Bayes and the Law: Supplementary Material

Norman Fenton, Martin Neil and Daniel Berger

n.fenton@qmul.ac.uk

4 March 2016

This is the supplementary material in support of the following article:

Fenton N.E, Neil M, Berger D, "Bayes and the Law", Annual Review of Statistics and Its Application, Volume 3, 2016, doi: 10.1146/annurev-statistics-041715-033428

1 Additional Material for Introduction

The use of statistics in legal proceedings (both criminal and civil) has a long, but not terribly well distinguished, history that has been very well documented in (Finkelstein, 2009; Gastwirth, 2000; Kadane, 2008; Koehler, 1992; Vosk and Emery, 2014).

Despite the wisdom of the dissenting opinion quoted in DePass v U.S 1983 (see (Koehler, 1992)) that

> "*Most knowledge, and almost all legal evidence is probabilistic*"

the opinion and practice of the vast majority of legal professionals is far better captured in the much earlier ruling of Virginia v. Hawk in 1908 (again see (Koehler, 1992)) which asserted:

> "*To allow a jury dispose of a case simply upon a weighing of probabilities is to turn them loose into the field of conjecture*"

2 Additional Material on Hypotheses (Section 2)

- A ***hypothesis*** is a statement (typically Boolean) whose truth value we seek to determine, but is generally unknown - and which may never be known with certainty. Examples include:
  - "Defendant is guilty of the crime charged" (this is an example of an ***offense level hypothesis*** also called the ***ultimate hypothesis***, since in many criminal cases it is ultimately the only hypothesis we are really interested in). More specific examples are:
    - "Defendant is guilty of the murder of person X"
    - "Defendant was over the legal alcohol limit when breathalysed"
    - "The chance of a female applicant being successful is no different to the chance of a male applicant being successful" (this is an example of an ultimate hypothesis in a civil case of sex discrimination)
  - "Defendant was the source of DNA found at the crime scene" (this is an example of what is often referred to as ***a source level hypothesis*** (Cook et al., 1998a))

3 Additional Material on Bayes (Section 2)

To help understand Bayes' and to show how an error like the prosecutors fallacy might be spotted through intuitive reasoning that tallies with the Bayes result, an approach like that in Figure 1 is typically used with the following reasoning:

> Out of the 9,999 other people than the defendant at the scene we expect about 10 to have the same DNA match (since it occurs in about 1 in 1000 people). So the DNA match tells us that the defendant is one of about 11 who match.

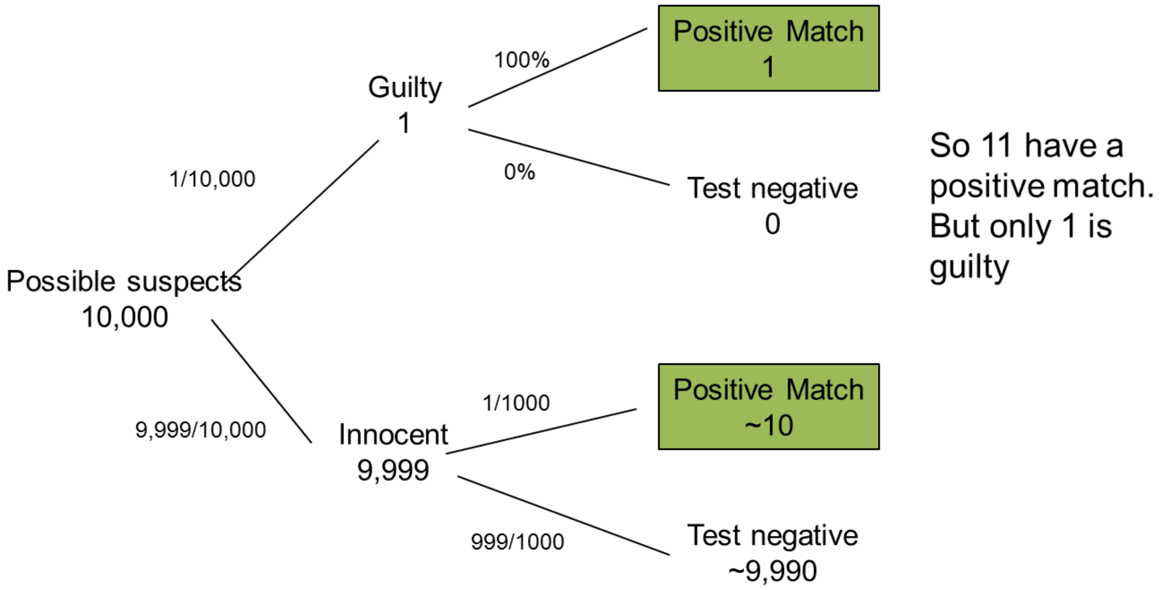

Figure 1 Visual explanation of Bayes. But in practice the problem is much more complex and this diagrammatic approach does not scale up

But one of the problems for lay people in understanding the Bayes result is that – in practice - there is rarely such a simple explanation (an issue discussed in Section 5). For example, a full DNA match will typically have a random match probability of 1 in a billion, so in such a case the reasoning Figure 1 no longer makes much sense, since it would require the notion of tiny fractions of a person.

4 Extended DNA Bayesian Network (Section 2)

To incorporate the possibility of contamination as well as the quality of the DNA traces there are further crucial hypotheses shown in the more complete version in Figure 2.

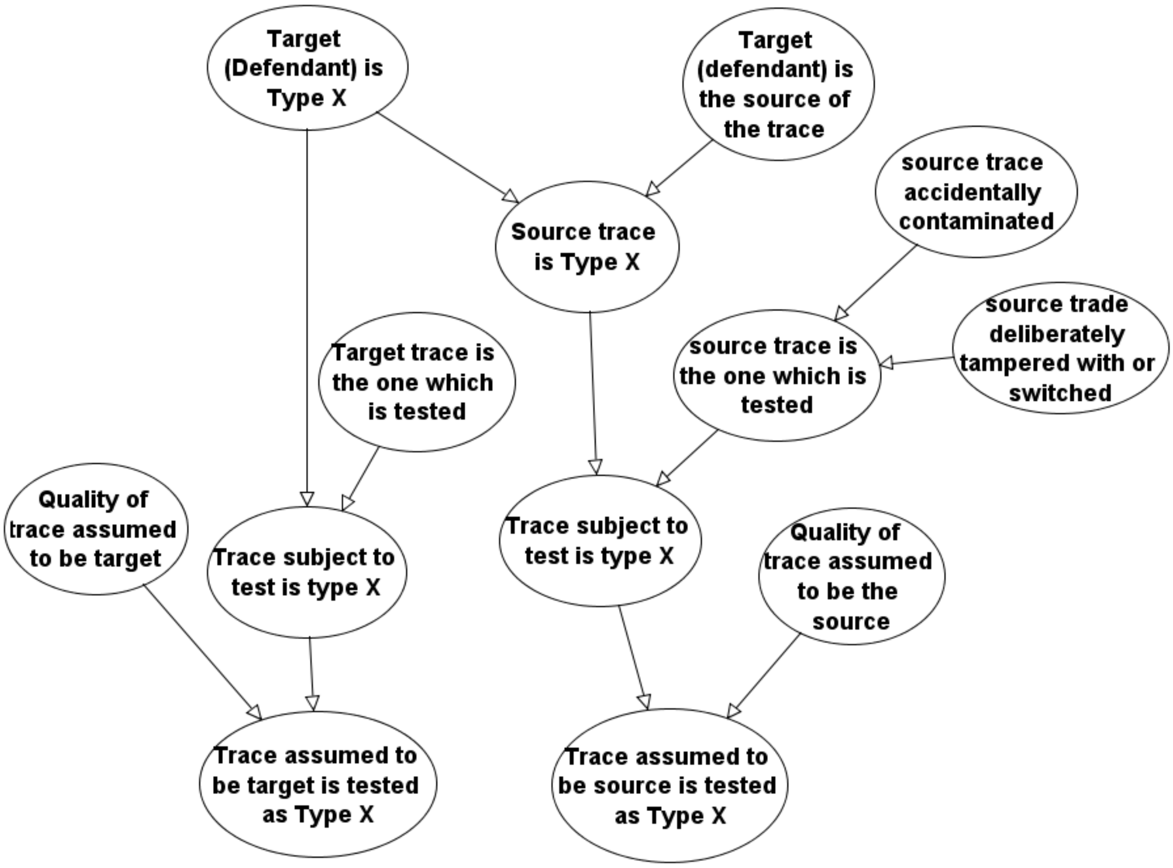

Figure 2 More complete BN model for simple DNA trace matching

Even this model does not include source level hypotheses like "Defendant was at the scene of the crime" and the ultimate hypothesis "Defendant committed the crime". These are only omitted here because, whereas the law might accept a statistical or forensic expert reasoning probabilistically about the source of the forensic evidence, it is presupposed that any probabilistic reasoning about the ultimate hypothesis is the province of the trier of fact, i.e., the judge and/or the jury.

5 Additional Material on Cohen problem (Section 2)

Moreover, in contrast to the argument in (Dawid, 1987), the BN approach is also scalable. Suppose, for example, we know that the truth or otherwise of A and B was dependent on some other fact C, i.e. that C is a common parent of A and B. Or suppose we know that witness 1 is more likely to lie if witness 2 lies. To perform the necessary Bayesian inference calculations by hand formulaically with such assumptions is no longer feasible. Indeed, in (Fienberg and Finkelstein, 1996) while discussing a similar problem involving a hypothesis H, two pieces of evidence E1 and E2 and two piece of background information I1 and I2 the authors avoid adding non-trivial assumptions about the relationship between the evidence because:

> "We have found the updating to be complex"

Yet in the BN it is a simple matter of either adding a node and/or adding a link between nodes and redefining the new NPTs. For example. to model the common cause parent of A and B we simply add a node C and define the NPTs of A and B conditioned on C. No other changes to the model are needed. If we assign the NPT of A given C to be 0.8 true when C is true and 0.8 false when C is false (and similarly for B given C) then the result of observing the same witness evidence as before is shown in the revised model in Figure 3.

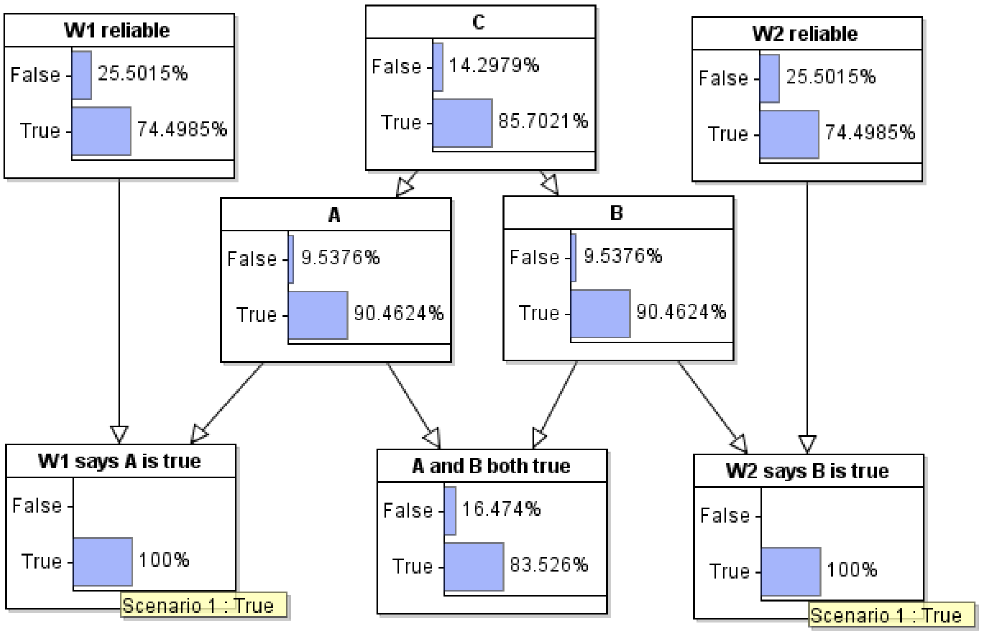

Figure 3 Revised BN model with common cause node C

Note that with the common cause node C, the probability that both A and B are both true now has increased to over 83%.

6 Additional Material on Context and Review (Section 3.1)

The vast majority of reported cases of explicit use of statistics in legal proceedings fall under this classification which, in (Fenton et al., 2015), we further sub-classify into:

- *Any form of class discrimination and bias* (such as discrimination against minority applicants for jobs/promotions or discrimination in criminal sentencing). Many such cases are discussed in (Bordman and Vining, 1983) (Fienberg and Finkelstein, 1996), (Finkelstein, 2009), (Gastwirth, 1997). An especially important example was New York City v Dept of Commerce 822 F. Supp 906 (E.D.N.Y 1993).. This case, which essentially tested the hypothesis that the 1990 decennial census discriminated between whites and minority groups, involved 14 expert statistical witnesses, although as reported in **Error! Reference source not found.** few used Bayesian methods. Bayes was, nevertheless explicitly mentioned. A commonly occurring issue in discrimination cases is Simpson's paradox (Bickel et al., 1975), which is best explained (and most easily avoided) from a causal, Bayesian perspective (Fenton and Neil, 2012; Pearl, 2000).

- *Any form of fraud/cheating* (including tax avoidance, price fixing, failure to disclose, cheating on exams etc). Many such cases are discussed in (Kadane, 2008). An especially important case is that of Marks v Stinson 1994 (discussed in detail in (Fienberg and Finkelstein, 1996)), which concerned the allegation of fraud (absentee ballots improperly obtained) in a special 1993 state senate election in Philadelphia. The Appeal Court judge relied heavily on the testimony of three statistical experts – one of whom used an explicitly Bayesian argument to compute the posterior probability that Marks had won the election based on a range of prior assumptions. Fienberg and Finkelstein assert that the judge misinterpreted the evidence provided by the experts and because the Bayesian expert had used a range of prior values to examine the sensitivity of assumptions, "his calculations were subject to an even greater misinterpretation that that of the other experts". So, although the judge did not dispute the validity of the use of Bayes, this was a clear example of how Bayesian reasoning can be misunderstood.

  Another important case was Minnesota v Philip Morris and six other tobacco companies and two tobacco trade groups 1994 (Gastwirth, 2000). This case alleged the defendants acted fraudulently, conspired to prevent development of less hazardous cigarettes and violated public trust to communicate accurately about the health effects of smoking. This case touches on the more general issue of epidemiological causation in the law (see (Loue, 2000) for a full discussion of related cases).

- *Possession of illegal materials/substances* (drugs, pornography etc) including illegal amounts of alcohol/drugs in the body while driving/working. Many such cases are covered in (Gastwirth, 2000), (Kadane, 2008). An especially important case was U.S v Shonubi, in which the key statistical issue was in determining the quantity of drugs involved as discussed (Izenman, 2000). The special statistical issues arising from alcohol and drug testing (including the potential for Bayes in resolving them) are covered extensively in (Vosk and Emery, 2014).

- *Having a physical or mental disability.* The most important issue here is in determining whether or not a person has an intellectual disability since this may determine whether or not the person can be tried in the first place and, in the extreme, executed. For example, in 2002, the Supreme Court in Atkins v. Virginia held it unconstitutional to execute people with an intellectual disability and left it to the states to define the term. Florida defined this as "significantly sub-average general intellectual functioning" where "subaverage" is an IQ of 70, two standard deviations below the mean of 100. The Florida Supreme Court interpreted the number 70 as an absolute cut off. The statistical problems with this approach (taking account of the uncertainty of IQ test measurement) are addressed in (Vosk and Emery, 2014). In 2104 the US Supreme Court ruled that it is unconstitutional to use an IQ score of 70 as a cutoff. The plaintiff in the case, convicted murderer Freddie Lee Hall, had been on death row in Florida for 35 years. He had taken multiple IQ tests, yielding scores ranging between 60 and 80, which Florida's Supreme Court ruled in 2012 to constitute an IQ above the 70 cut off and so too high to qualify for a reprieve from execution.

Depressingly few of the above cases involve explicit use of Bayes, and that where Bayes was used it was often misunderstood. Typically, where there is sufficient relevant data, classical statistical hypothesis testing rather than Bayes has been used to determine whether the null hypothesis of ‘no dubious behaviour’ can be rejected at an appropriate level of significance. This is despite the known problems of interpreting the resulting p-values and confidence intervals, which the use of the Bayesian approach to hypothesis testing avoids (Fenton and Neil, 2012; Press, 2002). The potential for misinterpretation is enormous (Vosk and Emery, 2014). For example, in (*U.S ex. Rel. DiGiacomo v Franzen, 680 F.2d 515 (7th Cir.)*, 1982) a forensic expert quoting a p-value from a study published in the literature interpreted it as the probability that each of the head hairs found on the victim were not the defendants. Sadly, there is a lack of awareness among statisticians that modern tools (such as those discussed in Section 2) make it possible to easily perform the necessary analysis for Bayesian hypothesis testing.

7 Additional Material on Section 3.4

The idea that different pieces of (possibly competing) evidence about a hypothesis H are combined in order to update our belief in H is central to all legal proceedings. As long ago as 1875 the Belhaven and Stenton Peerage case (described in detail in (Darroch, 1987)) established a precedent for combining different pieces of circumstantial evidence:

> *“My Lords, in dealing with circumstantial evidence, we have to consider the weight which is to be given to the united force of all the circumstances put together. You may have a ray of light so feeble that by itself it will do little to elucidate a dark corner. But on the other hand, you may have a number of rays, each of them insufficient, but all converging and brought to bear upon the same point, and, when united, producing a body of illumination which will clear away the darkness which you are endeavouring to dispel.”*

This approach is also central to Bayesian reasoning, which is easily able to cope with very diverse types of potentially competing evidence. Yet, it is difficult to find ***any*** well reported examples of the successful use of Bayes in combining diverse evidence in a real case. There are two reasons for this. One is to do with the lack of awareness of tools for building and running BN models that enable us to do Bayesian inference for legal arguments involving diverse related evidence. The second (not totally unrelated to the first) is due to the spectacular failure in one well publicised case where Bayes was indeed used to combine diverse competing evidence.

The case was that of (*R v Adams [1996] 2 Cr App R 467, [1996] Crim LR 898, CA and R v Adams [1998] 1 Cr App R 377*, 1996) already referred to above in connection to the misleading presentation of DNA evidence. This was a rape case (discussed in detail in (Donnelly, 2005)) in which the only prosecution evidence was that the defendant’s DNA matched that of a swab sample taken from the victim. The defence evidence included an alibi and the fact that the defendant did not match the victim’s description of her attacker. At trial the prosecution had emphasised the very low random match probability (1 in 200 million) of the DNA evidence. The defence argued that if statistical evidence was to be used in connection with the DNA evidence, it should also be used in combination with the defence evidence and that Bayes Theorem was the only rational method for doing this. The defence called a Bayesian expert (Prof Peter Donnelly) who explained how, with Bayes, the posterior probability of guilt was much lower when the defence evidence was incorporated. The appeal rested on whether the judge misdirected the jury as to the evidence in relation to the use of Bayes and left the jury unguided as to how that theorem could be used in properly assessing the statistical and non-statistical evidence in the case. The Appeal was successful and a retrial was ordered, although the Court was scathing in its criticism of the way Bayes was presented, stating:

> **"The introduction of Bayes' theorem into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task.**

While this statement is something that we are not totally unsympathetic to, the judge's subsequent statement is much more troubling (and it is a challenge we return to in Section 5):

> **The task of the jury is … to evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them"**

At the retrial it was agreed by both sides that the Bayesian argument should be presented in such a way that the jury could perform the calculations themselves (a mistake in our view). The jury were given a detailed questionnaire to complete to enable them to produce their own prior likelihoods, and calculators to perform the necessary Bayesian calculations from first principles. Adams was, however, again convicted. A second appeal was launched (based on the claim that the Judge had not summed up Donnelly's evidence properly and had not taken the questionnaire seriously). This appeal was also unsuccessful, with the Court not only scathing about the use of Bayes in the case but essentially ruling against its future use:

> *"..We do not consider that [the jury] will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion. We are very clearly of opinion that in cases such as this, lacking special features absent here, expert evidence should not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial."*

While the subsequent R v T ruling in 2010 dealt a devastating blow for the use of Bayes in presenting forensic (non DNA) evidence, the ruling against the use of Bayes in R v Adams is actually far more damaging. This is because it rules against the very use where Bayes has the greatest potential to simplify and clarify complex legal arguments. The fact that the complex presentation of Bayes in the case was (rightly) considered to be its death knell is especially regrettable given that in 1996 the tools for avoiding this complexity were already widely available.

8 Additional Material on Section 3.6

Bayes has also played an indirect (normally unreported) role in many cases. Based just on our own experience (and on that of colleagues) as expert consultants to lawyers, we know of dozens of cases in which Bayes was used to help lawyers in the preparation and presentation of their cases. Because of confidentiality (and sometimes sensitivity) this work normally cannot be publicised. There are rare exceptions such as (Kadane, 1990) who describes his proposed Bayesian testimony in an age discrimination case (settled before trial); some of the work we have been able to make public retrospectively is:

- R v Levi Bellfield 2007-08 (Fenton and Neil, 2011): a murder case in which we used BNs to identify and explain a range of fallacies (including several instances of the prosecutor's fallacy) in the prosecution opening. As a result of the defence barrister using our approach informally, the Judge instructed the prosecuting barrister not to repeat any of the fallacies identified in his closing arguments. Nethertheless, just days later in another murder case (R vs Mark Dixie, accused of murdering Sally-Anne Bowman) involving the same prosecuting QC several newspapers reported (on 12 Feb 2008) the following concerning evidence by a forensic scientist for the prosecution:

  > "Forensic scientist Julie-Ann Cornelius told the court the chances of DNA found on Sally Anne's body not being from Dixie were a billion to one."

  If the reports were accurate then this was again a blatant instance of the prosecutor fallacy.

- B v National Health Service 2005 (Fenton and Neil, 2010): a medical negligence case in which we provided a holistic BN to quantify the different prior risks involved in performing the actual invasive test that caused serious injury to the plaintiff and the standard non-invasive alternative that would normally have been used. The plaintiff's solicitor was able to use our argument semi-formally (it showed that with any reasonable range of prior assumptions the invasive test produced a worse outcome form a risk perspective) and the plaintiff was awarded substantial damages.

Because of the RvT ruling, most of our recent work cannot be made public, since judges may not welcome arguments that they may suspect are even informally based on Bayesian reasoning. We can say that we have advised lawyers in very high profile murder and rape cases and that much of the work involves exposing weaknesses in the statistical interpretation of DNA evidence (especially low-template DNA) including taking account of the potential for different types of errors. A very high-level explanation of the problem – showing why a full BN approach is required – is provided in (Fenton et al., 2014).

While our own experience suggests that the use of Bayesian methods (especially BNs) has been effective, in the absence of sufficient published studies it is impossible to state definitively whether this experience is widespread.

9 Additional Material to Section 4.1

There is a persistent attitude among some members of the legal profession that probability theory has no role to play at all in the courtroom. Indeed, the role of probability – and Bayes in particular – was dealt another devastating and surprising blow in a 2013 UK Appeal Court case ruling (Nulty & Ors v Milton Keynes Borough Council, 2013) (discussed in (Spiegelhalter, 2013)). The case was a civil dispute about the cause of a fire. Originally, it was concluded that the fire had been started by a discarded cigarette, although this seemed an unlikely event in itself because the other two explanations were even more implausible. The Appeal Court rejected this approach, effectively arguing against the entire Bayesian approach to measuring uncertainty by asserting essentially that there was no such thing as probability for an event that has already happened but whose outcome is unknown. Specifically Point 37 of the ruling asserted (about the use of such probabilities):

I would reject that approach. It is not only over-formulaic but it is intrinsically unsound. The chances of something happening in the future may be expressed in terms of percentage. Epidemiological evidence may enable doctors to say that on average smokers increase their risk of lung cancer by X%. But you cannot properly say that there is a 25 per cent chance that something has happened: Hotson v East Berkshire Health Authority [1987] AC 750. Either it has or it has not. In deciding a question of past fact the court will, of course, give the answer which it believes is more likely to be (more probably) the right answer than the wrong answer, but it arrives at its conclusion by considering on an overall assessment of the evidence (i.e. on a preponderance of the evidence) whether the case for believing that the suggested event happened is more compelling than the case for not reaching that belief (which is not necessarily the same as believing positively that it did not happen).

The bad news is that we know of many legal professionals who are similarly seduced by the notion that 'there is no such thing as probability'. As an eminent lawyer told us:

"*Look, the guy either did it or he didn't do it. If he did then he is 100% guilty and if he didn't then he is 0% guilty; so giving the chances of guilt as a probability somewhere in between makes no sense and has no place in the law*".

10 Additional Material on Section 4.3

It is also important to note that an additional danger in allowing H_d to be something different from "not H_p" is that, in practice, forensic experts may come up with an H_d that is not even mutually exclusive to H_p. This was also shown to be a real problem in the transcript of the R v Barry George appeal (Fenton et al., 2013a) where in considering the Firearm Discharge Residue evidence it appears that at least one expert witness was assuming that the 'alternative hypothesis to H_p ("BG was the man who shot JD") was H_d ("Integrity of BG coat was corrupted"). These two hypotheses are not mutually exclusive (since both may be true) and in such circumstances the LR is meaningless.

To ensure that these issues can be properly modelled in a BN the work in **Error! Reference source not found.** describes a method for:

- Defining the nodes of the BN in such a way as to carefully distinguish between those hypotheses that are mutually exclusive and exhaustive and those that are not. This ensures that experts are only ever required to consider pairs of mutually exclusive hypotheses when specifying likelihoods.
- Modelling the causal story between hypotheses and different pieces of evidence

11 Additional Material on Section 4.6

The UK Forensic Science Service Guide (Puch-Solis et al., 2012) recommends that, to help lawyers and jurors understand the significance of a LR, it should be presented on an equivalent verbal scale as shown in Figure 4.

Limited evidence to support	LR <1-10
Moderate evidence to support	LR 10-100
Moderately strong evidence to support	LR 100-1000
Strong evidence to support	LR 1000-10000
Very strong evidence to support	LR >10000

Figure 4 Verbal scale for LR

This recommendation (criticised in (Mullen et al., 2014)) contrasts with US courts that have advised against verbal scales and instead recommended that posterior probabilities should be provided based on a range of priors for the given LR. We believe that the US Courts approach is correct, but recognise that in the UK explicit use of numerical LRs are increasingly snubbed following the RvT judgment.

As part of our own legal advisory/expert witness work we have examined numerous expert reports in the last five years (primarily, but not exclusively from forensic scientists). These reports considered different types of match evidence in murder, rape, assault and robbery cases. The match evidence includes not just DNA, but also handprints, fibre matching, footwear matching, soil and particle matching, matching specific articles of clothing, and matching cars and their number plates (based on low resolution CCTV images). Although the DNA experts in many of these cases provided explicit probability statements (such as "the probability that the trace found came from a person unrelated to X is less than one in a billion") the other experts have invariably provided verbal quasi-probabilistic statements instead. Some of these used the FSS verbal scale as in:

> " the evidence provides moderate/strong/very strong support for the proposition that Y belongs to /comes from X"

But many have also used statements like:

> ".. the probability/chances that Y belongs to anybody other than X is so small that it can be discounted[1]"
> ".. the probability/chances that Y comes from anything/anywhere other than Z is so small that it can be discounted"

In all cases there was some kind of database or expert judgement on which to estimate frequencies and 'random match' probabilities, and in most cases there appears to have been some attempt to compute the LR. However, in all but the DNA cases, the explicit statistics and probabilities were not revealed in court – in several cases this was as a direct result of the RvT ruling which has effectively pushed explicit use of numerical LR 'underground'. Indeed,

[1] We also found lawyers who automatically assumed that evidence of fingerprint and DNA 'matches' were synonymous with 'identification'

we have seen expert reports that contained the explicit data being formally withdrawn as a result of RvT. This is one of the key negative impacts of RvT - we feel it is extremely unhelpful that experts are forced to suppress explicit probabilistic information.

12 Additional Material on Section 4.7

Despite the multiple publications applying BNs to legal arguments, even many Bayesian statisticians are either unaware of these breakthroughs or are reluctant to use the available technology. Hence, to avoid overly complex manual calculations they prefer instead to stick to overly simplified models that are not accurate representations of legal arguments. Consider the examples of DNA (and other types of) forensic match evidence that was represented in **Error! Reference source not found.** ad Figure 2. It turns out that, even in the simplest case, in practice expert witnesses are either not aware of the need to incorporate the possibility of errors in their analysis or they do not know how to do it. Indeed, on the basis of several dozen confidential reports from expert witnesses that we have been asked to scrutinize in the last 5 years, we believe that in practice proper analysis (i.e. accounting for possible testing errors) is not undertaken even in the simplest case. In fact ***in not one report did the experts make any attempt to incorporate into their explicit (or implicit) calculations the probabilistic uncertainty of match errors***. Where experts considered the possibility of match errors at all it was only in the context of cross-contamination, which in generic terms can be considered as the case where the trace being tested is not the same as the trace associated with the crime or crime scene. In ***all such cases*** the experts simply dismissed such a possibility as either "impossible" or "so small that it can be discounted".

If one tries to use Bayes theorem 'manually' to represent a legal argument one of the following results is inevitable:

1. *To ensure the calculations can be easily computed manually, the argument is made so simple that it no longer becomes an adequate representation of the legal problem*. This is essentially what happens with every example of the likelihood ratio used in a legal argument that attempts to capture the probative value of some evidence – the problem is reduced to a model involving just a single hypothesis H (such as 'guilty' or 'defendant is the source of the trace') and to assume either a single piece of evidence E or several pieces of evidence $E_1, \ldots, E_n$ that are completely independent, so that
$$P(E_1, \ldots, E_n | H) = P(E_1|H) * \ldots * P(E_n | H)$$
2. A non-trivial model is developed and the Bayesian calculations are written out and explained from first principles and the net result is to totally bemuse legal professionals and jurors. This was, of course, the problem in R v Adams. In (Fenton et al., 2015) we show other examples where statisticians provide unnecessarily complex arguments).

The manual approach is also not scalable since it would otherwise mean having to explain and compute one of the BN inference algorithms, which even professional mathematicians find daunting.

Since BNs (and efficient easy-to-use tools to do the inference calculations) have been around for nearly 25 years now, this manual approach with its built-in limitations is unnecessary. It is like trying to do an exceptionally difficult mathematical calculation (involving say, cube roots and division by numbers involving multiple decimal points) without a calculator. Since we have calculators there is no good reason not to use them.

13 Acknowledgements

This work was supported in part by European Research Council Advanced Grant ERC-2013-AdG339182-BAYES_KNOWLEDGE.

14 References

Agena Ltd, 2015. AgenaRisk.

Aitken, C., many other signatories, 2011. Expressing evaluative opinions: A position statement. Sci. Justice 51, 1–2.

Aitken, C.G.G., Taroni, F., 2004. Statistics and the evaluation of evidence for forensic scientists (2nd Edition). John Wiley & Sons, Ltd, Chichester.

Alberink, I., de Jongh, A., Rodriguez, C., 2014. Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. J. Forensic Sci. 59, 70–81.

Anderson, T., Schum, D.A., Twining, W., 2005. Analysis of Evidence. Cambridge University Press.

Balding, D.J., 2004. Weight-of-evidence for Forensic DNA Profiles. Wiley.

Balding, D.J., Donnelly, P., 1994. Prosecutor's fallacy and DNA evidence. Crim. Law Rev. October, 711–722.

Balding, D.J., Donnelly, P., 1996. Evaluating DNA profile evidence when the suspect is identified through a database search. J. Forensic Sci. 41, 599–603.

Berger, C.E.H., Buckleton, J., Champod, C., Evett, I., Jackson, G., 2011. Evidence evaluation: A response to the court of appeal judgement in R v T. Sci. Justice 51, 43–49.

Berger, D., 2014. Improving Legal Reasoning using Bayesian Probability Methods. Queen Mary University of London.

Bickel, P.J., Hammel, E.A., O'Connell, J.W., 1975. Sex bias in graduate admissions: Data from Berkley. Science (80-.). 187, 398–404.

Bordman, A.E., Vining, A.R., 1983. Statistics in employment discrimination cases. Law Contemp. Probl. 64.

Brown v. Smith, 137 N.C.App. 160, 526 S.E.2d 686, 2000.

Casscells, W., Graboys, T.B., 1978. Interpretation by physicians of clinical laboratory results. N. Engl. J. Med. 299 , 999–1001.

Cohen, L.J., 1977. The Probable and the Provable. Clarendon Press, Oxford.

Cole v. Cole, 74 N.C.App. 247, 328 S.E.2d 446, aff'd, 314 N.C. 660, 335 S.E.2d 897 (per curiam), 1985.

Collins, A., Morton, N.E., 1994. Likelihoof ratios for DNA identification. Proc Natl Acad Sci U S A 91, 6007–6011.

Cook, R., Evett, I.W., Jackson, G., Jones, P., Lambert, J.A., 1998a. A hierarchy of propositions: deciding which level to address in casework. Sci. Justice 38, 231–239.

Cook, R., Evett, I.W., Jackson, G., Jones, P.J., Lambert, J.A., 1998b. A model for case assessment and interpretation. Sci. Justice 38, 151–6.

Cosmides, L., Tooby, J., 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition 58 , 1–73.

Darroch, J., 1987. Probability and criminal trials: some comments prompted by the Splatt trial and the Royal Commission. Prof. Stat. 6, 3–7.

Dawid, A., Mortera, J., 1998. Forensic identification with imperfect evidence. Biometrika 85, 835–849.

Dawid, A.P., 1987. The Difficulty About Conjunction. J. R. Stat. Soc. Ser. D (The Stat. 36, 91–97.

Dawid, A.P., Musio, M., Fienberg, S.E., 2015. From Statistical Evidence to Evidence of Causality. Bayesian Anal.

Dawid, P., 2005. Statistics on Trial. Significance 2, 6–8.

Dawid, P., Mortera, J., 2008. Probability and Evidence. In: Rudas, T. (Ed.), Handbook of Probability: Theory and Applications. Sage Publications Inc, pp. 403–422.

Donnelly, P., 2005. Appealing Statistics. Significance 2, 46–48.

Edwards, W., 1991. Influence Diagrams, Bayesian Imperialism, and the Collins case: an appeal to reason. Cardozo Law Rev. 13, 1025–1079.

European Research Council, 2015. Advanced Grant ERC-2013-AdG339182-BAYES_KNOWLEDGE. Effective Bayesian Modelling with Knowledge Before Data [WWW Document]. URL http://bayes-knowledge.com/

Evett, I.W., 1995. Avoiding the transposed conditional. Sci. Justice 35, 127–131.

Evett, I.W., Foreman, L.A., Jackson, G., Lambert, J.A., 2000. DNA profiling: a discussion of issues relating to the reporting of very small match probabilities. Crim. Law Rev. May, 341–355.

Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J., Champod, C., 2002. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. J. Forensic Sci. 47, 520–530.

Evett, I.W., Weir, B.S., 1998. Interpreting DNA evidence : statistical genetics for forensic scientists. Sinauer Associates.

Faigman, D.L., Baglioni, A.J., 1988. Bayes' theorem in the trial process. Law Hum. Behav. 12, 1–17.

Fenton, N.E., 2011. Science and law: Improve statistics in court. Nature 479, 36–37.

Fenton, N.E., 2014. Assessing evidence and testing appropriate hypotheses. Sci Justice 54, 502–504.

Fenton, N.E., Berger, D., Lagnado, D.A., Neil, M., Hsu, A., 2013a. When "neutral" evidence still has probative value (with implications from the Barry George Case). Sci. Justice.

Fenton, N.E., Lagnado, D.A., Neil, M., 2013b. A General Structure for Legal Arguments Using Bayesian Networks. Cogn. Sci. 37.

Fenton, N.E., Neil, M., 2000. The Jury Fallacy and the use of Bayesian nets to simplify probabilistic legal arguments. Math. Today (Bulletin IMA) 36, 180–187.

Fenton, N.E., Neil, M., 2010. Comparing risks of alternative medical diagnosis using Bayesian arguments. J. Biomed. Inform. 43, 485–495.

Fenton, N.E., Neil, M., 2011. Avoiding Legal Fallacies in Practice Using Bayesian Networks. Aust. J. Leg. Philos. 36, 114–150.

Fenton, N.E., Neil, M., 2012. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Boca Raton.

Fenton, N.E., Neil, M., Berger, D., 2015. Bayes and the Law: Supplementary Material.

Fenton, N.E., Neil, M., Hsu, A., 2014. Calculating and understanding the value of any type of match evidence when there are potential testing errors. Artif. Intell. Law 22, 1–28.

Fienberg, E, S., 1989. The Evolving Role of Statistical Assessments as Evidence in the Courts. Springer Verlag.

Fienberg, S.E., 2011. Bayesian Models and Methods in Public Policy and Government Settings. Stat. Sci. 26, 212–226.

Fienberg, S.E., Finkelstein, M.O., 1996. Bayesian statistics and the law. In: Bernardo, J.M., Berge, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 5. Oxford University Press, pp. 129–146.

Fienberg, S.E., Kaye, D.H., 1991. Legal and Statistical Aspects of Some Mysterious Clusters. J. R. Stat. Soc. (series A) 154, 61–74.

Finkelstein, M.O., 2009. Basic Concepts of Probability and Statistics in the Law. Springer Science & Business Media.

Finkelstein, M.O., Fairley, W.B., 1970. A Bayesian Approach to Identification Evidence. Harv. Law Rev. 83, 489–517.

Finkelstein, M.O., Levin, B.A., 2001. Statistics for Lawyers. Springer.

Forrest, A.R., 2003. Sally Clark - a lesson for us all. Sci. Justice 43, 63–64.

Freckelton, I., Selby, H., 2005. Expert Evidence: Law, Practice, Procedure and Advocacy, 3rd ed. Lawbook Co., Australia.

Fung, W.K., 2003. User-friendly programs for easy calculations in paternity testing and kinship determinations. Forensic Sci. Int. 136, 22–34.

Gastwirth, J.L., 1997. Statistical evidence in discrimination cases. J R. Stat. Soc. A 160, 289–303.

Gastwirth, J.L., 2000. Statistical Science in the Courtroom.

Gigerenzer, G., 2002. Reckoning with Risk: Learning to Live with Uncertainty. Penguin Books, London.

Gill, R., 2014. How to become a convicted serial killer (without killing anyone). Justice Gap.

Good, I.J., 1995. When batterer turns murderer. Nature 375, 541.

Good, P.I., 2001. Applying Statistics in the Courtroom: A New Approach for Attorneys and Expert Witnesses. CRC Press.

Haigh, J., 2003. Taking Chances: Winning with Probability, 2nd ed. Oxford University Press.

Hepler, A.B., Dawid, A.P., Leucari, V., 2007. Object-oriented graphical representations of complex patterns of evidence. Law, Probab. Risk 6, 275–293.

Hill, R., 2005. Reflections on the cot death case. Significance 2.

Isaac Newton Institute Cambridge, 2015. FOS: Probability and Statistics in Forensic Science [WWW Document]. URL https://www.newton.ac.uk/event/fos

Izenman, A.J., 2000. Assessing the Statistical Evidence in the Shonubi Case. In: Gastwirth, J.L. (Ed.), Statistical Science in the Courtroom. Springer-Verlag New York Inc., pp. 415–443.

Jackson, G., Aitken, C., Roberts, P., 2013. PRACTITIONER GUIDE NO 4: Case Assessment and Interpretation of Expert Evidence.

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

James, S.H., Kish, P.E., Sutton, T.P., 2005. Principles of Bloodstain Pattern Analysis: Theory and Practice. CRC Press.

Jowett, C., 2001. Lies, Damned Lies, and DNA Statistics: DNA Match testing. Bayes' Theorem, and the Crminal Courts. Med Sci. Law 41, 194–205.

Kadane, J.B., 1990. A statistical analysis of adverse impact of employer decisions. J Amer. Stat. Assoc 85, 607–614.

Kadane, J.B., 2008. Statistics in the Law: A Practitioner's Guide, Cases, and Materials . Oxford University Press.

Kaye, D.H., 2001. Two Fallacies About DNA Databanks for Law Enforcement. Brooklyn Law Rev. 67, 179–206.

Kaye, D.H., 2007. Revisiting "Dreyfus": A More Complete Account of a Trial by Mathematics. Minn. Law Rev. 91, 825–835.

Koehler, J.J., 1992. Probabilities in the Courtroom: An Evaluation of the Objections and Policies. In: Kagehiro, D.K., Laufer, W.S. (Eds.), Handbook of Psychology and Law. Springer-Verlag, New York, pp. 167–184.

Koehler, J.J., 1993a. Error and Exaggeration in the Presentation of DNA Evidence at Trial. Jurimetrics 34, 21–39.

Koehler, J.J., 1993b. The Base Rate Fallacy Myth. Psycoloquy 4.

Langley, W., 2012. The case against DNA. Telegraph.

Lauritzen, S.L., Spiegelhalter, D.J., 1988. Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). J. R. Statis. Soc. B, 50, No 2, pp 157-224.

Loue, S., 2000. Epidemiological causation in the legal context: substance and procedures. In: Gastwirth, J.L. (Ed.), Statistical SCience in the Courtroom. Springer-Verlag New York Inc., pp. 263–280.

Lucy, D., Schaffer, B., 2007. Nature of legal evidence. The University of Edinburgh, Edinburgh. EH9 3JZ.

MARSHALL, T.C., SLATE, J., KRUUK, L.E.B., PEMBERTON, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7, 639–655.

Meester, R., Collins, M., Gill, R., van Lambalgen, M., 2007. On the (ab)use os statistics in the legal case against the nurse Lucia de B. Law, Probab. Risk 5, 233–250.

Meester, R., Sjerps, M., 2003. The Evidential Value in the DNA Database Search Controversy and the Two-Stain Problem. Biometrics 59, 727–732.

Meester, R., Sjerps, M., 2004. Why the Effect of Prior Odds Should Accompany the Likelihood Ratio When Reporting DNA Evidence. Law, Probab. Risk 3, 51–62.

Meier, P., Zabell, S., 1980. Benjamin Peirce and the Howland Will. J. Am. Stat. Assoc. 75, 497–506.

Morrison, G.M., 2012. The likelihood ratio framework and forensic evidence in court: a response to RvT. Int. J. Evid. Proof 16.

Mullen, C., Spence, D., Moxey, L., Jamieson, A., 2014. Perception problems of the verbal scale. Sci. Justice 54, 154–8.

Murphy, P., 2003. Evidence, Proof, and Facts: A Book of Sources. Oxford University Press.

National Research Foundation, 2009. Strengthening Forensic Science in the United States: A Path Forward.

Neil, M., Tailor, M., Marquez, D., 2007. Inference in hybrid Bayesian networks using dynamic discretization. Stat. Comput. 17, 219–233.

Nobles, R., Schiff, D., 2005. Misleading statistics within criminal trials: the Sally Clark case. Significance 2, 17–19.

Nordgaard, A., Hedell, R., Ansell, R., 2012. Assessment of forensic findings when alternative explanations have different likelihoods-“Blame-the-brother”-syndrome. Sci. Justice 52, 226–36.

North Carolina, 2009. Trial Judges’ Bench Book District Court Edition, Volume 1, Family Law. Paternity Chapter.

Nuemann, C., Evett, I.W., Skerrett, J., 2011. Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. J. R. Stat. Soc A 175, 1–26.

Nulty & Ors v Milton Keynes Borough Council, 2013. Nulty & Ors v Milton Keynes Borough Council EWCA Civ 15 (24 January 2013).

Pearl, J., 1988. Probabilistic reasoning in intelligent systems. Morgan Kaufmann , Palo Alto, CA.

Pearl, J., 2000. Causality: Models Reasoning and Inference. Cambridge University Press.

People v. Collins, 438 P. 2d 33 (68 Cal. 2d 319 1968) , 1968.

Press, S.J. (Ed.), 2002. Subjective and Objective Bayesian Statistics, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Puch-Solis, R., Roberts, P., Pope, S., Aitken, C., 2012. PRACTITIONER GUIDE NO 2: Assessing the Probative Value of DNA Evidence, Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Royal Statistical Society.

R v Adams [1996] 2 Cr App R 467, [1996] Crim LR 898, CA and R v Adams [1998] 1 Cr App R 377, 1996.

R v Alan James Doheny and Gary Adams, 1996. , Smith Bernal Reporting Ltd. Court of Appeal Criminal Division.

R v Deen, 1994.

R v T, 2010. R v T, EWCA Crim 2439, Case No 2007/03644/D2.

R. v. George, EWCA Crim 2722, 2007.

Redmayne, M., 1995. DNA evidence, probability and the courts. Crim. Law Rev. 464.

Redmayne, M., Roberts, P., Aitken, C., Jackson, G., 2011. Forensic Science Evidence in Question. Crim. Law Rev. 347–356.

Roberts, P., Aitken, C., Taroni, F., Biedermann, A., 2013. PRACTITIONER GUIDE NO 3: The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science.

Robertson, B., Vignaux, G.A., Berger, C.E.H., 2011. Extending the confusion about Bayes. Mod. Law Rev. 74, 444–455.

Robertson, B., Vignaux, T., 1995. Interpreting Evidence: Evaluating Forensic Science in the Courtroom. John Wiley and Son Ltd.

Robertson, B., Vignaux, T., 1997. Bayes’ Theorem in the Court of Appeal. Crim. Lawyer 4–5.

Robertson, B., Vignaux, T., 1998. Don’t teach Statistics to Lawyers! Proc. Fifth Int. Conf. Teach. Stat.

Royal Statistical Society, 2015. Statistics and the Law Working Group [WWW Document]. URL http://www.rss.org.uk/RSS/Influencing_Change/Statistics_and_the_law/RSS/Influencing_Change/Current_projects_sub/Statistics_and_the_law.aspx

Saks, M.J., Thompson, W.C., 2003. Assessing Evidence: Proving Facts. In: Carson, D., Bull, R. (Eds.), Handbook of Psychology in Legal Contexts (Second Edition). pp. 329–345.

Schneps, L., Colmez, C., 2013. Math on Trial Basic Books, 2013. Basic Books.

Sjerps, M., Berger, C., 2012. How clear is transparent? Reporting expert reasoning in legal cases. Law, Probab. Risk 11 , 317–329.

Spiegelhalter, D.J., 2013. Court of Appeal bans Bayesian probability (and Sherlock Holmes) [WWW Document]. URL http://understandinguncertainty.org/court-appeal-bans-bayesian-probability-and-sherlock-holmes

State of New Jersey v J.M. Spann, 130 N.J. 484: 617 A. 2d 247, 1993.

State v. Jackson, 320 N.C. 452, 358 S.E.2d 679, 1987.

Taroni, F., Aitken, C., Garbolino, P., Biedermann, A., 2014. Bayesian Networks and Probabilistic Inference in Forensic Science, 2nd ed. John Wiley & Sons, Chichester, UK.

Thompson, J., Danielson, S., Hartmann, E., Bucheger, D., Koplitz, J., 2001. Trial by Probability.

Thompson, W.C., 1996. DNA Evidence in the O.J. Simpson Trial. Color. Law Rev. 67, 827–857.

Thompson, W.C., Schumann, E.L., 1987. Interpretation of statistical evidence in criminal trials. Law Hum. Behav. 11, 167–187.

Thompson, W.C., Taroni, F., Aitken, C.G.G., 2003. How the probability of a false positive affects the value of DNA evidence. J. Forensic Sci. 48, 47–54.

Tillers, P., Gottfried, J., 2007. Case comment - United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y.2005): A collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable? Law, Probab. Risk 5, 5: 135–157.

Tillers, P., Green, E. (Eds.), 1988. Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism. Kluwer Academic Publishers.

Tribe, L.H., 1971. Trial by Mathematics. Harv. Law Rev. 84, 1329.

Triggs, C.M., Buckleton, J.S., 2004. Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. Law, Probab. Risk 3, 73–82.

U.S ex. Rel. DiGiacomo v Franzen, 680 F.2d 515 (7th Cir.), 1982.

University of Pittsburg, D.S.L., 2015. GeNIe: Graphical Network Interface.

Vosk, T., Emery, A.F., 2014. Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists. CRC Press.

Wagenaar, W.A., 1988. The Proper Seat: A Bayesian Discussion of the Position of Expert Witnesses. Law Hum. Behav. 12, 499–510.

Zeisel, H., Kaye, D., 1997. Prove It with Figures: Empirical Methods in Law and Litigation. Springer Verlag.

Ziliak, S.T., McCloskey, D.N., 2008. The Cult of Statistical Significance. The University of Michigan Press.