subdivided into two parts: the coincidence probability and the exclusion probability. A discussion of "natural frequencies," a concept introduced by Gigerenzer, will follow.[336]

2.1.1 Coincidence Probabilities

For this discussion, it is necessary to attempt to formalize this approach sufficiently. Use will be made of the following definition:

The coincidence approach proceeds to offer evidence against a proposition by showing that the evidence is unlikely if this proposition is true. Hence it supports the alternative proposition. The less likely the evidence under the proposition, the more support given to the alternative.

This is called the coincidence probability approach because either the evidence came from, say, the suspect or a "coincidence" has occurred.

There are many examples of evidence presented in this way:

- "Only 1% of glass would match the glass on the clothing by chance."
- "It is very unlikely to get this paint sequence match by chance alone."
- "Approximately 1 in a million unrelated males would match the DNA at the scene by chance."

We are led to believe that the event "match by chance" is unlikely and hence the evidence supports the alternative. At this stage let us proceed by assuming that if the evidence is unlikely under a particular hypothesis, then this supports the alternative.

This is strongly akin to formal hypothesis testing procedures in statistical theory. Formal hypothesis testing would proceed by setting up the hypothesis usually called the null, H_0. The probability of the evidence (or data) is calculated if H_0 is true. If this probability is small (say less than 5 or 1%), then the null is "rejected." The evidence is taken to support the alternative hypothesis, H_1.[305,579,612]

To set up a DNA case in this framework, we could proceed as follows. Formulate the hypothesis, H_0: the DNA came from a male not related to the suspect. We then calculate the probability of the evidence if this is true. We write the evidence as E, and in this context it will be something like:

E: The DNA at the scene is type α.

We assume that it is known that the suspect is also type α. We calculate the probability, Pr, of the evidence, E, if the null hypothesis H_0 is true, $\Pr(E|H_0)$. The vertical line, or conditioning sign, stands for the word "if" or "given."

Assuming that about 1 in a million unrelated males would have type α, we assign $\Pr(E|H_0)$ as 1 in a million. Since this is a very small chance, we

assume that this evidence suggests that H_0 is not true and hence is support for H_1. In this context, we might define the alternative hypothesis as:

H_1: The DNA came from the suspect.

Hence in this case, the evidence supports the hypothesis that the DNA came from the suspect. Later we are going to need to be a lot more careful about how we define hypotheses.

Hypothesis testing is a well-known and largely accepted statistical approach. The similarity between the coincidence approach and hypothesis testing is the former's greatest claim to prominence.

2.1.2 Exclusion Probabilities

The exclusion probability approach calculates and reports the exclusion probability. This can be defined as the probability that a random person would be excluded as the donor of this DNA, or the father of this child, or a contributor to this mixture. The details of these calculations will be discussed later. Again, the formal logic has not been defined; hence, it will be attempted here.

The suspect is not excluded. There is a probability that a random person would be excluded. From this it is inferred that it is unlikely that the suspect is a random person. Hence this evidence supports the alternative proposition that the suspect is the donor of the DNA. The higher the exclusion probability, the more support given to the alternative.

Examples are again common. For instance, the three phrases given previously can be reworked into this framework:

- "99% of windows would be excluded as a source of this glass."
- "It is very likely that a random paint sequence would be excluded as matching this sample."
- "Approximately 99.9999% of unrelated males would be excluded as the source of this DNA."

An advantage of the exclusion probability approach is that it can be easily extended beyond these examples to more difficult types of evidence such as paternity and mixtures:

- "Approximately 99% of random men would be excluded as the father of this child."
- "Approximately 99% of random men would be excluded as a donor to this mixture."

It was stated previously that the use of the frequentist approach in forensic science is related, but not identical, to the frequentist approach in probability theory. There are two common definitions of probability. These are

called the frequentist and the subjectivist definitions. It is not necessary to discuss these differences in any length here, as they have long been the cause of deep discussion in both philosophy and the theory of probability. Briefly, the frequentist approach treats probability as the expectation over a large number of events. For instance, if we roll a dice many times we expect about 1/6 of these rolls to be a "6." The subjectivist definition accepts that probability is a measure of belief, and that this measure will be conditional both on the information available and on the person making the assessment.[756] However, both the coincidence approach and the exclusion probability approach can be based on either frequentist or subjectivist probabilities. Proponents of the Bayesian or subjectivist school of probability criticize the frequentist definition. However, it is unfair to transfer this criticism of a frequentist probability to the frequentist approach to forensic evidence.

The coincidence and the exclusion probability approach do appear to be simple and have an intuitive logic that may appeal to a jury. Thompson[767] argued for their use in the O.J. Simpson trial apparently on the basis that they were conservative and more easily understood while accepting the greater power of likelihood ratios.

2.1.3 Natural Frequencies[d]

More recently, the argument has been taken up by Gigerenzer arguing that "*to be good it must be understood.*" He argues persuasively for the use of "natural frequencies." To introduce this concept, it is easiest to follow an example from Gigerenzer.[336]

> The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 10 of these men have a DNA profile that is identical with the trace recovered from the crime scene. If a man has this profile it is practically certain that a DNA analysis shows a match. Among the men who do not have this DNA profile, current DNA technology leads to a reported match in only 100 cases out of 10 million.[e]

Gigerenzer argues from his own research that this approach is more likely to be understood. He quotes that the correct understanding was achieved by 1% of students and 10% of professionals when using conditional probabilities. This rose to 40 and 70%, respectively, when "natural frequencies" were used.

[d] My thanks to Michael Strutt for directing me to this work.
[e] Gigerenzer is referring here to his estimate of error rates.

Of course, Gigerenzer's natural frequencies are nothing more than an example of the defense attorney's fallacy of Thompson and Schumann[769] or the recommendation of the Appeal Court regarding Doheney and Adams.[201,618]

I concede the seductive appeal of this approach. Let us accept at face value Gigerenzer's statement that they are more easily understood. I do, however, feel that this approach hides a number of serious issues.

First consider the assumption that N men could have been the perpetrator. Who is to make this decision? One would feel that the only people qualified and with the responsibility of doing this are the judge and jury. They have heard the non-DNA evidence and they can decide whether or not this defines a pool of suspects. Moreover, are we to assign an equal prior to all these men? Gigerenzer's approach has a tendency toward assigning equal priors to each of these men and to the suspect. This is a tenable assumption in some but not all circumstances. Essentially we have a partition of the population of the world into those "in" the pool of suspects and those "out" of it. Those "in" are assigned a prior probability of $1/N$. Those "out" are assigned a prior of 0.

What are we to do when the product of the match probability and the pool of possible suspects is very small? Let us take the case given above but reduce the match probability from 1 in a million to 1 in 10 million. This would lead to:

The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 1 of these men has a DNA profile that is identical with the trace recovered from the crime scene.

The witness will have to take great care that the jury understand this statement. There is a risk that they may assume that the suspect is this one man. What is needed is to explain that this is one man additional to the suspect and even then it is an expectation. There may be one man additional to the suspect, but there may also be 0, 2, 3, or more.

Let us take this case and reduce the match probability even further to 1 in a billion. This would lead to:

The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 0.01 of these men have a DNA profile that is identical with the trace recovered from the crime scene.

This will take some care to explain to the jury. Now suppose that the suspect has one brother in the set of 10 million men.

The expert witness testifies that there are about 10 million unrelated men and one brother who could have been the perpetrator. Approximately 0.01 of the unrelated men and 0.005 of the brother have a DNA profile that is identical with the trace recovered from the crime scene.

Taking the example further:

The expert witness testifies that there are about 10 million unrelated men and one brother who could have been the perpetrator. Approximately 0.002 of the unrelated Caucasian men, 0.004 of the unrelated African Americans, 0.004

of the unrelated Hispanics, and 0.005 of the brother have a DNA profile that is identical with the trace recovered from the crime scene.

If we accept the suggestion that it is more understandable, then it may have a use in those very simple cases where there is a definable pool of suspects, relatedness is not important, the evidence is certain under the prosecution hypothesis, and the product of the match probability times N is not small.

Outside this very restricted class of case, I would classify it in the "understood but wrong"[f] category even when it is understood. I really do doubt the usefulness of this approach. It is very difficult to see how to accommodate relatives, interpret mixtures, and report paternity cases within this framework. Gigerenzer has also subtly introduced the concept of 10 million replicate cases all with the same probability of error. This may be an acceptable fiction to lure the jury into a balanced view, but it would take a lot of thinking to reconcile it with my own view of probability. Even if we accept Gigerenzer's statement that natural frequencies are more easily understood and we decided to use this presentation method in court, it is important that forensic scientists think more clearly and exactly about what a probability is, what constitutes replication, and how probabilities may be assigned.

2.2 The Logical Approach

"We are all Bayesians in real day life." Bruce Budowle.[119]

"Bayes's theorem is a fundamental tool of inductive inference." Finkelstein and Levin.[300]

Frustrations with the frequentist approach to forensic evidence have led many people to search for alternatives.[105,258] For many, these frustrations stem from discussing multiple stains, multiple suspects, or from trying to combine different evidence types.[652,656] The foremost alternative is the logical approach (also called the Bayesian approach).[257,490,500,516,517,518] This approach has been implemented routinely in paternity cases since the 1930s.[255] It is however only in the later stages of the 20th century that it made inroads into many other fields of forensic science. It now dominates forensic literature, but not necessarily forensic practice, as the method of choice for interpreting forensic evidence.[6,8,170,171,173,214,334,585,659,663] Bär[47] gives an elegant review.

Let:

H_p be the hypothesis advanced by the prosecution,
H_d be a particular hypothesis suitable for the defense,
E represent the evidence, and
I represent all the background evidence relevant to the case.

[f] It is not the only error in this section by Gigerenzer. Professor Weir did not report likelihood ratios in the O.J. Simpson case and most laboratories and all accredited ones do undertake external QA trials.

The laws of probability lead to

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)} = \frac{\Pr(E|H_p,I)}{\Pr(E|H_d,I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \qquad (2.1)$$

This theorem is known as Bayes's theorem.[53] A derivation appears in Box 2.1. This theorem follows directly from the laws of probability. It can therefore be accepted as a logical framework for interpreting evidence.

To understand the workings of this formula, it is necessary to understand the workings of the conditioning sign. This is usually written as | and can be read as "if" or "given." This concept is little understood and is typically not taught well. The reader unfamiliar with it would be advised to work through the examples given in Evett and Weir.[267] A brief discussion is given in Box 2.2.

Equation (2.1) is often given verbally as

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds} \qquad (2.2)$$

The prior odds are the odds on the hypotheses H_p before DNA evidence. The posterior odds are these odds after DNA evidence. The likelihood ratio tells us how to relate these two. This would seem to be a very worthwhile thing to

Box 2.1 A Derivation of Bayes's Theorem

The third law of probability states:

$\Pr(a \text{ and } b|c) = \Pr(a,b|c) = \Pr(a|b,c)\Pr(b|c) = \Pr(b|a,c)\Pr(a|c)$

Rewriting this using H_p, H_d, E, and I

$\Pr(H_p,E|I) = \Pr(H_p|E,I)\Pr(E|I) = \Pr(E|H_p,I)\Pr(H_p|I)$

and

$\Pr(H_d,E|I) = \Pr(H_d|E,I)\Pr(E|I) = \Pr(E|H_d,I)\Pr(H_d|I)$

Hence

$$\frac{\Pr(H_p,E|I)}{\Pr(H_d,E|I)} = \frac{\Pr(H_p|E,I)\Pr(E|I)}{\Pr(H_d|E,I)\Pr(E|I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)}$$

Hence

$$\frac{\Pr(H_p|E,I)\Pr(E|I)}{\Pr(H_d|E,I)\Pr(E|I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)}$$

Cancelling $\Pr(E|I)$

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)} \qquad (2.1)$$

do, that is, to relate the odds before consideration of the evidence to those after the evidence. It also tells us how to update our opinion in a logical manner having heard the evidence.

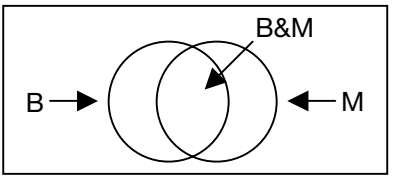
The prior odds, $\Pr(H_p|I)/\Pr(H_d|I)$, represent the view on the prosecution and defense hypothesis before DNA evidence is presented.[g] This view is something that is formed in the minds of the judge and jury. The information imparted to the jury is carefully restricted to those facts that are considered admissible and relevant. It is very unlikely that the prior odds are numerically expressed in the mind of the judge and jury and there is no need

Box 2.2. Conditional Probability

Several definitions are available for conditional probability. One proceeds from the third law of probability:

$$\Pr(a|b)=\frac{\Pr(a,b)}{\Pr(b)}$$

which can be interpreted quite well in set theory. For instance, evaluating $\Pr(a|b)$ involves enumerating the set of outcomes where event b is true and seeing in what fraction of these events a is also true.



Example: In a certain office there are ten men. Three men have beards (event B) and moustaches (event M). A further two have moustaches only. Say we were interested in $\Pr(B|M)$ we find the set of men where M is true: this has five members. Of these, three have beards. Hence $\Pr(B|M)=3/5$.

If we were interested in $\Pr(M|B)$ we find the set of men where B is true: this is three men. Of these, all three have moustaches. Hence $\Pr(M|B)=3/3=1$.[h]

[g] My wording is wrongly implying an order to events such as the "hearing of DNA evidence." In fact, the evidence can be heard in any order. The mathematical treatment will give the same result regardless of the order in which the evidence is considered.[659]

[h] In this simple example, we are making an assumption that each of the men is equally likely to be observed. This assumption may not be true in more general examples, but the principle behind the definition of the conditional probability remains valid.

that they should be numerical.[662,663] Strictly it is not the business of the scientist to form a view on the "prior odds" and most scientists would strictly avoid this (for a differing opinion, see Meester and Sjerps[543] and the subsequent discussion[220]). These odds are based on nonscientific evidence and it is the duty of judge and jury to assess this.[779,807]

The use of this approach typically reports only the likelihood ratio. By doing this the scientist reports the weight of the evidence without transgressing on those areas reserved for the judge and jury. This is the reason why the term "the logical approach"[i] has been used to describe this method. It has also been described elsewhere as "the likelihood ratio" approach. The term that is being avoided is "the Bayesian approach," which is the term used in most papers on this subject, including my own. This term is being avoided because, strictly, presenting a ratio of likelihoods does not necessarily imply the use of the Bayesian method. Most authors have intended the presentation of the likelihood ratio alone without necessarily implying that a discussion of Bayes' theorem and prior odds would follow in court. The intent was to present the scientific evidence in the context of a logical framework without necessarily presenting that framework.

However, the advantage of the logical approach is that the likelihood ratio can be put in a context of the entire case and in a consistent and logical framework. This advantage is somewhat lost if judge, jury, and scientist are reticent to use or even discuss Bayes' theorem in full.

Thompson[767] warns:

> Although likelihood ratios have appealing features, the academic community has yet fully to analyse and discuss their usefulness for characterising DNA evidence.

Pfannkuch et al.[616] describe their experiences teaching this material to undergraduate students:

> Bayes' theorem was the killer. There was an exodus of those mathematically unprepared and math-phobic students who were free to leave the course, supplemented by panic and agonised discussions with those who were trapped by their course requirements.

These professional scientists and teachers persisted and found good methods for teaching even math-phobic students because of the "*wealth of socially important problems*" that are best addressed by Bayes' theorem.

[i] I first had this distinction explained to me by Dr Christophe Champod.

Fenton and Neil[288] argue forcefully that Bayes' theorem is the method of choice for interpreting evidence, while giving the fair criticism that Bayesians have failed in their duty of communication. They quote the fact that many lawyers and other educated professionals misunderstand the subject.

Is there a lesson here? My own experience with practicing forensic scientists is that they can achieve an in-depth understanding of complex mathematical concepts and methods, especially when placed in a good learning environment and supported by colleagues and management. In this regard, I would like to commend the U.K. Forensic Science Service (FSS) practice of secluding scientists during training (in England we used excellent hotels in Evesham and the "Pudding club" somewhere south of Birmingham). The FSS also undertakes basic probability training and is considering putting in place a numerical competency in recruitment. This investment in people is repaid manyfold.

To gain familiarity with Equation (2.2), it is useful to consider a few results. What would happen if the likelihood ratio was 1? In this case, the posterior odds are unchanged by the evidence. Another way of putting this is that the evidence is inconclusive.

What would happen if the likelihood ratio was greater than 1? In these cases, the posterior odds would be greater than the prior odds. The evidence would have increased our belief in H_p relative to H_d. Another way of putting this is that the evidence supports H_p. The higher the likelihood ratio, the greater the support for H_p.

If the likelihood ratio is less than 1, the posterior odds would be smaller than the prior odds. The evidence would have decreased our belief in H_p relative to H_d. Another way of putting this is that the evidence supports H_d. The lower the likelihood ratio, the greater the support for H_d.

It has been suggested that a nomogram may be useful to help explain the use of this formulation. This follows from a well-known nomogram in clinical medicine. Riancho and Zarrabeitia[642] suggest the diagram that has been modified and presented in Tables 2.1 and 2.2. These tables are used by choosing a prior odds and drawing a line through the center of the *LR* value. The posterior odds may then be read directly. For example, assume that the prior odds are about 1 to 100,000 (against) and the likelihood ratio is 10,000,000; then we read the posterior odds as 100 to 1 (on).

The likelihood ratio (*LR*) is a numerical scale. One point can be hinged to words without argument; an *LR* of 1 is inconclusive. Other words may be attached to this scale to give a subjective verbal impression of the weight of evidence.[12,94,174,263,264] This association of words with numbers is subjective and necessarily arbitrary. One such scale used extensively in the FSS is given in Table 2.3.

The question of development of the prosecution and defense hypotheses was introduced above, but was not discussed in any depth. In fact, the defense are under no obligation to offer any hypothesis at all. An early discussion of

Table 2.1 Prosecutor's Nomogram

Prior		Likelihood Ratio	Posterior	
Probability	Odds		Odds	Probability
			100,000,000 to 1	99.999990%
0.001%	1 to 100,000		10,000,000 to 1	99.999989%
0.01%	1 to 10,000	10,000,000,000 1,000,000,000	1,000,000 to 1	99.9999%
0.1%	1 to 1000	100,000,000 10,000,000	100,000 to 1	99.999%
1%	1 to 100	1,000,000 100,000	10,000 to 1	99.99%
9%	1 to 10	10,000 1000	1000 to 1	99.9%
50%	1 to 1	100 10	100 to 1	99%
91%	10 to 1	1	10 to 1	91%
99%	100 to 1		1 to 1	50%

The prior and posterior probabilities associated with these odds are given next to the odds.

Reproduced and amended from Riancho and Zarrabeitia[642] with kind permission of the authors and Springer-Verlag who retain ownership of the copyright.

this appears in Aitken.[8] This is the subject of a large-scale project in the FSS called the Case Assessment Initiative.[193,194,272] The subject warrants separate treatment. Even though it has been introduced under the heading of "the logical approach," the development of propositions is actually universally important to evidence interpretation by any method (see Box 2.3).

2.3 The Full Bayesian Approach

The analysis given under the title of "the logical approach" works well if there are two clear hypotheses aligned with the prosecution and defense positions. However, regularly it is difficult to simplify a real casework problem down to two hypotheses.

To put this in context, consider a relatively simple STR case. We have a stain at the scene of a crime. Call this stain c and the genotype of this stain G_c, following the nomenclature of Evett and Weir.[267] A suspect comes to the attention of the police. Call this person s and the genotype G_s. The genotype of the suspect and the crime stain are found to be the same. We will write this as G_s=G_c.

Table 2.2 Defendant's Nomogram

Prior		Likelihood Ratio	Posterior	
Probability	Odds		Odds	Probability
0.1%	1 to 1000		100 to 1	99%
1%	1 to 100		10 to 1	91%
9%	1 to 10	10	1 to 1	50%
		1		
50%	1 to 1	1/10	1 to 10	9%
		1/100		
91%	10 to 1	1/1000	1 to 100	1%
		1/10,000		
99%	100 to 1	1/100,000	1 to 1000	0.1%
		1/1,000,000		
99.9%	1000 to 1	1/10,000,000	1 to 10,000	0.01%
		1/100,000,000		
99.99%	10,000 to 1	1/1,000,000,000	1 to 100,000	0.001%
			1 to 1,000,000	0.0001%

Reproduced and amended from Riancho and Zarrabeitia[642] with kind permission of the authors and Springer-Verlag who retain ownership of the copyright.

Table 2.3 A Verbal Scale

LR	Verbal Wording	
1,000,000+	Extremely strong	
100,000	Very strong	
10,000	Strong	Support for H_p
1000	Moderately strong	
100	Moderate	
10	Limited	
1	Inconclusive	
0.1	Limited	
0.01	Moderate	
0.001	Moderately strong	Support for H_d
0.0001	Strong	
0.00001	Very strong	
0.000001	Extremely strong	

Box 2.3 Which Way Up?

When we introduced Bayes's theorem, we wrote it as

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d,E|I)}=\frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)} \tag{2.1}$$

Why did we write it this way up? What was wrong with

$$\frac{\Pr(H_d|E,I)}{\Pr(H_p|E,I)}=\frac{\Pr(E|H_d,I)\Pr(H_d|I)}{\Pr(E|H_p,I)\Pr(H_p|I)}\ ?$$

This approach would work just as well. High numbers would be support for H_d, typically the defense hypothesis. Is the reason we defined it with H_p on top an indication of subconscious bias? Is this the reason Balding, Donnelly, and Nichols[42] wrote their *LR*'s up the other way? Were they trying to help us see something?

Under the coincidence approach, this would be the match that is caused by the suspect being the donor of the crime stain or by a coincidence. To make the comparison with hypothesis testing, we would formulate

H_0: The DNA came from a male not related to the suspect.

H_1: The DNA came from the suspect.

We then calculate the probability of the evidence if this is true. Let us write this as $\Pr(G_c|G_s, H_0)$, which can be read as the probability of the genotype of the crime stain if the crime stain came from a person unrelated to the suspect (and the suspect's genotype is G_s). This is often written as f,[j] and taken to be the frequency of the crime genotype (or the suspect's genotype since they are the same). We assume that this frequency is small and hence there is evidence against H_0 and for H_1.

Under the "logical approach," we simply rename these hypotheses:

H_p: The DNA came from the suspect.

H_d: The DNA came from a male not related to the suspect.

We then calculate the probability of the evidence under each of these hypotheses.[5,6,8,9,10,11,37,257,267] $\Pr(G_c|G_s, H_p)=1$ since the crime genotype will be G_c if it came from the suspect who is G_s. Again we take $\Pr(G_c|G_s, H_d)=f$. Hence

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

[j] This term will stand for two concepts in this text. This is unavoidable if we are to align with the published literature. In the context, it stands for the frequency of a profile. However, in population genetics f is often used for the within-population inbreeding parameter. When used in this latter context, it is synonymous with F_{IS}.

which is (typically) very much larger than 1 and hence there is evidence against H_d and for H_p.

But note that the following two hypotheses are not exhaustive:

H_p: The DNA came from the suspect.
H_d: The DNA came from a male not related to the suspect.

What about those people who *are* related to the suspect? Should they be considered? Genetic theory would suggest that these are the most important people to consider, and should not be omitted from the analysis. What we need is a number of hypotheses. These could be:

H_1: The DNA came from the suspect.
H_2: The DNA came from a male related to the suspect.
H_3: The DNA came from a male not related to the suspect.

Now consider H_2. What do we mean by "related"? Obviously there are many different degrees of relatedness. Suppose that the suspect has one father and one mother, several brothers, numerous cousins and second cousins, etc. We may need a multiplicity of hypotheses. In fact, we could envisage the situation where there is a specific hypothesis for every person on earth:

H_1: The DNA came from the suspect.
H_2: The DNA came from person 2, the brother of the suspect.
H_3: The DNA came from person 3, the father of the suspect.
⋮
H_i: The DNA came from person i related in whatever way to the suspect.
⋮
H_j: The DNA came from person j so distantly related that we consider the person effectively unrelated to the suspect.

What we need is a formulation that can handle from three to many hypotheses. Considering the enumeration given above, there would be about 6,000,000,000 hypotheses, one for each person on earth.

This is provided by the general form of Bayes' theorem (derived in Box 2.4).[37,38,42] This states that

$$\Pr(H_1|G_c, G_s) = \frac{\Pr(H_1)}{\sum_{i=1}^{N} \Pr(G_c|G_s, H_i)\Pr(H_i)}. \tag{2.4}$$

This equation is very instructive for our thinking but is unlikely to be directly useful in court, at least in the current environment. This is because the terms $\Pr(H_i)$ relate to the prior probability that the ith person is the source of the DNA. The introduction of such considerations by a forensic scientist is unlikely

Box 2.4

A comprehensive equation has been proposed[37] based on the general formulation of Bayes' rule. Following Evett and Weir:[267] for a population of size N, we index the suspect as person 1 and the remaining members of the population as 2, …, N. We will call the hypothesis that person i is the source of the DNA H_i. Since the suspect is indexed person 1, the hypothesis that the suspect is, in fact, the source of the DNA is H_1. The remaining hypotheses, H_2, …, H_N, are those hypotheses where the true offender is some other person. Before we examine the evidence, each person has some probability of being the offender $\Pr(H_i)=\pi_i$. Many factors may affect this, one of these being geography. Those closest to the scene may have higher prior probabilities while people in remote countries have very low prior probabilities. Most of the people other than the suspect or suspects will not have been investigated. Therefore, there may be little specific evidence to inform this prior other than general aspects such as sex, age, etc. The suspect is genotyped and we will call the genotype G_s. The stain from the scene is typed and found to have the genetic profile G_c, which matches the suspect. The remaining 2, …, N members of the population have genotypes G_2, …,G_N. These 2, …, N people have not been genotyped. We require the probability $\Pr(H_1|G_c, G_s)$. This is given by Bayes's rule as

$$\Pr(H_1|G_c,G_s)=\frac{\Pr(G_c|G_s,H_1)\Pr(G_s|H_1)\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s, H_i)\Pr(G_s|H_i)\Pr(H_i)}$$

Assuming that $\Pr(G_s|H_1)=\Pr(G_s|H_i)$ for all i, we obtain

$$\Pr(H_1|G_s,G_c)=\frac{\Pr(G_c|G_s,H_1)\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s,H_i)\Pr(H_i)}$$

We assume that the probability that the scene stain will be type G_c, given that the suspect is G_s and he contributed the stain, is 1. Hence,

$$\Pr(H_1|G_s,G_c)=\frac{\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s,H_i)\Pr(H_i)} \tag{2.4}$$

$$=\frac{1}{1+\sum_{i=2}^{N}\dfrac{\Pr(G_c|G_s,H_i)\Pr(H_1)}{\Pr(H_i)}}$$

(*continued*)

Box 2.4 (*continued*)

$$= \frac{1}{1+\sum_{i=2}^{N}\frac{\Pr(G_c|G_s,H_i)\pi_1}{\pi_i}}$$

Writing $\pi_i / \pi_1 = w_i$, we obtain

$$\Pr(H_1|G_s,G_c) = \frac{1}{1+\sum_{i=2}^{N}\Pr(G_c|G_s,H_i)w_i}$$

which is the equation given on page 41 of Evett and Weir. Here w_i can be regarded as a weighting function that expresses how much more or less probable the *i*th person is than the suspect to have left the crime stain based on only the non-DNA evidence.

to be permitted in court.[k] However, such an approach may be possible if the court supplies its view of the prior. For instance, the terms "forensically relevant populations"[131] and "relevant subgroup"[201] provide inadvertent references to such a prior. The time may come when courts countenance this type of consideration. We could envisage the situation where a court instructs the witness to consider only the subgroup "Caucasian sexually active males in the Manchester area," which is, in effect, setting a prior of zero outside this group.

In the likely absence of courts providing such priors, it is suggested that this unifying equation should be used to test various forensic approaches and to instruct our thinking. However, there is so much benefit in the use of this equation that research into how it could be used in court would be very welcome.

2.4 A Possible Solution

There is a "halfway house" between the likelihood ratio approach and the unifying equation that has neither been published previously nor tested, but has some considerable merit. Using the same nomenclature as above, we rewrite the likelihood ratio as

$$LR = \frac{\Pr(G_c|G_s,H_p)}{\sum_{i=2}^{N}\Pr(G_c|G_s,H_i,H_d)\Pr(H_i|H_d)} \tag{2.5}$$

where $H_2, \ldots, H_N$ is an exclusive and exhaustive partition of H_d (following Champod,[169] we will call these subpropositions). The advantage of this

[k] Meester and Sjerps[543] argue to the contrary.

approach is that it only requires priors that partition the probability under H_d. There is no requirement for the relative priors on H_p and H_d. This may be more acceptable to a court.

2.5 Comparison of the Different Approaches

The very brief summary of the alternative approaches given above does not do full justice to any of them. It is possible, however, to compare them. In the most simplistic overview, I would state that:

- The frequentist approach considers the probability of the evidence under one hypothesis.
- The logical approach considers the probability of the evidence under two competing hypotheses.
- The full Bayesian approach considers it under any number of hypotheses.

If we turn first to a critique of the frequentist approach, the most damning criticism is a lack of logical rigor. In the description given above, you will see that I struggled to define the frequentist approach and its line of logic with any accuracy. This is not because of laziness but rather that the definition and line of logic has never been given explicitly, and indeed it may not be possible to do so.

Consider the probability that is calculated. We calculate $\Pr(E|H_0)$ under the frequentist view. If it is small, we support H_1.

First note that because $\Pr(E|H_0)$ is small, this does not mean that $\Pr(H_0|E)$ is small. This is called the fallacy of the transposed conditional.[769]

Second note that simply because $\Pr(E|H_0)$ is small does not mean that $\Pr(E|H_1)$ is large. What if it was also small? Robertson and Vignaux[659] give a thought-provoking example adapted here slightly: Consider a child abuse case. Evidence is given that this child rocks and that only 3% of nonabused children rock. It might be tempting to assume that this child is abused since the evidence (R: rocking) is unlikely under the hypothesis (H_0: This child is nonabused). But we may be wrong to do so. Imagine that we now hear that only 3% of abused children rock. This would crucially alter our view of the evidence. We see that we cannot evaluate evidence by considering its probability under only one hypothesis. This has been given as a basic principle of evidence interpretation by Evett and Weir [267]and Evett et al.[281]

The logical flaws in the frequentist approach are what have driven many people to seek alternatives. Fortunately for justice and unfortunately for the advance of logic in forensic science, this flaw does not manifest itself in most simple STR cases. This is because the evidence is often certain under the

alternative H_1. In such cases, the frequentist approach reports f' and the logical approach LR = $1/f$. Critics of the logical approach understandably ask what all the fuss is about when all that is done in simple cases is calculate one divided by the frequency. Other criticisms have been offered. Effectively, these relate to reasonable criticisms of the difficulty of implementation and less reasonable criticisms arising largely from a lack of understanding of the underlying logic.[657,864] This brings us to a critique of the logical approach.

If we start with difficulty of implementation, one reasonable criticism of the logical approach is the ponderous nature of a statement involving a likelihood ratio. Contrast **A** with **B**:

A: The frequency of this profile among unrelated males in the population is less than 1 in a billion.
B: This evidence is more than a billion times more likely if the DNA came from the suspect than if it came from an unrelated male.

Many people would prefer **A** over **B**, and in fact studies have demonstrated that there are serious problems with understanding statements like **B**.[754,755] Some respondents described **B**-type statements as "*patently wrong.*" This is not to imply that there is no prospect of misunderstanding a frequentist statement because there clearly is, but rather to suggest that the likelihood ratio wording is more ponderous and will take more skill and explanation to present.

We next move on to the fact that the very advantage of the "logical approach" is that the likelihood ratio can be placed in the context of a logical framework. This logical framework requires application of Bayes' rule and hence some assessment of priors. However, the legal system of many countries relies on the "common sense" of jurors and would hesitate to tell jurors how to think.[201,618,660] Forcing jurors to consider Bayes's theorem would be unacceptable in most legal systems. It is likely that application of common sense will lead to logical errors, and it has been shown that jurors do not handle probabilistic evidence well. However, there is no reason to believe that these logical errors would be removed by application of a partially understood logical system, which is the most likely outcome of trying to introduce Bayes' theorem into court. If we recoil from introducing Bayes' theorem in court, then the likelihood ratio approach forfeits one of its principal advantages although it certainly retains many others in assisting the thinking of the scientist.

This is not a fatal flaw as likelihood ratios have been presented in paternity evidence since the mid-1900s. In this context, they are typically termed paternity indices and are the method of choice in paternity work.

Inman and Rudin[429] note that: "While we are convinced that these ideas are both legitimate and useful, they have not been generally embraced by the practising community of criminalists, nor have they undergone the refinement that only comes with use over time." This is fair comment from a U.S. viewpoint.

The considerations given above are real issues when applying the logical approach. There are a few more objections that arise largely from a misunderstanding of the underlying logic. These would include criticisms of conditioning on I and H_p and the arbitrariness of the verbal scale. For an elegant discussion, see Robertson and Vignaux.[652]

Forensic scientists are raised in a culture that demands that they should avoid any bias that may arise from ideas seeded into their minds by the prosecution (or anyone else). This has led to the interpretation that they should consider the evidence in isolation from the background facts or the prosecution hypothesis. This idea is a misconception or misreading of the use of the conditioning in the probability assessment. In essence all probabilities are conditional, and the more relevant the information used in the conditioning, the more relevant the resulting probability assignment. Failure to consider relevant background information would be a disservice to the court. An example given by Dr. Ian Evett considers the question: What is the probability that Sarah is over 5 feet 8 inches? We could try to assign this probability, but our view would change markedly if we were told that Sarah is a giraffe. Ignoring the background information (Sarah is a giraffe) will lead to a much poorer assignment of probability. This is certainly not intended to sanction inappropriate information and conditioning.

The second argument is a verbal trick undertaken in the legal context. Consider the numerator of the likelihood ratio. This is $\Pr(E|H_p, I)$, which can be read as: the probability of the evidence given that the prosecution hypothesis is correct and given the background information. The (false legal) argument would be that it is inconsistent with the presumption of innocence to "assume that the prosecution hypothesis is true." This again is a misconception or a misreading of the conditioning. When calculating the likelihood ratio we are not assuming that the prosecution hypothesis is true, which indeed would be bias. What we are doing is weighing the prosecution and defense hypotheses against each other by calculating the probability of the evidence *if* these hypotheses were true. This is an instance where the verbal rendering of Bayes' rule can be misconstrued (possibly deliberately) to give a false impression never intended in the logical framework.

I have also heard the following erroneous argument: If H_1 and H_2 are independent, then $\Pr(H_1 \text{ and } H_2)$ is less than $\Pr(H_1)$ or $\Pr(H_2)$.

This part of the statement is correct. It is actually correct whether or not the events are independent. However, sometimes it is extended to "in a trial in which the case for the prosecution involves many propositions that must be jointly evaluated the probability of the conjunction of these hypotheses will typically drop below .5, so it would seem that a probabilistically sophisticated jury would never have cause to convict anyone."[759]

Of course this is an erroneous attack on probability theory *per se*, not specifically Bayesian inference. But let us examine the argument. Suppose we have events:

B: The suspect had intercourse with the victim.
C: The intercourse was not consensual.

Let us assume that there is some evidence, *E*. We seek the probability of guilt, *G*, given the evidence, Pr(*G*|*E*). Usually a court would require both *B* and *C* to be very probable to conclude *G*. It is logically certain that Pr(*B* and *C*|*E*) is less than or equal to Pr(*B*|*E*) and it is also less than or equal to Pr(*C*|*E*). However rather than being logically worrying, this is actually the correct conclusion. If there is doubt about *B* or *C* or collectively doubt about *B* and *C*, then *G* is not a safe conclusion, and I would be very concerned about any inference system that did not follow these rules. Robertson and Vignaux[659,662,663] argue eloquently that any method of inference that does not comply with the laws of probability must be suspect.

However, I am unsure whether this was the point that was being advanced. Let us assume that the *propositions* are something like:

A: The blood on Mr. Simpson's sock is from Nicole Brown.
B: The blood on the Bundy walk is from Mr. Simpson.
C: LAPD did not plant the blood on the sock.
D: LAPD did not plant the blood on the Bundy walk.

Suppose that guilt is established if *A*, *B*, *C*, and *D* are true. Indeed then Pr(*A*, *B*, *C*, *D*) would be less than the probability of any of the individual events. However, in my view guilt may also be established if *A*, *B* and *C* are true but *D* is false (there are other combinations).

Let us assume that guilt is certain if one of the following combinations of events held:

True	**False**
A,B,C,D	
A,B,C	*D*
A,B,D	*C*
A,C	*B,D*
B,D	*A,C*

Guilt may also be true under other combinations that are not listed, but in such a case none of these events, *A, B, C,* or *D,* would be evidence for it.

Let us conservatively assign the probability of guilt under these alternatives as zero. Then

$$\Pr(G) = \Pr(A, B, C, D) + \Pr(A, B, C, \bar{D}) + \Pr(A, B, \bar{C}, D) + \Pr(A, \bar{B}, C, \bar{D}) + \Pr(\bar{A}, B, \bar{C}, D)$$

The probabilistic argument can be extended rather easily to any number of events or to include instances where guilt is not certain but probable given certain events. In fact rather than being problematic, I find the laws of probability rather useful.

Regarding the arbitrariness of the verbal scale, this point must be conceded except with reference to the point labeled inconclusive. However, any verbal scale, Bayesian or otherwise, is arbitrary. The problem really relates to aligning words that are fuzzy and have different meanings to different people to a numerical scale that possesses all the beauty that is associated with numbers. This problem will be alleviated in those rare cases where the logic and numbers are themselves presented and understood in court.

This brings us to the full Bayesian approach. There is little doubt that this approach is the most mathematically useful. Most importantly, it can accommodate any number of hypotheses, which allows us to phrase the problem in more realistic ways. It is the underlying basis of Bayes' nets, which will certainly play a prominent part in evidence interpretation in the future. However, it is impossible to separate out the prior probabilities from this formulation, and hence implementation would be possible only in those unlikely cases where the court was prepared to provide its prior beliefs in a numerical format. At this time, the approach must be considered as the best and most useful tool for the scientist to use, but currently not presentable in court. The unanswered question is whether the compromise approach given above is an acceptable solution to the courts.

When weighing these approaches against each other, the reader should also consider that the vast majority of the modern published literature on evidence interpretation advocates the logical or full Bayesian approaches. There is very little published literature advocating a frequentist approach, possibly because the lack of formal rigor in this approach makes publication difficult.

Throughout this book we will attempt to present the evidence in both a frequentist and a likelihood ratio method where possible. There are some situations, such as missing persons' casework, paternity, and mixtures, where only the likelihood ratio approach is logically defensible.

2.6 Evidence Interpretation in Court

2.6.1 The Fallacy of the Transposed Conditional

Initially I had not planned to write anything on the famous fallacies and especially the fallacy of the transposed conditional also known as the prosecutor's fallacy. What was left to say after so many publications on the subject?[7,35,58,97,148,201,239,261,267,288,291,334,399,506,618,637,638,654,655,658,659,660,661,686,769,841] However, I discovered in 2003 that there was still much uncertainty about the subject and indeed that groups of people with important responsibilities in the criminal justice system had never heard of the issue.

What can I add to a debate that is already well written about? I will again explain it here for those readers for whom the fallacies are new. I will also add a section that attempts to assess the mathematical consequences of this error and gives some tips on how to avoid making a transposition. Many of these tips come from my experiences working with colleagues at the Interpretation Research Group of the FSS in the U.K.: Champod, McCrossan, Jackson, Pope, Foreman, and most particularly Ian Evett. Few forensic caseworkers have written on the subject, although most have faced it.

The fallacy of the transposed conditional is not peculiar to the logical approach. It can occur with a frequentist approach as well. Opinion is divided as to whether the fallacy is more or less likely when using the logical approach. In essence, it comes from confusing the probability of the evidence given a specific hypothesis with the probability of the hypothesis itself. In the terms given above, this would be confusing $\Pr(E|H_p)$ with $\Pr(H_p)$, $\Pr(H_p|E)$, or $\Pr(H_p|E, I)$.

Following a publication by Evett,[256] we introduce the subject by asking "What is the probability of having four legs IF you are an elephant?" Let us write this as $\Pr(4|E)$ and we assign it a high value, say, 0.999.

Next we consider "what is the probability of being an elephant IF you have four legs?" Write this as $\Pr(E|4)$ and note that it is a very different probability and not likely to be equal to 0.999. This example seems very easy to understand both verbally and in the symbolic language of probability. But the fallacy seems to be quite tricky to avoid in court.

Imagine that we have testified in court along the lines of one of the statements given below:

- The probability of obtaining this profile from an unrelated male member of the New Zealand population is 1 in 3 billion.
- The frequency of this profile among members of the population of New Zealand unrelated to Mr. Smith is 1 in 3 billion.
- This profile is 3 billion times more likely if it came from Mr. Smith than if it came from an unrelated male member of the New Zealand population.

The first two are frequentist statements and the last is a statement of the likelihood ratio. Let us work with the first. We are quite likely in court to face a question along the lines: "In lay terms do you mean that the probability that this blood came from someone else is 1 in 3 billion?"

This is the fallacy of the transposed conditional. It has led to appeals and retrials. It appears to be very natural to make this transposition however incorrect. Every newspaper report of a trial that I have read is transposed and I suspect that many jurors and indeed judges make it.

How can a scientist who is testifying avoid this error? The answer involves training and thinking on one's feet. But I report here Stella's Spotting Trick (named after Stella McCrossan) and Ian's Coping Trick (named after Ian Evett).

Stella's spotting trick: The key that Stella taught was to ask oneself whether the statement given is a question about the evidence or hypothesis. Probabilistic statements about the hypothesis will be transpositions. Those about the evidence are likely to be correct. The moment that you notice the statement does NOT contain an IF or a GIVEN you should be cautious. Consider the sentence given above: "In lay terms do you mean that the probability that this blood came from someone else is 1 in a billion?" Is this a statement about a proposition or the evidence? The proposition here is that the blood came from someone else. And indeed the statement is a question about the probability of the proposition. Hence it is a transposition.

Ian's coping trick: The essence of this trick is to identify those statements that you are confident are correct and those that you are confident are incorrect. This is best done by memory. There will be a few standard statements that you know to be correct and a few transpositions that you know to be incorrect. Memorize these. Then there is the huge range of statements in between. These may be correct or incorrect. The prosecutor may have transposed in his/her head and is trying to get you to say what he/she thinks is a more simple statement. That is his/her fault not yours (if you are a forensic scientist reading this). He/she should have read and studied more. In this circumstance I suggest you say something like:

> I have been taught to be very careful with probabilistic statements. Subtle misstatements have led to appeals in the past. I am unsure whether your phrasing is correct or incorrect. However I can give some statements that I know are correct.

These will include the numerical statement of type 1, 2, or 3 given above or the verbal statements given in Table 2.3.

Of course, care by the scientist is no guarantee that the jury, judge, or press will not make the transposition themselves. For instance, Bruce Weir

had gone to great trouble with the wording in the report for his testimony in the O.J. Simpson case. Weir was careful and correct in his verbal testimony as well. As an example, he reported that there was a 1 in 1400 chance that the profile on the Bronco center console would have this DNA profile IF it had come from two people other than Mr. Simpson and Mr. Goldman. This was transposed by Linda Deutsh of the Associated Press (June 26, 1995) to "a chance of 1 in 1400 that any two people in the population could be responsible for such a stain." To quote Professor Weir: "It is incumbent on both the prosecution and defense to explain the meaning of a conditional probability of a DNA profile."[835]

I found another transposition in an interesting place. Horgan[415] was warning about errors in the Simpson case and went on to commit the prosecutor's fallacy while explaining the error of the defender's fallacy! "Given odds of 1 in 100,000 that a blood sample came from someone other than Simpson, a lawyer could point out that Los Angeles contains 10 million people and therefore 100 other potential suspects. That argument is obviously specious…" All the students in the 2003 (University of Auckland, New Zealand) Forensic Science class spotted the error when given it as an assignment!

Mathematical consequences of transposition: The transposition is of no consequence if the prior odds are in fact 1. This is because the answer arrived at by transposition and the "correct" answers are the same in this circumstance. The issue only occurs if the prior odds differ from 1. If the odds are greater than 1, then the transposition is conservative. Table 2.4 gives some posterior probabilities for differing prior probabilities. The table shows, as is known, that for a high likelihood ratio (a low match probability) the practical consequences are negligible. The practical consequences, if they occur at all, are for lower likelihood ratios and where there is little "other" evidence against the defendant or where there is evidence for the defendant.[95,833]

2.6.2 Establishing the Propositions

The concept of a hierarchy of propositions was first introduced by Aitken[8] and greatly developed by Cook et al.[193] and Evett et al.[272] Propositions are classified into three levels: offense, activity, or source. The top of the hierarchy is taken to be the offense level, where the issue is one of guilt or innocence. An example of this could be "the suspect raped the victim." It is often held that this level of proposition is for the courts to consider and above the level at which a forensic scientist would usually operate. The next level is taken to be the activity level. An example would be "the suspect had intercourse with the victim." This differs from the offense level in that it talks about an activity (intercourse) without making a comment about its intent (rape) that would need to consider other matters such as consent. The lowest level is taken to be

Table 2.4 Consequences of Transposing Assuming that DNA Evidence Gives a Match with Match Probability 1 in a Billion and My Subjective View of This

Prior Odds	Meaning	Posterior Probability with Transposition	Posterior Probability without Transposition	My Subjective View
4,000,000:1 against	The defendant is as likely as anyone else in New Zealand to be the donor	0.999999999	0.996015936	No practical consequence
4000:1 against	The defendant is more likely than a random person in New Zealand to be the donor	0.999999999	0.999996000	No practical consequence
1:1	The suspect is vastly more likely than a random person in New Zealand to be the donor	0.999999999	0.999999999	No practical consequence
Any odds on	The suspect is vastly more likely than a random person in New Zealand to be the donor			No practical consequence

the source level. At this level, we consider questions of the type "did this semen come from the suspect?" Considerations at this level do not directly relate to activity, in this example intercourse, which would involve issues such as from whence the sample was taken, drainage, and contamination.

It has become necessary to add another level below the source level. This has been termed sublevel 1. This has arisen because it is not always certain from what body fluid the DNA may have come. For instance, when considering the source level proposition "the semen came from the suspect," the sublevel 1 proposition would be "the DNA came from the suspect."

The further down the hierarchy the scientist operates, the more the responsibility for interpreting the evidence is transferred to the court.

It would be reasonable to leave the interpretation of such matters to the court if that were the best body to undertake this interpretation. However, if the matter requires expert knowledge regarding such matters as transfer and persistence, it would seem wise for the scientist to attempt interpretation at a higher level in the hierarchy, or at least to warn and equip the court to make such an attempt. The evidence must eventually be interpreted at the offense level by the court. If the evidence cannot be put in the context of the offense, then it is, in itself, irrelevant to the court.

Let us assume that the scientist can make a decision as to which level in the hierarchy the propositions should be formulated. The next step is to attempt to formulate one hypothesis for the prosecution and one for the defense. The defense are under no obligation to provide a hypothesis and, in fact, the defendant may have given a "no comment" interview. McCrossan et al. (in draft) ask:

> Is it the role of the forensic scientist to formulate the defense proposition when "no comment" is given?
> If the scientist does formulate a proposition on behalf of the defense, how should the implications of this action be highlighted/exposed in the statement?

One issue here is the consideration of the obvious alternative:

H_d: The suspect had nothing to do with the …(activity associated with the crime)

tends to maximize the *LR* and hence has a tendency to maximize the apparent weight of the evidence.

There is an issue as to whether the defense must choose only one proposition or whether they can have many. In fact, it is worthwhile considering what happens if the prosecution and defense hypotheses are not exhaustive. Let us assume that there could be three hypotheses H_1, H_2, and H_3. H_1 aligns with the prosecution view of the case, H_2 is the hypothesis chosen for the defense, and H_3 is any hypothesis that has been ignored in the analysis but is also consistent with innocence.

Set hypothetically:

Hypothesis H_i	**$\Pr(E\|H_i)$**
H_1	0.1
H_2	0.000001
H_3	1

Let us assume that we proceed with the "logical approach" and calculate

$$LR = \frac{\Pr(E|H_1)}{\Pr(E|H_2)} = \frac{0.1}{0.000001} = 100{,}000$$

which would be described as very strong support for H_1. Is this acceptable? Well, the answer is that it is only acceptable if the prior probability for H_3 is vanishingly small and if the three hypotheses exhaust all possible explanations. The approach of McCrossan et al. to hypothesis formation suggests

that all propositions for which there is a reasonable prior probability should be considered, either directly by the scientist or after the defense have made the scientist aware of such a possibility. Under these circumstances, there should be no risk of the likelihood ratio being misleading. The future may entertain a more comprehensive solution based on the general form of Bayes' theorem.

2.6.3 Errors in Analysis[1]

There have been justified complaints that most discussions, including our own, start from the premise that the typing has been completed in an error-free way.[57,201,501,509,583,618,665,765,767] Other than this brief section and a section in Chapter 8 on low copy number analysis, we will also assume that the analysis is error free.

However, there is clear evidence that errors do occur. For a brief review, see Thompson et al.[770] and the following correspondence.[179,200,771] The rate of such errors is probably low and quality assurance goes some way to reassuring the court and public that the error rate is not high. But it must be admitted that a good estimate of the rate is not available. Nor could one rate be applied fairly to different cases, different laboratories, or even different operators. There have been calls for monitoring of this rate (reviewed again in Thompson et al.; see also Chakraborty[159]). The error rate would be a very hard parameter to estimate and there are clear practical difficulties. This may have forestalled any large-scale effort to estimate this rate. A more likely explanation is the quite legitimate wish of forensic scientists that whenever an error is found, they do not want to count it; rather, they want to eliminate the possibility of its future reoccurrence. However, we endorse efforts to investigate the error rate. One reason for this is that all forensic scientists we know are honest, dedicated persons and any investigation such as this will be used primarily to improve methods.

Despite these barriers, there are modern collaborative exercises that take a very responsible approach to assessing the rate, the source of errors and that make suggestions for their reduction. Parson et al.[607] give the outcome of a very large mitochondrial DNA collaborative exercise. They report 16 errors. Ten of these errors were clerical, two were sample "mix-ups," one was assigned as contamination, and the remainder were assigned as arising from interpretational issues.

Errors can be of several types. Clearly, false exclusions and false inclusions have differing consequences. The most serious errors would be sample swapping or sample contamination. However, the most common "error" of which

[1] This section on error is provided by Christopher Triggs and John Buckleton.

we are aware is the assumption that a heterozygote is a homozygote because an allele is undetected. It is difficult to see how there could be too serious a consequence for this.

The presence of completely random contamination, from, say, plasticware, in a normal corroborative case is unlikely to lead to a false identification implicating the suspect. This type of contamination may be identified by the negative controls if the contamination is repeated. The same contamination in a database search case, if missed by the controls, could have far more serious consequences, for example implicating a worker in a plastic factory who is on the database. The risks of contamination from mortuary surfaces,[682] from scene officers,[683] and the presence of third-party DNA after simulated strangulation[681] have been discussed.

If a scene sample is contaminated with suspect DNA, then the suspect is at great risk. Forensic scientists are aware of these risks and treat them very seriously, but complacency should be rigorously opposed.

Other risks are run whenever subjective judgement is involved. This is slowly diminishing in forensic DNA work with the advent of automation but still remains in some areas. Risinger et al.[648] and Saks et al.[685] give a very well argued examination of the risks of observer effects in forensic science. Observer effects are errors in observation, recording, or decision making that are affected by the state of mind of even the most honest and diligent observer. Observers have been making this warning for some time:

> When you employ the microscope, shake off all prejudice, nor harbour any favourite opinions; for, if you do, 'tis not unlikely fancy will betray you into error, and make you see what you wish to see.[32]

A famous example is the count of human chromosomes. Early visualization techniques were rudimentary and counting was very difficult. In 1912, Hans von Winiwater reported 47 chromosomes in men and 48 in women (the Y chromosome is very small). In 1923, Theophilus Painter confirmed the count of 48 after months of indecision. This was despite his clearest views only showing 46. Despite improvements in the preparation and dyeing of chromosomes in the intervening 30 years, it was not until 1956 that Levan gave the correct count of 46. Levan was a plant biologist and did not "know" that humans had 48 chromosomes.[749]

Men generally believe quite freely that which they want to be true.[141]

Thompson et al. argue, correctly, that such effects are widely considered in other fields of science, and protocols to deal with them are in place.[648,685,770] These include such well-known experimental methods as the double blind testing mechanism in much medical research. Why not, then, in forensic science? We recommend the Risinger et al. and Saks et al. discussion as necessary reading for all forensic scientists and recommend that it be included in

their basic training as well as the relevant sections on bias, overinterpretation, and "how much should the analyst know" in Inman and Rudin[429] (for additional comments, see also USA Today[21] and King[464]).

Other possibilities for error include deliberate or accidental tampering from external persons. The FSS reported 156 security breaches in the year ending June 2002, a 37% decrease on the previous year. The report outlined two of these as examples. They involved theft of property such as computers and credit cards rather then evidence tampering.[78]

Without a good estimate of the error rate, we are left with speculation. The error rate is clearly greater than zero. No forensic scientists would claim that it is zero. This is obviously a legitimate avenue for defense examination, and we would recommend that all prosecution witnesses should treat it as a legitimate form of examination, and should not react in a hostile or defensive manner.

We now come to the issue of combining the error rate and the match probability. This has been suggested (see again Thompson et al.[770] for a review) but never, to our knowledge, applied. If we assume that both the error rate and the match probability are known and constant, then the mathematics are trivial. Below we reproduce the common form in which this is given, but either the full Bayesian approach (Equation (2.4)) or the compromise approach (Equation (2.5)) could handle this easily by introducing a subproposition of contamination. Taroni et al.[757] discuss the problem using Bayes' nets and demonstrate this point.

We have two profiles of interest: G_c, the true type of the profile recovered at the crime scene; and G_s, that of the suspect. We will assume that the profile G_s is always determined without error.

As usual, we have two hypotheses:

H_p: The suspect is the donor of the DNA in the crime sample.
H_d: The suspect is not the donor of the DNA.

We further consider the event, $\exists$, that the profile produced in the electropherogram is not a true representation of the type of the DNA in the crime sample; that is, an error in typing has occurred.

Its complementary event, $\overline{\exists}$, is that the profile produced in the electropherogram is a true representation of the type of DNA in the crime sample. We follow Thompson et al.[770] and assume that $\exists$ and $\overline{\exists}$ are not conditional on H_p or H_d. If we write the error rate as e, then we can take

$$\Pr(\exists) = e \text{ and } \Pr(\overline{\exists}) = (1-e)$$

$$\Pr(\exists | H_p) = \Pr(\exists | H_d) = \Pr(\exists) = e$$

$$\text{and } \Pr(\overline{\exists} | H_p) = \Pr(\overline{\exists} | H_d) = \Pr(\overline{\exists}) = 1-e$$

We have the four probabilities:

$\Pr(G_c \mid G_s, \overline{\exists}, H_p)$	1
$\Pr(G_c \mid G_s, \exists, H_p)$	The probability of a false positive match given that an error has occurred, k
$\Pr(G_c \mid G_s, \overline{\exists}, H_d)$	the match probability, f
$\Pr(G_c \mid G_s, \exists, H_d)$	The probability of a false positive match given that an error has occurred, k

The likelihood ratio becomes

$$LR = \frac{\Pr(G_c \mid G_s, \overline{\exists}, H_p)\Pr(\overline{\exists} \mid H_p) + \Pr(G_c \mid G_s, \exists, H_p)\Pr(\exists \mid H_p)}{\Pr(G_c \mid G_s, \overline{\exists}, H_d)\Pr(\overline{\exists} \mid H_d) + \Pr(G_c \mid G_s, \exists, H_d)\Pr(\exists \mid H_d)}$$

Thompson et al.,[770] Weir,[829] and Buckleton and Triggs (this text) give three different formulae for this likelihood ratio:

	LR
Buckleton and Triggs	$\dfrac{1-(1-k)e}{f(1-e)+ke}$
Thompson et al.	$\dfrac{1}{f+ke(1-f)}$
Weir	$\dfrac{1-ke}{f(1-2ke)+ke}$

Thompson et al. explicitly make the approximation that, in their notation, $\Pr[R|M] = 1$, a fuller treatment could take this probability as $1 - e + ke$. The formula for the Thompson et al. likelihood ratio would then agree with the Buckleton and Triggs formula.

We see that the Thompson et al. *LR* will always exceed the Buckleton and Triggs *LR* and for fixed values of the match probability, *f*, and error rate, *e*. The value of the false positive rate *k* that maximizes this difference depends on the relative magnitude of *f* and *e*. For those cases where the error rate *e* is much greater than the match probability *f*, the difference is maximized for values of *k* close to, but greater than 0. For example, if $f = 10^{-9}$ and $e = 10^{-4}$ the maximum difference between the two values of the likelihood ratio is 0.00994% and occurs when the false positive rate $k = 0.03152$.

While accepting that Thompson et al. have made an explicit approximation, it is instructive to look at the value of the likelihood ratio under certain limiting boundary conditions. We note the peculiar results for Thompson et al. and Weir in the fifth column of Table 2.5 when there is an unrealistically high error rate, *e*.

Table 2.5 Comparison of Approaches to Incorporate Error Rate by Buckleton and Triggs (BT), Thompson et al. (TTA), and Weir (W)

	LR	Profile common f→1	No error, $e \to 0$	Error certain, $e \to 1$	False positive probability $k=0$	False positive probability, $k=1$
BT	$\frac{1-(1-k)e}{f(1-e)+ke}$	1	$\frac{1}{f}$	1	$\frac{1}{f}$	$\frac{1}{f(1-e)+e}$
TTA	$\frac{1}{f+ke(1-f)}$	1	$\frac{1}{f}$	$\frac{1}{f(1-k)+k}$	$\frac{1}{f}$	$\frac{1}{f(1-e)+e}$
W	$\frac{1-ke}{f(1-2ke)+ke}$	1	$\frac{1}{f}$	$\frac{1-k}{f(1-2k)+k}$	$\frac{1}{f}$	$\frac{1-e}{f(1-2e)+e}$

To make any exploration of the likelihood ratio $(1-(1-k)e)/(f(1-e)+ke)$, we need to postulate an error rate. If this is larger than the match probability, then it will completely dominate the equation and hence $LR \approx 1 / ke$

This shows that the error rate and the match probability can be mathematically combined. But should they be? The arguments for and against have occurred in the literature (reviewed in Thompson et al. and indeed in court, e.g., *Regina v Galli*[637]). Those making the "for" argument would comment, correctly, that the jury may not be able to weigh the respective contributions of error rate and match probability. Those supporting the "against" argument would argue that an error rate is not known and hence the equation is not implementable. The error rate relates to many things. The arguments given above are phrased largely in the context of a single reference sample and a single stain sample. In many cases, there are multiple samples collected and perhaps typed at differing times. All of this would affect the probability of an error and that subset of errors that represent false inclusions. Lynch[526] makes the interesting point that eyewitness evidence is known to be fallible. Juries have been asked to evaluate this "eyewitness" risk on a case-by-case basis for a long time and no explicit combination is made of the error rate with the weight of evidence. Of course, eyewitness evidence is not presented numerically at all and this may be a fundamental difference.[526]

Our view is that the possibility of error should be examined by the judge and jury on a per case basis and is always a legitimate defense explanation for the DNA result. The two possible hypotheses that are consistent with "innocence" should be explored in court. This argument however does not answer the complaint that the jury may be unable to weigh the two hypotheses consistent with innocence (one numerical and the other not) and may give undue weight to the match probability.

Let us assume that we intend to develop a case-specific estimate of the probability of an error as Thompson et al. suggest, following Thompson:[768]

> … it makes little sense to present a single number derived from proficiency tests as THE error rate in every case, … I suggest that it be evaluated case-by-case according to the adequacy of its scientific foundation and its helpfulness to the jury.

Even assuming that this case-specific error rate is accurately estimated, there still is an objection to the combination of the probability of an error and that of a coincidental match. The likelihood ratio is uncertain in all cases because it is based on estimates and models. It is normal to represent this type of uncertainty as a probability distribution. If we add the possibility of error, then this distribution has a point mass at 1 and a continuous distribution around high values for the *LR*. In Figure 2.1 we give a hypothetical distribution of this sort. The Thompson et al. equation suggests we report the *LR* signified by the arrow. This value is in the void between the two modes, in a region where there is no density, and may be viewed by many as a very poor summary of the distribution. However, the large mode at the right of the figure, if reported without mention of error, could also be viewed as an equally poor summary.

The innocent man who has been implicated by an error or indeed by a coincidental match is at great risk of a false conviction and it is generally accepted that a false conviction is the worst outcome that can occur in the judicial system. The Thompson et al. formula, if applied, may very occasion-

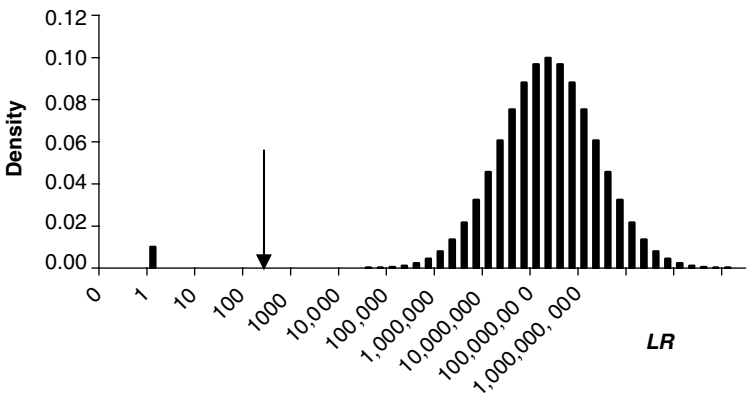


Figure 2.1 Hypothetical distribution for *LR*. The small mode at 1 represents the *LR* if an error has occurred. The larger mode centered at about 1,000,000,000 represents the *LR* if no error has occurred. The Thompson et al. equation would suggest that we report a value somewhere in the area signified by the arrow.

ally help such a man. In effect, the formula reduces the likelihood ratio and it may do so to the point where a jury will not convict on the DNA evidence alone. The reality, in our view, is that most often the wrongly implicated man will almost have to prove his innocence by establishing that an error HAS happened (it is difficult to see how unless alternative uncontaminated samples are available) or to produce very strong alibi evidence. Unless the wrongly accused man can produce considerable evidence in his favor, it is possible or even likely that he will be convicted. However, there is very little that statistics can do to help him. The reduction of the likelihood ratio affects both the correctly accused and the wrongly accused equally. We suspect that it is of some, but probably inadequate, help to the wrongly accused man and a false benefit to the correctly accused. The answer lies, in our mind, in a rational examination of errors and the constant search to eliminate them. The forensic community would almost universally agree with this.

Findlay and Grix[299] make the reasonable point that the very respect given to DNA evidence by juries places an obligation on scientists to maintain the highest standards and to honestly explain the limitations of the science in court.

It is appropriate to end this section with an appeal for higher standards of the already considerable impartiality in forensic laboratories. We recommend that all forensic scientists read the account by the father of the victim of a miscarriage caused by wrongful fingerprint evidence[541] or the call for standards by Forrest in his review of the Sally Clark case.[316] Most forensic scientists aspire to a position of impartiality but unconscious effects must be constantly opposed. In our view, language is one tool that can be utilized. The words "suspect" and "offender" have specific meanings but are often used interchangeably. In our view, both should be avoided. Both have too many emotional associations: Would you buy a "suspect" car? The preferable term is Mr. or Ms. We also object to the placing of the "suspect's" name in capitals as required by the procedures in some laboratories such as our own. Why is it "Detective Smith" but the suspect is termed "Mr. JONES?" All emotive terms or terms with unnecessary implications should be avoided.

The matter is one of culture. Everyone in a laboratory needs to cooperate in developing the culture of impartiality. People lose their life or liberty based on our testimony and this is a considerable responsibility.

2.6.4 The Absence of Evidence

Special attention is given in this section to interpreting the "absence of evidence." This is largely because of a widespread misunderstanding of the subject despite excellent writing on the matter (see, for instance, Inman and Rudin[429]). This misunderstanding has been fostered by the clever but false saying:

The absence of evidence is not evidence of absence.

We assume a situation where some evidence has been searched for but not found. Call this event $\bar{E}$. Bayes', theorem quickly gives us a correct way to interpret this evidence.

$$LR = \frac{\Pr(\bar{E}|H_p)}{\Pr(\bar{E}|H_d)}$$

The issue then is simply one of estimating whether the finding of no evidence was more or less likely under H_p than H_d. Unless some very special circumstances pertain, then the finding of no evidence will be more probable under H_d, and hence the absence of evidence supports H_d. Often, in real casework, this is only weak support for the hypothesis, H_d.

The special circumstances that could pertain would be those that made no evidence very likely under H_p but the finding of evidence likely under H_d. Situations involving such circumstances take a little bit of thinking to suggest.

This (correct) mathematical argument is not accepted by many forensic scientists and lawyers, but is universally accepted by interpretation specialists. The counter argument is that one can often think of an explanation for the absence of evidence. For instance, let us imagine that a fight has occurred where one person was stabbed and bled extensively. A suspect is found and no blood is found on his clothing. How is this to be interpreted? Many forensic scientists will observe that the suspect may have changed clothes, washed his clothes, or contact may have been slight in the first place. These observations are correct, but are more along the lines of explanations of the (lack of) evidence. It is better to look at this problem from the point of view of propositions. What was the probability that the suspect would have blood on him *if* he were the offender? Let us imagine that we do not know whether or not the suspect has changed or washed his clothes. Further, let us imagine that we have some information about the fight, but that this is inexact or unreliable. From this we must accept that it is uncertain whether we expect to find blood on the clothing or not, even if the suspect is, indeed, the offender. However, we must feel that this probability is not zero. There must have been some probability that we would have found blood on the clothing; why else were we searching for it? Only if this probability is essentially zero is the evidence inconclusive. Otherwise, if this probability is in any real way larger than zero, it will be larger than the probability if the suspect is not the offender, and hence the evidence will support the defense hypothesis.

Clearly this area is not well understood, nor is there widespread agreement. Further discussion in the literature would be most welcome. Research on transfer and persistence of evidence is also seen to be of great importance.

2.7 Summary

This chapter has reviewed options for a framework for interpretation. Subsequent chapters will focus on details of DNA interpretation. It is, however, very important to understand this fundamental structure for interpretation before proceeding to detailed analysis.

Additional reading: Inman and Rudin[429] give an elegant discussion of many aspects of evidence interpretation. This book would serve very well as part of all training courses in forensic science.

Robertson and Vignaux[651,656] consider both the legal concerns regarding this type of analysis, and more specifically the situation where the evidence itself is both multiple and uncertain. This is a level of complexity above and beyond anything considered in this chapter. They also introduce the useful concept of Bayesian networks that are being extensively researched as a method for forensic interpretation.