Evidence evaluation: A response to the court of appeal judgment in R v T

Charles E.H. Berger[a], John Buckleton[b], Christophe Champod[c], Ian W. Evett[d,*], Graham Jackson[e]

[a] *Netherlands Forensic Institute (NFI), PB 24044, the Hague, the Netherlands*

[b] *Environmental Science & Research, PB 92021, Auckland, New Zealand*

[c] *Ecole des Sciences Criminelles, Institut de Police Scientifique, Université de Lausanne, BCH, 1015 Lausanne-Dorigny, Switzerland*

[d] *Forensic Science Service, London Laboratory, 109 Lambeth Road, London SE1 7LP, United Kingdom*

[e] *University of Abertay Dundee, DD1 1HG, United Kingdom*

This is a discussion of a number of issues that arise from the recent judgment in R v T [1]. Although the judgment concerned with footwear evidence, more general remarks have implications for all disciplines within forensic science. Our concern is that the judgment will be interpreted as being in opposition to the principles of logical interpretation of evidence. We re-iterate those principles and then discuss several extracts from the judgment that may be potentially harmful to the future of forensic science. A position statement with regard to evidence evaluation, signed by many forensic scientists, statisticians and lawyers, has appeared in this journal [2] and the present paper expands on the points made in that statement.

* Corresponding author.

1. Introduction

The judgment[†] in R v T [1] (freely available at www.bailii.org) has caused considerable concern amongst forensic scientists within the UK and overseas. Apart from those aspects of the judgment in relation to the case at hand - on which we do not comment in any regard -remarks were made of a more general nature and we consider it essential that we respond to them in the open literature. Para 1 of the judgment states:

> *The appeal… raised an issue of some importance in relation to the use of likelihood ratios in the provision of an evaluative opinion where the statistical data available were uncertain and incomplete.*

And at para 15:

> *The real issue in relation to footwear marks… was the use of likelihood ratios in forming an evaluative opinion on the degree of likelihood that a mark had been made by a particular item of footwear.*

The Court formed the view that there were inadequate data available for use in forming an evaluative opinion and this led to a view that seems encapsulated in the following extract from para 86:

> *An attempt to assess the degrees of probability where footwear could have made a mark based on figures relating to distribution is inherently unreliable and gives rise to a verisimilitude of mathematical probability based on data where it is not possible to build that data in a way that enables this to be done;… We are satisfied that in the area of footwear evidence, no attempt can realistically be made in the generality of cases to calculate the probabilities. The practice has no sound basis.*

And, in para 95:

[†] The dictionary offers the two spellings, "judgment" and "judgement". It is convention to use the former in the context of "legal judgment". The latter seems preferred when talking more generally about "human judgement". In quotations, we have followed the spelling as it occurs in the judgment.

In our judgment, an expert footwear mark examiner can therefore in appropriate cases use his experience to express a more definite evaluative opinion where the conclusion is that the mark "could have been made" by the footwear. However no likelihood ratios or other mathematical formula should be used in reaching that judgement1 for the reasons we have given.

It is this last sentence that raises our principal concerns: it implies that a likelihood ratio may not underpin the evaluation of a forensic scientist unless the two key probabilities can be assigned by purely statistical methods. In spite of the fact that the judgment states (para 32) that "We are solely concerned in this appeal with the use (of likelihood ratios) in relation to footwear evidence", nevertheless it seems to us inevitable that it will be cited to support court submissions in relation to other evidence types.

The reason for this paper is to explain that the evaluation of evidence for a court of law is not just a matter of "using likelihood ratios" but one of working to a set of principles that are founded on logic. To deny scientists the contemplation of the likelihood ratio -whether quantitative or qualitative - is to deny the central element of this logical structure.

The phrase "likelihood ratio" and the role of Bayes' theorem in establishing a logical framework for forensic inference now appear so widely in the forensic science literature that we do not consider it necessary to explain the mathematical background. In the next section, we give only a brief overview of the logical structure and we urge any reader who is unfamiliar with its basis to refer to standard texts - particularly Robertson and Vignaux [3], Aitken and Taroni [4] and Aitken, Roberts and Jackson [5]. Whereas we have chosen to address the readership of Science and Justice, papers dealing with the same judgment have been submitted to legal journals by Redmayne et al. [6] and by Robertson, Vignaux and Berger [7].

After outlining the basic principles we briefly explain how what has come to be known as the "Bayesian approach" has evolved. We then explain a little of the notion of probability and the relationships between probability, data, statistics, knowledge and experience. This leads to a discussion of presenting an evaluation to a court. Finally, we make remarks that seem relevant to us about related issues such as: the transposed conditional; the distinction between the Bayesian approach and Bayes' Theorem; and exposure to wider debate.

2. The logic of evaluating forensic science evidence in court

We emphasise that we are considering the evaluation of evidence at court; it often happens that a scientist is required to advise on the influence that his observations might have on the conduct of a police investigation. The requirements of this latter investigative function (described in [8]) are not considered in this paper.

We consider the case where a scientist is required to assist a court of law by providing an evaluative opinion with regard to the assistance that a set of scientific observations might provide in addressing matters that the court is deliberating. We confine ourselves to criminal law but believe that the principles extend analogously to the civil law. For the purpose of simplification we consider the situation where a single defendant is on trial for a stated offence. Then the scientist's evaluation should be governed by the following principles:

1. Interpretation must be carried out within a framework of circumstances. It is necessary for the scientist to make clear her understanding of those aspects of the circumstances that are relevant to the evaluation.
2. It is necessary for the scientist to consider her observations in the light of propositions that represent both the prosecution and defence views. The propositions must be clearly stated and are subject to change at the direction of prosecution, defence or the court.
3. It is necessary for the scientist to consider the probability that the observations would have been made given that the prosecution proposition were true; and the probability that the observations would have been made given that the defence proposition were true. The relative magnitude of these two probabilities, known as the likelihood ratio, determines the assistance provided by the observations in weighing the two propositions against each other.

The logical reasoning for these principles is founded on Bayes' theorem. The approach also invokes notions relating to the nature of probability that we explain in the next section. We note that these principles are the foundation of the standards for evidence interpretation that have recently been adopted in the UK by the Association of Forensic Science Providers [9].

In the UK and the USA, the development of Bayesian methods for interpreting forensic science evidence was initially inspired by a paper by Finkelstein and Fairley in the Harvard Law Review of 1970 [10]. From the latter half of the 20th century and the first decade of the 21st the papers that employ Bayesian reasoning in some form or another for forensic science problems number in the hundreds. However, research at the University of Lausanne, published in 1998 [11] demonstrated that this kind of reasoning had been employed some ninety years earlier. A central element in the notorious conviction for treason in 1894 of Alfred Dreyfus was a document that was known as the Bordereau. An opinion with regard to the authenticity of the Bordereau was given by Bertillon, regarded by many as one of the founding fathers of forensic science, who concluded that it was a forgery. As part of a review of the conviction of Dreyfus, the French Academy of Sciences set up a committee, composed of Darboux, Appell and Poincaré, which concluded in 1908 that the approach which Bertillon followed in forming his opinion had been inherently flawed. Although couched in language that today appears arcane, it is perfectly clear that their reasoning was what we would now call Bayesian - though that term was not in use then.

A similar perspective has been provided by the eminent legal commentators Bender, Nack and Treuer [12] in the German system of justice. They argue (§592) that there are three key questions that should guide any judge for the assessment of any piece of evidence:

(i) How common is the evidence if the main proposition were true?
(ii) How common is the evidence if the alternative proposition (converse of the main proposition) were true?
(iii) In which of the above cases is the evidence more likely? The asymmetry between both assessments provides the weight of evidence.

Although expressed in completely different terms, the connexion between these three questions and the principles that we stated earlier should be obvious. A similar position is adopted in McCormick [13], which explains (§185) how the likelihood ratio indicates the probative value of circumstantial evidence.

3. Probability

3.1. Aleatory and epistemic probabilities

Perhaps the most important notion to recognise in understanding probability is that all probabilities are conditional. To assign a probability for the truth of a proposition, one needs information. Even if one considers something as simple as the toss of a coin, to assign a probability to the outcome one needs to know something about the physical nature of the coin and how it is to be tossed. In general, the tossing mechanism (such as before the start of a sporting event between two teams) is chosen so as to maximise the uncertainty of the outcome. Given such information, most would be content with the statement that the probability of a head is 0.5. In another kind of uncertain situation one might, as the Court was clearly aware, study an appropriate collection of data to condition a probability. The Court was content with the notion that this is properly and reliably done in the case of calculating a DNA match probability.

However, there are still further situations where the notion of using a database appears unrealistic. Consider, for example, the question: What is the probability that Mary, Queen of Scots[‡], knew of the plot to murder her husband?

Clearly, to address such a question one would wish to address historical sources. These may be extremely extensive, complicated and of varying degrees of veracity. It is not surprising, then, that different people would give different answers to the question. For the “best” answer one would wish to look for the historian who is considered the most knowledgeable on the life of Mary. Similarly, one would not be surprised to see different historians having different views on the matter. It is this kind of example that makes it clear to us that probability is personal.

Probabilities such as that in the coin tossing example are known as aleatory: they arise in situations where the uncertainty derives solely from the randomness of the conditions that pertain. On the other hand, the latter example is of an epistemic probability: the uncertainty derives solely from limitations of knowledge.

A DNA match probability, when it is calculated from undisputed genetic principles and using a database that is indisputably relevant to the case at hand, could be argued to be a function of purely aleatory uncertainty (though, as we see in the next

‡ Readers unfamiliar with this queen will find a short biography at http://en.wikipedia.org/wiki/Mary_Queen_of_Scots.

section, this is a matter for debate). On the other hand, when a handwriting expert is considering the comparison of features between two sets of handwriting then any probabilities assigned to variations in features are purely epistemic. In more general usage, it is customary to call aleatory probabilities "objective", in the sense that their magnitudes are independent of human judgement. Epistemic probabilities, on the other hand, are called "subjective" in the sense that they are personal, depending as they do on human judgement. Any reader who seeks a comprehensive non-mathematical introduction to the nature of probability could do no better than the book Understanding Uncertainty by Lindley [14] who introduces the concept as a measure of "degree of belief".

Whereas much of the logic of statistical inference in the 20th century concerned itself entirely with aleatory probabilities, there grew an increasingly substantial school of thought that was founded on the realisation that it was possible to apply all of the richness of statistical and probabilistic reasoning to epistemic, or subjective, probabilities. Gradually, and for reasons that are not entirely clear (see, in particular, Fienberg [15]), this came to be known as the Bayesian approach.

3.2. Probability and statistics

The Court in R v T appeared to take the view that assigning probabilities is a matter of statistics alone. For example, the judgment says, at para 77:

> *It is clear that in DNA cases, there has been for sometime a sufficient statistical basis that match probabilities can be given.*

And, at para 78:

> *However, no case was drawn to our attention which suggests that a mathematical formula is appropriate where it has no proper statistical basis.*

And, at para 80:

> *An approach based on mathematical calculations is only as good as the reliability of the data used. The acceptance of a mathematical approach to the calculation of a match probability in DNA cases is based on the*

reliability of the statistical database, though an element of judgment is required (emphasis added).

These extracts emphasise the notion of a probability being assigned from data and a statistical calculation of some kind. However, even here - as the Court recognised in the italicised clause in the third extract - there is an essential contribution to be made by personal judgement. Underlying the calculation of a DNA match probability is a statistical/population genetic model that cannot be anything other than an approximation to the complexities of the real world. The application of the model will incorporate estimates of unknown parameters that are themselves approximations. Even with DNA the "element of judgment" might be wider than the Court appreciated. For example, if there is a full ten locus match between the DNA profile from a crime sample and the DNA profile of a known individual, in the UK the scientist will report a match probability of one in a billion. In the USA it would not be surprising to find the match probability, for exactly the same profile, reported as one in a trillion - even if the same database were used. There is much more to this issue than a simple statistical calculation, see, for example, Evett et al. [16] and Buckleton [17]. Indeed, it takes judgement to decide how to model a phenomenon and individuals may disagree about the appropriateness of any particular model.

Para 1 of the judgment (see Introduction) refers to data that were "uncertain and incomplete". The Court clearly felt this to be so in relation to the case at hand, but the essential point to recognise here is that whenever we are making an inference from a sample the data are always an incomplete representation of the full picture; furthermore, their relevance is a matter of judgement and the uncertainty that concerned the Court is an unavoidable feature of such inference. The probability that is quoted then will inevitably be a personal probability and the extent to which the data influence that probability will depend on expert judgement. This is not a process that can be governed strictly by mathematical reasoning but this does not make it any less "scientific": scientists are called on to exercise personal judgement in all aspects of their several pursuits. See Taroni et al. [18] for a cogent discussion of this.

In relation to footwear evidence, the judgment says, at para 86:

> *…we have concluded that there is not a sufficiently reliable basis for an expert to be able to express an opinion based on the use of a mathematical formula.*

It is our understanding of the judgment that the formula referred to here was one that was used in the individual case to calculate a relative frequency. As we have said, we intend no reference to the specifics of the case itself. More generally, it is our position that the scientist considers both of the probabilities that constitute the likelihood ratio and, if data exist that are of some relevance then it is unreasonable to deny her the facility to take account of that data by exercising reasonable scientific judgement. This is not a matter of a "mathematical probability" (which we take to mean an aleatory probability) but an epistemic probability that is conditioned by all of the data and information that are available to the expert. The judgment in R v Weller [19] is in favour of this kind of view. The recognition of subjective probabilities as a sound basis for offering expert opinion should not be misconstrued as being an open door to unstructured reasoning and ad hoc "guesses". Any appraisal offered, based on the spectrum of knowledge an expert may call upon, should be transparent and subject to peer review and challenge. We stress that the probability may be informed not only by systematic research and data but also by expert judgement, though we recognise that the notion of what constitutes "expert judgement" would justify a separate paper entirely.

In addition to its position with regard to assigning probabilities by statistics, the Court expressed concerns about database size. Para 84 says:

> *Use of the FSS's own database could not have produced reliable figures as it had only 8122 shoes whereas some 42 million are sold every year.*

Yet the Court was clearly convinced of the robustness of DNA statistics that yield match probabilities of the order one in a billion based on databases that are typically made up from a few hundred individuals. Is a database of 8122 shoes too small? We don't think so - one thing that seems inescapable is that it is larger than a database of zero. Any knowledge is better than no knowledge and the Court surely could not be suggesting that the data should be ignored. Of course, it must be treated with caution and its limitations should be recognised - but that is what scientists are trained to do. Para 86 of the judgment says that "in the area of footwear evidence, no attempt can realistically be made in the generality of cases to use a formula to calculate the probabilities." But we hope that we have explained that probabilities are assigned rather than calculated: that means using whatever data are available sensibly and exercising judgement with regard to limitations of size and representativeness.

3.3. The likelihood ratio

We have seen that the Court of Appeal is content with the notion of likelihood ratios when they perceive them, as with DNA, as deriving from purely mathematical methods applied to reliable databases. Although we have explained that this view is illusory, for the sake of discussion, we will call this "Type A" scientific evidence: an unattainable ideal where the probabilities that form the evaluation are aleatory.

At the other extreme is scientific evidence of the kind that the Court considered in reference to the previous judgment in R v Atkins and Atkins [20] which concerned facial comparison evidence. We will call this extreme "Type C" scientific evidence. Para 93 of the judgment includes:

> *The court was again holding that evaluative evidence could be given; there was no issue as to the use of a likelihood ratio, as it was clear that there was no statistical database.*

The inescapable conclusion to draw from this is that the Court saw the likelihood ratio as something that derives only from a calculation of some sort. However, as we have explained, the likelihood ratio is the ratio of two probabilities and there is nothing in the logical framework that we have described that demands those two probabilities to be aleatory. Indeed, the foundation of the Bayesian paradigm (from which the forensic "Bayesian approach" derives) is that the logic works equally well with purely epistemic probabilities as it does with aleatory probabilities: the depth and richness of the Bayesian paradigm derives entirely from this profound tenet. This is what is "Bayesian" about the Bayesian approach.

With regard to this type C evidence we would still consider it necessary for the expert to adhere to the logical principles stated in section 1: if she did otherwise then she would have abandoned logic. The principles lead inevitably to the two key probabilities and we entirely accept that the expert will not be able to assign precise values to them. Nevertheless, it is essential for the expert to address qualitatively their relative magnitudes. Under which of the two propositions is the evidence more probable? That is the proposition which is better supported by the scientific observations. The likelihood ratio is still central, even if the probabilities can only be qualitatively assessed though it is worth reiterating that the expert needs to be able to state the basis for such an assessment. We fully endorse the notion of transparency in this regard: it is not acceptable merely to cite years of service, numbers of previous cases and to expect the Court to accept ipse dixit.

Between the extremes of types A and C we have evidence types where the scientist's opinion will depend in part on data, part on published material and part on knowledge and judgement. Not surprisingly, we call this type B and it is to this category that footwear evidence belongs along with, in particular, fibres and glass evidence. In relation to the latter, the judgment cited R v Abadom [21] and said, at para 99:

> *The decision of the court in Abadom as long ago as 1982 explained the importance of referring to all the material so that the cogency and probative value of the conclusions can be tested and evaluated by reference to it.*

Whereas the judgment is addressing a disclosure issue here, it is a recognition of the multi-faceted approach to a type B evidential assessment. However, the judgment also says at para 91:

> *It is important to note as happened in R v Abadom… that an expert can give an opinion using a statistical database by simply using that database and expressing an opinion by reference to it, without recourse to the type of mathematical formula used in this case or to any form of Bayes theorem.*

The "mathematical formula" that the judgment refers to appears to be a relative frequency calculation that was put on the file for this particular case and we repeat that we make no comment on that case. However, we are not comfortable with the lack of clarity as to how the expert might, without performing any calculations, use data to inform an opinion. The idea that an expert might just look at a database and, by some undefined mental process, form an opinion does not seem right to us and seems counter to the requirement for transparency that is rightly called for elsewhere in the judgment. At the very least, simple calculations of proportions would seem desirable.

4. Presenting an evaluation at court

4.1. The verbal scale

Having followed the principles of evidence evaluation and arrived at a likelihood ratio, whether qualitative or quantitative, there remains the question of how this is to be presented to a court. The notion that evidence might support a particular proposition is commonly encountered in the courtroom. For example, these are just a small sample from the Court of Appeal judgments:

In R v Denis John Adams [22]

> *None of these provides any support for Mr Thwaites' proposition.*

In R v Gilfoyle [23]

> *…the material before him demonstrated convincing support for the deceased having taken her own life.*

In R v Puaca [24]

> *But what they did say, and say most forcefully, was that there was no pathological evidence to support Dr Heath's view.*

If the probability of the observations is greater given the prosecution proposition than it is given the defence proposition it is intuitively reasonable and logically justifiable to state that the observations support the former. The notion of something being added to one of the scales of justice is apt. However, we are well aware that some items of evidence might contribute to the resolution of the two propositions more than others: we tend to think of DNA evidence as having the potential to be extremely strong; whereas evidence from a basic visual comparison of hairs, for example, might be relatively weak.

In those cases where a quantitative likelihood ratio has been calculated, the interpretation of R v Doheny and Adams [25] currently in England and Wales suggests that it is the number alone that should be put to the jury. However, the solution for all those cases where the likelihood ratio is qualitative is the notion of using some kind of verbal qualifier to convey the strength of the support that the observations bring. The approach is endorsed by Kaye et al. [26] (§15.6). Forensic scientists have always recognised the desirability of maintaining some sort of consistency of language between experts, disciplines and organisations; this, in turn, inspires the wish to standardise, as far as reasonable, on a small number of qualifiers. This notion of a verbal scale was recognised as valid in R v Atkins and Atkins [20]. R v T accepts the principle but denies that it has anything to do with the likelihood ratio (para 93). On the contrary, it is our position that the concept of the likelihood ratio should be central to every logical evaluation, irrespective of the extent to which the expert's judgement has been informed by data.

A verbal scale has been in use in the Forensic Science Service (FSS) for many years and the version that was current at the time of the trial is reproduced in para 31 of the judgment. The expressions of support range from "weak support", through "strong support" to "extremely strong support". Because of the desire to achieve a measure of consistency across all evidence types, broad ranges of values of the likelihood ratio have been assigned to the various expressions. But this should not be taken to mean (as implied at para 31 of the judgment) that a numerical value must be derived in a given case in order that a verbal expression might be selected - even if that were preferable whenever possible. In the case of type C assessments, the verbal scale restricts the expert to the specified range of qualifiers and provides a rough ordering that, at best, implies that "strong" is stronger than "moderate" but not as strong as "very strong". We recognise that there is much scope for improvement - but the limitations derive not from the verbal scale per se, rather from the complexities of the observations and the lack of data for interpreting them.

Where type B assessments are concerned the extent to which the likelihood ratio is quantified will vary from case to case, depending on the expert's assessment of the quality of whatever data are available. But it is not the verbal scale that is driving things. The assessment comes first, the decision about a verbal qualifier comes later.

4.2. The use of "could have"

The Court's endorsement of the phrase "could have" runs counter, in our view, to what is said in R v Puaca [24]. We quote from para 42 of that judgment:

> *Whereas "inconsistency" is often probative, the fact of consistency is quite often of no probative value at all… We consider that there is a very real danger in adducing before a jury dealing with a case such as the present evidence of matters which are "consistent" with a conclusion, at least unless it is to be made very clear to them that such matters do not help them to reach the conclusion. If it is introduced in evidence, and particularly if it is given some emphasis, a jury may well think that it assists them in reaching a conclusion: for why otherwise are they being told about it?*

In our opinion, the same observations may be made in relation to the use of "could have". To say that a shoe "could not have" made a mark has clear probative value. To say that a shoe "could have" made a mark is unbalanced unless one adds that it is also true that some other shoe could have made the mark. In effect, one is addressing two propositions but giving no indication as to which is supported by the evidence. The judgment says at para 73:

> *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".*

On the contrary, our view is that "could have made" is valueless for expressing evidential weight and this view is shared by others. For example, Lyon and Koehler advocate the use of a likelihood ratio instead of poorly informative terms such as "consistent with" for expert witnesses offering opinions in sexual abuse cases [27].

In most situations "could have" is no more than a statement of the blindingly obvious such as: The DNA in the crime sample has the same profile as that of the suspect, therefore it could have come from him.

The latter part of this sentence does not warrant the status of an expert opinion; whereas the jury will understand it, it does not convey any appreciation of the weight of evidence that the observations provide to assist them in their task. "Moderate support" is an attempt, however imperfect, to convey some impression of weight of evidence.

In a public inquiry that followed the miscarriage of justice against Guy Paul Morin in Canada, the Kaufman report [28] revealed major shortcomings in laboratory work and interaction between experts, police and judges with regard to hair and fibre evidence. The report noted that the use by experts of vague conclusions such as “the hair might come from…”, “the fibres are consistent with…” or “the fibres match”, which lend themselves to a range of interpretations, was one of the causes of the attribution of disproportionate probative value to evidence which actually made only a limited contribution to the establishment of the facts.

The recent report of the US National Academy of Sciences [29] also drew attention to the dangers of using tendentious, or even fallacious, terminology, to the extent that its authors felt the need to make a specific recommendation on the subject. The report specifically referred to Aitken and Taroni [4], Evett et al. [30] and Evett [31] as “providing the essential building blocks for the proper assessment and communication of forensic findings”.

The warmth with which the Court endorsed a “could have” statement is in stark contrast with many other rulings in the DNA field (e.g. [32,33] Section III and more generally in Kaye et al. ([26] (§15.4)). These rulings hold that a statement of non-exclusion, which we equate with the “could have” statement, is not admissible unless a more precise expression of weight of evidence is available.

4.3. The transposed conditional

The second logical principle, in a footwear case for example, would direct the scientist to address two questions, one of which would be: What is the probability that those observations would have been made if the suspect’s shoe made the crime mark?

For an understanding of the logic of evidence evaluation, it is necessary to recognise that this is not the same as: What is the probability that the suspect’s shoe made the crime mark given that those observations were made?

Failure to recognise the fundamental difference between those two probabilities (and the answers that they generate) is known to statisticians as the “fallacy of the transposed conditional”. Within the context of forensic science, it was dubbed the “prosecutor’s fallacy” by Thompson and Schuman [34]. See also [35].

The dangers of the fallacy have been recognised by the Court of Appeal, within the context of DNA evidence, in the judgments in R v Deen [36] and R v Doheny and

Adams [25]. However, it appears not to have been recognised by the Court in R v T as may be seen from the following example in para 33:

> *In the present case it was expressed (i.e. the likelihood ratio) as the probability that the Nike trainers owned by the appellant had made the marks discovered at the scene divided by the probability that the Nike trainers had not made the marks.*

This is not right. The likelihood ratio in this context is the probability of the observations given that the Nike trainers owned by the appellant had made the marks; divided by the probability of the observations given that the Nike trainers had not made the mark. The likelihood ratio is always a function of the probabilities of the observations; the quotation is expressed in terms of the probabilities of the propositions.

Para 91 of the judgment says:

> *If there was a sufficient database in footwear cases an expert might be able to express a view reached through a statistical calculation of the probability of the mark being made by the footwear, very much in the same way as in the DNA cases…*

This is yet another of the many instances in the judgment where a transposed conditional is implied. The scientist should not address the "probability of the mark being made by the footwear" rather, the probability of the observations given that the mark was made by the footwear and, of course, the probability of the observations given the appropriate defence proposition.

4.4. The US examiners verbal scale

The judgment (para 64) cites a verbal scale that is followed by shoe and toolmark examiners in the USA; this is quite different from that described in Section 4.1. A report from the marks working group of the European Network of Forensic Science Institutes (ENFSI) [37] includes a similar scale.

The US examiners' scale embodies phrases and notions that have been abandoned in the UK because of their potential to confuse. In particular, one of the phrases is based on the use of "could have", which we have already dealt with in the previous section. Another category is to say that the shoe "probably made" the mark.

But this is clearly a statement about the probability of the truth of a proposition. There are two ways in which a statement can be arrived at. The first, as we have seen is to commit the fallacy of the transposed conditional. The second is to follow a path that has been widespread practice in paternity testing for many years - and that is to assume that, based on all of the evidence other than the scientific observations, the prosecution proposition is just as likely to be true as the defence proposition. The "probably made" conclusion then follows from an application of Bayes' Theorem in a manner that was rejected in R v Denis John Adams [22]. More details on this kind of approach can be found in [38] and [39].

Indeed, were a statement of the kind "probably came from" used in a DNA case, it would violate the judgment in R v Doheny and Adams [25]. It is, perhaps, surprising that legal practitioners in the USA have apparently not realised that the "probably made" style of opinion is susceptible to challenge either on the basis of logic or on the basis of inappropriate assumptions about the other evidence in the case.

Another category in the US examiners' scale is a categoric opinion of the kind "this shoe made that mark". Such opinions are, indeed, expressed by UK examiners, but they do not form part of the verbal scale, which is designed only for corroborative weight of evidence.

5. Other issues

5.1. The Court of Appeal and Bayesian inference

Para 90 of the judgment includes:

> *It is quite clear therefore that outside the field of DNA (and possibly other areas where there is a firm statistical base), this court has made it clear that Bayes theorem and likelihood ratios should not be used.*

Para 89 explains that this view derives from R v Doheny and G Adams [25] and R v Denis John Adams [22]. However, we should point out that in the Doheny and G Adams trials, the weight of the DNA evidence was conveyed by means of a match probability (the phrase "random occurrence ratio", which appears to have been coined for reasons that are far from clear has, thankfully, not become general use) not a likelihood ratio nor was Bayes' theorem used in any sense in either case.

In R v DJ Adams, defence encouraged the jury to assign probabilities to each item of non-scientific evidence in the case (geography, eye-witness, alibi etc.) and to combine them by means of Bayes' theorem. In no sense was Bayes' theorem used to evaluate the scientific evidence in the case; nor was a likelihood ratio quoted - again the evaluation of the DNA evidence was based on a match probability.

It is a matter of concern to us that an exercise where Bayes' theorem was employed for the non-scientific evidence in a trial has been the cause of a direction against the notion of applying the principles of logic to the evaluation of scientific evidence.

It may be useful to address a point that may have been a source of confusion: there is a distinction to be made between using "Bayes theorem" and following a "Bayesian approach". To do the former, it is necessary to assign prior probabilities and manipulate them, via the likelihood ratio, to arrive at posterior probabilities. To follow the Bayesian approach, however, one only has to follow the three principles detailed earlier - principles that can, in fact, be rationalised without using Bayes theorem. As we have explained, the adjective "Bayesian" has derived from the subjective view of probability, rather than the mechanics of using Bayes' theorem.

5.2. Exposure to wider debate

Para 52 of the judgment refers to the adoption of the Bayesian approach by the Association of Forensic Science Providers in the UK, adding:

> *Despite inquiries made by us, it is not clear to what extent, if any, it was subject to wider debate outside the forensic science community.*

Apart from the substantial literature that has appeared in journals in the forensic science, legal and statistical fields, the FSS has carried out various consultative exercises. Considerable progress was made towards implementing the principles of the Bayesian approach within the FSS through the Case Assessment and Interpretation (CAI) project and the work was described in a series of papers, including, for example, Evett et al. [30]. As the project developed, the Crown Prosecution Service (CPS) was consulted across England and Wales about its implication for casework. The new statement format that is described in [30] was not finalised until agreement with the CPS. Starting from around 2000 and up to 2006 a lecture, primarily relating to DNA but from a Bayesian perspective, was given regularly to Continuing

Education Seminars of the Judicial Studies Board. In this way, several hundred judges from England and Wales were exposed to the ideas and the feedback was positive.

As we said in the Introduction, papers involving the Bayesian approach to forensic science number in the several hundreds. Whereas most will be in the forensic science literature, others have been published in legal journals and others in the statistical literature - most notably the journals of the Royal Statistical Society.

Following the appeals in R v Clark [40], the Royal Statistical Society set up a working group to deliberate on the use of statistics in legal trials. The group comprises statisticians, forensic scientists, academic lawyers, barristers and a judge. A report by Aitken et al. [5] explains the Bayesian approach in some detail.

The eighth International Conference on Forensic Inference and Statistics is to be held in Seattle, USA, in July 2011. This series, started in 1990, and alternating its venues between the USA and Europe, aims to bring together forensic scientists, statisticians and lawyers to discuss matters relating to the interpretation of evidence, particularly in criminal proceedings.

6. Conclusion

From the viewpoint of public perception, technological advances have epitomised the evolution of forensic science over the last 20 years. However, the development of its logical foundations and their practical application that have taken place over a comparable period will, we believe, be of even greater importance. The main thrust of the evolution of what has come to be known as the Bayesian approach has centred on the UK, Europe and Australasia. This is a movement that seeks to contribute to improvement, innovation and a search for how science can best inform the proceedings of a criminal justice system.

Forensic science is moving forward (in some countries more than in others) from a craft towards a science, and an increasing level of professionalism is required of those acting in it. Every claim of improvement inevitably entails an admission that things were not as good before. This however should never impede the possibility of improvement. The Court rightly observed that the adoption of the ideas of the Bayesian approach and, more specifically the CAI model, has been patchy in those countries that are leading the movement. Furthermore, practitioners in the USA seem to be largely unaffected by it. It is encouraging that the Board of ENFSI - an organisation with 58 member laboratories in 33 countries - has engaged itself to implement the principles outlined in the position statement [2] in its member laboratories. It is regrettable that the judgment in R v T will inevitably be seen to be, and used as, a weapon for defending the status quo or even a return to pre-scientific notions.

The judgment sees the likelihood ratio as something that inevitably is founded on statistical calculations and data. We hope that we have adequately explained that this is not the case. The likelihood ratio is a central element in a framework for evaluative opinion that is both logical and balanced. Furthermore, the probabilities that the scientist is directed to address are always founded (even with DNA) on personal judgement. This is not a bad thing, it is an inescapable feature of science and it is why science is still a human endeavour. It is also why there is always scope for honest disagreement between experts; as we have seen, probabilities are personal, even when informed by data. The Bayesian approach does not give the expert the right to make up numbers as she pleases, but rather the duty to make sure the probabilities involved obey the laws of probability. Overall, the duty of transparency prevails and the expert should articulate the basis for her probabilistic assignments in a given case.

In R v Denis John Adams [22] the Court of Appeal came out strongly against the use of Bayes' theorem by the jury for evaluating non-scientific evidence. There is a serious danger that the ruling in R v T will be seen as a proscription of the Bayesian approach to evaluating scientific evidence.

Acknowledgement

The authors are grateful to David Kaye and Paul Roberts for comments on a draft of the paper.

References

[1] R. v T, Court of Appeal - Criminal Division, EWCA Crim 2439, 2010.

[2] Expressing Evaluative, Opinions: a position statement, Science & Justice 51 (2011) 1-2.

[3] B. Robertson, G.A. Vignaux, Interpreting Evidence - Evaluating Forensic Science in the Courtroom, John Wiley & Sons Inc., Chichester, 1995.

[4] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd ed ., John Wiley & Sons Ltd., Chichester, 2004.

[5] C. Aitken, P. Roberts, G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. London: The Royal Statistical Society, in press. http:// www.rss.org.uk/statsandlaw.

[6] M. Redmayne, P. Roberts, C.G.G. Aitken, G. Jackson, Forensic Science Evidence in Question. Criminal Law Review (2011) 347-356.

[7] B. Robertson, G.A. Vignaux, C.E.H. Berger, Extending the Confusion About Bayes, Modern Law Review (2011) 444-455.

[8] G. Jackson, S. Jones, G. Booth, C. Champod, I.W. Evett, The nature of forensic science opinion - a possible framework to guide thinking and practice in investigations and in court proceedings, Science & Justice 46 (2006) 33-44.

[9] Association of Forensic Science Providers, Standards for the formulation of evaluative forensic science expert opinion, Science & Justice 49 (2009) 161-164.

[10] M.O. Finkelstein, W.B. Fairley, A Bayesian approach to identification evidence, Harvard Law Review 83 (1970) 489-517.

[11] F. Taroni, C. Champod, P. Margot, Forerunners of Bayesianism in early forensic science, Jurimetrics Journal 38 (1998) 183-200.

[12] R. Bender, A. Nack, W.-D. Treuer, Tatsachenfeststellung Vor Gericht - glaubwürdigkeits- Und Beweislehre Vernehmungslehre, 3rd ed.Verlag C.H. Beck, München, 2007.

[13] K.S. Broun, R.P. Mosteller, G.E. Dix, E.F. Roberts, I.E. J, J.W. Strong, D.H. Kaye, E. Swift, Mccormick on Evidence, Thompson/West Publishing Co., Eagan, MN, 2006.

[14] D.V. Lindley, Understanding Uncertainty, John Wiley & Sons, Inc., Hoboken, 2006.

[15] S.E. Fienberg, When did Bayesian inference become “Bayesian”? Bayesian Analysis 1 (2006) 1-40.

[16] I.W. Evett, L.A. Foreman, G. Jackson, J.A. Lambert, DNA profiling: a discussion of issues relating to the reporting of very small match probabilities, Criminal Law Review (May 2000) 341-355.

[17] J.S. Buckleton, Population genetic models, in: S.J. Walsh, C.M. Triggs, J.S. Buckleton (Eds.), Forensic DNA Evidence Interpretation Methods and Interpretation, CRC Press, Boca Raton, 2004, pp. 65-122.

[18] F. Taroni, C.G.G. Aitken, P. Garbolino, De Finetti’s subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists, Science & Justice 41 (2001) 145-150.

[19] R. v Peter Weller, UK Court of Appeal, EWCA Crim 1085, 2010. [20] R. v Dean Atkins, Atkins Michael, Court of Appeal - Criminal Division, EWCA Crim 1876, 2009.

[20] R. v Dean Atkins, Atkins Michael, Court of Appeal - Criminal Division, EWCA Crim 1876, 2009.

[21] R. v Abadom, Court of Appeal - Criminal Division, [1983] 1 Wlr 126, [1983] 1 All Er 364.

[22] R. v Denis John Adams, Court of Appeal - Criminal Division, EWCA Crim 222, 1996.

[23] R. v Norman Edward Gilfoyle, Court of Appeal - Criminal Division, EWCA Crim 81, 2000.

[24] R. v Steven Puaca, Court of Appeal - Criminal Division, EWCA Crim 3001, 2005.

[25] R. v Alan James Doheny, R. v Gary Adams, Court of Appeal - Criminal Division, EWCA Crim 728, 1996.

[26] D.H. Kaye, D.E. Bernstein, J.L. Mnookin, The New Wigmore, a Treatise on Evidence: Expert Evidence, 2nd ed., Aspen Publishers Inc., New York, 2011.

[27] T.D. Lyon, J.J. Koehler, The relevance ratio: evaluating the probative value of expert testimony in child sexual abuse cases, Cornell Law Review 82 (1996) 43-78.

[28] F. Kaufman, Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin, Ministry of the Attorney General, Ontario, Toronto, 1998.

[29] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington, D.C., 2009

[30] I.W. Evett, G. Jackson, J.A. Lambert, S. McCrossan, The impact of the principles of evidence interpretation on the structure and content of statements, Science & Justice 40 (2000) 233-239.

[31] I.W. Evett, The theory of interpreting scientific transfer evidence, Forensic Science Progress 4 (1990) 141-179.

[32] State of Western Australia v Dair, WADC 157, 2006.

[33] W.C. Thompson, Motion to Exclude DNA Evidence in United States V Cuff, 37 F. Supp. 2d 279 (S.D.N.Y. 1999), BNA Criminal Practice Guide (January 12, 2000) 14-18.

[34] W.C. Thompson, E.L. Schumann, Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy, Law and Human Behavior 11 (1987) 167-187.

[35] D.J. Balding, P. Donnelly, The prosecutor's fallacy and DNA evidence, Criminal Law Review (1994) 711-721.

[36] R v Deen, Court of Appeal - Criminal Division, The Times, January 10th.

[37] ENFSI Expert Working Group Marks Conclusion Scale Committee, Conclusion scale for shoeprint and toolmarks examination, Journal of Forensic Identification 56 (2006) 255-280.

[38] A. Biedermann, F. Taroni, P. Garbolino, Equal prior probabilities: can one do any better? Forensic Science International 172 (2007) 85-93.

[39] A. Biedermann, F. Taroni, C.G.G. Aitken, Letter to the editor - re: conclusion scale for shoeprint and toolmarks examinations, J. For. Ident. 56 (2) (2006) 685-6938 (with response).

[40] R. v Sally Clark, Court of Appeal - Criminal Division, EWCA Crim 1020, 2003.