Discussion paper: Hard cases make bad law—reactions to \( R_v T \)

WILLIAM C. THOMPSON

Department of Criminology, Law & Society and School of Law, University of California, Irvine, CA 92697, USA

In the case of \( R_v T \) (2010), the Court of Appeal for England and Wales rejected the testimony of an expert who had used likelihood ratios to assess the probative value of shoe-print evidence. Because likelihood ratios are widely used in forensic science, and their use has been actively promoted by leaders in the field (Association of Forensic Science Providers, 2009; Cook et al., 1998; Evett, 1998; Robertson and Vignaux, 1995), the court’s opinion has understandably caused consternation in the forensic science community (Berger et al. 2011; Robertson et al. 2011; Redmayne et al. 2011). Recognizing the importance of the issues involved, the editors of \( \textit{Law, Probability and Risk} \) have devoted this issue to articles commenting on the case. At the invitation of Editor Colin Aitken, it is my privilege and honour to review and respond to those articles.

In the interest of full disclosure, I will say at the outset that I think \( R_v T \) is an inept judicial opinion that creates bad law. The opinion went awry because the justices who wrote it misunderstood a key aspect of the evidence they were evaluating. The justices sought to achieve laudable goals, but their misunderstanding of basic principles of inductive logic, and particularly Bayes’ theorem, led them to exclude a type of expert evidence that, in general, is helpful and appropriate in favour of an alternative type of expert evidence that is fundamentally inconsistent with the goals the court sought to achieve. The case has already received severe criticism\(^1\) and will inevitably come to be seen for what it is—a judicial blunder.

1. Laudable goals

Although the commentaries in this issue differ sharply over whether and how experts should use likelihood ratios, there is substantial agreement among the commentaries, and between the commentators and the court, on the basic goals that expert testimony should seek to accomplish. As we consider the merits of using likelihood ratios in forensic assessments it will be important to keep those shared goals in mind.

First, all agree that the role of the forensic science expert should be limited to evaluating the strength of the particular evidence that the expert was assigned to assess. As American shoe-print expert William Bodziak (this issue) explains, the results of the examination must depend on ‘the examination of the impression only’ and should not be influenced by ‘data such as whether the suspect had a motive or an opportunity to commit the crime, or had a past criminal record, and other similar facts and

\(^1\)Email: william.thompson@uci.edu

\(^1\) Robertson et al. (2011) see the case as potentially catastrophic, saying it ‘has the potential to halt progress in forensic science and to reduce forensic scientific evidence to unanalysable impressionistic conclusory statements. . . .’ Redmayne et al. (2011) take a more hopeful view, but admit that their Pollyanna-like reading of the case is more faithful to what they suppose to have been the justices’ intentions than to what the opinion actually said.
speculation...’. The court agreed, stating: ‘it is not [the expert’s] function to evaluate the other evidence in the case’ (para. 75).

This essential point has, unfortunately, not always been recognized by forensic experts. Some forensic experts see no problem with relying on other evidence in the case, thinking it can help them correctly interpret the evidence before them (Thompson, 2011). Others deny relying on broader evidence in the case but fail to take basic preventative measures (such as blinding and sequential unmasking procedures) to prevent their knowledge of such evidence from unintentionally influencing their scientific judgements (Risinger et al., 2002; Krane et al., 2008; Thompson et al., 2011; Thompson, 2009).

Courts understandably take a dim view of experts who stray beyond their assigned domain and draw conclusions based on a broader assessment of the case. Using rather militant imagery, legal scholars say that such experts have ‘invaded the province of the jury’ and ‘usurped the jury’s function’ because assessing the broader evidence in the case is the job of the jury, not the expert. It appears that the court’s rejection of likelihood ratios in R v T was based partly on the impression that likelihood ratio analysis entails or requires consideration of factors outside the expert’s domain. Mr. Bodziak (this issue) appears to share that impression. I agree with Biederman and Taroni (this issue) that this impression is simply incorrect.

To compute a likelihood ratio the expert needs to know the relevant propositions (hypotheses) to be compared—in R v T one proposition, obviously, was that the print was made by defendant’s shoe; the alternative proposition was that it was made by another shoe. Beyond that, the expert may need to know certain facts about the case, such as where the print was found, in order to estimate the conditional probability of observing that particular print under the relevant propositions. But there is no need for the expert to consider, or even know about, other evidence unrelated to the print that might incriminate or exculpate the defendant. Contrary to what Mr. Bodziak asserts, no one has argued that forensic scientists ‘should’ consider non-forensic evidence when estimating likelihood ratios. Moreover, it seems quite clear from the court’s description of the evidence in R v T that the likelihood ratios computed by the shoe-print expert, Mr. Ryder, were properly designed to reflect solely the value of the shoe-print evidence. There is no indication that Mr. Ryder gave consideration to non-forensic evidence when estimating likelihood ratios.

A second point on which all commentators appear to agree is that there should be a valid scientific basis for the expert’s opinion. The court in R v T was concerned about whether the shoe-print expert had an adequate basis for his likelihood ratio calculations given that ‘the statistical data available were uncertain and incomplete’ (para. 1). Several of the articles in this issue expand perceptively on this issue, pointing out various uncertainties involved, noting the difficulty of identifying relevant databases and some of the inferential errors experts can make in drawing statistical conclusions (Bodziak, this issue; Biederman and Taroni, this issue; Nordgaard and Rasmussen, this issue). These problems are not limited to shoe-print examination. As the U.S. National Academy of Sciences has noted, there

---

2 For example, if people who commit crimes favour different types of shoes than the general population, it would be helpful for the expert to know whether the shoe print was found in an area where it could only have been made by a criminal perpetrator, as opposed to an area where it could have been made by others, because this information would affect the conditional probability of observing the print under the alternative proposition. See, Champod et al. (2004).

3 It is possible, of course, that Mr. Ryder was unintentionally biased when making his likelihood ratio estimates by his knowledge of the investigative facts of the case. But that is a problem all forensic scientists face when they fail to use blind interpretive procedures; it is not an inherent problem with the use of likelihood ratios. And this problem can easily be solved by use of a case manager to screen the analyst from domain-irrelevant information (Thompson, 2011).
are deficiencies in the empirical validation of many forensic science disciplines, particularly those involving pattern matching (National Research Council, 2009). But banning likelihood ratios does nothing to solve these problems. Indeed, I will argue below that one of the advantages of likelihood ratios is that they help identify and illuminate validation issues, while alternative approaches tend to hide such issues.

A third point on which all commentators agree is the importance of transparency. The basis for the expert’s opinion, and the expert’s underlying logic, should be made clear. If the expert makes assumptions, relies on empirical data, or makes intermediate inferences, it should all be noted and acknowledged in the case notes (at least), if not also in the expert’s report. The expert should strive, as far as possible, to render his or her inferential process open to critical scrutiny.

A major problem with the report and testimony of Mr. Ryder, the shoe-print examiner in \( R \text{ v } T \), was his failure to report explicitly that his conclusions rested in part on a likelihood ratio calculation. The court declared that ‘it is simply wrong in principle for an expert to fail to set out the way in which he has reached his conclusion in his report’ (para. 108(ii)). There is a strong suggestion in the opinion that the justices thought Mr. Ryder was trying to hide his use of a likelihood ratio from the court because he knew or suspected that a conclusions based on a likelihood ratio would be inadmissible. I believe the justices were mistaken in their assessment of admissibility. As discussed below, the law does not (and should not) preclude the admissibility of likelihood ratios of the kind computed by Mr. Ryder. Nevertheless, it is understandable that the justices would be outraged at the thought of an expert hiding something from the court, and their negative reaction to Ryder’s evidence may rest partly on this perception.

If lack of transparency is the problem, however, is banning likelihood ratios the solution? I will argue below that the expert’s judgemental process is more transparent when the expert explicitly computes likelihood ratios than when the expert makes the kind of global judgements favoured by Mr. Bodziak.

A fourth point on which all commentators agree is the importance of communicating expert conclusions in a manner that jurors (and lawyers) can easily comprehend. Favouring comprehensibility is easier than knowing how to achieve it, however. While some of the commentators believe likelihood ratios can be made comprehensible through the use of verbal scales of equivalence (Biederman and Taroni, this issue; Nordgaard and Rasmussen, this issue), legal scholars Andrew Ligertwood and Gary Edmond (this issue) argue that likelihood ratios will inevitably be misunderstood and misused. Moreover, they raise questions about whether expert testimony framed in terms of likelihood ratios is compatible with jurors’ duty to assess the evidence in accordance with traditional standards of proof. As an alternative they suggest that experts either present statistical conclusions in the form of frequencies or that experts opinie on source probabilities. For example, the expert might say that only one shoe in 1000 could have made a particular mark, or that the probability that the defendant’s shoe made the mark is 99.9%. Before considering the merits of this proposal, however, I need to say more about the confusion over Bayesian statistics that arose in \( R \text{ v } T \).

2. A fundamental misunderstanding

The first thing that students learn about Bayes’ theorem is the meaning of its basic components—the prior odds, the posterior odds, and the likelihood ratio. The justices who wrote \( R \text{ v } T \) clearly did not understand this first thing about Bayes’ theorem. In particular, they confused the likelihood ratio with a posterior probability. They took the expert’s statements about the relative probability of the evidence
under the relevant propositions to be statements about the probability that the propositions are true. As a result, they mistakenly interpreted the expert’s likelihood ratio as an opinion about the probability of source.\textsuperscript{4}

Confusion of likelihood ratio and source probability is seen repeatedly in the opinion. For example, the court incorrectly said that a likelihood ratio expresses ‘the degree of likelihood . . . the marks “could have been made by a particular shoe”’ (para. 30). In the very next paragraph, the court again incorrectly said that a likelihood ratio reflects ‘the degree of likelihood that the marks were made by a particular shoe’ (para. 31). Then, after quoting a correct definition of the likelihood ratio, the court went on to say, incorrectly, that in the case at hand the likelihood ratio ‘was expressed as the probability that the Nike trainers . . . had made the marks discovered at the scene divided by the probability the Nike trainers had not made the marks’ (para. 33i). Later, the court again incorrectly stated that a likelihood ratio provides ‘the degree of probability that a mark had been made by a particular shoe’ (para. 50).\textsuperscript{5}

Having confused likelihood ratios with posterior probabilities, the court also appeared to confuse (or conflate) the likelihood ratios computed by Mr. Ryder, the shoe-print expert, with Bayesian computations of the type performed by Professor Peter Donnelly in \textit{R v Adams} (1996, 1997). \textit{Adams} was a rape case with sharply conflicting evidence. On one hand, the defendant and perpetrator shared a DNA profile; the profile was rare—the random match probability was estimated to be about 1 in 200 million. On the other hand, the defendant had a strong alibi, the victim did not recognize him, and he did not match the victim’s description of the rapist. Professor Donnelly, a prominent statistician, was called by the defendant to comment on how jurors might integrate this conflicting evidence. He explained the use of Bayes’ theorem and showed jurors how to use it to combine the DNA evidence with the non-DNA evidence in order to compute the posterior probability of the defendant’s guilt. The thrust of his testimony was that, notwithstanding the impressive random match probability, there might still be a significant chance that the defendant was innocent if the jurors thought the non-DNA evidence established a low prior probability of his guilt.

After Adams was convicted, the Court of Appeal expressed ‘very grave doubt as to whether that evidence was properly admissible, because it trespasses on an area peculiarly and exclusively within the province of the jury, namely the way in which they evaluate the relationship between one piece of evidence and another.’ Of particular concern to the court was ‘the theorem’s methodology’ that required ‘that items of evidence be assessed separately according to their bearing on the accused’s guilt, before being combined in the overall formula’, a process that the court considered ‘simply inappropriate to the jury’s task’. By commenting on how the jury might evaluate non-scientific evidence, and how the jury might combine the scientific and non-scientific evidence, the expert arguably ‘invaded the province of the jury’ as discussed above. On retrial the defence again presented Donnelly’s testimony and, following conviction, the Court of Appeal again expressed doubts about whether such testimony was proper (\textit{R v Adams-No 2}, 1997), as had the Court of Appeal in \textit{R v Doheny} (1996).

Even if Professor Donnelly’s testimony did not ‘invade the province of the jury’ it suffers from another serious problem—it is likely to bias the jurors against the defendant by requiring them to make

\textsuperscript{4} Robertson \textit{et al} (2011) identify this error as a version of the prosecutor’s fallacy and note multiple instances in the court’s opinion.

\textsuperscript{5} Adding insult to error, the court attributed the false statement to I.W. Evett, J.A. Lambert and J.S. Buckleton, three scholars who would never make such an error.
separate probability assessments based only on parts of the trial evidence, after informing them about the other parts of the evidence (Thompson, 1997). Jurors cannot be expected to make unbiased estimates of the prior probability (i.e. the probability of guilt based on the non-DNA evidence) after being informed of the DNA evidence. It is predictable that knowledge of the DNA evidence will cause jurors to overestimate their priors, which will lead to double-counting of the DNA evidence in the Bayesian calculation. Thus, there are a number of good reasons for courts to reject the kind of Bayesian testimony offered by Professor Donnelly in *Adams*.6

But those reasons do not apply to the computation of likelihood ratios by a shoe-print expert like Mr. Ryder. Mr. Ryder did not invite the jurors to make any Bayesian calculations. He offered no instructions on how to combine his testimony about the shoe prints with other evidence. Indeed, he made no comment whatsoever about any aspect of the case other than the shoe prints, saying simply that his analysis found moderate scientific support for the conclusion that the shoe print was made by the defendant’s Nike trainer. His notes suggested that this conclusion was based, in part, on computation of a likelihood ratio, but he did not present the likelihood ratio to the jury, and even if he had done so it is difficult to see how it could have created the problems associated with Professor Donnelly’s testimony in *Adams*.

Failing to see how very different Ryder’s testimony was from Donnelly’s, the court in *R v T* mistakenly assumed that the judicial rejection of Bayesian calculations in *Adams* and *Doheny* was also a rejection of likelihood ratios. The court repeatedly states that ‘Bayes theorem and likelihood ratios should not be used’ acknowledging no distinction whatsoever between the two. The fact that likelihood ratios have been widely used in connection with DNA evidence, without anyone claiming that this practice violates the principles established in *Adams* and *Doheny*, should perhaps have given the justices pause and alerted them to this flaw in their analysis. But the justices found a different basis on which to distinguish the DNA cases, citing the ‘sound statistical database’ that exists for DNA statistics. Thus, in the court’s view, the principles established by *Adams* and *Doheny* applied only ‘outside the field of DNA (and possibly other areas where there is a firm statistical base)’ (para. 90).

This position is rather awkward, however, given that both *Adams* and *Doheny* were cases that involved DNA evidence. The court failed to address the obvious and important question of how the Bayesian testimony that was found to be unacceptable in *Adams* and *Doheny* differed from the likelihood ratio testimony that is accepted in other DNA cases. Had the justices understood the difference (between testimony about Bayesian computations and testimony about likelihood ratios), I believe they would have seen that the likelihood ratio evidence in *R v T* does not pose the same problems as the Bayesian testimony in *Adams* and *Doheny*. Of course the justices’ failure to understand the difference between likelihood ratios and Bayesian posteriors is what got them into difficulty in the first place.

3. **Should forensic experts use likelihood ratios?**

Although the court’s legal analysis was faulty, some of the commentators in this issue support the court’s bottom line conclusion that ‘likelihood ratios should not be used’. When assessing this claim it is helpful to distinguish the explicit quantification of conditional probabilities for the purpose of making likelihood ratio calculations from the implicit use of likelihood ratios in thinking about

---

6 Whether *Adams* established a precedent that requires rejection of Bayesian testimony is debatable, however. See Robertson *et al.* (2011), who argue that *Adams* ‘is not authority...for rejection of the Bayesian approach’ because ‘the professor’s evidence was not a matter raised in the appeal’ and ‘the relevant matters were not argued’.
evidence. It would be foolish, in my view, to contend that experts should ‘never’ use likelihood ratios. The only logical way to assess the strength of forensic evidence for supporting a hypothesis, such as the hypothesis that the defendant’s shoe made a particular mark, is to consider the probability of observing such a mark if it was made by the defendant’s shoe, relative to the probability of observing such a mark if it was made by some other shoe, which is, of course, a likelihood ratio. If the expert is not at least implicitly using likelihood ratios to think about the value of evidence, the expert is either thinking illogically or relying on inappropriate factors to make the assessment (Buckleton, 2005).

Whether experts should make their reasoning explicit by quantifying their estimates of conditional probability is a separate question. One advantage of explicit quantification is that it renders the expert’s judgement transparent by laying out clearly what factors the expert considered and how the expert weighed those factors. The likelihood ratios that Mr. Ryder calculated in \( R \) vs. \( T \) are an excellent example. He considered four factors when comparing the shoe print at the crime scene to defendant’s shoes: the sole pattern, shoe size, degree of wear and the absence (in defendant’s shoe) of some damage marks apparent in the shoe print. One can readily see from his calculations exactly what conclusions he reached about the relevant conditional probabilities and how much weight he assigned to each factor in his overall evaluation. It is clear, e.g. that he judged that if the shoe print was made by a shoe other than the defendant’s that there was a 20% chance it would have the same sole pattern, a 10% chance it would have the same size, and a 50% chance it would have a compatible degree of wear. It is also clear that he assigned relatively little weight to the failure to find in the defendant’s shoe the damage marks seen in the shoe print. Any of these judgements might, of course, be contested. The great advantage of explicit quantification is that the underlying judgements are laid out in a manner that opens each of them to scrutiny and evaluation. One can raise questions about Mr. Ryder’s judgement about the expected frequency of size 11 Nike trainers, as Mr. Bodziak (this issue) does, precisely because Mr. Ryder’s laboratory notes specified what his judgement was on this question. We can also see and evaluate the sensitivity of Mr. Ryder’s overall evaluation to possible variations in his underlying conditional probability estimates. This is the essence of transparency.

By contrast, the approach favoured by Mr. Bodziak and most American forensic scientists is opaque. Mr. Bodziak would eschew any explicit quantification of conditional probabilities. He would instead make a global evaluative judgement by focusing on similarities and differences between the shoe print and defendant’s shoe. Relying on his training and experience he would endeavour to determine whether similar features are class or individual characteristics and, for class characteristics, how rare or common they are. On that basis he would determine whether the shoe print could have been made by defendant’s shoe and, if so, whether there is sufficient similarity to conclude that it probably or certainly was made by that shoe. But the evaluative process would occur entirely within his mind. Most forensic experts who follow the American approach make record neither of the individual factors that they consider, nor of how they weigh each factor in the overall assessment. It is difficult if not impossible to evaluate (or contest) the accuracy of the underlying judgements that factor in the expert’s evaluation because the expert’s report provides no indication of what those judgements were. Hence, I think that Sjerps and Berger (this issue) are entirely correct when they say: ‘an expert who argues: “the shoe mark was probably left by the shoe, because based on my long-term experience I can tell that there are enough corresponding acquired features to justify this conclusion”’ is absolutely

\[7\] Redmayne et al. (2011) make the same point: ‘Does [the court’s rejection of likelihood ratios] mean that experts should disavow…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.’
non-transparent: the argumentation is a logical black box.’ One of the most dismaying features of the R v T ruling is that, while trumpeting the importance of transparency, the court rejected an approach that allows a relatively high level of transparency in favour of this logical black box.

The court was justifiably concerned about the absence of a sufficient database to support Mr. Ryder’s likelihood ratio estimates. But forbidding experts to quantify their judgements does nothing to solve concerns about validation. An expert like Mr. Bodziak, if he is thinking about the evidence logically, will still need to consider the relative likelihood of the evidence under alternative hypotheses. If there is an insufficient scientific basis for making such determinations, then Mr. Bodziak’s judgements will be as invalid as Mr. Ryder’s. Validation problems are not solved by eschewing explicit quantification, they are merely hidden. A court that accepts the testimony of a Bodziak, while rejecting the testimony of a Ryder is not demanding that forensic testimony be valid; it is demanding that deficiencies in validation not be made too obvious.8

The American approach may also hide an expert’s reliance on inappropriate factors to make the judgement. There is growing evidence that forensic science experts can be influenced by investigative facts that point to the guilt or innocence of a suspect, even when these facts are not pertinent to the scientific questions before the expert (Dror and Rosenthal, 2008). And this influence may occur without the expert being aware of it (Thompson, 2011; Risinger et al., 2002). Knowing a suspect has confessed, e.g. may bolster the expert’s assessment of the strength of the scientific evidence against the suspect. Contextual biases of this sort are generally more influential when people make global judgements in the absence of explicit guidelines (e.g. Uhlmann and Cohen, 2005). Consequently, I suspect that an expert who makes global, experience-based judgements without making the underlying factors explicit and without quantification will be more susceptible to contextual bias than an expert who employs the likelihood ratio approach.

4. Should forensic experts present likelihood ratios to juries?

When assessing the court’s conclusion that ‘likelihood ratios should not be used’ (para. 90), it is also helpful to separate the expert’s own use of likelihood ratios to evaluate the evidence from the expert’s use of likelihood ratios to communicate conclusions to the jury. One use does not require the other. For example, an expert might make explicit use of likelihood ratios, as Mr. Ryder did, when evaluating the strength of the evidence at the laboratory but then choose to use a different formulation for presenting his conclusions to the jury.

Andrew Ligertwood and Gary Edmond (this issue) offer two arguments against ‘presenting’ forensic science conclusions in court in the form of likelihood ratios. Their first and most sweeping claim is that likelihood ratios are fundamentally incompatible with our system of legal proof because juries are ‘not directed to find a criminal matter proved when it acquires a certain degree of mathematical probability’ but rather to find guilt only if there is no reasonable explanation for the evidence consistent with innocence. In other words, they claim that jurors are unable to evaluate a case properly under a reasonable doubt standard when expert witnesses frame their conclusions in terms of likelihood ratios.

8 Professor Risinger (in press) has argued that validation problems can also be hidden when experts compute likelihood ratios based on subjective estimates of conditional probabilities: ‘there is something about the generation of likelihood ratios with numbers from nowhere that tends to cover up the weakness of the ingredients.’ Even so, I believe the problem is worse with the form of testimony recommended by Mr. Bodziak.
I find this aspect of their argument unconvincing. Legal scholars have, for years, used likelihood ratios to model the concept of legal relevance and to describe the strength of evidence (Lempert, 1977). If likelihood ratios are meaningful in scholarly discourse about the value of legal evidence, it is difficult to see why they would be inappropriate for communicating the value of such evidence to jurors. I believe problems of the kind Ligertwood and Edmond envision will arise only if experts present source probabilities rather than likelihood ratios or if jurors misunderstand likelihood ratios to be source probabilities.

Hence, I am more persuaded by the second argument that Ligertwood and Edmond advance: that jurors are likely to misunderstand likelihood ratios, and to misunderstand them in ways that undermine proper decision making. ‘Is there not a risk’, they ask, ‘that the jury will simply (and of course illogically) equate the likelihood ratio with the likelihood of guilt?’ Undoubtedly the answer to this question is yes. Fallacious misuse of probabilities has been demonstrated in social science studies (Thompson and Schumann, 1987; Koehler, 2001). Perhaps the best evidence that likelihood ratios are likely to be misunderstood, however, is that they have been misunderstood in exactly the way Ligertwood and Edmond fear by distinguished jurists, like those who wrote the opinion in R v T, and by distinguished legal scholars (including, it appears, Ligertwood and Edmond themselves).

Forensic experts sometimes try to reduce the chances that their conclusions will be misunderstood by presenting verbal characterizations of the likelihood ratio rather than the number itself. A scale of verbal equivalents to the likelihood ratio, initially proposed by Ian Evett (1998), has been widely used in the UK. On this scale, a likelihood ratio between 10 and 100 is characterized as ‘moderate support’ for the favoured hypothesis. According to Mr. Ryder’s calculations, the likelihood ratio for the shoe print in R v T was slightly less than 100. Consequently, he stated in his report and testimony that there was ‘a moderate degree of scientific evidence to support the view that the [Nike trainers recovered from defendant] had made the footwear mark’. (R v T, 2010, para. 24). Nordgaard and Rasmusson (this issue) discuss an alternative scale used by the Swedish National Laboratory of Forensic Sciences on which the strength of the likelihood ratio is characterized by a number ranging from −4 to +4. It is difficult to judge whether jurors will understand such testimony and how they might respond to it, but I doubt that use of these scales will solve the problem. As Ligertwood and Edmond point out, the numerical likelihood ratio will pop up in testimony anyway if lawyers inquire of the expert how he determined that there is ‘a moderate degree of support’ for a proposition, or ask what the expert means when she says that the strength of the evidence is ‘+2’. These scales have also been criticized on grounds that they lack precision as ‘different people may interpret words like “moderate” and “strong” in very different ways’ (Morrison, 2011).

9 Among evidence scholars there has been some fascinating discussion of whether likelihood ratios provide an adequate basis for a theory of legal relevance (Pardo and Allen, 2008). Michael Pardo (forthcoming) has recently argued that the value of evidence as legal proof can vary in ways not adequately captured by likelihood ratio models. My own view is that Pardo has demonstrated the inadequacy of some specific likelihood ratio models rather than an inherent limitation of likelihood ratios per se for modelling relevance. But even if Pardo’s broader critique of likelihood ratios is correct, his point is one that speaks to epistemological theory; it is debatable whether it would have any practical implications for how best to communicate scientific findings to lay jurors.

10 Ligertwood and Edmond report incorrectly that Mr. Ryder calculated ‘a final likelihood that the shoe-print in question originated from a shoe worn by the accused’. In fact, Mr. Ryder did no such thing. He computed the probability of observing the shoe print under alternative hypotheses about its origin. It appears that Ligertwood and Edmond mistakenly equated this likelihood ratio with a source probability, which ironically helps to make their point: if judges and legal scholars misinterpret likelihood ratios so readily, how likely is it that jurors will understand them correctly? This misunderstanding may also underlie their unpersuasive argument that likelihood ratios are incompatible with our system of legal proof.

11 Empirical studies of this issue would be helpful although, as far as I know, none has yet been conducted.
5. Should experts present frequencies rather than likelihood ratios?

Ligertwood and Edmond propose that experts use frequencies rather than likelihood ratios to characterize the strength of forensic evidence. Using DNA evidence as an example, they suggest that it is better to tell juries that the expected frequency of the matching DNA profile in some reference population is one in a million (a random match probability) than to tell the jury that the observed profile is a million times more likely if it came from the accused than from a randomly selected suspect (a likelihood ratio). They argue that the random match probability is easier for jurors to understand and easier for them to integrate with other evidence when judging whether the traditional standard of proof has been met.

I agree that, compared with likelihood ratios, random match probabilities are easier for jurors to understand and use, and less susceptible to fallacious misuse. The problem with Ligertwood’s and Edmond’s proposal is that forensic evidence often cannot be characterized correctly using random match probabilities. Random match probabilities are suitable only when forensic scientists evaluate evidence in accordance with what we might call the match-frequency paradigm in which the expert sets criteria for a match, determines whether the items being compared meet those criteria, and if so makes an estimate of how frequently other members of a reference population would also meet those criteria. This paradigm can be used for DNA evidence, but only in simple cases in which a match is certain to occur under the prosecution hypothesis. It cannot be used in cases involving mixtures, uncertainty about number of contributors, and the possibility of undetected alleles. In these more complex cases, analysts are forced to use likelihood ratios because use of a single random match probability would be profoundly misleading (Curran and Buckleton, 2010; Gill et al., 2006; Evett and Weir, 1998).

The same problem arises in many other areas of forensic science. Geoffrey Morrison (2011) has argued cogently that ‘[f]or continuously valued data with inherent variability there can be no meaningful concept of a “match”, the numerator of the likelihood ratio cannot be either 0 or 1, and only a full likelihood ratio can be calculated.’ Morrison is most concerned with forensic voice comparison analysis, but he correctly points out that ‘continuously valued data with inherent variability’ are common in forensic science. Because results in these areas cannot be neatly dichotomized (as match or no-match), the strength of findings in these areas cannot be measured with a simple random match probability. The expert ‘must’ use a likelihood ratio to assess the strength of the evidence (Morrison, 2010).

But is it necessary for the expert to use a likelihood ratio to explain the evidence to the jury? If Ligertwood and Edmond are correct in arguing that jurors can more readily understand frequencies, experts might consider converting likelihood ratios into equivalent frequencies when presenting their results in court. I am not proposing that experts present random match probabilities (RMPs) but rather that they present what I will call ‘random match equivalents’ (RMEs). A DNA expert who computed the likelihood ratio in a mixture case to be 10 000 might, e.g. tell the jury that his analysis suggest that the evidence is ‘comparable in strength’ to finding a match on a characteristic found in only one person in 10 000 (from the relevant reference population). A shoe-print expert who computed the likelihood ratio to be approximately 100 might tell the jury that his analysis suggests that the evidence ‘is comparable in strength’ to finding a match on a characteristic found in approximately one shoe in 100 (from a relevant reference population). Although the RME could be criticized as artificial, I believe this approach would convey the value of the evidence to the court and the jury more effectively than likelihood ratios (or verbal equivalents) with less risk that the jurors would draw fallacious conclusions.
6. Should forensic experts ‘opine on the probability of source’?

The court in *R v T* suggested that Mr. Ryder should have limited his testimony to saying that the shoe print ‘could have been made’ by the defendant’s shoe. Testimony of this type suggests that information contained in the print was sufficient to allow the expert to divide shoes into two categories—those that ‘could’ have made the print and those that ‘could not’. But the conclusion ‘could have made’ is impossible to evaluate without information about the relative size of the two categories. It is nonsense to claim, as the court does, that ‘the term ‘could have made’ is a more precise statement of the evidence’ than testimony about a likelihood ratio (para. 73). A statement that might mean almost anything can hardly be considered precise. Courts in the USA generally do not allow experts to testify that a DNA sample ‘could have come’ from a particular individual without providing statistics on the frequency of the matching characteristics. As one court explained, ‘[w]ithout the probability assessment, the jury does not know what to make of the fact that the patterns match: the jury does not know whether the patterns are as common as pictures with two eyes, or as unique as the Mona Lisa.’ (*U.S. v. Yee*, 1991, p. 181). The same problem obviously applies to the use of ‘could have made’ testimony in connection with shoe prints and any other type of forensic evidence.

In *R v T* the court said that an expert might ‘in appropriate cases use his experience to express a more definitive evaluative opinion. . .’ (para. 74). Although the court failed to specify what form that opinion should take, I suspect that the court was alluding to what Ligertwood and Edmond call ‘opining on the probability of source’, which occurs when a shoe-print expert, e.g. expresses the opinion that a particular shoe ‘was likely to be the source’ of a print. Ligertwood and Edmond say they support this form of testimony, as does Mr. Bodziak. In the USA, ‘opining on the probability of source’ is a common occurrence in forensic science testimony, but I regard it as a bad practice that forensic scientists should avoid and courts should forbid.

It is impossible for a forensic expert to draw conclusions about source probability based solely on the forensic evidence. In order to ‘opine on the probability of source’ the expert must take account of (or take a position on) the strength of other evidence outside the expert’s domain, which requires the expert to invade the province of the jury. Consider a shoe-print expert who finds strong similarities between a defendant’s shoe and a print at the crime scene. Suppose that she concludes that only one shoe in 10 000 would be so similar. Can she conclude that the defendant’s shoe ‘probably made the print’? She certainly cannot do so without considering the strength of other evidence against the defendant. Although defendant’s shoe ‘matches’ the print, it is clear that other shoes would also match, and without additional evidence no conclusions can be drawn about the relative probability it was defendant’s shoe, rather than one of the others. To conclude that defendant’s shoe ‘probably made the print’ the expert must implicitly draw conclusions or make assumptions about the strength of the other evidence against the defendant, which invades the jury’s province. In my view, courts should never allow such testimony.13

---

12 Specifically, the expert must conclude (based on something other than the forensic evidence) that the prior odds of the defendant’s shoe making the print exceed the value $x$, where $x$ is the minimum value which, when multiplied by the likelihood ratio, produces a product greater than one.

13 The only possible exception would be instances in which the expert concludes, based solely on scientific evidence, that two items share one-of-a-kind or unique features. In such cases experts sometimes opine with certainty that the two items have a common source—which is tantamount to concluding that the source probability is 100% (*Thompson and Cole*, 2007; *Murphy and Thompson*, 2010). But serious questions have been raised about the scientific validity of such conclusions in the pattern-matching disciplines of forensic science (*National Research Council*, 2009). Moreover, as *Cole* (2009) has noted,
7. Conclusions

According to a famous legal maxim ‘hard cases make bad law’. \( R \nu T \) was unquestionably a hard case. The evidence was difficult to understand and the behaviour of the government’s expert, Mr. Ryder, was troubling. Ryder failed to provide a complete report of his analysis, leaving out the fact that he had based his conclusions in part on a likelihood ratio. His likelihood ratio, when examined by the court, had a ‘rough-and-ready’ quality (Morrison, 2011) that led the justices to question its scientific validity. And the justices mistakenly thought he was invading the jury’s province by relying on non-scientific evidence to reach his conclusions. It is understandable, then, that the justices reacted with disdain to his evidence. They wanted expert testimony that was transparent, understandable, valid and based solely on scientific analysis, and they believed that Mr. Ryder’s testimony met none of those goals. In rejecting the use of likelihood ratios, however, the court has unquestionably made bad law—law that threatens the very goals that the court was trying to achieve.

Likelihood ratio calculations (like those of Mr. Ryder) are far more transparent than the intuitive, experience-based judgements advocated by Mr. Bodziak. If the expert’s judgement rests on a weak scientific foundation, that fact becomes more apparent when the expert explicitly computes likelihood ratios, as Mr. Ryder did, than when the expert makes the kind of global evaluative judgement favoured by most forensic scientists in the USA. Additionally, while Mr. Ryder’s testimony did ‘not’ invade the province of the jury, allowing experts to ‘opine on the probability of source’ (as American experts often do) surely would invade the jury’s province. By rejecting the use of likelihood ratios in favour of the American approach, the court moved away from two of the goals it sought to achieve—and that is the tragedy of the case.

For forensic experts who recognize that likelihood ratios are an essential component of good science, the ruling in \( R \nu T \) is exasperating. But the case has already inspired a great deal of helpful commentary. In a common law system, bad law, once recognized as such, can be corrected in subsequent appellate rulings. \( R \nu T \) will not be the last word on these issues.

REFERENCES


‘a broad consensus in the scholarly literature rejects the notion that individualization and uniqueness constitute the proper conceptual foundation for the forensic sciences’.


