

Bayes and the Law

Norman Fenton, Martin Neil and Daniel Berger

*Submitted by invitation to
Volume 3 of the Annual Review of Statistics and Its Application*

n.fenton@qmul.ac.uk

27 January 2015

Abstract

Although the last forty years has seen considerable growth in the use of statistics in legal proceedings, it is primarily classical statistical methods rather than Bayesian methods that have been used. Yet the Bayesian approach avoids many of the problems of classical statistics and is also well suited to a broader range of problems. This paper reviews the potential and actual use of Bayes in the law and explains the main reasons for its lack of impact on legal practice. These include misconceptions by the legal community, over-reliance on the use of the likelihood ratio and the lack of adoption of modern computational methods. We argue that Bayesian Networks (BNs), which automatically produce the necessary Bayesian calculations, provide an opportunity to address most concerns about using Bayes in the law.

Keywords: Bayes, Bayesian networks, statistics in court, legal arguments

1 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 [74] [78], [95], [99] [78], [95] [148]. The historical reticence to accept statistical analysis as valid evidence is, sadly, not without good reason. When, in 1894, a statistical analysis was used in the Dreyfus case [98] it turned out to be fundamentally flawed. Not until 1968 was there another well documented case, *People v Collins* [1], in which statistical analysis played a key role. In that case another flawed statistical argument further set back the cause of statistics in court. The *Collins* case was characterised by two errors:

- 1) It underestimated the probability that some evidence would be observed if the defendants were innocent by failing to consider dependence between components of the evidence; and
- 2) It implied that the low probability from the calculation in 1) was synonymous with innocence (the so-called ‘prosecutors’ fallacy).

Since then the same errors (either in combination or individually) have occurred in well reported cases such as *R v Sally Clark* [87], *R v Barry George* [10], *Lucia de Berk* [110]. Although original ‘bad statistics’ used in each case (presented by forensic or medical expert witnesses without statistical training) was exposed through ‘good statistics’ on appeal, it is the ‘bad statistics’ which leaves an indelible stain. Yet, the role of legal professionals (who allow expert witnesses to commit the same well known statistical errors repeatedly) is rarely questioned.

Hence, although the last 40 years has seen considerable growth in the use of statistics in legal proceedings, its use in the courtroom has been mostly restricted to a small class of cases where classical statistical methods of hypothesis testing using p-values and confidence intervals are used for probabilistic inference. Yet, even this type of statistical reasoning has severe limitations as discussed extensively in [134] [57][150], including specifically in the context of legal and forensic evidence [74] [148]. In particular:

- The use of p-values can also lead to the prosecutor’s fallacy since a p-value (which says something about the probability of observing the evidence given a hypothesis) is interpreted as being the same as the probability of the hypothesis given the evidence [78].
- Confidence intervals are almost invariably misinterpreted since their proper definition is both complex and counter-intuitive (indeed it is not properly understood even by many trained statisticians) [57].

The poor experience – and difficulties in interpretation - with classical statistics means that there is also strong resistance to any alternative approaches. In particular, this resistance extends to the Bayesian approach, despite the fact that the Bayesian approach is especially well suited for a broad range of legal reasoning [71].

Although the natural resistance within the legal profession to a new statistical approach is one reason why Bayes has, to date, made only minimal impact, it is certainly not the only reason. Many previous papers have discussed the social, legal and logical impediments to the use of Bayes in legal proceedings [56] [70][144] [146] and in more general policy decision making [71]. However, we contend that there is another rarely articulated but now dominant reason for its continued limited use: it is because most examples of the Bayesian approach have oversimplified the underlying legal arguments being modelled in order to ensure the computations can be carried out manually. While such an approach may have been necessary 20 years ago it no longer is with the advent of easy-to-use and efficient Bayesian network algorithms and tools [57]. This continuing over-simplification is an unnecessary and debilitating burden.

The paper is structured as follows: Section 2 describes the basics of Bayes (and Bayesian Networks) for legal reasoning, covering the core notion of the likelihood ratio. Section 3 reviews the actual use of Bayes in the law. Section 4 describes why, in our view, Bayes has made such a minimal impact, while Section 5 explains what can be done to ensure it gets the exploitation we feel it deserves. We draw throughout primarily on English language sources and most cases cited are in English speaking countries (normally with a jury system). Most are from the UK and USA.

2 Basics of Bayes for legal reasoning

We start by introducing some terminology and assumptions that we will use throughout:

- 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 innocent 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)
 - “Defendant was present at the crime scene” (this is an example of what is often referred to as a ***source level hypothesis***)
- The ***alternative hypothesis*** is a statement which is the negation of a hypothesis.
- A piece of ***evidence*** is a statement that, if true, lends support to one or more hypotheses.

The relationship between a hypothesis H and a piece of evidence E can be represented graphically as in the example¹ in Figure 1 where we assume that:

- The evidence E is a DNA trace found at the scene of the crime that must have come from someone in the world population who committed the crime (for simplicity we assume the crime was committed on an island with 10,000 people so that the relevant ‘world population’ here is 10,000)
- The defendant was arrested and some of his DNA was sampled and analysed

¹ This is actually a very simple example of a Bayesian Network.

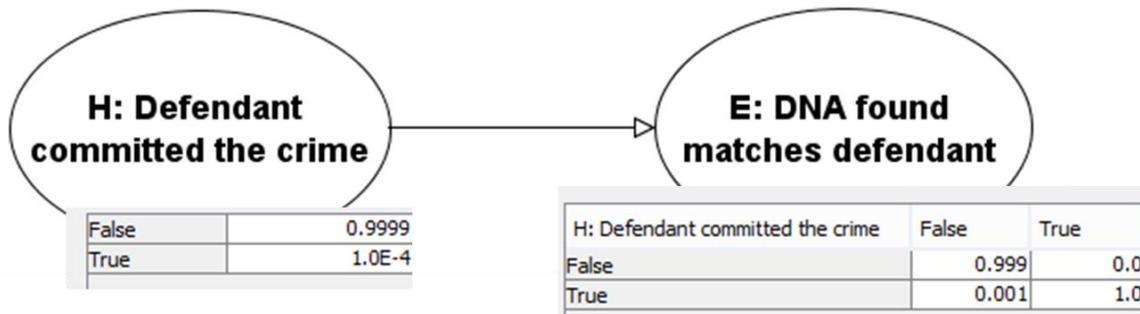


Figure 1 Causal view of evidence, with prior probabilities shown in tables

The direction of the causal structure makes sense here because H being true (resp. false) can **cause** E to be true (resp. false), while E cannot ‘cause’ H . However, inference can go in **both** directions. If we observe E to be true (resp. false) then our belief in H being true (resp. false) increases. It is this latter type of inference that is central to all legal reasoning since, informally, lawyers and jurors normally use the following widely accepted procedure for reasoning about evidence:

- Start with some (unconditional) prior assumption about the ultimate hypothesis H (for example, the ‘innocent until proven guilty’ assumption equates to a belief that “the defendant is no more likely to be guilty than any other member of the population”).
- Update our prior belief about H once we observe evidence E . This updating takes account of the *likelihood* of the evidence.

This informal reasoning is a perfect match for Bayesian inference where the prior assumption about H and the likelihood of the evidence E are captured formally by the probability tables shown in Figure 1. Specifically, these are the tables for the **prior probability** of H , written $P(H)$, and the conditional probability of E given H , which we write as $P(E | H)$. Bayes’ theorem provides the formula for updating our prior belief about H in the light of observing E to arrive at a **posterior belief about H** which we write as $P(H | E)$. In other words Bayes calculates $P(H | E)$ in terms of $P(H)$ and $P(E | H)$. Specifically:

$$P(H | E) = \frac{P(E | H)P(H)}{P(E)} = \frac{P(E | H)P(H)}{(E | H)P(H) + (E | \text{not}H)P(\text{not}H)}$$

The first table (the probability table for H) captures our knowledge that the defendant is one of 10,000 people who could have committed the crime. The second table (the probability table for $E | H$) captures the assumptions that:

- The probability of correctly matching a DNA trace is one (so there is no chance of a false negative DNA match). This probability $P(E | H)$ is called the **prosecution likelihood** for the evidence E .
- The probability of a match in an innocent person (the ‘random DNA match probability’) is 1 in 1,000. This probability $P(E | \text{not}H)$ is called the **defence likelihood** for the evidence E .

With these assumptions, it follows from Bayes’ theorem that, in our example, the posterior belief in H after observing the evidence E being true is about 9%, i.e. our belief in guilty

moves from a prior of 1 in a 10,000 to a posterior of 9%. Alternatively, our belief in innocence moves from a prior of 99.99% to a posterior of 91%.

Note that the posterior probability of innocence is very different from the random match probability of 1 in 1,000. The incorrect assumption that the two probabilities $P(\text{not } H | E)$ and $P(E | \text{not } H)$ are the same characterises what is known as the *prosecutor's fallacy* (or the error of the transposed conditional). A prosecutor might state, for example, that “the probability the defendant was not at the scene given this evidence is one in a thousand”, when actually it is 91%. This simple fallacy of probabilistic reasoning has affected numerous cases [24][62], but can always be avoided by a basic understanding of Bayes' Theorem. A closely related, but less common, error of probabilistic reasoning is the *defendant's fallacy*, whereby the defence argues that since $P(\text{not } H | E)$ is still low after taking into account the prior and the evidence, the evidence should be ignored.

Unfortunately, people without statistical training – and this includes most highly respected legal professionals – find Bayes' theorem both difficult to understand and counter-intuitive [33][38][62]. Legal professionals are also concerned that the use of Bayes' requires us to assign prior probabilities. In fact, an equivalent formulation of Bayes (called the ‘odds’ version of Bayes’) enables us to interpret the value of evidence E without ever having to consider the prior probability of H . Specifically, this version of Bayes' tells us:

the posterior odds of H are the prior odds of H times the likelihood ratio

where the Likelihood Ratio (LR) is simply the prosecution likelihood of E divided by the defence likelihood of E , i.e.

$$\frac{P(E | H)}{P(E | \text{not } H)}$$

In the example in Figure 1 the prosecution likelihood for the DNA match evidence is 1, while the defence likelihood is 1/1,000. So the LR is 1,000. This means that, whatever the prior odds were in favour of guilt, the posterior odds must increase by a factor of 1,000 as a result of seeing the evidence. In general, if the LR is bigger than 1 then the evidence results in an increased posterior probability of H (with higher values leading to the posterior probability getting closer to 1), while if it is less than 1 it results in a decreased posterior probability of H (and the closer it gets to zero the closer the posterior probability gets to zero). If the LR is equal to 1 then E offers no value since it leaves the posterior probability of guilt unchanged.

The LR is therefore an important and meaningful measure of the probative value of evidence. In our example the fact that the DNA match evidence had a LR of 1000 meant the evidence was highly probative in favour of the prosecution. But as impressive as that sounds, whether or not it is sufficient to convince you of which hypothesis is true still depends entirely on the prior $P(H)$. If $P(H)$ is, say 0.5 (so the prior odds are evens 1:1), then a LR of 1000 results in posterior odds of 1000 to 1 in favour of H . That may be sufficient to convince a jury that H is true. But if $P(H)$ is very low - as in our example (9999 to 1 against) - then the same LR of 1000 results in posterior odds that still strongly favour the defence hypothesis by 10 to 1.

Notwithstanding this observation (and other problems with the LR that we discuss in Section 4) the fact that it does determine the probative value of evidence and can be calculated without reference to the prior probability of H , has meant that it has become a potentially

powerful application of Bayesian reasoning in the law. Indeed, the CAI (Case Assessment and Interpretation) method [37][90] has the LR at its core. By forcing expert witnesses to consider both the prosecution and defence likelihood of their evidence – instead of just one or the other – it also avoids most common cases of the prosecutor’s fallacy.

While Bayes Theorem provides a natural match to intuitive legal reasoning in the case of a single hypothesis H and a single piece of evidence E , practical legal arguments normally involve multiple hypotheses and pieces of evidence with complex causal dependencies. For example, even the simplest case of DNA evidence strictly speaking involves 3 unknown hypotheses and two pieces of evidence with the causal links shown in Figure 2 [42] [64] once we take account of the possibility of different types of DNA collection and testing errors [100][141].

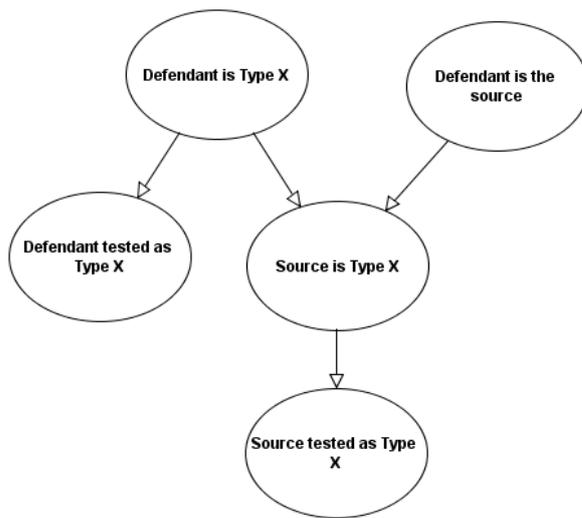


Figure 2 Bayesian network for DNA match evidence. Each node has states *true* or *false*

This is an example of a Bayesian Network (BN). As in the simple case of Figure 1, to perform the correct Bayesian inference once we observe evidence we need to know the prior probabilities of the nodes without parents and the conditional prior probabilities of the nodes with parents. Assuming that it is possible to obtain suitable estimates of these prior probabilities, the bad news is that, even with a small number of nodes, the calculations necessary for performing correct probabilistic inference are far too complex to be done manually. Moreover, until the late 1980’s there were no known efficient computer algorithms for doing the calculations. This is the reason why, until relatively recently, only rather trivial Bayesian arguments could realistically be used in legal reasoning.

However, algorithmic breakthroughs in the late 1980s made it possible to perform correct probabilistic inference efficiently for a wide class of Bayesian networks and tools [105] [123]. These algorithms have subsequently been incorporated into widely available toolsets that enable users without any statistical knowledge to build and run BN models [57]. Moreover, in the last 5 years further algorithmic breakthroughs have enabled us to model an even broader class of BNs, namely those including numeric nodes with arbitrary statistical distributions [115]. These breakthroughs are potentially crucial for modelling legal arguments. The supporting material for this paper [66] provides detailed examples of the use of BNs including how they simplify and rationalise previous Bayesian legal arguments.

3 Context and Review of Bayes in legal proceedings

To put the role of Bayesian reasoning in legal proceedings in context, we consider actual uses of Bayes in court according to a classification of cases that involve:

- Hypothesis testing of specific ‘dubious’ behaviour (discussed in Section 3.1)
- Determining extent to which ‘trait’ evidence helps with identification (Section 3.2)
- Combining multiple pieces of statistical evidence (Section 3.3)
- Combining multiple pieces of diverse evidence (Section 3.4)

Additionally, Section 3.5 considers cases where Bayesian solutions have been used indirectly or retrospectively.

3.1 Hypothesis testing of specific ‘dubious’ behaviour

The vast majority of reported cases of explicit use of statistics in legal proceedings fall under this classification which, in [66], we further sub-classify into:

- Any form of class discrimination and bias
- Any form of fraud/cheating
- Possession of illegal materials/substances

The review of such cases in [66] shows that depressingly few involve explicit use of Bayes, and that where Bayes was used it was often misunderstood. For example, in *Marks v Stinson* 1994 (discussed in detail in [86]) 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 than 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.

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 [57][124]. The potential for misinterpretation is enormous [148]. For example, in [2] 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 [14]) make it possible to easily perform the necessary analysis for Bayesian hypothesis testing.

3.2 Determining extent to which ‘trait’ evidence helps with identification

This classification refers to all cases in which statistical evidence about ‘traits’ in the broadest sense (as defined in [64]) is used. So traits range from ‘forensic’ physical features like DNA (coming from different parts of the body), fingerprints, or footprints, through to more basic

features such as skin colour, height, hair colour or even name. But it could also refer to non-human artefacts (and their features) related to a crime or crime scene, such as clothing and other possessions, cars, weapons, glass, soil etc.

Any statistical use of trait evidence requires some estimate (based on sampling or otherwise) of the trait incidence in the relevant population. Much of the resistance to the use of such evidence is due to concerns about the rigour and validity of these estimates – with the errors made in the case of *People v Collins* [1] still prominent in many lawyers' minds.

Nevertheless, the rapid growth of forensic statistics in the last 25 years has led to a corresponding increase in the use of statistical trait evidence. This includes not just DNA, but also glass, paint, fibres, soil, fingerprints, handwriting, ballistics, bite marks, earmarks, and footwear. It is not surprising, therefore, that almost all of the publicised use of Bayes in legal proceedings relates to this class of evidence.

Its use in presenting trait evidence (as recommended, for example in [16][22][35] [46][53][54] [89] [90]) has been most extensive in respect of DNA evidence for determining paternity [16] [35] [77] [107]. For example (see [117]), North Carolina's court of appeals and supreme court have upheld the admissibility of genetic paternity test results using Bayes theorem with a 50% prior, non-genetic probability of paternity, citing cases: [3] [4] [9]. In contrast to the 50% prior recommended by North Carolina, in *Plemel v Walter* 1987 (discussed in [71]) the need to present results against a range of priors was recommended; indeed this was also recommended in the criminal case [13].

Outside of paternity testing the use of the LR in presenting DNA evidence has primarily been in the sense of 2 above (i.e. exposing the prosecutor's fallacy in the original statistical presentation of a random match probability). Notable examples in the UK of precisely this scenario are the rape cases of *R v Deen* [5], *R v Doheny and Adams G* [6] [131], *R v Adams D* [8]; in each case an Appeal accepted the Bayesian argument showing that there was the potential to mislead in the way the DNA match evidence against the defendants had been presented by the prosecution at the original trial. Similar, uses of the LR for presenting DNA evidence being accepted as 'valid' have been reported in New Zealand and Australia [66][140].

It is worth noting that the interpretation of DNA match probabilities is critically dependent on the context for the match and, in particular, serious errors of probabilistic reasoning may occur in cases (such as *People v Puckett*, USA 2008) where the match arises from a database search. In such cases Bayesian reasoning can again avoid errors [23] [108].

It is difficult to determine the extent to which the LR for presenting forensic evidence other than DNA has been used in courts. Although mostly unreported, we know of cases involving glass, fibres and soil matches, while articles now promote its use in fingerprint evidence [18] [116][118]. The most impressive well-publicised success concerns its use in relation to firearm discharge residue (FDR) in *R v Barry George* [10]; the use of the LR in the appeal – showing that the FDR evidence had 'no probative value' was the reason granted for a retrial (in which George was found not guilty) with the FDR evidence deemed inadmissible. The appeal court ruling in this case (along with our concerns about the over-simplistic use of the LR that we return to in Section 4) is discussed in detail in [63].

However, while the R v George appeal judgment can be considered a major success for the use of Bayes, the 2010 UK Court of Appeal Ruling (known as RvT [11]) dealt it a devastating blow. The ruling quashed a murder conviction in which the prosecution had relied heavily on footwear matching evidence presented using Bayes and the LR. Specifically, even though there was recognition that it was used for this purpose in Netherlands, Slovenia and Switzerland) points 86 and 90 of the ruling respectively assert:

“..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”.

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

Given its potential to change the way forensic experts analyse and present evidence in court, there have been numerous articles criticising the ruling [15][111][25][120][137][127][133]. However, as we explain in Section 4, the ruling is beginning to have a devastating impact on the way some forensic evidence is presented with experts deliberately concealing or obfuscating their calculations.

While most reported cases of the prosecutor’s fallacy relate to poor presentation of trait evidence, it is not the only way it occurs. In R v Sally Clark, Clark was convicted for the murder of her two young children who had died one year apart [75][119]. To counter the hypothesis that the children had died as a result of Sudden Infant Death Syndrome (SIDS) the prosecution’s expert witness Roy Meadow stated that there was “only a 1 in 73 million chance of both children being SIDS victims”. This figure wrongly assumed two SIDS deaths in the same family were independent and was presented in a way that may have led the jury into the prosecutor’s fallacy. These errors were exposed by probability experts during the appeal and they also used Bayes to explain the impact of failing to compare the prior probability of SIDS with the (also small) probability of double murder [45] [87]. Clark was freed on appeal in 2003.

3.3 Combining multiple pieces of statistical evidence

Where there are multiple pieces of statistical evidence (such as trait evidence) involved in a case there is a need for careful probabilistic analysis taking account of potential dependencies between different pieces of evidence. Bayes is ideally suited to such an analysis, but sadly the failure to recognise this has led to many very well publicised errors. For example, in the Collins case the original statistical analysis failed to account for dependencies between the various components of the trait evidence. Similarly, in R v Clark the above mentioned figure of 1 in 73 million for the probability of two SIDS deaths in the same family was based on an unrealistic assumption that two deaths would be independent events. More generally courts have failed to properly understand what [71] refers to as:

‘...cases involving a series of events or crimes and a related legal doctrine for the admissibility of evidence on ‘similar events’.

The Lucia de Berk case [110] is a classic example of this misunderstanding, whereby a prosecution case is fitted around the fact that one person is loosely connected to a number of related events. People die in hospitals and it is inevitable that there will be instances of individual nurses associated with much higher than normal death rates in any given period of time. The extremely low probability $P(\text{evidence} \mid \text{not guilty})$ – which without any real evidence that the defendant has committed a crime tells us very little – is used to drive the prosecution case. Another very similar case (currently subject to an appeal based on precisely these concerns) is that of Ben Geen who was also a nurse convicted of multiple murders and attempted murder [80].

Fienberg and Kaye [68] describe a number of earlier cases and note that, despite the fact that Bayes is especially well suited since $P(H \mid \text{evidence})$ increases as more similar events are included in the evidence, in none of them was a Bayesian perspective used.

3.4 Combining multiple pieces of diverse evidence

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. Yet, although Bayes is the perfect formalism for this type of reasoning, 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 (Denis)* [8] already referred to above in connection to the misleading presentation of DNA evidence. This was a rape case (discussed in detail in [48]) 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.

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.

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.

3.5 Cases where Bayesian solutions have been used indirectly or retrospectively

While there is a shortage of well publicised success stories for the use of Bayes in actual legal proceedings, there is no shortage of retrospective studies of real cases where statistics may or may not have been used in the actual trial or appeal, but where it has been argued that a Bayesian approach would have been useful to improve understanding (and expose flaws) [16][22][83][95][106][128][136][139]. Certain cases have, unsurprisingly, attracted numerous retrospective Bayesian analyses, notably the case of Collins [49][72][46], O J Simpson [82] [143] and Sacco and Vanzetti [85] [61].

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 [93] 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 [62] and B v National Health Service 2005 [58].

4 Why Bayes has had minimal impact in the law

The review in Section 3 leaves little doubt that the impact of Bayes in the Law has been minimal. To understand *how* minimal, it is illustrative to note that, for all its welcome advice, the 350-page National Research Council report on “Strengthening Forensic Science in the United States [114] *does not contain a single mention of Bayes, and it contains just one footnote reference to likelihood ratios.* Although, in contrast in the UK, a number of published guidelines recommend the use of LRs extensively for forensic evidence

presentation [90][125], the use of such methods is actually likely to *decrease* following rulings such as RvT.

In this section we explain what we consider to be the reasons for continued resistance to Bayes in the law.

4.1 Standard resistance to the use of Bayes

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 [12] (discussed in [138]). 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 ... Either it has or it has not.

Although it is easy to show the irrationality of this viewpoint (see, for example [57]) supporters of it often point to the highly influential paper by Tribe [146], which was written as a criticism of the prosecutor’s presentation in *People v Collins* [1]. While Tribe’s article did not stoop to the level of the ‘no such thing as probability’ argument, it is especially sceptical of the potential use of Bayes’ Theorem since it identifies the following concerns [26] [71] that are especially pertinent for Bayes:

- That an accurate and/or non-overpowering prior cannot be devised.
- That in using statistical evidence to formulate priors jurors might use it twice in reaching a posterior
- That not all evidence can be considered or valued in probabilistic terms.
- That no probability value can ever be reconciled with “Beyond A Reasonable Doubt”
- That due to the complexity of cases and non-sequential nature of evidence presentation, any application of Bayes would be too cumbersome for a jury to use effectively and efficiently.
- That probabilistic reasoning is not compatible with the law, for policy reasons. In particular, that jurors are asked to formulate an opinion of the defendant’s guilt during the prosecutor’s case, which violates the obligation to keep an open mind until all evidence is in.

Although many of Tribe’s concerns have long been systematically demolished in [49], [99] and more recently in [26], [145], the arguments against are far less well known among legal professionals than those in favour.

4.2 LR models are inevitably over-simplified

We believe that much of the recent legal resistance to Bayes is due to confusion, misunderstanding, over-simplification and over-emphasis on the role of the LR. The issues (dealt with in depth in the recent papers [63][64][65]) are summarised in what follows.

The simplest and most common use of the LR – involving a single piece of forensic trace evidence for a single source level hypothesis – can actually be very complex [84]. Even if we completely ignore much of the context (including issues of reliability of trace sample collection/storage and potential testing errors) the LR may still be difficult or even impossible to elicit because somehow we have to factor in to H_d every person other than the defendant who could have been the source of the trace (potentially every other person in the world) [22][120]. For example, in the case of DNA for a close relative, H_r , and a totally unrelated person, H_u , $P(E | H_r)$ is much higher than $P(E | H_u)$. This means that, in reality, H_d is made up of multiple hypotheses that are difficult to articulate and quantify. The standard pragmatic solution (which has been widely criticised [22]) is to assume that H_d represents a ‘random person unrelated to the defendant’. But not only does this raise concerns about the homogeneity of the population used for the random match probabilities, it also requires separate assumptions about the extent to which relatives can be ruled out as suspects.

In the light of these observations it is perhaps not surprising that Wagnenaar [149] had difficulties in getting the judge to interpret his LR descriptions.

4.3 The mutual exclusivity of hypotheses is not respected in all LR uses

The most powerful assumed benefit of the LR is that it provides a valid measure of the probative value of the evidence E in relation to the alternative hypotheses H_p (prosecution hypothesis) and H_d (defence hypothesis). But when we say the evidence E “supports the hypothesis H_p ” what we mean is that the “posterior probability of H_p after observing E is greater than the prior probability of H_p ” and “no probative value” means the “posterior probability of H_p after observing E is equal to the prior probability of H_p ”. The proof of the meaning of probative value in this sense relies both on Bayes’ theorem and the fact that H_p and H_d are mutually exclusive and exhaustive – i.e. are negations of each other (H_d is the same as “not H_p ”) [63]. If H_p and H_d are mutually exclusive but not exhaustive then the LR tells us absolutely nothing about the relationship between the posterior and the priors of the individual hypotheses.

We already noted that it may be very difficult in practice to elicit $P(E | H_d)$ where H_d is simply the negation of H_p because H_d may encompass too many options. The inevitable temptation is to ‘cherry pick’ an H_d which seems a reasonable alternative to H_p for which $P(E | H_d)$ can be elicited. The arguments in [63] show the dangers in such an approach. Indeed, it was shown that the “ $LR = 1$ implies evidence has no probative value” argument that was the key in the *R v Barry George* appeal may have been compromised by precisely this problem. Similarly, it has also been noted that the failure to choose appropriate (mutually exclusive and exhaustive) hypotheses in the arguments used by Bayesians to expose the flaws in the *R v Sally Clark* case may have exaggerated the scale of the problem [65].

It is especially confusing (and even disturbing) that the Guidelines [90] [129] specify that the LR requires two mutually exclusive but not necessarily exhaustive hypotheses. While this may make life more convenient for its primary target audience (i.e. forensic scientists) it can severely compromise how lawyers and jurors should interpret the results.

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 [63]. Two hypotheses that are not mutually exclusive may both be true and in such circumstances the LR is meaningless.

4.4 The ‘prior misconception’

The LR is popular with forensic experts precisely because it can be calculated without having to consider any prior probabilities for the hypotheses [124]. But this is something of a misconception for two reasons. First, the LR actually tells us nothing about the probability that either hypothesis is true, no matter how high or low it is. We can only make conclusions about such (posterior) probabilities if we know the prior probabilities. Although this observation has been well documented [60][109][147] this issue continues to confound not just lawyers, but also forensic experts and statisticians. An indication of the extent of the confusion can be found in one of the many responses by the latter community to the RvT judgement. Specifically, in the otherwise excellent position statement [15] (signed by multiple experts) is the extraordinary point 9 that asserts:

“It is regrettable that the judgment confuses the Bayesian approach with the use of Bayes’ Theorem. The Bayesian approach does not necessarily involve the use of Bayes’ Theorem.”

By the “Bayesian approach” the authors are specifically referring to the use of the LR, thereby implying that the use of the LR is appropriate, while the use of Bayes’ Theorem may not be.

The second reason why it is a misconception is because it is impossible to specify the likelihoods meaningfully without knowing something about the priors. This is because the priors are all conditioned on some background knowledge K and it follows that, without an agreed common understanding about this background knowledge, we can end up with vastly different LRs associated with the same hypotheses and evidence [44]. Examples are provided in [66]

4.5 The LR for source level hypotheses tells us nothing about offense level hypotheses

Even when hypotheses are mutually exclusive and exhaustive, there remains the potential during a case to confuse what in [36] were referred to as source-level and offence-level hypotheses. Sometimes one may mutate into another through slight changes in the precision with which they are expressed. A LR for the source-level hypotheses will not in general be the same as for the offence-level hypotheses.

In [63] it is shown (using BNs that model dependencies between different hypotheses and evidence) that it is possible that an LR for a piece of evidence *E* that strongly favours one side for the source-level hypotheses can actually strongly favour the *other* side for the offence-level hypotheses even though both pairs of hypotheses seem very similar. Similarly, an LR that is neutral under the source-level hypotheses may actually be significantly non-neutral under the associated offence-level hypotheses. Once again these issues were shown to be a real problem in the transcript of the *R v Barry George* appeal [63].

4.6 Confusion about LR being expressed on a verbal scale

The UK Forensic Science Service Guides [129] [125] recommends that, to help lawyers and jurors understand the significance of a LR, it should be presented on an equivalent verbal scale. This recommendation (criticised in [112]) 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. 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, 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.

4.7 Continued use of manual methods of calculation

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. Yet, 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.

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 [66] we show a number of other examples where statisticians provide unnecessarily complex arguments (such as countering the so-called paradoxes of Bayesian reasoning in the law [34]).

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.

5 The way forward

The scale of the challenge for Bayes and the law is well captured by the following statement made at a recent legal conference (to the first author of this paper²) by one of the UK's most eminent judges:

“No matter how many times Bayesian statisticians try to explain to me what the prosecutors fallacy is I still do not understand it and nor do I understand why there is a fallacy.”

This is despite the fact that, in response to the cases described in Section 3.2, guidance to UK judges and lawyers specifies the need to be aware of expert witnesses making the prosecutor's fallacy and to make such evidence inadmissible. Yet, based on informal evidence from cases we have been involved in, the prosecutor's fallacy (and other related probabilistic fallacies) continues to be made regularly by both expert witnesses and lawyers. Indeed, one does not need an explicit statement of probability for the fallacy to be made. For example, a statement like:

“the chances of finding this evidence in an innocent man are so small that you can safely disregard the possibility that this man is innocent”

is a classic instance of the prosecutor's fallacy that is frequently used by lawyers. Based on our own experiences as expert witnesses, we believe the reported instances are merely the tip of the iceberg.

That such probabilistic fallacies continue to be made in legal proceedings is a sad indictment of the lack of impact made by statisticians in general (and Bayesians in particular) on legal practitioners. This is despite the fact that the issue has been extensively documented by multiple authors including [20] [24], [57], [49], [52], [59], [92], [97], [101], [128], [142] [52][76][94][106][113] [100] and has even been dealt with in populist books such as [79][84]. There is almost unanimity among the authors of these works that a basic understanding of Bayesian probability is the key to avoiding probabilistic fallacies. Indeed, Bayesian reasoning is explicitly recommended in works such as [73][76][52][83][89][128][132][135], although there is less of a consensus on whether or not

² The first author had not himself had an opportunity to explain the fallacy to the judge but did later send his own explanation which the judge thanked him for without indicating if this helped his understanding!

experts are needed in court to present the results of all but the most basic Bayesian arguments [132].

We argue that the way forward is to use BNs to present probabilistic legal arguments since this approach avoids much of the confusion surrounding both the over-simplistic LR and more complex models represented formulaically and computed manually. Unfortunately, it is precisely because BNs are assumed by legal professionals to be ‘part of those same problems’ that they have made little impact. The use of BNs for probabilistic analysis of forensic evidence and more general legal arguments is by no means new. As long ago as 1991 Edwards [49] provided an outstanding argument for the use of BNs in which he said of this technology:

“I assert that we now have a technology that is ready for use, not just by the scholars of evidence, but by trial lawyers.”

He predicted such use would become routine within “two to three years”. Unfortunately, Edwards was grossly optimistic. In addition to the reasons already identified above there is the major challenge in getting legal professionals to accept the validity of subjective probabilities that are an inevitable part of Bayes. Yet, ultimately, any use of probability – even if it is based on frequentist statistics – relies on a range of subjective assumptions. The objection to using subjective priors may also be calmed by the fact that it may be sufficient to consider a range of probabilities, rather than a single value for a prior. BNs are especially suited to this since it is easy to change the priors and do sensitivity analysis [57].

A basic strategy for presenting BNs to legal professionals is described in detail in [62] and is based on the *calculator analogy*. This affirms that since we now have efficient and easy-to-use BN tools there should be no more need to explain the Bayesian calculations in a complex argument than there should be any need to explain the thousands of circuit level calculations used by a regular calculator to compute a long division.

Only the simplest Bayesian legal argument (a single hypothesis and a single piece of evidence) can be easily computed manually; inevitably we need to model much richer arguments involving multiple pieces of possibly linked evidence. While humans must be responsible for determining the prior probabilities (and the causal links) for such arguments, it is simply wrong to assume that humans must also be responsible for understanding and calculating the revised probabilities that result from observing evidence. The Bayesian calculations quickly become impossible to do manually, but any BN tool, such as [14] enables us to do these calculations instantly.

The results from a BN tool can be presented using a range of assumptions including different priors. What the legal professionals (and perhaps even jurors if presented in court) should never have to think about is how to perform the Bayesian inference calculations. They do, of course, have to consider the prior assumptions needed for any BN model. Specifically, they have to consider:

- whether the correct nodes and causal links have been modelled
- what are the suitable values (or range of values) required for the priors and conditional priors
- whether the hidden value K for ‘context’ has been considered and communicated throughout the model

But these are precisely what have to be considered in weighing up any legal argument. The BN simply makes this all explicit rather than hidden, which is another clear benefit of the approach.

We recognise that there are significant technical challenges to overcome to make the construction of BNs for legal reasoning easier. Recent work addresses the lack of a systematic, repeatable method for modelling legal arguments as BNs by using common idioms and an approach for building complex arguments from these [85][61]. We also have to overcome various technical constraints of existing BN algorithms and tools such as forcing modellers to specify unnecessary prior probabilities (this is being addressed in work such as [50]). However, the greater challenges are cultural and presentational and these issues are being addressed in projects like [21][50][88][134] which involve close interaction with legal professionals and empirical studies on how best to present arguments.

One of the greatest opportunities offered by BNs is to provide a much more rigorous and accurate method to determine the combined impact of multiple types of evidence in a case (including, where relevant, absence of evidence). In particular, this applies urgently to any case in which DNA evidence with match probabilities is presented (hence avoiding the kind of errors highlighted in [104]). As was argued (and accepted) in *R v Adams*, if probability is to be used for one piece of evidence, it is unfair to exclude its use for other types of evidence.

BNs can also expose other features of DNA evidence (such as those relating to homogeneity of population samples) that clearly demonstrate the illogical and irrational decision in *RvT* to elevate DNA evidence to its privileged position of being the only class of evidence suitable for statistical analysis. Match probabilities for all types of trait evidence – including DNA – involve uncertainties and inaccuracies related to sampling and subjective judgment and inaccuracies in testing. Moreover, the situation with respect to low-template DNA evidence [22][51] is especially critical since a ‘DNA match’ in such cases may have low probative value (especially if the potential for secondary and tertiary transfer is considered) but still has a powerful impact on lawyers and jurors; we are currently involved in appeal cases where some low-template DNA evidence was assumed to ‘trump’ all the opposing evidence in the case. BN models demonstrate that the impact of the DNA evidence has been exaggerated (examples are provided in [66]).

6 Conclusions

Proper use of Bayesian reasoning has the potential to improve the efficiency, transparency and fairness of criminal and civil justice systems. It can help experts formulate accurate and informative opinions; help courts determine admissibility of evidence; help identify which cases should be pursued; and help lawyers to explain, and jurors to evaluate, the weight of evidence during a trial. It can also help identify errors and unjustified assumptions entailed in expert opinions.

Yet, despite the impeccable theoretical credentials of Bayes in presenting legal arguments, and despite the enormous number of academic papers making the case for it, our review of the impact in practice of Bayes in the Law is rather depressing. It confirms that its impact on legal practice has been minimal outside of paternity cases. While it has been used in some well-known cases to expose instances of the prosecutor's fallacy, that very fallacy continues to afflict many cases. And while the LR has been quite widely used to present the impact of forensic evidence, we have explained the multiple problems with it and the danger of allowing it to be seen as synonymous with Bayes and the law. The fact that these problems were *not* the reason for the RvT judgment in 2010 which effectively banned the use of the LR (outside of DNA evidence) makes the judgment even more of a travesty than many assume it to be.

In addition to the RvT judgment, we have seen the devastating appeal judgments in *R v Adams* (effectively banning the use of Bayes for combining 'non statistical evidence') and in [12] which went as far as banning the use of probability completely when discussing events that had happened but whose outcome was unknown. Even when an appeal judgment was favourable to Bayes (*R v George* [10]) subsequent research has discovered weaknesses in the LR reasoning [63]. Similar weaknesses have been recently exposed in other 'successful' Bayesian arguments such as those used in the successful Sally Clark appeal [65].

The good news is that the difficulties encountered in presenting Bayesian arguments can be avoided once it is accepted that over-simplistic LR models and inference calculations from first principles are generally unsuitable. We have explained that BNs are the appropriate vehicle for overcoming these problems since they enable us to model the correct relevant hypotheses and the full causal context of the evidence. Unfortunately, too many people in the community are unaware of (or reluctant to use) the tools available for easily building and running BN models.

We differ from some of the orthodoxy of the statistical community in that we believe there should never be any need for statisticians or anybody else to attempt to provide complex Bayesian arguments from first principles in court. The *Adams* case demonstrated that statistical experts are not necessarily qualified to present their results to lawyers or juries in a way that is easily understandable. Moreover, although our view is consistent with that of Robertson and Vignaux [132] in that we agree that Bayesians should not be presenting their arguments in court, we do not agree that their solution (to train lawyers and juries to do the calculations themselves) is feasible. Our approach, rather, draws on the analogy of the electronic calculator.

There is still widespread disagreement about the kind of evidence to which Bayesian reasoning should be applied and the manner in which it should be presented. We have suggested methods to overcome these technical barriers, but there are still massive cultural

barriers between the fields of science and law that that will only be broken down by achieving a critical mass of relevant experts and stakeholders, united in their objectives. We are building toward this critical mass of experts in projects such as [21] [50][88][134].

Finally, our review has exposed the irrational elevation of DNA, by the legal community – and some members of the mathematical community – to a type of evidence (over and above that of other types) which has almost unique probabilistic qualities. In fact BNs offer a tremendous opportunity to radically improve justice in all cases that involve DNA evidence (especially low-template DNA), by properly capturing its impact in relation to all other evidence.

7 Acknowledgements

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

8 References

- [1] (1968). *People v. Collins*, 438 P. 2d 33 (68 Cal. 2d 319 1968)
- [2] *U.S ex. Rel. DiGiacomo v Franzen*, 680 F.2d 515 (7th Cir. 1982)
- [3] *Cole v. Cole*, 74 N.C.App. 247, 328 S.E.2d 446, *aff'd*, 314 N.C. 660, 335 S.E.2d 897 (1985) (per curiam)
- [4] *State v. Jackson*, 320 N.C. 452, 358 S.E.2d 679 (1987)
- [5] *R v Deen* 1994
- [6] (1996). *R v Alan James Doheny and Gary Adams. Smith Bernal Reporting Ltd*, Court of Appeal Criminal Division. <http://www.bailii.org/ew/cases/EWCA/Crim/1996/728.html>
- [7] (1997). *R v Doheny and Adams* 1 Crim App R 369.
- [8] *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).
- [9] *Brown v. Smith*, 137 N.C.App. 160, 526 S.E.2d 686 (2000)
- [10] (2007). *R. v. George*, EWCA Crim 2722
- [11] (2010). *R v T*. EWCA Crim 2439
<http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf>
- [12] *Nulty & Ors v Milton Keynes Borough Council* EWCA Civ 15 (24 January 2013) (2013)
<http://www.bailii.org/ew/cases/EWCA/Civ/2013/15.html>.
- [13] *State of New Jersey v J.M. Spann*, 130 N.J. 484: 617 A. 2d 247 (1993).
- [14] AgenaRisk software (2013). www.agenarisk.com
- [15] Aitken, C. and many other signatories (2011). "Expressing evaluative opinions: A position statement." *Science and Justice* 51(1): 1-2.
- [16] Aitken, C. G. G. and F. Taroni (2004). *Statistics and the evaluation of evidence for forensic scientists* (2nd Edition), John Wiley & Sons, Ltd.

- [17] Aitken, C., Taroni, F., and Garbolino, P., A graphical model for the evaluation of cross-transfer evidence in DNA profiles. *Theoretical Population Biology*, 2003. 63: p. 179-190.
- [18] Alberink, I., de Jongh, A., & Rodriguez, C. (2014). Fingerprint evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. *Journal of Forensic Sciences*, 59(1), 70–81.
- [19] Allen, R. J. and Miller, J. S. (1993). "The Common Law Theory of Experts: Deference or Education?" *Northwestern Law Review* 87: 1131-1147.
- [20] Anderson, T., Schum, D. A. and Twining, W. (2005). *Analysis of Evidence*, Cambridge University Press.
- [21] Bayes and the Law. (2015). <https://sites.google.com/site/bayeslegal/>
- [22] Balding, D. J. (2005). *Weight-of-Evidence for Forensic DNA Profiles*, Wiley.
- [23] Balding, D.J. and Donnelly, P., Evaluating DNA profile evidence when the suspect is identified through a database search, in *Journal of Forensic Science* 41, 599-603. 1996.
- [24] Balding, D.J. and Donnelly, P., Prosecutor's fallacy and DNA evidence, in *Criminal Law Review*, 711-722. 1994.
- [25] Berger, C. E. H., J. Buckleton, C. Champod, I. Evett and G. Jackson (2011). "Evidence evaluation: A response to the court of appeal judgement in R v T." *Science and Justice* 51: 43-49.
- [26] Berger, D (2014) *Improving Legal Reasoning using Bayesian Probability Methods*, PhD Thesis, Queen Mary University of London
- [27] Bex, F. J., P. J. van Koppen, H. Prakken and B. Verheij (2010). "A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence " *Artificial Intelligence and Law* 18(2): 123-152.
- [28] Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkley. *Science*, 187, 398–404.
- [29] Bordman, A. E., & Vining, A. R. (1983). Statistics in employment discrimination cases. *Law and Contemporary Problems*, 64(4).
- [30] Broeders, T. (2009). Decision-Making in the Forensic Arena. In "Legal Evidence and Proof: Statistics, Stories and Logic". (Eds H. Kaptein, H. Prakken and B. Verheij, Ashgate) 71-92.
- [31] Buckleton, J. S., Walsh, S. and Harbison, S. A. (2001). "The fallacy of independence testing and the use of the product rule." *Sci Justice* 41(2): 81-84.
- [32] Buckleton, J., C. M. Triggs and S. J. Walsh (2005). *Forensic DNA Evidence Interpretation*, CRC Press.
- [33] Casscells, W., Schoenberger, A., and Graboys, T.B., Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 1978. 299 p. 999-1001.
- [34] Cohen, L. J. (1977). *The Probable and the Provable*. Oxford: Clarendon Press.
- [35] Collins, A., & Morton, N. E. (1994). Likelihood ratios for DNA identification. *Proc Natl Acad Sci U S A*, 91(Medical Sciences), 6007–6011.
- [36] Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA. (1998). A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*, 38: 231-239.
- [37] Cook, R., Evett, I. W., Jackson, G., Jones, P. J., & Lambert, J. A. (1998). A model for case assessment and interpretation. *Science & Justice : Journal of the Forensic Science Society*, 38(3), 151–6.
- [38] Cosmides, L. and Tooby, J. (1996). "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty." *Cognition* 58 1-73

- [39] Cowell RG, S L Lauritzen and J Mortera (2011), 'Probabilistic expert systems for handling artifacts in complex DNA mixtures', *Forensic Science International:Genetics*, 5(3), p.202-209
- [40] Curran, J. N. (2005). An introduction to Bayesian credible intervals for sampling error in DNA profiles. *Law, Probability and Risk*, 4, 115–126.
- [41] Darroch, J. (1987). "Probability and criminal trials: some comments prompted by the Splatt trial and the Royal Commission." *Professional Statistician* 6: 3-7.
- [42] Dawid, A., & Mortera, J. (1998). Forensic identification with imperfect evidence. *Biometrika*, 85(4), 835–849.
- [43] Dawid, A. P. and Evett, I. W. (1997). "Using a graphical model to assist the evaluation of complicated patterns of evidence." *Journal of Forensic Sciences* 42: 226-231.
- [44] Dawid, A.P., (2004) Which likelihood ratio (Comment on 'Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, *Law, Probability and Risk* 3(1):65–71.
- [45] Dawid, P., *Statistics on Trial. Significance*, 2005. 2(1): p. 6-8. 10.1111/j.1740-9713.2005.00075.x
- [46] Dawid, P., & Mortera, J. (2008). Probability and Evidence. In T. Ruda (Ed.), *Handbook of Probability: theory and applications* (pp. 403–422). Sage Publications Inc.
- [47] Dawid, P., Musio, M., & Fienberg, S. E. (2013). From Statistical Evidence to Evidence of Causality. Retrieved from <http://arxiv.org/abs/1311.7513>
- [48] Donnelly, P. (2005). "Appealing Statistics." *Significance* 2(1): 46-48.
- [49] Edwards, W. (1991). "Influence Diagrams, Bayesian Imperialism, and the Collins case: an appeal to reason." *Cardozo Law Review* 13: 1025-1079.
- [50] European Research Council Advanced Grant ERC-2013-AdG339182-BAYES_KNOWLEDGE. (2014). Effective Bayesian Modelling with Knowledge Before Data. https://www.eecs.qmul.ac.uk/~norman/projects/B_Knowledge.html
- [51] Evett IW, Gill PD, 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.* May;47(3):520-30.
- [52] Evett, I. W. (1995). "Avoiding the transposed conditional." *Science and Justice* 35(3): 127-131
- [53] Evett, I. W. and B. S. Weir (1998). *Interpreting DNA evidence : statistical genetics for forensic scientists*, Sinauer Associates.
- [54] Evett, I. W., L. A. Foreman, G. Jackson and J. A. Lambert (2000). "DNA profiling: a discussion of issues relating to the reporting of very small match probabilities." *Criminal Law Review* (May) 341-355.
- [55] Evett, I. W., L. A. Foreman, G. Jackson and J. A. Lambert (2000). "DNA profiling: a discussion of issues relating to the reporting of very small match probabilities." *Criminal Law Review* (May) 341-355.
- [56] Faigman, D.L. and Baglioni, A.J., Bayes' theorem in the trial process. *Law and Human Behavior*, 1988. 12(1): p. 1-17. <http://dx.doi.org/10.1007/BF01064271>
- [57] Fenton N., and Neil, M. *Risk Assessment and Decision Analysis with Bayesian Networks*, CRC Press, 2012.
- [58] Fenton, N. and Neil, M. (2010). "Comparing risks of alternative medical diagnosis using Bayesian arguments." *Journal of Biomedical Informatics* 43: 485-495.
- [59] Fenton, N. and Neil, M., *The Jury Fallacy and the use of Bayesian nets to simplify*

- probabilistic legal arguments. *Mathematics Today (Bulletin of the IMA)*, 2000. 36(6): p. 180-187.
- [60] Fenton, N. E. (2011). "Science and law: Improve statistics in court." *Nature* 479: 36-37
- [61] Fenton, N., Neil, M., Lagnado, D.A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science*, 37, 61–102.
- [62] Fenton, N.E. and Neil, M. (2011), 'Avoiding Legal Fallacies in Practice Using Bayesian Networks', *Australian Journal of Legal Philosophy* 36, 114-151
- [63] Fenton, N. E., D. Berger, D. Lagnado, M. Neil and A. Hsu, (2014). "When 'neutral' evidence still has probative value (with implications from the Barry George Case)", *Science and Justice*, 54(4), 274-287
- [64] 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. *Artificial Intelligence and Law*, 22, 1–28.
- [65] Fenton, N. E. (2014). Assessing evidence and testing appropriate hypotheses. *Sci Justice*. 54(4), 319–20.
- [66] Fenton, N. E. Neil, M., & Berger, D (2015) "Bayes and the Law: Supplementary Material".
- [67] Fienberg, S. E., & Schervish, M. J. (1986). Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decision making, *The Symposium. Boston University Law Review*, 66, 771–798.
- [68] Fienberg, S. E., & Kaye, D. H. (1991). Legal and Statistical Aspects of Some Mysterious Clusters. *Journal of the Royal Statistical Society (series A)*, 154(1), 61–74.
- [69] Fienberg, S. E. (2005). To Tell the Truth: On the Probative Value of Polygraph Search Evidence. *Jurimetrics Journal*, 46, 107–116.
- [70] Fienberg, S. E. (2011). Bayesian Models and Methods in Public Policy and Government Settings. *Statistical Science*, 26(2), 212–226.
- [71] Fienberg, S. E., & Finkelstein, M. O. (1996). Bayesian statistics and the law. In J. M. Bernardo, J. O. Berge, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 5* (pp. 129–146). Oxford University Press.
- [72] Finkelstein, M. O., & Fairley, W. B. (1970). A Bayesian Approach to Identification Evidence. *Harvard Law Review*, 83(3), 489–517.
- [73] Finkelstein, M. O. and Levin, B. A. (2001). *Statistics for Lawyers*, Springer.
- [74] Finkelstein, M. O. (2009). *Basic Concepts of Probability and Statistics in the Law* (p. 186). Springer Science & Business Media.
- [75] Forrest, A. R. (2003). "Sally Clark - a lesson for us all." *Science & Justice*: 43: 63-64.
- [76] Freckelton, I. and Selby, H. (2005). *Expert Evidence: Law, Practice, Procedure and Advocacy*. Australia, Lawbook Co.
- [77] Fung, W. K. (2003). User-friendly programs for easy calculations in paternity testing and kinship determinations. *Forensic Science International*, 136(1-3), 22–34.
- [78] Gastwirth, J. L. (2000). *Statistical Science in the Courtroom*. New York: Springer-Verlag.
- [79] Gigerenzer, G. (2002). *Reckoning with Risk: Learning to Live with Uncertainty*. London, Penguin Books.
- [80] Gill, R. (2014). How to become a convicted serial killer (without killing anyone). *The Justice Gap*. Retrieved from <http://thejusticegap.com/2014/08/become-convicted-serial-killer-without-killing-anyone/>

- [81] Gilovich, T., Griffin, D., and Kahneman, D., *Heuristics and Biases: The Psychology of Intuitive Judgment*. . 2002, Cambridge: Cambridge University Press.
- [82] Good, I.J., When batterer turns murderer. *Nature*, 1995. 375: p. 541.
- [83] Good, P. I. (2001). *Applying Statistics in the Courtroom: A New Approach for Attorneys and Expert Witnesses*, CRC Press.
- [84] Haigh, J. (2003). *Taking Chances: Winning with Probability*, Oxford University Press.
- [85] Hepler, A. B., Dawid, A. P., & Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1-4), 275–293.
- [86] Hepler, A. B., & Weir, B. S. (2004). Using Bayesian Networks for Paternity Calculations: Adding an Evolutionary Perspective. *JSM (Joint Statistical Meetings) 2004*. Toronto, Canada.
- [87] Hill, R. (2005). "Reflections on the cot death cases." *Significance* 2(13-15).
- [88] Isaac Newton Institute for Mathematical Sciences, U. of C. (2015). FOS: Probability and Statistics in Forensic Science. <https://www.newton.ac.uk/event/fos>
- [89] Jackson, G., Jones, S., Booth, G., Champod, C. and Evett, I. W. (2006). "The nature of forensic science opinion-a possible framework to guide thinking and practice in investigations and in court proceedings." *Science and Justice* 46(1): 33-44
- [90] Jackson, G., Aitken, C., & Roberts, P. (2013). *PRACTITIONER GUIDE NO 4: Case Assessment and Interpretation of Expert Evidence*. Royal Statistical Society
- [91] Jamieson, A., & Bader, S. (2011). Got a match, Guv? *The Barrister*, (49). Retrieved from <http://www.barristermagazine.com/archive-articles/issue-49/got-a-match,-guv.html>
- [92] Jowett, C., Lies, Damned Lies, and DNA Statistics: DNA Match testing. Bayes' Theorem, and the Criminal Courts, in *Med Sci. Law* 41(3), 194-205. 2001.
- [93] Kadane, J. B. (1990). A statistical analysis of adverse impact of employer decisions. *J Amer. Statist. Assoc*, 85, 607–614.
- [94] Kadane, J. B. and Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, John Wiley & Sons
- [95] Kadane, J. B. (2008). *Statistics in the Law : A Practitioner's Guide, Cases, and Materials: A Practitioner's Guide, Cases, and Materials* (p. 472). Oxford University Press.
- [96] Kahneman, D., Slovic, P., and Tversky, A., *Judgement Under Uncertainty: Heuristics and Biases*. 1982: New York: Cambridge University Press.
- [97] Kaye, D. H. (2001). "Two Fallacies About DNA Databanks for Law Enforcement." *Brooklyn Law Review* 67(1): 179-206.
- [98] Kaye, D. H. (2007). "Revisiting 'Dreyfus': A More Complete Account of a Trial by Mathematics." *Minnesota Law Review* 91(3): 825-835.
- [99] Koehler, J. J. (1992). Probabilities in the Courtroom: An Evaluation of the Objections and Policies. In *Handbook of Psychology and Law*. Kagehiro, D. K. and Laufer, W. S. New York, Springer-Verlag: 167-184.
- [100] Koehler, J. J. (1993). "Error and Exaggeration in the Presentation of DNA Evidence at Trial." *Jurimetrics* 34: 21-39
- [101] Koehler, J. J. (1993). "The Base Rate Fallacy Myth." *Psychology* 4(49).
- [102] Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios and error rates. *University of Colorado Law Review*, 67, 859–886.
- [103] Korb K, Nicholson A. (2010), "Bayesian Artificial Intelligence", 2nd Edition, Chapman & Hall/CRC Computer Science & Data Analysis.

- [104] Langley, W. (2012, March 6). The case against DNA. Telegraph. Retrieved from <http://www.telegraph.co.uk/news/science/9115916/The-case-against-DNA.html>
- [105] Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *J. R. Statist. Soc. B*, 50, No 2, pp 157-224.
- [106] Lucy, D. and Schaffer, B. (2007). Nature of legal evidence, The University of Edinburgh, Edinburgh. EH9 3JZ. www.maths.ed.ac.uk/~cgga/nature-of-evidence.pdf
- [107] MARSHALL, T. C., SLATE, J., KRUIK, L. E. B., & PEMBERTON, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7(5), 639–655.
- [108] Meester, R., & Sjerps, M. (2003). The Evidential Value in the DNA Database Search Controversy and the Two-Stain Problem. *Biometrics*, 59(3), 727–732.
- [109] Meester, R. and Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, *Law, Probability and Risk* 3 (1): 51-62.
- [110] Meester, R., M. Collins, R. Gill and M. van Lambalgen (2007). "On the (ab)use of statistics in the legal case against the nurse Lucia de B." *Law, Probability & Risk* 5: 233-250.
- [111] Morrison, G. M. (2012). "The likelihood ratio framework and forensic evidence in court: a response to RvT." *International Journal of Evidence and Proof* 16(1).
- [112] Mullen, C., Spence, D., Moxey, L., & Jamieson, A. (2014). Perception problems of the verbal scale. *Science & Justice : Journal of the Forensic Science Society*, 54(2), 154–8.
- [113] Murphy, P. (2003). *Evidence, Proof, and Facts: A Book of Sources*, Oxford University Press.
- [114] National Research Council. (2009). *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press.
- [115] Neil, M., Taylor, M., & Marquez, D. (2007). Inference in hybrid Bayesian networks using dynamic discretization. *Statistics and Computing*, 17(3), 219–233.
- [116] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Sciences*, 52(1), 54–64.
- [117] North Carolina. (2009). *Trial Judges' Bench Book District Court Edition, Volume 1, Family Law. Paternity Chapter*.
- [118] Nuemann, C., Evett, I. W., & Skerrett, J. (2011). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J. Royal Statistical Soc A*, 175(2), 1–26.
- [119] Nobles, R. and Schiff, D. (2005). "Misleading statistics within criminal trials: the Sally Clark case." *Significance* 2(1): 17-19.
- [120] Nordgaard, A. Hedell, R., and Ansell, R. (2012). Assessment of forensic findings when alternative explanations have different likelihoods-"Blame-the-brother"-syndrome. *Science and Justice* 52: 226-236.
- [121] Park, R. C. (1996). Character Evidence Issues in the O.J. Simpson Case - Or, Rationales of the Character Evidence Ban, with Illustrations from the Simpson Case. *University of Colorado Law Review*, 67.
- [122] Park, R. C., P. Tillers, et al. (2010). "Bayes Wars Redivivus — An Exchange." *International Commentary on Evidence* 8(1): 1-38.
- [123] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Palo Alto, CA, Morgan Kaufmann
- [124] Press, S. J. (Ed.). (2002). *Subjective and Objective Bayesian Statistics*. Hoboken, NJ, USA:

John Wiley & Sons, Inc.

- [125] Puch-Solis, R., P. Roberts, S. Pope and C. Aitken (2012). PRACTITIONER GUIDE NO 2: Assessing the Probative Value of DNA Evidence, Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society
<http://www.rss.org.uk/uploadedfiles/userfiles/files/Practitioner-Guide-2-WEB.pdf>
- [126] Redmayne, M. (2001). *Expert Evidence and Criminal Justice*, Oxford University Press.
- [127] Redmayne, M., P. Roberts, C. Aitken and G. Jackson (2011). "Forensic Science Evidence in Question." *Criminal Law Review* (5): 347-356.
- [128] Redmayne, M., DNA evidence, probability and the courts, in *Criminal Law Review*, 464. 1995.
- [129] 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.
- [130] Robertson, B. and T. Vignaux (1995). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, John Wiley and Son Ltd.
- [131] Robertson, B. and Vignaux, T., Bayes' Theorem in the Court of Appeal. *The Criminal Lawyer*, 1997: p. 4-5.
- [132] Robertson, B. and Vignaux, T., Don't teach Statistics to Lawyers!, in *Proceedings of the Fifth International Conference on Teaching of Statistics*. 1998: Singapore. p. 541-547.
- [133] Robertson, B., G. A. Vignaux and C. E. H. Berger (2011). "Extending the confusion about Bayes." *The Modern Law Review* 74(3): 444-455.
- [134] Royal Statistical Society. (2015.). *Statistics and the Law Working Group*.
http://www.rss.org.uk/RSS/Influencing_Change/Statistics_and_the_law/RSS/Influencing_Change/Current_projects_sub/Statistics_and_the_law.aspx
- [135] Saks, M.J. and Thompson, W.C., *Assessing Evidence: Proving Facts*, in *Handbook of Psychology in Legal Contexts (Second Edition)*, Carson, D. and Bull, R., Editors. 2003.
- [136] Schneps, L., & Colmez, C. (2013). *Math on Trial Basic Books*, 2013. Basic Books.
- [137] Sjerps, M. and C. Berger (2012). "How clear is transparent? Reporting expert reasoning in legal cases." *Law, Probability and Risk* 11 (4): 317-329.
- [138] Spiegelhalter, D. J. (2013). *Court of Appeal bans Bayesian probability (and Sherlock Holmes)*. Retrieved from <http://understandinguncertainty.org/court-appeal-bans-bayesian-probability-and-sherlock-holmes>
- [139] Taroni, F., Aitken, C., Garbolino, P., & Biedermann, A. (2014). *Bayesian Networks and Probabilistic Inference in Forensic Science (2nd ed.)*. Chichester, UK: John Wiley & Sons.
- [140] Thompson, J., Danielson, S., Hartmann, E., Bucheger, D., & Koplitz, J. (2001). *Trial by Probability*. Retrieved from <https://www.uwstout.edu/mscs/upload/Trial-by-Probability.pdf>
- [141] Thompson, W. C., F. Taroni and C. G. G. Aitken (2003). "How the probability of a false positive affects the value of DNA evidence." *Journal of Forensic Sciences* 48(1): 47-54.
- [142] Thompson, W.C. and Schumann, E.L., Interpretation of statistical evidence in criminal trials. *Law and Human Behavior*, 1987. 11(3): p. 167-187. <http://dx.doi.org/10.1007/BF01044641>
- [143] Thompson, W.C., DNA Evidence in the O.J. Simpson Trial. *Colorado Law Review*, 1996. 67(4): p. 827-857.
- [144] Tillers, P., & Green, E. (Eds.). (1988). *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism*. Kluwer Academic Publishers.
- [145] Tillers, P. and J. Gottfried (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, Probability & Risk* 5: 135-157.
- [146] Tribe, L. H. (1971). "Trial by Mathematics." *Harvard Law Review* 84: 1329.
- [147] Triggs C.M. and J.S. Buckleton, Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, *Law, Probability and Risk* (2004) 3(1): 73-82
- [148] Vosk, T., & Emery, A. F. (2014). *Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists* , CRC Press.
- [149] Wagenaar, W. A. (1988). The Proper Seat: A Bayesian Discussion of the Position of Expert Witnesses. *Law and Human Behavior*, 12(4), 499-510
- [150] Ziliak, S. T., & McCloskey, D. N. (2008). *The Cult of Statistical Significance*. The University of Michigan Press.