The logical foundations of forensic science: towards reliable knowledge

Ian Evett
Principal Forensic Services Ltd, 34 Southborough Road, Bromley BR1 2EB, UK

The generation of observations is a technical process and the advances that have been made in forensic science techniques over the last 50 years have been staggering. But science is about reasoning—about making sense from observations. For the forensic scientist, this is the challenge of interpreting a pattern of observations within the context of a legal trial. Here too, there have been major advances over recent years and there is a broad consensus among serious thinkers, both scientific and legal, that the logical framework is furnished by Bayesian inference (Aitken et al. Fundamentals of Probability and Statistical Evidence in Criminal Proceedings). This paper shows how the paradigm has matured, centred on the notion of the balanced scientist. Progress through the courts has not been always smooth and difficulties arising from recent judgments are discussed. Nevertheless, the future holds exciting prospects, in particular the opportunities for managing and calibrating the knowledge of the forensic scientists who assign the probabilities that are at the foundation of logical inference in the courtroom.

1. Introduction

The public perspective of forensic science is shaped by popular culture, in particular film and TV. ‘Send it to f’rensic’ barks the harassed DCI, investigating the latest gruesome murder. Within minutes, it seems, his sergeant reports ‘I’ve got f’rensic on the line, Guv: they got DNA and it belongs to Gnasher Briggs’. ‘Great’, says the DCI, ‘we got him bang to rights’. Reality, of course, is considerably more nuanced than this but the example does reflect a wide perception that ‘forensic’ is all about rapid testing: indeed, we are accustomed to seeing and hearing the adjective divorced from science as in ‘forensic officer’, ‘forensic department’, ‘forensic examination’.

Fast, accurate and informative testing of evidential material is crucial to crime investigation. But this paper is about what happens after the investigation has culminated in the arrest and charge of one or more suspects. It is very useful to make a distinction between what can be conveniently labelled as the investigative and the evaluative phases of forensic science: it is the latter phase that we enter once legal proceedings are initiated against a suspect who, in many cases, later becomes the defendant in a criminal trial at a court of law.

As things progress from the investigative to the evaluative stages, so the number of participants in the process increases. The police, working with the Crown Prosecution Service, continue to have a major role, while the suspect is advised and represented in his defence by his solicitor; then, for a trial, we have barristers to present the prosecution and defence cases, a judge to oversee the process and a jury to provide a verdict. It is the role of the forensic scientist in this process that we consider here.

2. A judicial perspective: evidence of possibility

A person’s DNA profile consists of a series of peaks that represent alleles at a number of different loci, collectively known as a genotype. If a DNA sample from a crime scene is adequate in quality and quantity, then it will reveal the genotype of the person who left it. These genotypes are highly variable between unrelated individuals—we say that they have high discriminating power, in the
sense that, if the genotypes of two different individuals are compared it is highly probable that they will be different (identical twins excepted). If the DNA profile of a crime sample yields a different set of peaks from those in the profile from a possible suspect, then this is a valuable tool for excluding that person from further investigation. If, however, as in the example in the first paragraph, the profile from the crime sample yields the same set of peaks as a profile from a suspect (in that case, the infamous Gnasher Briggs), the investigator has powerful reasons for believing that the suspect is the person who left the crime sample.

Moving on to the evaluative phase, the important question arises of how this evidence may be presented to jurors in a manner that best assists them in coming to a good verdict. It is intuitively reasonable that the extent to which the genotype from the crime profile is unusual is relevant here, and this notion is at the basis of calculations that are carried out by forensic scientists in DNA profiling cases that are taken to court. However, it will not come as a surprise to learn that offenders in crime cases do not go to great trouble to leave DNA samples of pristine quality and adequate quantity for straightforward analysis. One of the challenges arises when DNA from a crime scene is a mixture of the DNA from two or more people, the analysis being further complicated by: the contributors’ DNA being present in differing quantities; poor quantity and quality of the sample; and analytical artefacts. In such a case, it might be difficult for a forensic scientist to present a quantitative assessment of evidential weight in relation to the outcome of a comparison between a defendant and crime sample. This was the position in the case of R v Dlugosz and, to cut a long story short, the matter was eventually considered, along with two other cases, at the Central Court of Criminal Appeal [1]. This is from the judgment:

There was no dispute ... that DNA evidence from a mixed profile could be used simply to establish that the defendant might have been a contributor or could not have been a contributor. It was accepted that it is often useful for a jury simply to know that fact without any further elaboration.

Consider the second part of the first sentence of that extract. If the scientist had been able to establish from a comparison between the crime profile and the genotype of the defendant (formerly the suspect, of course) that he could not have been a contributor to it, then that is a simple outcome that the jury will have no difficulty in understanding. However, consider the converse: the judgment is saying that, provided the observations justify the statement, it is useful for the jury to be told that the defendant might have been a contributor to the mixture. Indeed, in the next sentence, the judgment says that it is often useful for the jury to be told that without further elaboration. Let us see how that works out at court. The scientist, called to give evidence by the prosecution, explains that the profile from the mixture is complicated and that it has not been possible to carry out a calculation of evidential weight; nevertheless, the jury are told, ‘the defendant might have been a contributor to the mixture’. Later, in cross-examination, defence counsel puts it to the scientist that it is also true that the defendant might not have been a contributor to the mixture. The scientist must agree to this because it is obviously true.

However, we are forced to ask whether this is a good way for the scientist’s evidence to be adduced and here we come to a distinction between the two roles we met earlier.

At the investigative phase, we expect to see the scientist working closely with the investigator in the search for the offender; at the evaluative phase, however, we expect to see the scientist as someone who is there to help the court and this requires no partisanship towards the prosecution view of things. So, rather than a scientist who is perceived as making concessions to the defence, we would advocate that both perspectives are addressed throughout the evaluation. The balanced scientist, following this view would say then: ‘the defendant might, or might not, have been a contributor to the mixture’.

An analogy will convey the usefulness of this sentence. You are driving down a country lane, unsure of your precise location and, with your satnav not working, you are seeking the village of Much Muckle. At a fork in the road, you have the choice between going left or right. There is no signpost but there is someone who appears to be a local, leaning over a nearby gate. Pointing down the right fork you ask ‘is that the road to Much Muckle?’ The friendly local replies ‘well, it might be, but then again it might not’. Helpful? Hardly: it is as just as informative as if he had said ‘I don’t know’.

Following the view of the Appeal Court, the scientist has the option of making a simple statement which is biased towards the prosecution or of expanding the statement to make it unbiased—but uninformative. Fortunately, the science of evidence evaluation has advanced much further than this.

3. Evolution of the balanced scientist

Historically and also in popular media, we have seen the forensic scientist as performing a key role in the investigation of a serious crime. Once a suspect has been arrested and charged the investigator will have a strong conviction—and considerable enthusiasm—for the view that the suspect is the person who committed the crime: if that conviction influences the evaluation of the scientist, then there is a real danger of the kind of miscarriage of justice that has happened in the past where the scientist has allowed investigative enthusiasm to spill over into evaluative assessment. This is not just a historical and fictional danger—it exists today and Dror [2], speaking at this discussion meeting, calls it contextual bias.

I spent most of my career as a member of the Home Office Forensic Science Service (FSS) until its closure in 2010, and I was fortunate to witness various measures that were taken towards alleviating the problem of prosecution bias. The most important changes flowed from the time in the 1990s that the FSS was set up as a government agency, with greater financial independence than before and a cultural change towards serving the need of our customers. As part of the process of change, a project was set up to consider case assessment and interpretation (CAI). Most of the revenue of the FSS came from the police forces and in that sense they were certainly our biggest customers. One can see great sense in that at the investigative stage: quick tests, discriminating and cheap, supplemented by rapid opinions, often speculative, expressed simply and, ideally, unequivocally. But at the evaluative stage, it was clear to us in the CAI project team that our overriding responsibility was to the court. As the project developed, and we talked to police officers, it became obvious that some of them did not see things that way. There was a long established view that it was our job to assist the prosecution to make the case against the defendant stick. While this cultural
difference led to some tension, we were allowed to press on with the project and there is no doubt that the outcomes contributed substantially to the paradigm shift that has occurred in forensic science evidence evaluation.

But the ground source of the paradigm shift lies in profound changes which took place from the middle of the last century in the understanding of the nature of probability. To continue the story of the evolution of the balanced scientist, we need to digress briefly.

4. The nature of probability

Most of the scientists that I know, introduced to statistics at undergraduate level, found it a perplexing and forbidding discipline. Based on the notion of large numbers of repeatable experiments, such as tossing coins or drawing balls from urns, it led to the notions of the significance level and the confidence interval—the latter, in particular, being a peculiar construct which is widely misunderstood. From this perspective, probability is seen as the limiting value of the outcome of a very long run of repeatable and identical experiments. This is generally referred to as the frequentist view of statistics and it is the discipline which I was brought up in as a statistician from undergraduate to postgraduate.

But this view of probability is extremely limiting. It is not possible to visualize the interesting uncertainties of life as the culmination of long runs of repeatable experiments. British readers, and any others familiar with the works of Shakespeare, will have heard of the murders of the 'Princes in the Tower'. Consider the proposition that the princes were murdered under the orders of their uncle, King Richard III, the alternative being that the murders were carried out without Richard's foreknowledge. Clearly, the issue is shrouded in uncertainty because, in the first instance, of the secrecy of the incident and subsequently because of the passage of time and the apparent unreliability of early, politically biased accounts. Given that uncertainty, the truth of the matter is not clearly apparent. To satisfy my own curiosity, I might take some trouble to carry out a study of the mystery of what happened to the two princes in The Tower (the Wikipedia entry is an excellent place to start!). When I expand my knowledge of the subject, it would not be surprising if my probability changed. We could write it as

\[ \Pr(R|K_{\text{un}}). \]

This highlights another property of probability: it can change in the light of new information. Everyday examples of this are all around us. For example, our decision to hold a barbecue for the neighbours on Saturday may need to be reconsidered when we hear the weather forecast on Friday.

5. A case example

Let us consider an entirely hypothetical case. Mr X, a well-known local villain, was dining in a fairly crowded restaurant when a man entered, walked up to X's table, drew a handgun and killed him with a single shot to the head. During the ensuing confusion, the gunman ran out of the restaurant and was driven away at high speed. A few days later, police arrested a Mr Y, also a local man, and charged him with the murder of Mr X. Mr Y denied all involvement with the crime, said that he had never been to the restaurant and was driven away at high speed. A few days later, police arrested a Mr Y, also a local man, and charged him with the murder of Mr X. Mr Y denied all involvement with the crime, said that he had never been to the restaurant and was driven away at high speed. A few days later, police arrested a Mr Y, also a local man, and charged him with the murder of Mr X. Mr Y denied all involvement with the crime, said that he had never been to the restaurant and was driven away at high speed.

When a firearm is discharged, microscopic particles, known as gunshot residue (GSR) are ejected into the air and may be deposited on nearby surfaces, such as the skin and clothing of anyone in the vicinity. Using high-powered microscopy, a forensic scientist carried out an examination of samples taken from a jacket belonging to Mr Y, recovered from his home because it appeared similar to that worn by the gunman, and also of swabs which had been taken from the head and shoulders of the deceased. The scientist made the following observations:

A number of GSR particles were found on the right sleeve of Mr Y's jacket. Many GSR particles were found on the swabs from the deceased. The GSR particles from the jacket were of a chemical composition indistinguishable from that of the GSR particles from the deceased.

There is a range of chemical components of firearm ammunition but the GSR particles found in this case were similar to those from a high proportion of modern ammunition.

Mr Y is brought to trial. The prosecution's case is that Mr Y murdered Mr X and, in support of that case, they present a series of witnesses who create a framework of circumstantial evidence, including: Mr X and Mr Y have conflicting business interests and there is a history of antagonism between them; two eyewitnesses picked out Mr Y from a set of photographs as resembling the gunman. The defence
position is that Mr Y had nothing to do with the incident and two of his friends testify that he was playing cards with them and several others over a time period that straddled the time of the murder. It is helpful if we consider the position of a juror who is faced with the circumstantial evidence alone. Let us encapsulate the prosecution and defence positions in a pair of propositions:

- \( H_p \): Mr Y is the man who shot Mr X.
- \( H_d \): Mr Y had nothing to do with the shooting of Mr X.

Let us use \( I \) to summarize all of the circumstantial evidence, then it is reasonable to encapsulate the juror’s uncertainty with regard to the truth of these two propositions by two probabilities:

\[
Pr(H_p|I) \quad \text{and} \quad Pr(H_d|I).
\]

This is a little like betting on a two-horse race: if one of the two propositions is true, the other must be false. If, for example, the juror considers, on the basis of the circumstantial evidence that he has a 0.9 probability \( (\text{or, in common parlance, a 90\% chance}) \) that the prosecution proposition is true then, provided he is rational, he has a 0.1 probability that the defence proposition is true. For what follows, it is really helpful to consider the ratio of these two probabilities, which are called the \textit{odds} in favour of the prosecution proposition: we would say that the juror’s odds at this juncture are 9 to one in favour of the prosecution proposition. Because we are about to update the juror’s uncertainty by providing him with new evidence, we call these the \textit{prior odds} and we can write them out as

\[
Pr(H_p|I) \quad \text{and} \quad Pr(H_d|I).
\]

The scientist presents the scientific observations which we will summarize by the letter \( E \). Now the juror has additional information on which to base his probabilities so we can talk about the \textit{posterior odds}, which we write as

\[
Pr(H_p|E, I) \quad \text{and} \quad Pr(H_d|E, I).
\]

The posterior odds differ from the prior odds only by the addition of \( E \) to the conditioning. Unless \( E \) is uninformative, the additional conditioning will have either increased or decreased the juror’s belief in the prosecution proposition; in other words, the posterior odds will be either greater or smaller, respectively, than the prior odds. The key to the process is provided by a fundamental result of probability theory known as \textit{Bayes’ theorem}. In the present context, the theorem gives us:

\[
Pr(H_p|E, I) = \frac{Pr(H_p|I) \cdot Pr(E|H_p, I)}{Pr(H_d|I) \cdot Pr(E|H_d, I)}
\]

This tells us that the posterior odds (on the left) are equal to the prior odds (first term on the right), multiplied by the ratio of two probabilities (second term on the right). It is this second term that is the key to the process. It is called by some writers the \textit{Bayes factor} but, for reasons we do not need to go into, it is much more widely known as the \textit{likelihood ratio}. This tells us the fundamental difference between the role of the juror and the forensic scientist. The juror is concerned with the probabilities of propositions given evidence, such as \( Pr(H_p|E, I) \), whereas the scientist is concerned with the probability of evidence, given propositions, such as \( Pr(E|H_p, I) \). The importance of this distinction cannot be overstated. Bayes’ theorem shows us that the scientist must address two questions:

What is the probability that I would have made those observations if the prosecution proposition were true? a

What is the probability that I would have made those observations if the defence proposition were true?

Thus, we establish the role of the balanced scientist encapsulated in the following principles:

1. It is necessary for the scientist to address a pair of propositions: one representing the prosecution position and one representing the defence proposition.
2. It is necessary for the scientist to consider the probability of the scientific observations given that each, in turn, of the propositions is true.
3. The two probabilities, as well as being conditioned by the two propositions, are also conditioned by the case circumstances (\( I \)).
4. The ratio of the two probabilities determines the weight of evidence to be assigned.

Let us see how this formulation applies to the case example. The scientist considers:

- What is the probability that I would have found a number of GSR particles, of the same composition as those found on the deceased, on the right sleeve of Mr Y’s jacket if he is the person who shot Mr X?

Note how, in accord with principle 3 above, several aspects of the circumstances of the incident are relevant, including:

- The extent to which GSR particles are transferred to the clothing of a person who discharges a handgun. Presumably, the scientist will have access to papers in the literature that give the results of appropriate experiments; indeed, the scientist may have carried out such experiments with colleagues.
- The extent to which any transferred GSR particles would be expected to remain on the clothing of the firmer. The nature of the clothing is relevant here (tweed, for example, may be more retentive of fine particles than leather) as well as the time interval between the crime and the examination of the garment. Again, it would be useful if the scientist had access to publications describing relevant experimentation.
- The GSRs were found on the right sleeve. Is Mr Y right or left handed? Did the eyewitness information include anything about the handedness of the man who shot Mr Y?
- The actual number of GSRs recovered. We have been vague about this in the description of the case but, in practice, we must recognize that the scientist’s view might be quite different if a large number were found as opposed to maybe just one or two.

The second question for the scientist is:

What is the probability that I would have found a number of GSR particles, of the same composition as those found on
the deceased, on the right sleeve of Mr Y’s jacket if he had nothing to do with the shooting of Mr X?

Issues to be considered here include:

— What would we expect to find if we looked for GSR particles on the clothing of a person who is unconnected with firearms? Again, the scientist would be expected to have a comprehensive grasp of any relevant surveys that had been reported in the literature. The scientist’s laboratory may have data from its own surveys.

— Is there anything in the suspect’s personal circumstances which could have led to GSR particles on his clothing? Perhaps he legitimately possesses a firearm.

— Have appropriate procedures been followed to minimize the probability of contamination?

Given all of these considerations in relation to the two key questions, it is not surprising that the scientist’s task is far from trivial and the discussion shows how much the knowledge and understanding of the scientist is critically important to the proceedings.

There are some evidence types, most notably DNA, where the scientist will be able to carry out calculations from reliable data to furnish the court with numerical values for the two key probabilities and hence a quantitative likelihood ratio. There are many more where, such as in the present example, data are sparse, poorly structured and of limited relevance to the case at hand and the scientist’s assessment of the weight to be assigned to the observations will necessarily be qualitative. The choice of wording to convey such qualitative opinions is not a straightforward subject but the notion of support for a given proposition is not only familiar to the court but also has a respectable scientific pedigree.

Let us, for example, imagine that in this particular case that 10 or so GSR particles had been recovered from the sleeve of Mr Y’s jacket. The scientist, based on specialist knowledge and understanding of the relevant issues, considers that the probability of such a set of observations is relatively large if Mr Y were truly the man who shot Mr X. On the other hand, again considering a complex set of issues, the scientist considers that the probability of making such a set of observations is extremely small if Mr Y had nothing to do with the incident. So the likelihood ratio in this case is qualitatively large and, on this basis, the scientist reports to the court that there is strong support for the prosecution proposition.

There may be other cases where the circumstances are not as clear cut as in the preceding example: no eyewitnesses, uncertainty about the time and manner of the incident, uncertainties relating to the personal circumstances of the defendant, and so on. If these conditions pertained in the case of the murder of Mr X, then the scientist may be unable to assign probabilities to address propositions of the kind addressed above. The only thing that the scientist can evaluate for the court is the evidence that the chemical composition of the GSRs provides in relation to a different pair of propositions:

$H_0$: the GSR from Mr Y’s jacket came from the same weapon as the GSR found on Mr X.

$H_A$: the GSR from Mr Y’s jacket came from some other weapon.

Previously, we considered propositions that related to activity: now we are considering propositions relating to source. The observation that the chemical compositions of the two sets of GSRs were indistinguishable provides some evidence to support the former proposition but the GSRs are of a very common type and so the evidence is weak. Not only that, but now the jury are left with the task of using the evidence relating to these two source-level propositions to address the activity-level propositions that are much more closely relevant to the task they have to carry out. However, if the scientist, with a specialist knowledge and understanding of the issues, is unable to address questions at activity level in the given case it would seem unreasonable to expect a juror, completely unqualified in the practice of evaluating GSR evidence, to do so. In such an event, there are some quite difficult questions to be addressed about the admissibility of the scientific observations.

6. R v Gjikokaj

The case of Gjikokaj [3] was similar in some respects to the case example discussed in the previous section. Mr Gjikokaj was accused of the murder of someone who had been shot with a firearm. Two GSR particles were found in Mr Gjikokaj’s car that were found to be of the same chemical composition as GSRs recovered from the scene of the crime. Mr Gjikokaj was convicted of murder at Crown Court and later appealed. One of the grounds of appeal was whether or not the GSR evidence should have been put to the jury.

At the original trial, the scientist called by the prosecution was unable to address what we have called ‘activity-level’ propositions here:

Dr M… would not give an evaluative opinion in relation to … .

Where only two particles were found, it was the policy of (name of company) that an evaluative opinion could not be given, as two particles were insufficient for that purpose. He was bound by that policy and he could therefore not give an opinion that evaluated the possibilities.

The Appeal Court are using ‘evaluative opinion in relation to … ’ in the context where we would talk about activity-level propositions. The question then was whether it was useful for the jury to know about the source-level evidence: the observations on the chemical compositions of the two sets of GSRs. This is what the judgment says:

It was admissible to show, in a case where the evidence was circumstantial, that it was not open to the appellant to say there was an absence of scientific evidence connecting him with the crime. The scientific evidence was consistent with the appellant being there and he could not therefore claim that the absence of forensic evidence showed he could not have been there and fired the shot. The primary scientific evidence was therefore plainly admissible for that purpose.

A little more translation is needed: by ‘primary scientific evidence’ the court are referring to the observations on the GSR.

The first sentence of this extract is deeply puzzling—the court needed to be told of the evidence so that the defendant could not say that there was no scientific evidence. The next sentence is an elaboration of this. ‘The scientific evidence was consistent with the appellant being there’—this, regretfully is a prosecution view of the evidence—because the scientific evidence is also ‘consistent with’ other explanations.

The judgment ruled that the GSR evidence was admissible, even though the evaluation of its weight could only be addressed at the source level. Recall that the scientist was unable to advise the jury on what the observations meant in relation to the activity-level questions that they must
address in order to decide on the guilt of the defendant. This brings the difficult question of admissibility that I raised in the example discussed in the preceding section.

7. R v T
In another Appeal Court hearing [4], the Bayesian approach to footwear marks came under scrutiny. This case deserves a paper of its own [5] and there is no possibility of doing justice here to many interesting issues that it raised. But there is one small point that illustrated again the matter of balance. Para 73:

An opinion that a shoe ‘could have made the mark’ is not in our view the same as saying that ‘there was moderate [scientific]’ support for the prosecution case’. The use of the term ‘could have made’ is a more precise statement of the evidence; it enables a jury better to understand the true nature of the evidence than the more opaque phrase ‘moderate [scientific]’ support.

The judgment here is promoting a statement of ‘evidence of possibility’ through the phrase ‘could have’, in the same way as ‘might have’ is used in R v Dlugosz (ibid) and the problem is the same. For the scientist to say ‘the shoe could have made the mark’ is prosecution biased unless it is also said that ‘the mark could have been made by another shoe’. Taken together, the two sentences convey no useful information.

8. Twisted thinking
There is a very good way of assessing the extent to which one understands the new paradigm: it is the extent to which one understands a very common mistake that is known as the fallacy of the transposed conditional. Consider the following exchange:

Scientist: If the DNA had come from someone other than the defendant, then the probability I would have observed these components in the crime profile is one in a million.
Counsel: So, given that you have observed those components in the crime profile the probability that the DNA profile came from someone other than the defendant is only one in a million?

When it is written out this way, it is fairly easy to see that counsel has twisted the sentence around. Each is concerned with a proposition (the DNA came from someone other than the defendant) and a set of observations. The first is the probability of the observations given the proposition; the second is a probability of the proposition given the observations. In the first, the proposition is the ‘conditional’; in the second, the observations are the ‘conditional’. Hence the name for the fallacy. When written out carefully the fallacy is fairly easy to spot, but in the rough and tumble of courtroom exchanges things are more difficult. The following well-known example is from the trial of R v Doheny [6].

Counsel: What is the combination, taking all those [bands in the DNA profile] into account?
Scientist: Taking them all into account, I calculated the chance of finding all of those bands and the conventional blood groups to be about 1 in 40 million
Counsel: The likelihood of it being anybody other than Alan Doheny
Scientist: Is about 1 in 40 million
Counsel: You deal habitually with these things, the jury have to say, of course, on the evidence, whether they are satisfied beyond doubt that it is he. You have done the analysis, are you sure that it is he?
Scientist: Yes.

We can see that the scientist had calculated the probability of making his observations if the DNA had come from someone other than the defendant—it is the next question that invited him into transposing that probability into one that directly addressed the proposition. Instead of the probability of the evidence being small, the jury are suddenly hearing that the probability of the proposition (that the DNA is that of anyone other than the defendant) is small. The next question compounds the problem because counsel leads the scientist into giving his own opinion about something that is none of his business.

The scientist in the preceding example was led into the fallacy by prosecuting counsel and, indeed, this is often called the ‘prosecutor’s fallacy’ but, alas, it is not just prosecutors who make this mistake. Newspaper reports of DNA profiling are almost guaranteed to embody the error as in: ‘Mr J told the court that the probability that the DNA did not come from the defendant is less than one in a billion’.

9. Paradigm shift
The changes in thinking with regard to the nature of forensic science have followed from the paradigm shift that has taken place in the process of reasoning in the face of uncertainty. We stand on the shoulders of giants, which include Savage, Lindley, de Finetti, Turing, Good and Jaynes. The phrase ‘Bayesian inference’ is founded on the notion that probability is subjective: this, in turn, invokes the notion that Bayes’ theorem provides the logical means of updating uncertainty in the light of new information. Research at Lausanne [7] showed that this view contributed to the report that Poincaré provided as part of the process of reversing the notorious conviction of Alfred Dreyfus at the end of the nineteenth century. But it was not until the 1970s that Bayesian reasoning really started to influence forensic science thinking, to a large extent motivated by a paper in the Harvard Law Review by Finkelstein & Fairley [8]. Recently, the Royal Statistical Society, with sponsorship from the Nuffield Foundation, set up a working group of scientists, statisticians and lawyers which has produced an invaluable set of monographs on evidence evaluation; that by Aitken et al. [9] gives an excellent introduction to the subject.

As I explained earlier, the CAI project within the FSS developed this thinking in the 1990s within the notion of serving a customer. The project crystallized thinking with regard to the following particular aspects.

(a) Investigator/evaluator
We have seen that the scientist has a vital role during the investigation: assisting the police in working out what happened and in the search for suspects. Once at least one suspect is charged, then the scientist’s role becomes that of the evaluator. The CAI project established the differences in behaviour and thinking between the two roles. The present paper is concerned almost solely with the evaluative role.

(b) Pre-assessment
This is the notion that the scientist should consider the examination strategy at the earliest stages of undertaking any work
on a case. The formulation of propositions and the consideration of expectations, though often an iterative process, should be at the forefront of the scientist's mind through the entire process.

(c) Hierarchy of propositions
The formulation of a pair of propositions to address is the key to evidence interpretation. The CAI project did much to clarify this process. We have seen that the juror will be concerned with propositions at the level of the offence itself (e.g. Mr S murdered Mr T, Mr Q raped Ms Z); to assist in that process, the scientist should endeavour to address propositions that relate to activities close to those that concern the jury (e.g. Mr K punched Mr J, Mr R had sexual intercourse with Ms T). There will be cases where, generally because of circumstantial limitations, the scientist will not be able to address activity-level propositions and will retreat to source-level propositions (e.g. The fibres on Mr N came from the sweater of Mr U. The glass found on Mr P’s jacket came from the window X). In this last eventuality, the task of relating the source-level evaluation of the scientist to the offence level issues must be passed to the jurors. This raises the question of admissibility that we met earlier.

(d) Statement writing
The principles of evaluation and the developed concepts of CAI led naturally to a structure for the scientist’s evidence for court proceedings. Essential elements of a statement include a synopsis of those aspects of the case circumstances that are relevant to the evaluation and a clear expression of a pair of propositions. In some cases, the scientist will be able to express the weight of evidence in relation to the propositions quantitatively but, outside the field of DNA, it will in most cases be a qualitative statement of some level of support for one of the two propositions.

10. Categorical opinions
There is another aspect of the classical view of the forensic scientist which is still very much the pillar of some disciplines—most notably fingerprints. The bastion of fingerprint comparison evidence is the notion of ‘positivity’. The fingerprint specialist carries out a comparison between a finger mark left at the scene of a crime with a fingerprint, taken under controlled conditions, from a known person and comes to one of three conclusions: inconclusive, elimination or identification. In the case of this third alternative, the specialist will tell a jury that in his/her opinion it is completely certain that the crime mark was left by the person who provided the print. No uncertainty, no probability: positivity.

Following the modern paradigm, we see two key issues here. First, the logical model shows that the role of the forensic scientist is to consider the probabilities of the observations given two competing propositions. We saw in the discussion of the Doheny and Adams judgment (ibid) an example where the scientist mistakenly took the probability of the evidence and turned it into a probability statement about a proposition: the fallacy of the transposed conditional. In the DNA field, this error has been the basis of successful appeals. Yet, fingerprint experts are permitted to make statements with regard to the proposition that a given print/mark pair were made by the same person. There is a paradox here. Second, we must ask whether there is ever a logical basis for a statement of complete certainty in a situation where the inference is necessarily inductive. The issue is discussed in some detail by Champod [10] in his paper at this discussion meeting.

Classically, the fingerprint specialist makes and records personal observations of similarities and differences between the mark and the print. At some stage during this process, the specialist reaches an inner conviction that the two have been made by the identical region of skin surface. This is a psychological process which defies logical analysis. The specialist will defend the reliability of the conclusion by reference to the robustness of the comparison procedure, quality systems, his/her training, certification and length of experience. I should stress that I do not consider this to be necessarily a bad system; indeed, I have defended it in the past [11] and it has served the criminal justice system well for over a century. Nevertheless, high profile mistakes have been made [12] and, worldwide, there is a perceived need for progress. Dror [2] is speaking at this discussion meeting on his approach, in particular on dealing with the issue of contextual bias. My own work in the field has consisted of contributing to development within the FSS of a mathematical model for forensic fingerprint comparison [13] following the Bayesian paradigm. This leads, for a given comparison and pair of source-level propositions, to a quantitative likelihood ratio for expressing evidential weight. We did not see this as a mechanism for replacing conventional fingerprint specialists. On the contrary, we saw it as a means for scientific advancement within the culture, particularly through its use as a training medium and, most interestingly, as a means for calibrating opinion. Although this work necessarily stopped with the closure of the FSS, it is good to see the ideas being further developed overseas, particularly in Switzerland, The Netherlands and the USA.

11. Judgement, knowledge, data and experience
The first role of the scientist is to carry out observations. These may be of all different kinds, for example: shapes of letters in a suspected forged signature; colour and composition of fibres recovered from a crime scene; measurements of refractive index on glass fragments; minute points of detail in a fingerprint; audio patterns in the recording of a telephone voice; or, as in the earlier example, the number and composition of GSR particles. In some cases, the observations will be quantitative (e.g. glass refractive index, a DNA profile), in others qualitative (e.g. handwriting style and letter shapes, facial characteristics) and, more generally, a combination of both (such as in the GSR example).

But the scientist needs to do more than make observations. After all, it would be quite unrealistic to expect the average juror to interpret the GSR observations without guidance—he probably would not even have heard of GSR particles before that particular trial. So the scientist has a second role which is as important as the first: it consists of guiding the jury towards an understanding of what the observations mean within the context of all of the other evidence they have heard and the issues that they are required to address. The classical view of the forensic scientist was of someone who was qualified to advise on weight of evidence on the basis of vast experience—years spent on the job and numbers...
of previous cases done. The earlier quotation from the Doheny trial (ibid) illustrates this: ‘you deal habitually with these things ... you have done the analysis, are you sure it is he?’. Not only do we have the implication that doing lots of cases is a qualification but also the idea that the scientist may dispose on something (the issue of whether or not the defendant’s DNA is present) that is rightly the province of the jury.

It is for the jury to decide on the extent to which they are prepared to be guided by the scientist: what are the factors which should influence the jury in this regard? This question is as old as forensic science itself. The classic vision of the ‘forensic expert’ is someone (male, of course!) of mature age; physical presence; eloquence; extensive academic credentials; and great experience. Unfortunately, history is littered with examples of scientists/medics of great stature and long experience providing court opinions that led to what subsequently proved to be miscarriages of justice. Experience is valuable if it enhances understanding but it is a dangerous tool for increasing knowledge, and the words of Baum [14] neatly summarize the problem:

The trouble with ‘experience’ as a way of approximating to reliable knowledge is that all of us tend to reinterpret each individual experience in the light of a previously held conceptual framework.

And here is a lovely anecdote from Popper [15]:

Once, in 1919, I reported to (Adler) a case which to me did not seem particularly Adlerian, but which he found no difficulty in analysing in terms of his theory of inferiority feelings, although he had not even seen the child. Slightly shocked, I asked him how he could be so sure. ‘Because of my thousandfold experience’, he replied: whereupon I could not help saying: ‘And with this new case, I suppose, your experience has become one thousand-and-one-fold’.

In court, it is quite commonplace for jurors to hear evidence from experts called by prosecution and defence who present quite different opinions. Take a handwriting case, for example: a court may hear an expert called by prosecution present an opinion that the observations support the proposition that the defendant wrote a particular passage of handwriting, whereas an expert called by defence presents the opinion that the observations support the proposition that someone else wrote the handwriting. Each expert would claim credibility by telling the jury of length of their experience—counted by years in the job and numbers of cases done. But is this good enough? Most certainly not. The true conditions of what happened in any particular case are never known with certainty and, just because a jury have clearly accepted an expert opinion, that does not mean that it was particularly reliable. Experience is useful for gaining familiarity with how courts work and how one effectively interacts with the various players: but it cannot alone be a foundation of reliable knowledge.

We have seen that in the new paradigm, the appropriate means by which the scientist should form an opinion in relation to a given pair of propositions is to consider the probabilities of the observations given each proposition in turn. The simplest notion of informing such a probability is that of referring to a database of appropriate size and representativeness. DNA profiling is a fine example of where genotype probabilities may be assigned numerically by reference to quality databases using population genetic models that are widely held to be respectable. This is why one is accustomed to seeing in the newspapers that the weight of evidence to be conveyed by the DNA in a particular case was assigned by means of a number. However, things are quite different in almost all other forensic science fields. We have seen an example in the GSR case: it would be good to have an extensive database of results from controlled discharges of firearms in a range of circumstances that include those relevant to the case at hand. However, the experiments to create such data are inherently dangerous, of complex design and of substantial cost. The expert will need to do the best that can be done from research reported in the literature, possibly supplemented by small-scale local experiments directed to the specific circumstances of the case in hand. If we turn to the subject of handwriting comparison, it would be tempting to believe that a large-scale database of handwriting features and styles could be constructed but the sheer complexity of the variables and their interactions have ensured that past exercises have only been on a relatively small scale. There is no such thing as a comprehensive database for handwriting comparison and little prospect of one for the foreseeable future.

There is a view of the ideal forensic scientist as one who assigns probabilities solely by reference to databases or ‘objectively’. However, the reality is that across the entire spectrum of the discipline, probabilities are assigned to a greater or lesser degree by subjective judgement. Toolmarks, footwear, bullet striations, blood patterns, earprints, fibres, facial comparison are evidence types for which the assessment of evidential weight is largely a subjective process, particularly at the activity level, and will continue thus for many years. There is a quite widely held view that equates ‘subjective’ to ‘unscientific’ but this is a misconception: subjectivity is at the core of scientific endeavour—just listen to the debates at any scientific conference! There is nothing unscientific about subjective judgement provided it is exercised with discipline within a logical framework.

Thus the scientist’s probabilities must be assigned from: a thorough knowledge of the particular evidence type; a deep understanding of the relevant mechanisms and issues; full awareness of all literature and current developments in the field; sound judgement; and an acute awareness of the boundaries of one’s own knowledge. The probabilities that follow from this are necessarily subjective but this, as we have seen, is the intrinsic nature of probability. Whereas individual knowledge is the key to the evaluative process, the future offers the prospects for doing much more.

(a) Knowledge-based systems

One of the foundations of science is the sharing of knowledge: we would expect a scientist to be able to draw on as broad a pool of knowledge as possible in an individual case. This brings us to considering how knowledge can be shared effectively. Clearly, the literature is the most important organ for doing this and the expert should be the conduit for drawing on all of the most relevant aspects of the literature to advise the jury in an individual case. But there is a way of amplifying and complementing this process. Every case is an exercise in the application of knowledge and it is an enormous waste to lock the fruits of that process away in case files. At some time in the future, faced with a particular evaluative issue, it will help the scientist to reflect on whether other colleagues have faced a similar problem in the past and, if so, how their thoughts crystallized. This notion was the inspiration for a knowledge-based system that was developed by two colleagues of mine in the 1980s [16] in the context of glass evidence in a program called CAGE and this is a sketch of how it worked.
In a glass case, the scientist may assign probabilities to the issues of glass transfer from broken windows to nearby individuals and its persistence on clothing. CAGE developed the idea of recording those probabilities and storing them in a structured database, ordered according to the casework conditions: this, with time, would become a knowledge base for the use of all of the experts who shared and contributed to the system. In the future a scientist may consult the knowledge base and select from previous cases where the conditions and circumstances were as close as possible to those in the present case. The probabilities assigned by scientists in those previous cases may then prove helpful for the evaluation of the case at hand. Once that case had been completed, the scientist would be asked to add the new assessments to the knowledge base and so it would grow.

There is, of course, a potential flaw in this vision. Bear in mind that case conditions are quite different from a controlled experiment and this was one of the reasons why we object to the notion of a scientist who gains knowledge solely from doing cases. Indeed, there is a real danger that this could foster the growth of an ignorance base. An essential corollary then is that the knowledge base should be calibrated. Processes must be in place for continuous review of the content of the knowledge base through experimentation under controlled conditions. For example, regular snapshots of selected portions of the knowledge base could be taken to see how the experts’ assessment represents a good understanding of the true situation. The knowledge base can then be updated appropriately and all of the experts who have contributed to it may learn of the strength and limitations of opinions they have expressed in the past.

Recall that the scientist in the case of the shooting of Mr X needed to address probabilities relating to the transfer and persistence of GSR—a field that would obviously benefit from a well-structured knowledge base. And there are many other areas including, for example, blood pattern analysis, fibres and DNA transfer. All of these could be set up through international collaboration via the Internet.

(b) Bayesian networks

In the GSR example, we saw that there were several issues that the scientist needed to address in formulating the two key probabilities. Each of the issues has its own set of uncertainties and so there is something of a network of probabilities to be brought together. Much work has been done in the development of sophisticated computer programs for construction of these ‘Bayesian networks’, and their implementation for the solution of forensic science problems is illustrated in many different situations by Taroni et al. [17]. Such methods will contribute substantially to the construction and development of knowledge-based systems.

(c) Calibration

I do not wish to decry completely the value of experience but it will be clear from several of the points I have made that, in my view, the reliability of scientists’ opinions should be judged, at least in part, by their performance under controlled conditions. For me, the most impressive work that has been done thus far across all fields of forensic science has been a major collaborative exercise among the handwriting specialists of Australasia.

The catalysts for this work have been Bryan Found and Doug Rogers of the Victoria Police Forensic Services Centre and LaTrobe University in setting up ‘barrage tests’ [18]; large numbers of prepared comparisons to be carried out blind by all participants. The notion then is that every participant in the scheme carries to court a dossier that details his/her performance under known conditions. This approach would be applicable to all areas of evaluation.

Evidence types that incorporate properties that may be physically measured (e.g. toolmarks, earprints, fingerprints, bullet striations) lend themselves to mathematical modelling. I have already mentioned the work that was carried out on fingerprint comparisons in the FSS. No-one seriously expects such methods to lead to automated expertise but they offer exciting opportunities for calibration of expert opinion. The analogy of the glass evidence knowledge base described earlier then becomes a vast repository of mark/print comparisons, calibrated quantitatively and available for training, certifying and monitoring fingerprint specialists nationwide—or, perhaps, even worldwide, one day.

In the security world, there has been much investment for biometric systems for reliable recognition of individuals. More recently, mathematical methods of calibration have been developed and extended to provide evaluative approaches to facial comparison [19] and to speech comparison [20,21]. There is much scope for applying these methods to other areas where the comparison process is complex, such as the examples given in the previous paragraph.

To reiterate the argument made earlier: subjective assignments of probability are central to the forensic science paradigm but the driving principle for progress is that they should be conditioned not by casework experience, but by calibration under controlled conditions.

12. Conclusion

The closure of the FSS had two immediate consequences. First, police forces—particularly the larger ones—increased the scale of their in-house scientific and technical facilities; second, all of the remaining provision for forensic science services was privatized. It is tempting to believe that a consequence of increased privatization is increased scientific independence but, sadly, that is not the case. The police are still the customers and award their work to forensic providers through complex block contracting procedures. Contracts are renewed and renegotiated every few years and companies can go to the wall if they are beaten by their competitors in a process dominated by considerations of costs and speed.

I have explained how forensic science faces conflicting demands: fast, cheap, accurate testing at the investigative phase; careful, reflective, probabilistic analysis at the evaluative phase. I do not claim these to be dichotomous—indeed, there is a large measure of overlap—nevertheless, I believe that there is a strong case for placing the responsibility for investigative forensic science directly under police control and that for the provision of evidence for court purposes under the control of a body that answers only to the judicial system. This is one of the challenges for the future. However, to meet that challenge, we need to create an awareness among the judiciary of the modern paradigm. I have given examples of judgments that embody old-fashioned thinking with regard to evaluation: each of these has made our task of developing the vision of the balanced logical scientist that little bit harder.
Epistemology is the theory of knowledge: the critical study of its acquisition, validity, structure and its scope. The nature of forensic science is now firmly rooted in the Bayesian paradigm. The future is epistemology: establishing the structure for managing the knowledge that informs the probabilities that are central to the logical evaluation of scientific evidence. A world without ‘could have(s)’, ‘might have(s)’ and ‘consistent with(s)’; without either prosecution or defence bias; but logical, rational assessment of evidential value based on calibrated knowledge within a transparent set of circumstances and clearly stated propositions.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

Acknowledgements. I wish to thank my friends and colleagues Graham Jackson, Sue Pope, Angela Shaw, Christophe Champod, Charles Berger and Cedric Neumann who provided helpful comments. I also thank the two anonymous reviewers whose suggestions led to improvements in the paper.

Endnotes

1I hasten to emphasise that (i) I have never encountered a real-life investigator who fitted this stereotype. (ii) Guasher Briggs is a product of my imagination and has no resemblance to any real person.

References