

Forensic Science Evidence in Question

By

**Mike Redmayne; Paul Roberts; Colin
Aitken and Graham Jackson**

Reprinted from Criminal Law Review
Issue 5, 2011

Sweet & Maxwell
100 Avenue Road
Swiss Cottage
London
NW3 3PF
(Law Publishers)

SWEET & MAXWELL

Forensic Science Evidence in Question

Mike Redmayne

Professor of Law, LSE

Paul Roberts

Professor of Criminal Jurisprudence, University of Nottingham

Colin Aitken

Professor of Forensic Statistics, University of Edinburgh

Graham Jackson

Professor of Forensic Science, Abertay University

☞ Forensic evidence; Forensic science; Statistical evidence

How should forensic scientists and other expert witnesses present their evidence in court? What kinds and quality of data can experts properly draw on in formulating their conclusions? In an important recent decision in *R. v T*¹ the Court of Appeal revisited these perennial questions, with the complicating twist that the evidence in question incorporated quantified probabilities, not all of which were based on statistical data. Recalling the sceptical tenor of previous judgments addressing the role of probability in the evaluation of scientific evidence,² the Court of Appeal in *R. v T* condemned the expert's methodology and served notice that it should not be repeated in future, a ruling which rapidly reverberated around the forensic science community causing consternation, and even dismay, amongst many seasoned practitioners.³ At such moments of perceived crisis it is essential to retain a sense of perspective.

There is, in fact, much to welcome in the Court of Appeal's judgment in *R. v T*, starting with the court's commendable determination to subject the quality of expert evidence adduced in criminal litigation to searching scrutiny. English courts have not consistently risen to this challenge, sometimes accepting rather too easily the validity of questionable scientific techniques.⁴ However, the Court of Appeal's reasoning in *R. v T* is not always easy to follow, and there are certain passages in the judgment which, taken out of context, might even appear to confirm forensic scientists' worst fears. This article offers a constructive reading of *R. v T*,

¹ *R. v T* [2010] EWCA Crim 2439; [2011] 1 Cr. App. R. 9.

² *Adams (No.1)* [1996] 2 Cr. App. R. 467 CA (Crim Div); [1996] Crim. L.R. 898; *Adams (No.2)* [1998] 1 Cr. App. R. 377 CA (Crim Div). For discussion, see P. Roberts and A. Zuckerman, *Criminal Evidence*, 2nd edn (Oxford: OUP, 2010), pp.159–163.

³ See I. Evett et al, "Expressing Evaluative Opinions: A Position Statement" (2011) 51 *Science & Justice* 1; C. Berger, J. Buckleton, C. Champod, I. Evett and G. Jackson, "Evidence Evaluation: A Response to the Appeal Court Judgment in *R v T*" (2011) 51 *Science & Justice* (forthcoming).

⁴ Compare, e.g. *Dallagher* [2002] EWCA Crim 1903; [2003] 1 Cr. App. R. 12; [2002] Crim. L.R. 821, relating to the admissibility of ear-print evidence.

emphasising its positive features whilst rejecting interpretations which threaten, despite the Court of Appeal's best intentions, to diminish the integrity of scientific evidence adduced in English criminal trials and distort its probative value.

What was the evidence?

T was convicted of murder partly on the basis of footwear mark ("shoeprint") comparison evidence provided by the Forensic Science Service (FSS). Having considered the rarity of the sole pattern, shoe size, and indications of wear and damage, the FSS expert concluded that there was "a moderate degree of scientific support" for the proposition that the marks at the crime scene were made by T's Nike trainers. At trial, the expert stated under cross-examination that about 786,000 pairs of trainers with the relevant sole pattern had been distributed by Nike in the period 1996–2006, of which about 3 per cent would be size 11—T's shoe size. As a point of comparison, the expert explained, about 300 million pairs of sports shoes had been sold in the United Kingdom during the last seven years.

In order to arrive at the conclusion stated in his report, indicating "moderate support" for the prosecution's contention, the expert had used a reasoning methodology employing likelihood ratios, in accordance with standard FSS operating procedure.⁵ A likelihood ratio is a way of measuring the strength of evidence. Likelihood ratios compare two probabilities conditioned on a pair of alternative assumptions: for example, the probability of finding a matching shoeprint if T's shoes made the marks, and the probability of finding a matching shoeprint if T's shoes had *not* made the marks. Among other things, likelihood ratios reflect the insight that, even if there is a perfect match between the crime scene mark and a suspect's shoe, *it does not follow that the suspect's shoe made the mark*. Imagine that the crime scene print was made by a brand new shoe fresh out of the box. The suspect has a new pair of shoes of the relevant type, but so do many others who have just purchased the same popular shoe. The evidential value of the match depends on just how popular this type of shoe is and, correspondingly, how many pairs are currently in circulation. The likelihood ratio is able to quantify this vital dimension of the evidence.

In *R. v T*, the expert assumed that, if the prosecution's proposition was true and T's shoe had made the questioned marks, the probability of a matching pattern was 1, i.e. the pattern on the shoe's sole would definitely match the pattern observed in the crime scene marks. The probability of a matching pattern if the related defence proposition was true reflects the relative frequency of the sole pattern in the relevant population. In the present case, the expert relied on a database indicating that 20 per cent of shoes examined in FSS laboratories during the period 2005–2007 bore the matching sole pattern. Thus, the probability of finding a matching pattern if T's shoes had *not* made the mark was calculated as 0.2. The ratio of these two probabilities is $1/0.2$, giving a likelihood ratio of 5 for the pattern evidence.⁶ Spelling out the implications of the likelihood ratio, this means that the odds in favour of the proposition that T's shoes left the marks would be five times greater after finding a match between the pattern on T's trainers and the crime

⁵ For an accessible introduction to the likelihood ratio approach to forensic science evidence, see M. Redmayne, *Expert Evidence and Criminal Justice* (Oxford: OUP, 2001), Ch.3.

⁶ Because $1/0.2 = 5$.

scene marks than they were before the pattern evidence was considered. The expert then calculated three further likelihood ratios relating to the size, wear, and damage sustained by the Nike trainers recovered from T. For example, T's trainers and the ones that left the crime scene marks had similar patterns of wear. On the assumption that about half of the relevant trainers would share this characteristic, the likelihood ratio for wear was $1/0.5 = 2$. Finally, the expert combined the four likelihood ratios according to the simple formula $\text{Pattern} \times \text{Size} \times \text{Wear} \times \text{Damage}$, to produce a composite likelihood ratio of roughly 100.⁷ In other words, the footwear mark evidence taken as a whole increased the odds that T's trainers had left the marks at the scene one hundred-fold.

None of these calculations appeared in the expert's written report or were presented in his oral testimony at trial. Instead, the expert utilised a conventional scale developed by the FSS to translate likelihood ratios into a verbal summary of the strength of the evidence which, it is thought, is easier for juries to understand than a raw number. According to this scale, a likelihood ratio between 10 and 100 should be expressed as "moderate support" for the proposition that T's shoes had left the marks. Had the likelihood ratio been slightly larger, the expert would have been able to say that the footwear mark evidence provided "moderately strong" support for the prosecution case.

What did the Court of Appeal say?

Allowing T's appeal against conviction, the Court of Appeal found fault with several aspects of the footwear mark evidence adduced at T's trial. Most obviously, there was an apparent discrepancy between the (undisclosed) likelihood ratio calculations relied on by the expert in producing his written report and the Nike distribution figures quoted by the expert during cross-examination at trial. The figures quoted in court, and repeated during the trial judge's summing up, were seemingly less favourable to the defence because they emphasised the comparative rarity of the accused's size 11 Nike trainers as a percentage of all shoes distributed in the country. The Court of Appeal was understandably concerned that the jury might have relied on figures potentially exaggerating the strength of the prosecution's case. More generally, the court felt that, albeit without personal fault, the expert in *R. v T* had adopted a reporting protocol which obscured the basis on which he had formed his conclusions. "[I]t is simply wrong in principle", fumed the court, "for an expert to fail to set out the way in which he has reached his conclusion in his report".⁸ There was, in other words, an elementary and catastrophic failure of transparency, entirely at odds with the legal requirement that expert reports must spell out the "details of any literature or other information which the expert has relied on in making the report".⁹

⁷ The details of the calculation are given in *R. v T* [2010] EWCA Crim 2439 at [36]. It appears that the expert's formula attempted to take account of relationships of dependency between individual factors, e.g. between wear and size. For further explanation and discussion of the significance of dependency in probabilistic reasoning, see C. Aitken, P. Roberts and G. Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses* (Royal Statistical Society, 2010).

⁸ *R. v T* [2010] EWCA Crim 2439 at [108].

⁹ Criminal Procedure Rules 2010 r.33.3.

However, the Court of Appeal trained most of its critical fire on the use of likelihood ratios by expert witnesses in circumstances where, in the opinion of the court, the expert lacked adequate data to support any probabilistic calculations. As the court observed, “[a]n approach based on mathematical calculations is only as good as the reliability of the data used”.¹⁰ Whilst probabilistic calculations are acceptable and well-established in relation to DNA profiling evidence, the data sources relied on by the expert in *R. v T* to calculate likelihood ratios for sole pattern and shoe size were branded unreliable and imprecise.¹¹ Even the national sales figures quoted in court were merely approximations which excluded counterfeit shoes and those sold through the Foot Locker retail chain, for which records were apparently unavailable. Accordingly, the Court of Appeal concluded, resort to probabilistic calculations in relation to footwear mark comparisons “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 use a formula to calculate the probabilities. The practice has no sound basis”.¹² The court added, for good measure, that “outside the field of DNA (and possibly other areas where the practice has a firm statistical base), this court has made it clear that ... likelihood ratios should not be used”.¹³

If experts are precluded from expressing footwear mark comparisons in terms of likelihood ratios, what are the remaining alternatives? The judgment in *R. v T* is frustratingly ambiguous on this vital question. In some situations, it seems, the court might prefer to restrict experts in footwear cases to saying no more than that the defendant’s shoes “could have made” the crime scene marks. “The use of the term ‘could have made’”, explained the court, “is a more precise statement of the evidence; it enables the jury better to understand the true nature of the evidence than the more opaque phrase ‘moderate [scientific] support’”.¹⁴ However, in particular cases there may be additional factors, such as “unusual size or pattern”, that enable the expert “to go further than ‘could have made’ and express, on the basis of such factors, a more definite evaluative opinion”.¹⁵ Later in the judgment, the court expands on these observations, invoking the facial imaging case of *Atkins*¹⁶ by way of positive example. Significantly, in *Atkins* “there was no issue as to the use of a likelihood ratio”.¹⁷ By analogy, “an expert footwear mark examiner can therefore in appropriate cases use his experience to express a more definite

¹⁰ *R. v T* [2010] EWCA Crim 2439 at [80].

¹¹ Cf. M. Saks and J. Koehler, “The Coming Paradigm Shift in Forensic Identification Science” (2005) 309 Science 892 (arguing that overtly probabilistic DNA profiling evidence may soon become the normative methodological paradigm for all forensic identification sciences).

¹² *R. v T* [2010] EWCA Crim 2439 at [86].

¹³ *R. v T* [2010] EWCA Crim 2439 at [90]. Contrary to the court’s suggestion, its previous ruling in *Adams (No. 2)* [1998] 1 Cr. App. R. 377 CA (Crim Div), lends no support whatever to the decision in *T*. The *Adams* Court of Appeal objected to the use of likelihood ratios by the jury to evaluate non-scientific evidence, employing Bayes’ Theorem to calculate a “posterior probability” of guilt. Whilst there may have been good reasons for that decision, *Adams* says nothing about whether experts themselves should use likelihood ratios in forming their opinions.

¹⁴ *R. v T* [2010] EWCA Crim 2439 at [73].

¹⁵ *R. v T* [2010] EWCA Crim 2439 at [74].

¹⁶ *Atkins* [2009] EWCA Crim 1876; [2010] 1 Cr. App. R. 8; [2010] Crim. L.R. 141.

¹⁷ *R. v T* [2010] EWCA Crim 2439 at [93].

evaluative opinion where the conclusion is that the mark ‘could have been made’ by the footwear. However no likelihood ratios or other mathematical formulae should be used in reaching that judgment”.¹⁸

What did *R. v T* decide?

It should be clear from the foregoing case summary that, strictly speaking, the ratio of *R. v T* is limited to expert evidence of footwear marks. The Court of Appeal explicitly accepted the use of probabilistic calculations in relation to DNA profiling evidence (and “possibly other areas in which there is a firm statistical base”)¹⁹ and made no attempt to evaluate the use of likelihood ratios in relation to other types of scientific evidence which did not fall for consideration in the instant appeal. That being said, many of the court’s specific observations as well as its general approach to probabilistic reasoning might easily be extrapolated to other areas of forensic science and expert witness testimony. So even though *R. v T* creates a formal legal precedent only in relation to footwear comparison evidence, the stakes are high in striving to ensure that the judgment is not misinterpreted or misapplied more broadly.

On the positive side of the equation, the Court of Appeal’s concern to ensure that expert report writing is transparent, that experts’ conclusions are properly supported by adequate data, and that juries are not misled by statistical evidence can only be applauded. However, we also need to address the potential implications of passages in the judgment which might be taken to impose general constraints on how forensic scientists and other expert witnesses should arrive at their conclusions and present their evidence in contested criminal trials.

One possible interpretation of the court’s remarks is that, absent better data or unusual features, an expert should not go beyond saying that a shoe “could have made” a particular mark. But this would be an extraordinarily unhelpful limitation on the scope of expert witness testimony. After all, to say that a particular shoe “could have made” a mark is logically consistent with any other shoe in the world having made the mark. It is a bit like saying that the accused “could have been” the person that the eyewitness saw running from the scene of the crime on the basis that, like the perpetrator, the accused has two legs and therefore cannot be eliminated from the pool of bipedal suspects! Moreover, the really objectionable feature of “could have” testimony is not that its logical probative value is slight, but rather that it presents a very tangible risk of causing the jury to overestimate its probative value. Jurors will reasonably assume the salience of the information being presented to them at trial—“why else would the prosecution be telling us this, if it wasn’t important?”—and might easily infer that expert testimony to the effect that “the accused’s shoe could have made the mark” effectively means “the accused’s shoe probably made the mark”, or “almost certainly did make the mark”,

¹⁸ *R. v T* [2010] EWCA Crim 2439 at [95].

¹⁹ *R. v T* [2010] EWCA Crim 2439 at [90].

etc.²⁰ Behavioural science data support this speculation.²¹ Consider also the situation in *George*,²² where it transpired on appeal that the small particle of firearm discharge residue (FDR) found in Barry George's pocket was equally as likely to have been there if George *had* fired the gun that killed Jill Dando (the prosecution proposition) as it was if he *had not* fired the gun (the defence proposition). In the absence of any hard data on the prevalence, transfer or retention of FDR—an equivalent situation to *R. v T*, on the Court of Appeal's analysis—should prosecution experts really have testified, e.g., that the FDR “could have come” from George's having fired the gun and left it at that? Surely not. The Court of Appeal in *George* thought that the way in which the FDR evidence had been led at trial might well have caused the jury to overestimate its true probative value. As the new scientific evidence presented on appeal brought to light, the FDR evidence was effectively neutral—in other words, irrelevant to the proceedings. The Court of Appeal's insistence in *R. v T* that “could have made” is a more precise statement of scientific evidence than “moderate support” simply does not stand up to logical, let alone empirical, scrutiny.

An alternative interpretation of *R. v T*, more faithful to the spirit of the Court of Appeal's intentions if not necessarily to every paragraph of the judgment, is that experts may still offer evaluative opinions using verbal summaries of probative value such as “moderate support” and “strong support”, even in the absence of robust statistical data, provided that they steer well clear of “likelihood ratios and other mathematical formulae”. Whilst this is arguably preferable to encouraging experts to express meaninglessly vague and dangerously misleading opinions, such a stricture would nonetheless be unwise. There are many reasons why experts might justifiably adopt a probabilistic approach, in footwear mark analysis no less than in other areas of forensic science. One set of considerations relates to the consistency and comparability of experts' conclusions and their meaningfulness for the jury. If experts are going to continue to give evaluative opinions—i.e. to say more than “could have”—some attempt at quantification is the most obvious way of trying to achieve a measure of consistency in the use of phrases such as “moderate” or “strong” support. Otherwise, there is likely to be a forensic free-for-all, with one (cautious) expert's “moderate support” equating to another (cavalier) expert's “very strong support”, predictably leading jurors astray or leaving them entirely baffled. Quantification also promotes transparency. When expert evidence is broken down into its individual elements and each element is assigned a likelihood ratio (i.e. precisely the methodology adopted by the footwear mark expert in *R. v T*), it is easier for a second expert to check the work of the first

²⁰ These very concerns were expressed by the Court of Appeal in *Atkins* [2009] EWCA Crim 1876 at [23]–[25]. It is conceivable that in *R. v T* [2010] EWCA Crim 2439 at [66], the court intended to endorse the scale of expressions used by footwear mark examiners in the United States, where “could have made” lies between “probably made” and “inconclusive”. While this would clarify the meaning of “could have made” (supposing that this understanding was widely shared and could be communicated to fact-finders effectively), the court should also be aware that “probably made” involves a transposed conditional (the so-called “prosecutor's fallacy”), a logical mistake rightly condemned by the Court of Appeal in *Doherty* [1997] 1 Cr. App. R. 369; [1997] Crim. L.R. 669 CA (Crim Div) and in *Deen*, *The Times*, January 10, 1994, and by the Privy Council in *Pringle v R* [2003] UKPC 9. *Cf R. v T* [2010] EWCA Crim 2439 at [91], where the Court of Appeal itself commits the fallacy. For further discussion of both the “could have” terminology and transposed conditionals, see Aitken, Roberts and Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*, 2010.

²¹ See, e.g. D. McQuiston-Surrett and M. Saks, “Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact” (2008) 59 *Hastings LJ* 1159.

²² *George* [2007] EWCA Crim 2722.

(as in fact was done in T's case). This also facilitates scrutiny and challenge by an opposing expert—which is a structural expectation of adversarial criminal proceedings, and something that courts should especially welcome in an era of proactive trial management in which judges are expected to manage and clarify expert disagreement through pre-trial processes.²³ Last but surely not least, a probabilistic approach employing likelihood ratios is an application of logical thinking. It encourages analytical rigour and balance in the production of forensic science evidence, specifically because it requires the expert to think in terms of competing defence and prosecution explanations and to take account of all aspects of the evidence relevant to forming a scientific opinion. To the extent that a variable such as wear or damage makes it less likely that the accused's shoes made the suspicious marks, for example, this should be factored into an overall evaluation of the strength of the evidence, regardless of the fact that the expert may have been instructed by the police and is called to testify at trial by the prosecution.²⁴

This brings us to the most difficult aspect of the Court of Appeal's judgment in *R. v T*. The court rightly demanded full transparency in the writing of expert reports, but then said that footwear mark experts should not employ likelihood ratios in formulating their "evaluative opinions". Does this mean that experts should disavow familiar forensic tools of logical reasoning, and thereby eschew a careful and balanced analysis of the evidence? The Court of Appeal cannot have intended to issue that irrational directive, regardless of what it might have said. Even on the wild assumption that the Court of Appeal meant to instruct forensic scientists how to think, how could such an injunction ever be enforced on forensic scientists working, as it were, in the privacy of their own laboratories? The best that might be imagined would be a policy of "don't ask, don't tell", whereby experts formulated their conclusions according to their good faith understanding of scientific protocol but carefully concealed their "deviant" probabilistic reasoning from legal scrutiny. This is hardly in keeping with the Court of Appeal's ringing endorsement of transparency,²⁵ and offers little more than a third-best solution to a second-rate compromise. Better all round if experts are simply encouraged to give their best evidence as clearly, intelligibly, honestly, and transparently as they are able, and in accordance with the precepts of scientific validity, logic and reason.

Lurking behind the Court of Appeal's rejection of likelihood ratios appears to be a concern that, in the absence of hard data, numerical quantification could produce an almost fraudulent "verisimilitude of mathematical probability". "It cannot be right", fretted the court, "to seek to achieve objectivity by reliance on data which does not enable this to be done".²⁶ Now, "objectivity" is a difficult and

²³ See Judge Roderick Denyer QC, *Case Management in the Crown Court* (Oxford: Hart, 2008); *Reed* [2009] EWCA Crim 2698; [2010] 1 Cr. App. R. 23; [2010] Crim. L.R. 716 at [128]–[134].

²⁴ The court did suggest that, as in *Abadam* [1983] 1 W.L.R. 126 CA (Crim Div); (1983) 76 Cr. App. R. 48; [1983] Crim. L.R. 254 (a case involving comparison of glass fragments), "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": *R. v T* [2010] EWCA Crim 2439 at [91]. However, confining an expert opinion to a simple frequency of occurrence has numerous drawbacks, including its potential (as we have already noted) to ignore aspects of the evidence favourable to the accused. For criticism of this approach in glass comparison casework, see Redmayne, *Expert Evidence and Criminal Justice*, 2001, pp.37–41.

²⁵ *R. v T* [2010] EWCA Crim 2439, see especially at [108]: "the practice of using a Bayesian approach and likelihood ratios to formulate opinions placed before a jury without that process being disclosed and debated in court is contrary to principles of open justice".

²⁶ *R. v T* [2010] EWCA Crim 2439 at [87].

frequently abused concept.²⁷ Suffice it here to emphasise that numerical quantification does *not* necessarily imply any claim to objectivity. When a weather forecaster speaks of a 20 per cent chance of snow, it should be clear that, rather than expressing any verifiably objective fact about the empirical world, this is a statement of probability regarding future events informed by appropriate meteorological data and expertise. The footwear mark expert in *R. v T* did not appear to make any extravagant claims to objectivity in the sense of presenting his findings as unarguable or as involving no element of judgement. It was obvious that evaluative judgement loomed large in his assessment of the significance of wear and damage to T's shoes. Admittedly, numerical quantification sometimes flirts with the sin of "false precision"—imagine a weather forecaster predicting a "23 per cent chance of snow". But in *R. v T* the expert's ultimate conclusion was couched in broad terms: the evidence gave "moderate support" to the prosecution's case.

It is tempting to trace the root cause of the court's quantification-anxiety to its evident dissatisfaction with the statistical data relied on by the expert in calculating his four likelihood ratios. Indeed, the most coherent and justifiable interpretation of these difficult passages of the judgment is that they express an objection to unreliable data, rather than to a particular reasoning methodology (likelihood ratios) or formula for presenting evidence to the fact-finder (degree of support). In the court's reckoning:

"Use of the FSS's own database could not have produced reliable figures as it had only 8,122 shoes whereas some 42 million are sold every year."²⁸

While there is much truth and wisdom in the court's insistence that "an approach based on mathematical calculations is only as good as the reliability of the data", the adequacy of data to support a conclusion depends crucially on the nature of that conclusion. It is a question of fitness *for purpose*, not fitness at large (which is meaningless).²⁹ The right question to ask, then, is: were the data utilised by the expert in *R. v T* good enough to justify his—quite restrained—conclusion of "moderate support"? Perhaps a database comprising "only" 8,122 sets of measurements (a sample size, by the way, that most social scientists can only dream about) would have been insufficient to justify a conclusion of "very strong" support for the prosecution case, but the expert's conclusions in *R. v T* were far less emphatic. As for the fact that the FSS database contains only a small proportion of all the shoes sold in the United Kingdom, this is probably beside the point. The pertinent question is: which database is likely to provide the best comparators, relative to the task in hand?³⁰ Shoe choice is influenced by social and cultural

²⁷ For further elucidation, see I. Evett, "Expert Evidence and Forensic Misconceptions of the Nature of Exact Science" (1996) 36 *Science & Justice* 118; M. Redmayne, "Objective Probability and the Law of Evidence" (2003) 2 *Law, Probability & Risk* 275.

²⁸ *R. v T* [2010] EWCA Crim 2439 at [84]. As a general matter, concerns about the size of the sample relative to the target population may be misconceived. Opinion polls in general elections, for example, frequently sample only about 1,000 voters from an electorate of tens of millions, but such data can still supply a basis for reliable predictions so long as the samples are chosen appropriately and the results interpreted carefully.

²⁹ Cf. the Court of Appeal's observation in *Benn* [2004] EWCA Crim 2100 at [44]: "the question of the validity of a database depends upon the purpose which is to be served".

³⁰ For a detailed analysis of the difficult question as to which databases should be used in analysis of trace evidence, see C. Champod, I. Evett and G. Jackson, "Establishing the Most Appropriate Databases for Addressing Source Level Propositions" (2004) 44 *Science & Justice* 153.

factors. Middle-aged university lecturers presumably buy different trainers to teenage schoolboys. And the shoes making the marks most often found at crime scenes are not the general public's most popular purchases.³¹ Consequently, the FSS database—comprising shoes owned by those who, like T, have been suspected of committing offences—may well be a more appropriate source of comparison data than national sales figures. In *R. v T* itself, the difference was significant and was favourable to the defence: the switch from the FSS data used in the expert's report to the national sales figures quoted at trial was rightly viewed by the Court of Appeal as a cogent ground for allowing T's appeal against conviction. Even had complete national data (incorporating counterfeits and Foot Locker sales) been available, it would arguably have been unfairly prejudicial to the defence to use those data in preference to the FSS's suspect-specific database.

It is unnecessary here to venture any firm view on the adequacy of the expert's data-conclusion nexus in *R. v T*. We only wish to say that the Court of Appeal's conclusory observations fail to do justice to the difficulties and nuances of selecting appropriate data for comparison purposes in forensic science practice. While the court was concerned that there "are, at present, insufficient data for a more certain and objective basis for expert opinion on footwear marks",³² it is rarely helpful to talk about "objective" data in forensic contexts. Choice of data always involves a degree of judgement about whether a particular dataset is fit for purpose. At the same time, reliance on data is ubiquitous and inescapable. When the medical expert witness testifies that "I have never encountered such a case in forty years of clinical practice", he is utilising data but he calls them "experience", relying on memory rather than any formal database open to probabilistic calculations. It would be just as foolish to maintain that memory and experience are never superior to quantified probabilities in criminal litigation, as it would be to insist that memory and experience are always preferable to and should invariably displace empirical data and quantified probabilities in the courtroom.

What is to be done?

R. v T raises some of the most fundamental, timely and difficult issues surrounding the production of forensic science evidence and its communication to fact-finders in criminal proceedings. The Court of Appeal's emphatic directives that expert report writing must be transparent, that experts' conclusions should be properly supported by adequate data, and that juries must not be bamboozled by statistical evidence, are valuable reminders of basic principles to which all forensic scientists and criminal justice professionals should pay heed.

However, this article has taken issue with much of the Court of Appeal's reasoning in *R. v T*. It is hard to believe that the court really wanted to restrict experts in fields such as footwear mark analysis to vague—and potentially misleading—conclusions stating only that the accused's shoes (or whatever) "could have made" the questioned marks. It is even harder to believe that the court wanted to prohibit the careful and balanced approach to expert evidence associated with

³¹ See T. Napier, "Scene Linking Using Footwear Mark Databases" (2002) 42 *Science & Justice* 39; M. Tonkin and J. Bond, "Step by Step: How Shoe Prints Could Help Police Officers Home in on Suspects" *Police Review*, February 12, 2010, 26.

³² *R. v T* [2010] EWCA Crim 2439 at [87].

likelihood ratios, a methodology which has been gaining ground within the forensic science community³³ and is now endorsed by the UK's Association of Forensic Science Providers.³⁴ If the real objection is to numerical quantification, the court should have explained why quantification is problematic even in the absence of unwarranted claims to objectivity or false precision. In promoting critical scrutiny of the quality of data utilised in expert reports, *R. v T* makes a positive contribution to English law. However, the Court of Appeal went off on the wrong track when it neglected to ask the most pertinent questions about the nature of the relationship between data and the particular inferential conclusions being advanced in the instant case.

The strict ratio of *R. v T* is that existing data are legally insufficient to permit footwear mark experts to utilise probabilistic methods involving likelihood ratios when writing reports or testifying at trial. For the reasons identified in this article, we hope that the Court of Appeal will reconsider this ruling at the earliest opportunity. In the meantime, we are concerned that some of the Court's more general statements could frustrate the jury's understanding of forensic science evidence, and even risk miscarriages of justice, if extrapolated to other contexts and forms of expertise. There is no reason in law why errant obiter dicta should be permitted to corrupt best scientific practice. We venture to suggest that it would defeat the Court of Appeal's deeper purposes and its own clear statements of principle if forensic scientists and the lawyers who instruct them were to adopt, through confusion or misplaced obedience to authority, any of the fallacious readings of *R. v T* exposed and rejected in this article.

³³ See G. Jackson, "Understanding Forensic Science Opinions" in J. Fraser and R. Williams (eds.), *Handbook of Forensic Science* (Willan, 2009); B. Robertson and G. A. Vignaux, *Interpreting Evidence: Evaluating Forensic Science in the Courtroom* (Wiley, 1995).

³⁴ AFSP, "Standards for the Formulation of Evaluative Forensic Science Expert Opinion" (2009) 49 *Science & Justice* 161. The Association of Forensic Science Providers aims to "represent the common interests of the providers of independent forensic science within the UK and Ireland with regard to the maintenance and development of quality and best practice in forensic science and expert witness in support of the Justice System from scene to court, irrespective of commercial pressures associated with the competitive forensic marketplace".