Avoiding Probabilistic Reasoning Fallacies in Legal Practice using Bayesian Networks

Norman Fenton and Martin Neil

RADAR (Risk Assessment and Decision Analysis Research)
School of Electronic Engineering and Computer Science
Queen Mary (University of London)
London E1 4NS

and

Agena Ltd
www.agena.co.uk
11 Main Street, Caldecote
Cambridge CB23 7NU

norman@eecs.qmul.ac.uk, martin@eecs.qmul.ac.uk

29 June 2011

Abstract

Probabilistic fallacies, such as the prosecutor fallacy, have been widely documented, yet these fallacies continue to occur in legal practice. This paper considers how best to avoid them, drawing on our experience as expert witnesses/advisors in recent trials. Although most fallacies are easily avoided by applying Bayes' Theorem, attempts to explain this to lawyers using the standard mathematical formulas seem doomed to failure. For simple arguments it is possible to explain common fallacies using purely visual presentation alternatives to the Bayes formulas (such as event probability trees) in ways that are fully understandable to lay people. However, as the evidence (and dependence between different evidence) becomes more complex, these visual approaches become infeasible. We show how Bayesian networks can be used to address the more complex arguments in such a way that it is not necessary to expose the underlying complex Bayesian computations. We demonstrate this new approach in explaining well known fallacies and a new fallacy that arose in a recent major murder trial. We also address the barriers to more widespread take-up of these methods within the legal profession, including the need to 'believe' the correctness of Bayesian calculations and the common reluctance to consider subjective prior probabilities.

Keywords: legal fallacies, probability, Bayes

Word count: 11,871

1 Introduction

The issue of probabilistic reasoning fallacies in legal practice (hereafter referred to simply as *probabilistic fallacies*) is one that has been well documented, being dealt with even in populist books such as [53][59]. An excellent overview of the most common fallacies and some of the more famous (primarily UK-based) cases in which they have occurred can be found in [10], with further explanatory material in [39]. Other works that deal in general terms with one or more fallacy include [19][20][37][38][45][52][65][70][73][81][89][90][97][99][108], while discussions of fallacies in the context of reasoning about evidence can be found in [8][9][17][38][51][68][80][84]. An extensive account of fallacies affecting dozens of US cases is provided in [74] while [48] describes four further relevant US cases. More detailed analyses of important individual cases include those of: Sally Clark (covered in [50][61][67][86]); O.J Simpson (covered in [55], [109]); Denis John Adams (covered in [33] [31]); nurse Lucia de B (covered in [83]); and Doheny and Adams (covered in [2] [94]). While all these particular cases have occurred within the last 20 years, the phenomenon is by no means new. The Collins case [1], which is the subject of many of the above cited studies, took place in 1968 and some well-documented cases date back to the 19th century; these include the 1875 Belhaven and Stenton Peerage case (described in detail in [29]) and the 1894 Dreyfus case (described in detail in [71]).

For the purposes of this paper an *argument* refers to any reasoned discussion presented as part of, or as commentary about, a legal case. Hence, we are talking about probabilistic fallacies occurring in arguments. There are other classes of fallacies that have occurred frequently in arguments, such as cognitive fallacies (including most notably confirmation bias [15][34][52]), but these are outside the scope of this paper.

There is almost unanimity among the authors of the works cited in the first paragraph that a basic understanding of Bayesian probability is the key to avoiding probabilistic fallacies. Indeed, Bayesian reasoning is explicitly recommended in works such as [14] [39][49][51][38][56][64][90][95][96][100] [114], although there is less of a consensus

on whether or not experts are needed in court to present the results of all but the most basic Bayesian arguments [95][96].

Yet, despite the many publications and other publicity surrounding them, and despite the consensus (within the probability and statistics community) on the means of understanding and avoiding them, probabilistic fallacies continue to proliferate legal arguments. Part of the problem can be attributed to a persistent attitude among some members of the legal profession that probability theory has no role to play at all in the courtroom; supporters of this viewpoint often point to a highly influential paper by Tribe in 1971 [113]. However, Tribe's arguments have long been systematically demolished by the likes of Koehler [72] and Edwards [37], and more recently by Tillers and Gottfried [111]; in any case Tribe's arguments in no way explain or justify the errors that have been made. Informed by our experience as expert witnesses on a number of recent high-profile trials (both criminal and civil) we seek to address this problem by proposing a different approach to detect, explain and avoid fallacies. Our approach, which can actually be applied to all types of reasoning about evidence, exploits the best aspects of Bayesian methods while avoiding the need for non-mathematicians to understand mathematical formulas.

Central to our approach is a recognition that members of the legal profession cannot be expected to follow even the simplest instance of Bayes Theorem in its formulaic representation. This explains why, even though many lawyers are aware of the fallacies, they struggle to understand and, hence, avoid them. Instead of continuing the struggle to pursuade non-mathematicians to understand mathematics we propose an alternative approach and demonstrate how it has already been applied with some effect on real cases.

The paper is structured as follows.

- In Section 2 we provide an overview of the most common fallacies within a new classification framework that is conceptually simpler than previous approaches. We also compare the formulaic and visual versions of Bayes theorem in explaining key fallacies like the prosecution fallacy.

- Section 3 identifies why Bayes theorem is not just the means of avoiding fallacies but also, paradoxically, the reason for their continued proliferation. Specifically, this is where we explain the limitations of using purely formulaic explanations. We explain how, in simple cases, alternative visual explanations such as event trees enable lay people to fully understand the result of a Bayesian calculation without any of the maths or formulas. In order to extend this method to accommodate more complex Bayesian calculations we integrate the event tree approach with Bayesian networks, explaining how the latter can be used in a way that is analogous to using an electronic calculator for long division.

- Section 4 brings the various threads of the paper together by showing how our proposed approach has been used in practice.

Although the paper addresses the issue of what we might reasonably expect of experts and juries in the context of probabilistic reasoning, it does not address this issue in any general way (a relevant comprehensive account of this can be found in [13]).

The paper provides a number of original contributions: a classification of fallacies that is conceptually simpler than previous approaches; a new fallacy; and most importantly a new approach/method of fully exploiting Bayes in legal reasoning. Much of the material is drawn from our involvement in the R v Bellfield case [6] (a case resulting in the conviction of Levi Bellfield for the murders of Marsha McDonnell and Amnelie Delagrange and the attempted murder of Kate Sheedy). The proposal to use Bayesian networks for legal reasoning and evidence evaluation is by no means new (see, for example, [12][37][106]), but what is new is our approach to the way this kind of reasoning is presented.

2 Some probabilistic fallacies

2.1 From hypothesis to evidence and back: the transposed conditional

Probabilistic reasoning of legal evidence often boils down to the simple causal scenario shown in Figure 1: we start with some hypothesis *H* (such as the defendant was or was not present at the scene of the crime) and observe some evidence *E* (such as blood type at the scene of the crime does or does not match the defendant's).

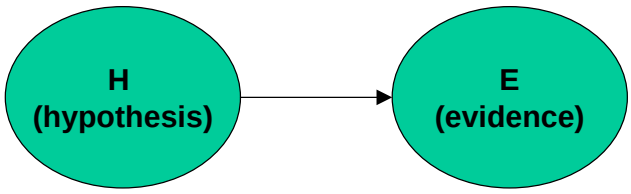


Figure 1 Causal view of evidence

The probability of *E* given *H*, written P(*E*|*H*), is called the *conditional probability*. Knowing this conditional probability enables us to revise our belief about the probability of *H* if we observe *E*.

Many of the most common fallacies of reasoning arise from a basic misunderstanding of conditional probability. An especially common example is to confuse:

the probability of a piece of evidence (*E*) given a hypothesis (*H*)

with

the probability of a hypothesis (*H*) given the evidence (*E*).

In other words P(*E*|*H*) is confused with P(*H*|*E*). This is often referred to as the ***fallacy of the transposed conditional*** [38]. As a classic example, suppose that blood type matching the defendant's is found at the scene of the crime (this is *E*) and that this blood type is found in approximately one in every thousand people. Then the statement:

the probability of this evidence given the defendant is not the source is 1 in 1000

(i.e. P(*E*|*H*)=1/1000 where we are assuming *H* is the statement 'defendant is not the source'*)*

is logically and mathematically correct. However, it is a *fallacy* to conclude that:

> *the probability the defendant is not the source given this evidence is 1 in 1000 (i.e.* $P(H|E)=1/1000$*)*

In this context, the transposed conditional fallacy is sometimes also called the *prosecutor's fallacy*, because the claim generally exaggerates the prosecutor's case; it suggests that there is an equally small probability that the defendant is not the source as there is the probability of observing the match in a random person.

A definitive explanation of the fallacy is provided by Bayes Theorem. However, as explained in [112] the usual Bayes formulation is notoriously difficult for lay people to grasp. It is therefore instructive (and important for what follows) to consider first an alternative very simple and informal visual explanation.

First suppose (Figure 2) that, in the absence of any other evidence, there are 10,000 people who could potentially have been the source of the blood (indeed, it is important to note that the failure to take account of the size of the potential source population is an instance of another fallacy, called the base-rate fallacy [69]).

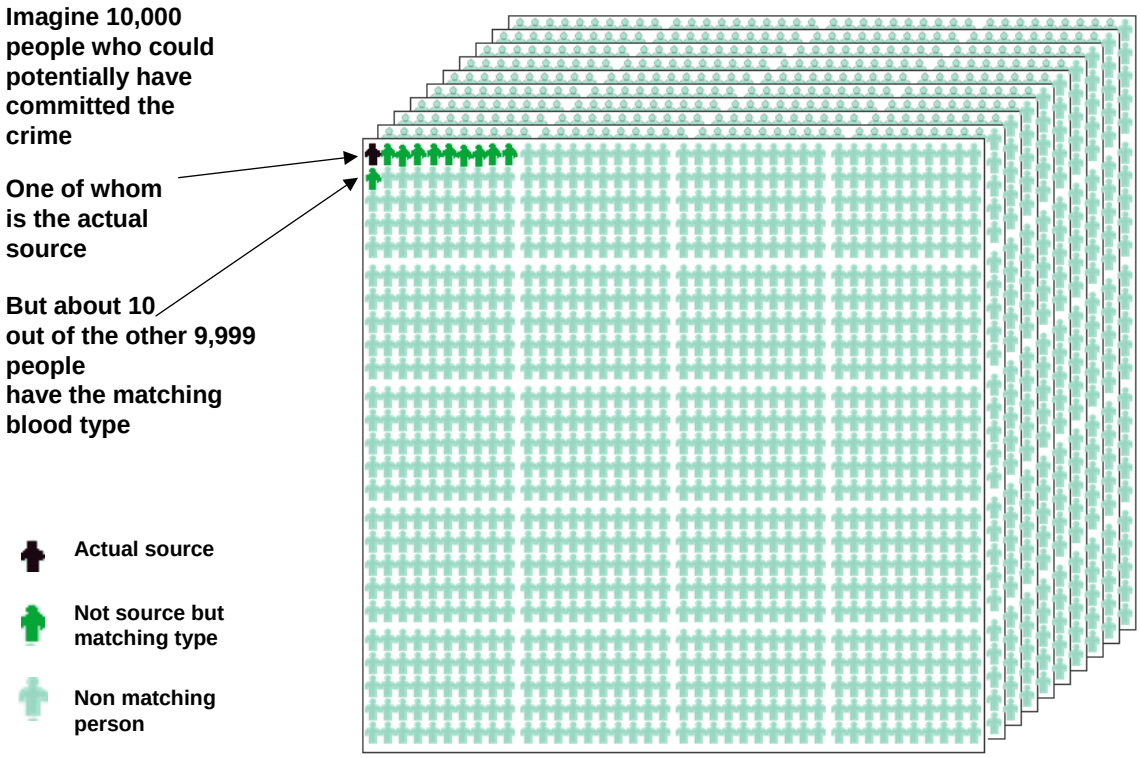


Figure 2 The potential source population

Of course only one is the actual source. But, because of the 1 in 1000 blood match probability, about 10 out of the other 9,999 people have the matching blood type.

This means there is a probability of 10/11 (i.e. about 91% chance) that a person with the matching blood type is not the source. **In other words P(*H*|*E*) is 0.91 (very likely) and not 1 in a thousand (highly unlikely) as claimed by the prosecution.**

In contrast to the above visual explanation of the fallacy, the calculations can be done formally with Bayes theorem, which provides a simple formula for updating our prior belief about *H* in the light of observing *E*. In other words Bayes calculates P(*H*|*E*) in terms of P(*E*|*H*). Specifically:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} = \frac{P(E \mid H)P(H)}{(E \mid H)P(H) + (E \mid notH)P(notH)}$$

So, using the same assumptions as above with 10,000 potential suspects and no other evidence, the prior P(*H*) is equal to 9,999/10,000. We know that P(*E*|*H*)=1/1000. For the denominator of the equation we also need to know P(*E*|not *H*) and P(not *H*). Let us assume that if the defendant *is* the source then the blood will certainly match so

P(*E*|not *H*)=1. Also, since P(*H*)=9,999/10,000 it follows that P(not *H*)= 1/10,000. Substituting these values into Bayes Theorem yields

$$P(H \mid E) = \frac{\frac{1}{1000} \cdot \frac{9,999}{10,000}}{\frac{1}{1000} \cdot \frac{9,999}{10,000} + 1 * \frac{1}{10,000}} = \frac{9,999}{10,999} \approx 0.91$$

While mathematicians and statisticians inevitably prefer the conciseness of the formulaic approach it turns out that most lay people simply fail to understand or even believe the result when presented in this way, as demonstrated empirically in [24] Although some authors, such as [25][53], claim it is the use of abstract probabilities and formulas, rather than the underlying concept, that acts as a barrier to understanding, as soon as we introduce more complexity into the problem, (such as a second piece of evidence) there is no approach that will enable non-mathematicians to understand (let alone perform) the calculations properly.

2.2 Many fallacies, but a unifying framework

The previous particular example of a transposed conditional fallacy is just one of a class of such fallacies that have been observed in arguments. In the context of DNA evidence Koehler [74] defined a range of such fallacies by considering the following chain of reasoning:

Match report → True Match →Source →Perpetrator

The direction of reasoning here is not causal (as in Figure 1) but deductive. Specifically, a reported match is suggestive of a true match, which in turn is suggestive that the defendant is the source. This in turn is suggestive that the defendant is the actual perpetrator (i.e. is guilty of the crime). Of course, it is erroneous to consider any parts of this chain of deductions as following automatically. Errors in the DNA typing process can result in a reported match where there is no true match [110]. A true match can be coincidental if more than one member of the

population shares the DNA features recorded in the sample; and finally even if the defendant was the source he/she may not be the perpetrator since there may be an innocent reason for their presence at the crime scene.

Koehler's analysis and classification of fallacies can be generalised to apply to most types of evidence by considering the *causal* chain of evidence introduced in Figure 3.

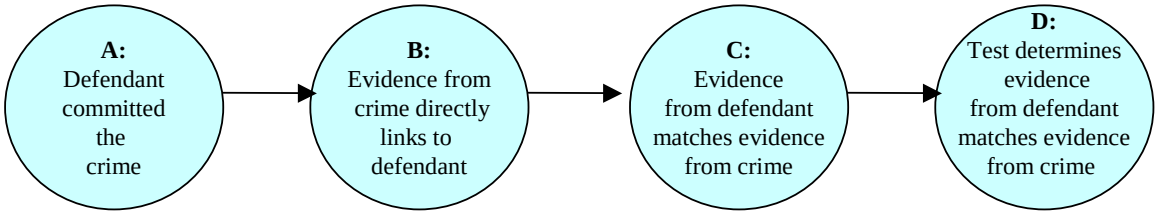


Figure 3 Causal chain of evidence

We will show that this schema allows us to classify fallacies of reasoning that go beyond Koehler's set.

In this schema we assume that evidence could include such diverse notions as:

- anything revealing a DNA trace (such as semen, saliva, or hair)
- a footprint
- a photographic image from the crime scene
- an eye witness statement (including even a statement from the defendant).

For example, if the defendant committed the crime (*A*) then the evidence may be a CCTV image showing the defendant's car at the scene (*B*). The image may be sufficient to determine that the defendant's car is a match for the one in the CCTV image (*C*). Finally, experts may determine from their analyses of the CCTV image that it matches the defendant's car (*D*).

Koehler's approach is heavily dependent on the notion 'frequency of the matching traits', denoted *F*(traits). This is sometimes also referred to as the 'random match probability'. In our causal framework *F*(traits) is equivalent to the more formally defined

$$P(C \mid \text{not } B)$$

i.e the probability that a person NOT involved in the crime, coincidentally provides evidence that matches.

> **Example**: In the case [6] the prosecution relied heavily on a blurred CCTV still image of a car at the scene of one of the murders that was claimed to match the defendant's. Image analyst experts were able to reveal the type of the car and 3 of the 7 digits on the number plate. Hence P(*C* | not *B*) was determined by the number of vehicles of the same type whose number plates matched in those 3 digits.

With this causal framework, we can characterise a range of different common fallacies resulting from a misunderstanding of conditional probability, thus extending the work of Koehler (using Koehler's terminology wherever possible). Full details are provided in an earlier version of this work [42], but the following are especially important examples:

1. '*Source probability error':* This is where we equate P(*C* | not *B*) with P(not *B* | *C*). Many authors (see, for example [19][82][97][99]) refer to this particular error as the prosecutor fallacy.

2. '*Ultimate issue error'* This is where we equate P(*C* | not *B*) with P(not *A* | *C*). This too has been referred to as the *Prosecutor fallacy* [108]: it goes beyond the source probability error because it can be thought of as compounding that error with the additional incorrect assumption that P(*A*) is equal to P(*B*).

3. *P(Another Match) Error*: This is the fallacy of equating the value P(*C* | not *B*) with the probability (let us call it *q*) that at least one innocent member of the population has matching evidence. The effect of this fallacy is usually to grossly exaggerate the value of the evidence *C*. For example, Koehler [74] cites the case in which DNA evidence was such that P(*C* | not *B*) = 1/705,000,000, but where the expert concluded that the probability that

another person (other than the defendant) having the matching DNA feature must also be equal to 1/705,000,000. In fact, even if the 'rest of the population' was restricted to, say, 1,000,000, the probability of at least one of these people having the matching features is equal to:

$$1 - (1 - 1/705{,}000{,}000)^{1{,}000{,}000}$$

and this number is approximately 1 in 714, not 1 in 705,000,000 as claimed.

Where the evidence *B* is DNA evidence, the match probability P(*C* | not *B*) can be very low (millions or even billions to one), which means that the impact of this class of fallacies can be massive since the implication is that this very low match probability is equivalent to the probability of innocence.

2.3 Avoiding fallacies using the likelihood ratio (advantages and disadvantages)

What is common across all of the fallacies described above and in the Appendix of [42] is that ultimately the true utility of a piece of evidence is presented in a misleading way – the utility of the evidence is either exaggerated (such as in the prosecutor fallacy) or underestimated (such as in the defendant fallacy). Yet there is a simple probabilistic measure of the utility of evidence, called the ***likelihood ratio***. For any piece of evidence *E*, the likelihood ratio of *E* is the probability of seeing that evidence if the defendant is guilty divided by the probability of seeing that evidence if the defendant is not guilty. It follows directly from Bayes Theorem that if the likelihood ratio is bigger than 1 then the evidence increases the probability of guilt (with higher values leading to higher probability of guilt) while if it is less than 1 it decreases the probability of guilt (and the closer it gets to zero the lower the probability of guilt). An equivalent form of Bayes Theorem (called the 'odds' version of Bayes) tells us that the posterior odds of guilt are the prior odds times the likelihood ratio. If the likelihood ratio is equal to or close to 1 then *E* offers no real value at all since it neither increases nor decreases the probability of guilt.

Evett and others have argued [38] that many of the fallacies are easily avoided by focusing on the likelihood ratio. Indeed, Evett's crucial expert testimony in the appeal case of R v Barry George [5] (previously convicted of the murder of the TV presenter Gill Dando) focused on the fact that the forensic gunpowder evidence that had led to the original conviction actually had a likelihood ratio of about 1. This is because both P(*E* | Guilty) and P(*E* | not Guilty) were approximately equal to 0.01. Yet only P(*E* | not Guilty) had been presented at the original trial (a report of this can be found in [11]).

Another advantage of using the likelihood ratio is that it removes one of the most commonly cited objections to Bayes Theorem, namely the obligation to consider a prior probability for a hypothesis like 'guilty' (i..e. we do not need to consider the prior for nodes like A or B in Figure 3). For example, in the prosecutor fallacy example above, we know that the probability of seeing that evidence if the defendant is not guilty is 1/1000 and the probability of seeing that evidence if the defendant is guilty is 1; this means the likelihood ratio is 1000 and hence, *irrespective of the 'prior odds'*, the odds of guilt have increased by a factor of 1000 as a result of observing this evidence. Hence, the use of the likelihood ratio goes a long way toward allaying the natural concerns of lawyers who might otherwise instinctively reject a Bayesian argument on the grounds that it is intolerable to assume prior probabilities of guilt or innocence.

While we strongly support the use of the likelihood ratio as a means of both avoiding fallacies and measuring the utility of evidence, in our experience lawyers and lay people often have similar problems understanding the likelihood ratio as they do understanding the formulaic presentation of Bayes. As we shall see in the next section, this problem becomes acute in the case of more complex arguments.

3 The fallacies in practice: why Bayes hinders as much as helps

3.1 The fallacies keep happening

The specific rulings on the prosecutor fallacy in the case of R v Deen (see [17]) and R v Doheny/Adams (see [3]) should have eliminated its occurrence from the courtroom. The same is true of the rulings in relation to the dependent evidence fallacy in the case of People vs Collins and Sally Clark. Indeed the Sally Clark case prompted the President of the Royal Statistical Society to publish an open letter to the Lord Chancellor regarding the use of statistical evidence in court cases [58] (we shall return to this letter in Section 3.2).

Unfortunately these, and the other fallacies described in [42], continue to occur frequently. Moroever, in addition to the many examples cite in Section 1 and in [42], it is also important to note that one does not need an explicit statement of probability to fall foul of many of the fallacies. 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 prosecution fallacy (see [39]). Indeed, based on examples such as these and our own experiences as expert witnesses, we believe the reported instances are merely the tip of the iceberg.

For example, although this case has not yet been reported in the literature as such, in R v Bellfield 2007 [6] the prosecution opening contained instances of many of the fallacies described in Section 2, plus a number of new fallacies (one of which is described in Section 4 below). When our report [43] was presented by the defence to the prosecutor and judge, it was agreed that none of these fallacies could be repeated in the summing-up. Nethertheless, just days later in another murder case (R vs Mark Dixie, accused of murdering Sally-Anne Bowman) involving *the same prosecuting QC* a forensic scientist for the prosecution committed a blatant instance of the prosecutor fallacy, as reported by several newspapers on 12 Feb 2008:

> "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."

3.2 The problem with Bayesian explanations

What makes the persistence of these fallacies perplexing to many statisticians and mathematicians is that all of them can be easily exposed using simple applications of Bayes Theorem and basic probability theory as shown in Section 2.

Unfortunately, while simple examples of Bayes Theorem are easy for statisticians and for mathematically literate people to understand, the same is not true of the general public. ***Indeed, we believe that for many people – and this includes from our own experience highly intelligent barristers, judges and surgeons, any attempt to use Bayes theorem to explain a fallacy is completely hopeless.*** They simply switch-off at the sight of a formula and fail to follow the argument.

Moreover, the situation is even more devastating when there are multiple pieces of possibly contradictory evidence and interdependencies between them. For example, there is a highly acclaimed half-page article by Good [56] that uses Bayes theorem with 3 pieces of related evidence to expose a fallacy in the OJ Simpson trial. Yet, because of its reliance on the formulaic presentation, this explanation was well beyond the understanding of our legal colleagues. Even more significant from a legal perspective was the landmark case of R vs Adams (discussed in [31] and [33]). Although Donelly [33] highlights the issue of the prosecutor fallacy, this fallacy was not an issue in court. The issue of interest in court was the use of Bayesian reasoning to combine the different conflicting pieces of evidence shown in Figure 4 (to put this in the context of Figure 3 the node “Adams Guilty” corresponds to node A while each of the other nodes corresponds to instances of node B).

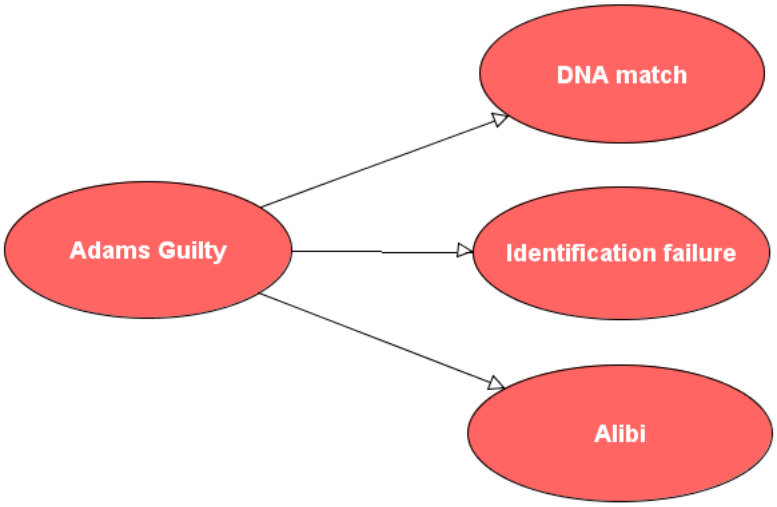


Figure 4 Hypothesis and evidence in the case of R v Adams

An example of part of the Bayesian calculations (using the likelihood ratio) required to perform this analysis are shown in Figure 5). While, again, this is simple for statisticians familiar with Bayes, arguments like this are well beyond the comprehensibility of most judges and barristers, let alone juries.

$$V = \frac{Pr(H_p \mid E, I_1, I_2)}{Pr(H_d \mid E, I_1, I_2)}$$
$$= \frac{Pr(E \mid H_p)}{Pr(E \mid H_d)} \times \frac{Pr(I_1 \mid H_p)}{Pr(I_1 \mid H_d)} \times \frac{Pr(I_2 \mid H_p)}{Pr(I_2 \mid H_d)} \times \frac{Pr(H_p)}{Pr(H_d)}$$

Figure 5 Likelihood ratio calculation for Adams case taken from [9].

Yet, in court, the defence expert (Donelly) presented exactly such calculations, assuming a range of different scenarios, from first principles (although Donelly states in [33] that this was at the insistence of the defence QC and was not his own choice). The exercise was, not surprisingly, less than successful and the appeal judge ruled

> **"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 generally sympathetic to, his subsequent statement is much more troubling:

> **"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"**

This statement characterises the major challenge we face. At an empirical level, the statement is deeply concerning because the extensive literature on fallacies discussed in Section 2 and in the Nobel prize-winning work of Kahneman and Tversky [69] confirms that lay people *cannot* be trusted to reach the proper conclusion when there is probabilistic evidence. Indeed, experts such as forensic scientists and lawyers, and even professional statisticians, cannot be trusted to reach the correct conclusions.

Where we differ from some of the orthodoxy of the statistical community is 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. In some respects this puts us at odds with the President of the Royal Statistical Society whose letter [58] (the background to which was described above) concluded:

> The Society urges you to take steps to ensure that statistical evidence is presented only by appropriately qualified statistical experts, as would be the case for any other form of expert evidence.

The problem with such a recommendation is that it fails to address the real concerns that resulted from the Adams case, namely that statistical experts are not actually 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 [95][96] 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 for long division.

3.3 Bayes and the long division (calculator) analogy

The point about a long division calculation is that, like a Bayes calculation, it is almost impossible to do it accurately by hand. Instead, it is accepted that we would use an electronic calculator. If somebody presented in court the result of a long division from a calculator you might, as a precaution, check the result by running the same calculation in another calculator. But you would certainly not expect the person to provide a first principles explanation of all the thousands of circuit level calculations that take place in his particular calculator in order to justify the result that was presented, especially if he also had to explain from first principles the fact that no completely accurate result was possible due to recurring decimals. Not only would that be unreasonable, but the jury would also surely fail to understand such an explanation. This might lead the judge to conclude (perfectly reasonably) that

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

If, additionally, the judge were to conclude (as in the Adams case) that

> **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"**

then the long division result that was presented should be disregarded and members of the jury should be allowed to use common sense (and certainly not a calculator) to come up with their own result for the long division. Common sense would inevitably lead jury members to deduce different values and, based on the known difficulty of performing such calculations purely intuitively, these may be inaccurate.

While the above scenario seems ludicrous we claim it is *precisely analogous* of the expectations when Bayes theorem (rather than long division) has to be used. ***Our proposal is that there should be no more need to present Bayesian arguments from first principles than there should be a need to explain the underlying circuit calculations that take place in a calculator for the division function***.

This is because Bayesian analysis share the following properties of long division:

1. *Most people can understand and do it from scratch in very simple cases.* When a very simple Bayesian argument is presented visually (as in Section 2) using concrete frequencies people not only generally understand it well [15], but they can construct their own correct simple calculations. To emphasize this point, we have used both the formulaic and visual explanations presented in Section 2 to numerous lay people, including lawyers and barristers. Whereas they find it hard both to 'believe' and reconstruct the formulaic explanation, they inevitably understand the visual explanation.

   The particular visual explanation presented can actually be regarded as simply an animated version of an event tree (also called decision tree, probability tree or frequency tree) [101]. The equivalent event tree is shown in Figure 6.

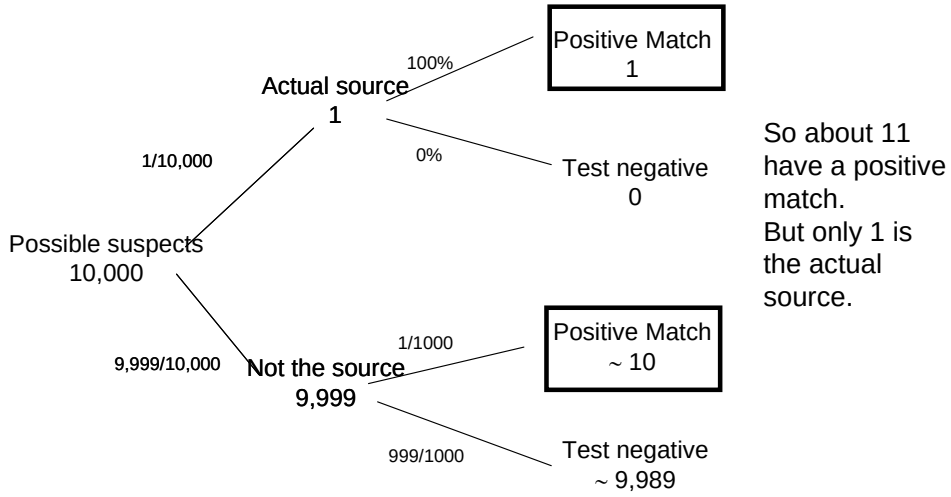


Figure 6 Event/decision tree representationof Bayesian argument

To emphasize the impact of such alternative presentations of Bayes, we were recently involved in a medical negligence case [47] where it was necessary to quantify the risks of two alternative test pathways. Despite the statistical data available, neither side could provide a coherent argument for directly comparing the risks of the alternative pathways, until a surgeon (who was acting as an expert witness) claimed that Bayes Theorem provided the answer. The surgeon's calculations were presented formulaically but neither the lawyers nor the other doctors involved could follow the argument. We were called in to check the surgeon's Bayesian argument and to provide a user-friendly explanation that could be easily understood by lawyers and doctors sufficiently well for them to argue it in court themselves. It was only when we presented the argument as an event tree that everything became clear. Once it 'clicked' the QC and surgeons felt sufficiently comfortable with it to present it themselves in court. In this case, our intervention made the difference between the statistical evidence on risk being used and not being used.

The important point about the visual explanations is not just that they enable lay people to do the simple Bayes calculations themselves, but they also provide confidence that the underlying Bayesian approach to conditional probability and evidence revision makes sense.

Ideally, it would be nice if these easy-to-understand visual versions of Bayesian arguments were available in the case of multiple evidence with interdependencies. Unfortunately, in such cases these visual methods do not scale-up well. But that is even *more* true of the formulaic approach, as was clear from the Adams case above. And it is just as true for long division.

2. *Scientists have developed algorithms for doing it in the general case*: Think of the 'general' case as a causal network of related uncertain variables. In the simple case there are just two variables in the network as shown in Figure 1. But in the general case there may be many such variables as in Figure 4. Such networks are called Bayesian networks (BNs) and the general challenge is to compute the necessary Bayesian calculations to update probabilities when evidence is observed. In fact, no computationally efficient solution for BN

calculation is known that will work in all cases. However, a dramatic breakthrough in the late 1980s changed things. Researchers such as Lauritzen and Spiegelhalter [77] and Pearl [88] published algorithms that provided efficient calculation for a large class of BN models. The algorithms have indeed been tested and validated within the community and are described in sufficient detail for them to be implemented in computer software. Note also that, as in the case of long division, the algorithms do not need to be understood by lay people.

3. *There are computers that implement the algorithms to acceptable degrees of accuracy*. In fact, there are many commercial and free software tools that implement the calculation algorithms and provide visual editors for building BNs. See [46] for an extensive review. The algorithms and tools are now sufficiently mature that (with certain reasonable assumptions about what is allowed in the models) the same model running in a different tool will provide the same result. As with using a calculator for long division the accuracy of the result depends on the quality of the inputs and assumptions. For division we only have to decide what the numerator is and what the denominator is. For a BN we have to make assumptions about:

    - Which variables are dependent on which others (i.e. what is the topology of the BN)
    - What are the prior probabilities for each variable; for variables with parents this means agreeing the probability of each state conditioned on each combination of parent states.

    We believe that making these assumptions is typically not as onerous as has often been argued [46]. In fact, the assumptions are normally already implicit somewhere in the case material. Moreover, where there are very different prior assumptions (such as the contrast between the prosecution assumptions and the defence assumptions) in many cases, the prior assumptions turn out not to be very critical since the same broad conclusions follow from a wide-range of different assumptions; the technique for using a range of different assumptions is called sensitivity analysis (which is easily performed in a BN). For example,

> Ward conducted a BN analysis of the Collins evidence [37], and found that, despite the fallacies committed at the trial, running the model with a wide range of different assumptions always led to a very high probability of guilt. In the medical case [47] described above we used the very different 'priors' of the claimant and the defence but for both the result came down firmly in favour of the claimant case. Also, one of the benefits of the BN approach is that all the assumptions are actually forced into the open.

It is important to note that the notion of using BNs for legal reasoning is not new, although we may be the first to have used them to help lawyers understand the impact of key evidence in real trials. As discussed above, Edwards produced an outstanding paper on the subject in 1991 [37], while Kadane and Schum [68] used the approach retrospectively to analyse the evidence in the Sacco and Vanzetti case. Other important contributions have been made by Aitken and colleagues [12] and Taroni et al [106] whose book describes the potential for BNs in the context of forensic evidence; other work on BNs in the forensic evidence space includes [26][27][28][30][76][87]. We specifically used BNs to explain the jury fallacy in [45] and recommended more general use of BNs in legal reasoning – an idea taken up by a practicing barrister [65][66]. More generally, the idea of graphical, causal type models for reasoning about evidence dates back as far as 1913 [116], while other relevant modern approaches can be found in [21], [115].

We recognise there *are* some challenging issues, which remain open research questions. The most important of these is the fact that the BN approach forces us to make assumptions that are simply not needed, such as the prior probability of state combinations that are impossible in practice, and (in some cases) the prior probability of nodes like 'guilty' which are not needed when the likelihood ratio is used.

3.4 So how might a complex Bayesian argument be presented in court?

The Bayesian approach is most relevant in cases where we need to combine different types of evidence, in particular where some evidence may be considered more

‘objective’ than other evidence. For example, it is accepted that evidence of a DNA match can be turned into a probability statement about the defence hypothesis as follows (for simplicity here we will assume that ‘not guilty’ is the same as ‘not the source of the DNA’):

Probability (“DNA match” given “Defendant not guilty”) = 1 in 20 million

where 1 in 20 million is the random match probability. Such a probability is considered to be ‘objective’ and ‘statistically sound’. However, courts are less likely to accept probabilities for more ‘subjective’ evidence. So if, for example, it was known that the defendant was not selected by the victim from an identification parade, then it would be considered contentious to make a statement like

The probability (“identification failure” given “Defendant guilty”) = 1 in 10

However, Bayesians argue that *both* of the statements are subjective probabilities (and of course in the Bayesian calculations there is no conceptual difference between them). It is beyond the scope of this paper to engage in the debate about when evidence is ‘sufficiently statistically sound’ (many book and papers have been written about this topic and a recent ruling suggesting that DNA is more statistically sound that other types of evidence [7] has generated fierce and well argued opposition [20] [92] [93]). However, note that the difference in the above two statements, with respect to their level of ‘statistical soundness’, is by no means as clear-cut as some may believe. In practice, there is great uncertainty – and a lot of subjective judgement involved – in arriving at a DNA match probability. For example, as described in [23] and [40], DNA experts in the UK and the USA report very different random match probabilities for the same person (often many orders of magnitude different such as one in a billion compared to one in a trillion) and most match probabilities do not even take account of the potential for testing errors to massively change the results [110]. In comparison, the ‘subjective’ probability of the identification failure, may be much more reliable (with lower error bounds) if it is provided by an expert who has seen many such parades and their outcomes.

If, therefore, a DNA match probability is used to support the prosecution hypothesis, it seems reasonable to also use a probability for the identification failure to support the defence hypothesis, especially if we do the calculations using a range of values to represent the potential errors in judgement. This is precisely what happened in the Adams case introduced in Section 3.2. In this case the DNA match was the only evidence against the defendant, but it was also the only evidence that was presented in probabilistic terms in the original trial, even though there was actually great uncertainty about the value (the match probability was disputed, but it was accepted to be between 1 in 2 million and 1 in 200 million [31]) This had a powerful impact on the jury. The other two pieces of evidence favoured the defendant – failure of the victim to identify Adams and an unchallenged alibi. In the Appeal the defence argued that it was wrong to consider the impact of the DNA probabilistic evidence alone without combining it with the other evidence.

So what the defence did was to perform the Bayesian analysis incorporating all three pieces of evidence under a range of different prior probability assumptions (ranging from those most favourable to the prosecution to those most favourable to the defence). In principle this was a very reasonable and logical thing to do. Unfortunately, as described in Section 3.2, the defence's strategy was to get the jury to understand and perform the calculations themselves using the Bayesian formulas in Figure 5. This strategy was doomed and led directly to the disastrous ruling already discussed.

Using this case as an example, our proposed alternative strategy is to start with the simple intuitive example and then simply present the results of running the BN model. Figure 4 already presented the structure of the BN required for the Bayesian argument. However, any Bayesian argument should NOT begin with the BN model but rather would present either the population diagram approach of Figure 2 or event-tree explanation like Figure 6 of the impact of a single piece of evidence. In this case we would start, therefore, by explaining the impact of the DNA match. But we could start with 10 million suspects and the DNA match probability of 1 in 2 million. This would lead us to conclude that, in addition to the actual source, about 5 other suspects are just as likely to be guilty as the defendant in the absence of any other evidence. This would mean that a prior probability of guilt of 1 in 10 million becomes 1 in 6. If (as was actually argued in the case) the realistic set of possible suspects was 200,000

rather than 10 million, then the number of positive matches is a fraction (one tenth) of a person rather than a set of people. This would lead us to revise the prior probability of guilt from 1/200,000 to a posterior of 1/1.1 which is equal to approximately 0.91 or 91%. Obviously using a DNA match probability figure of 1 in 200 million results in an even higher posterior probability of guilt.

At this point the lawyer would be able to say something like the following:

> "What we have demonstrated to you is how we revise our prior assumption when we observe a single piece of evidence. Although we were able to explain this to you from scratch, there is a standard calculation engine (accepted and validated by the mathematical and statistical community) which will do this calculation for us without having to go through all the details. In fact, when there is more than a single piece of evidence to consider it is too time-consuming and complex to do the calculations by hand, but the calculation engine will do it instantly for us. This is much like relying on a calculator to do long division for us. You do not have to worry about the accuracy of the calculations; these are guaranteed. All you have to worry about is whether our original assumptions are reasonable. But we can show you the results with a range of different assumptions"

The lawyer could then present the results from a BN tool. To confirm what has already been seen the lawyer could show two results. One (Figure 7) the results of running the tool with no evidence entered and the second (Figure 8) the results of running the tool with the DNA match entered. The lawyer would emphasize how the result in the tool exactly matches the result presented in the event tree.

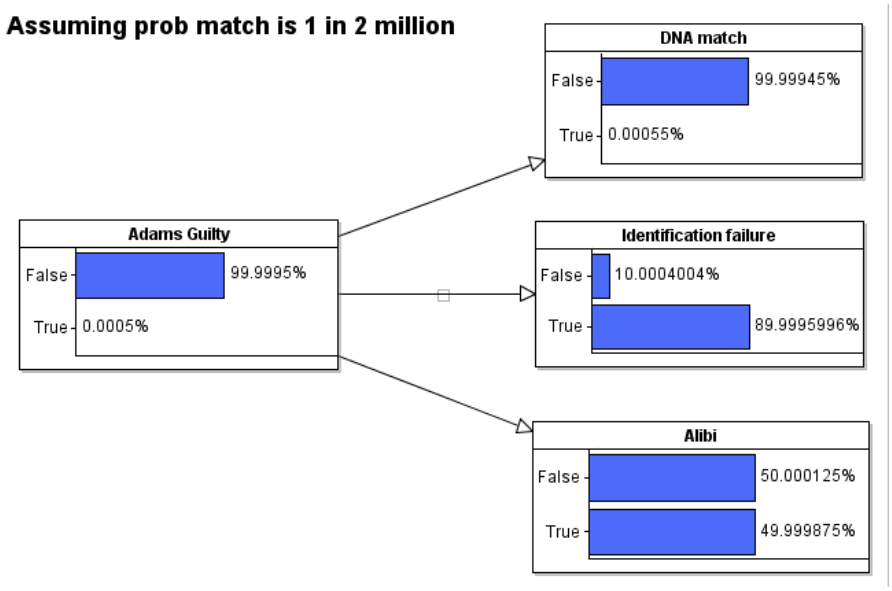


Figure 7 Model with prior marginal probabilities (assumes 200,000 possible suspects)

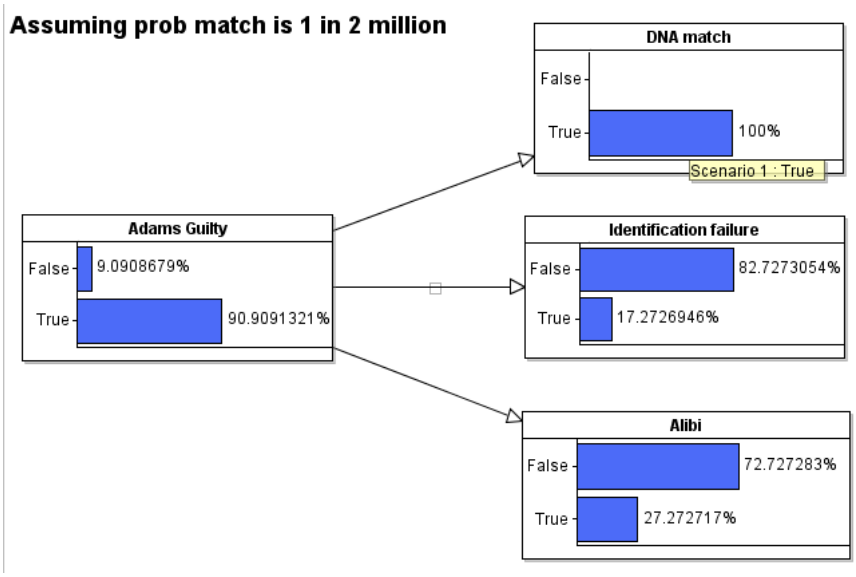


Figure 8 Result of entering DNA match

Next the lawyer would present the result of additionally entering the ID failure evidence (Figure 9). The lawyer would need to explain the P(*E*|*H*) assumption (in one of the scenarios the probability of ID failure given guilt was 0.1 and the probability of ID failure given innocence was 0.9). The result shows that the probability of guilt swings back from 91% to 52%.

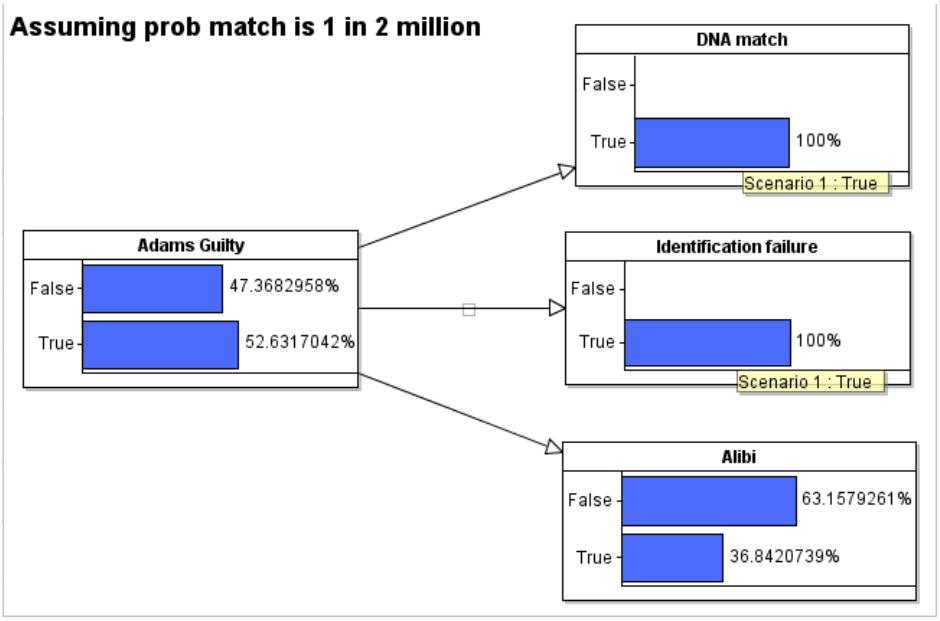


Figure 9 Identification failure added

In the same way the result of adding the alibi evidence is presented (Figure 10). With this we can see that the combined effect of the three pieces of evidence is such that innocence is now more likely than guilt.

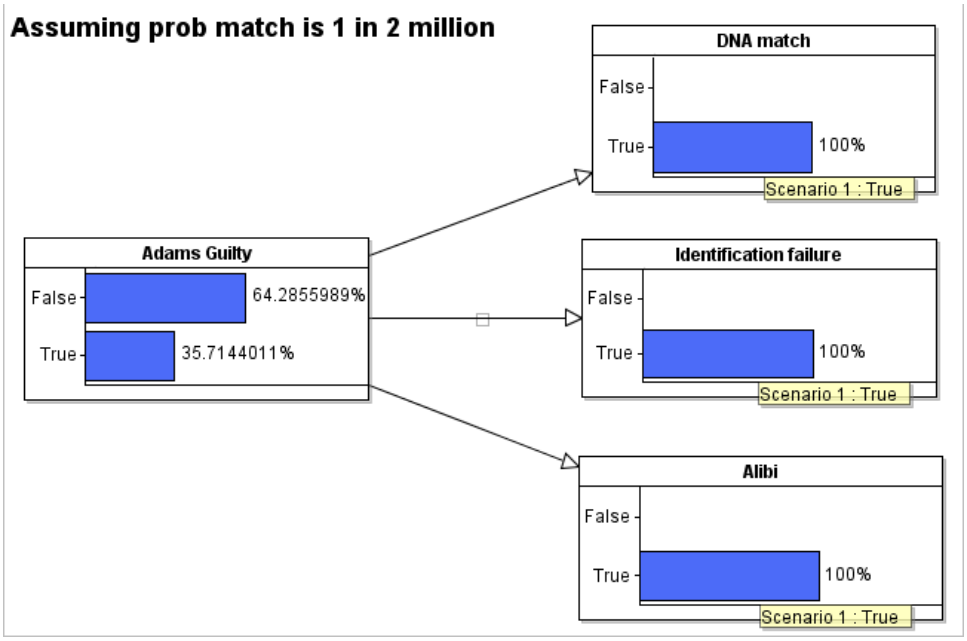


Figure 10 Alibi evidence entered

Finally, we can rerun the model with the other extreme assumption about the match probability of 1 in 200 million. Figure 11 shows the result when all the evidence is entered. In this case the probability of guilt is much higher (98%). Of course it would

be up to the jury to decide not just if the assumptions in the model are reasonable, but whether the resulting probability of guilt leaves room for doubt. What the jury would certainly NOT have to do is understand the complex calculations that have been hidden in this approach but were explicit both in the case itself and also in the explanation provided in both [31] and [9]. In this respect the judge's comments about the jury's task:

> **"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"**

does not seem so unreasonable to a Bayesian after all.

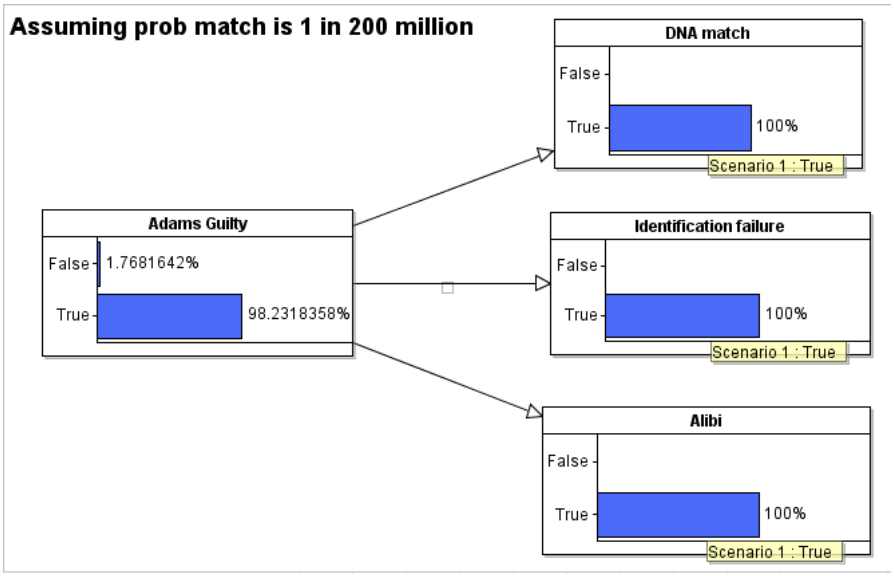


Figure 11 Effect of all evidence in case of 1 in 200 million match probability

We used this method in the medical negligence case [47] discussed in Section 3. Event trees were used for the basic argument and also for gaining trust in Bayes. Having gained this trust with the lawyers and surgeons we were able to present the results of a more complex analysis captured in a BN, without the need to justify the underlying calculations. Although the more complex analysis was not needed in court, its results provided important insights for the legal and medical team.

4 The proposed framework in practice

To bring all the previous strands together we now return to the R v Bellfield case [6]. Levi Bellfield was charged with two murders (Amelie Delagrange and Marsha Macdonnell) and three attempted murders. A key piece of evidence presented by the prosecution against Bellfield was a single blurred CCTV image of a car at the scene of the Marsha Macdonnell murder (the bulk of the evidence was otherwise largely circumstantial). The prosecution claimed that this car was Bellfield's car. The Prosecution used two vision experts to narrow down the number of potentially matching number plates in the image. We were originally brought in to determine if the statistical analysis of number plate permutations was correct. In fact, we believed that the image evidence had been subject to confirmation bias [34]. We used a BN to draw conclusions about the number of potentially matching number plates (and hence vehicles) that may not have been eliminated from the investigation.

Following on from the first piece of work the defence asked us to review the entire prosecution opening. Having discussed the well-known legal fallacies with them they sensed that the prosecution had introduced a number of such fallacies. Hence, we produced a report [43] that analysed the Prosecution Opening statement and identified several explicit and implicit instances of probabilistic fallacies that consistently exaggerated the impact of the evidence in favour of the prosecution case. These fallacies included one instance of the transposed conditional, several instances of impossibility of false negatives, several instances of base rate neglect, at least one instance of the previous convictions fallacy, and many instances of both the dependent evidence fallacy and the coincidences fallacy. We used Bayesian reasoning, with examples of simple BNs, to confirm some of the fallacies. The informal versions of the arguments in the report were used as a major component of the defence case.

We can present an example of the work done in completely general terms, using the approach proposed in this paper. This example involves a new fallacy that we have called the “Crimewatch” fallacy. Crimewatch is a popular TV programme in which

the public are invited to provide evidence about unsolved crimes. The fallacy can be characterised as follows:

> Fact 1: Evidence X was found at the crime scene that is almost certainly linked to the crime.
> Fact 2: Evidence X could belong to the defendant
> Fact 3: Despite many public requests (including, e.g, one on Crimewatch) for information for an innocent owner of evidence X to come forward and clear themselves, nobody has done so

The fallacy is to conclude from these facts that:

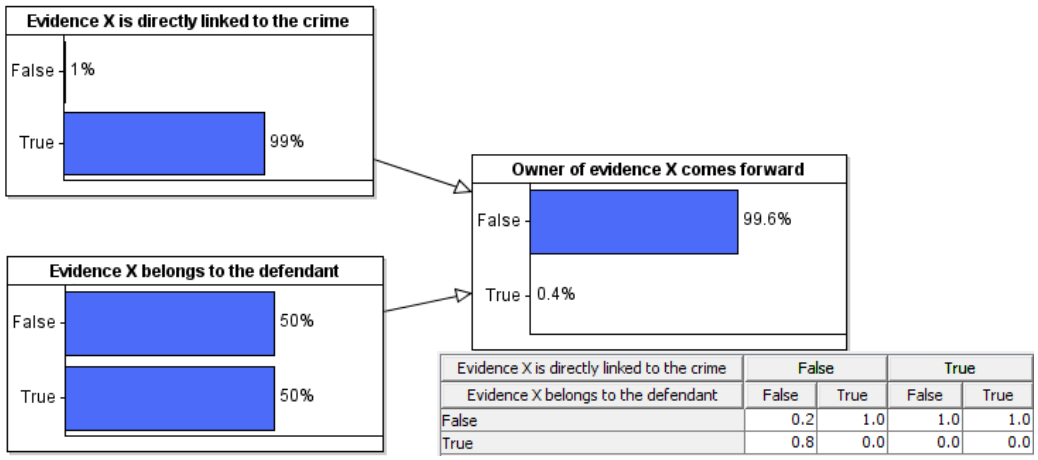
> ***It is therefore highly improbable that evidence X at the crime scene could have belonged to anybody other than the defendant.***

The prosecution opening contained an explicit example of this fallacy. The evidence X was the CCTV still image of the car that the police believed was driven by Marsha Macdonnell's attacker. Fact 2 was the prosecution hypothesis that this was a car belonging to Levi Bellfield. Fact 3 and the conclusion were the exact statements made in the prosecution opening.

The conclusion is a fallacy. Intuitively, we can explain the fallacy as follows: because fact 2 is 'almost' certain (the police were already convinced that the car was driven by the attacker) we can predict fact 3 will almost certainly be true even before the Crimewatch programme is screened (the attacker, whether it is the defendant or somebody else) is almost certainly ***not*** going to come forward as a result of the Crimewatch programme. Hence, when fact 3 is confirmed (nobody does come forward) it has negligible impact on the prior probability of guilt. In other words the Crimewatch UK evidence, far from proving that it is "highly improbable that the car could have belonged to anybody other than the defendant" actually tells us almost nothing than we already knew or assumed.

But to explain this properly we need a Bayesian argument. Once again the presentation of the Bayesian argument is initiated with a visual explanation of the

impact of a single piece of evidence. We then present a BN (as shown in Figure 12) that captures the whole problem. In this BN we start with priors that are very favourable to the prosecution case. Thus, we assume a very high probability, 99%, that evidence X was directly linked to the crime.



Evidence X is directly linked to the crime	False		True	
Evidence X belongs to the defendant	False	True	False	True
False	0.2	1.0	1.0	1.0
True	0.8	0.0	0.0	0.0

Figure 12: Crimewatch UK Priors

Looking at the conditional probability table we assume generously that if the owner of X was innocent of involvement then there is an 80% chance he/she would come forward (the other assumptions in the conditional probability table are not controversial).

What we are interested in is how the *prior probability of the evidence X being the defendant's* changes when we enter the fact that no owner comes forward. The prosecution claim is that it becomes almost certain. The key thing to note is that, with these priors, there is already a very low probability (0.4%) that the owner comes forward.

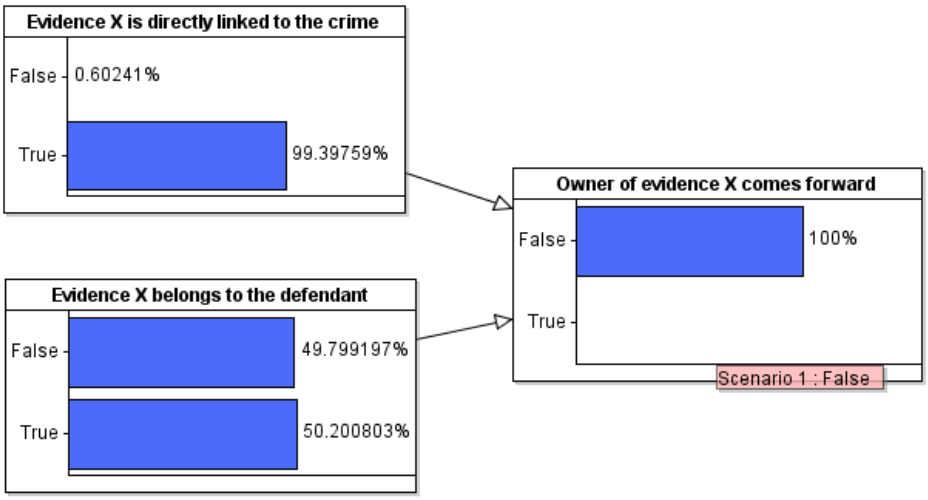


Figure 13 Now we enter the fact 3 (nobody comes forward)

Consequently, when we now enter the fact that nobody comes forward (Figure 13) we see that the impact on the probability that X belongs to the defendant is almost negligible (moving from 50% to 50.2%).

This demonstrates the Crimewatch fallacy and that the evidence of nobody coming forward is effectively worthless despite what the prosecution claims.

In fact, the only scenarios under which the evidence of nobody coming forward has an impact are those that contradict the heart of the prosecution claim. For example, let us assume (Figure 14) that there is only a 50% probability that the evidence X is directly linked to the crime

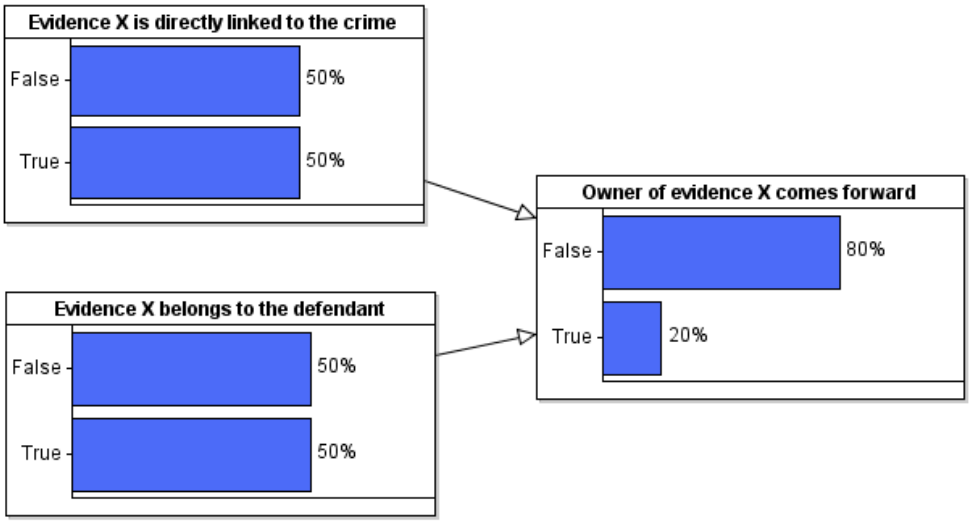


Figure 14 Different priors

Then when we enter the fact that nobody comes forward (Figure 15) the impact on our belief that X is the defendant's is quite significant (though still not conclusive) moving from 50% to 62.5%. But, of course, in this case ***the priors contradict the core of the prosecution case***.

Note that we could, instead of the BN presentation, have presented an equivalent formulaic argument deriving the likelihood ratio of the Crimewatch evidence. This would have shown the likelihood ratio to be close to 1, and hence would also have shown that the utility of the evidence is worthless. However, as has been stressed throughout this paper, the BN presentation proved to be more easily understandable to lawyers.

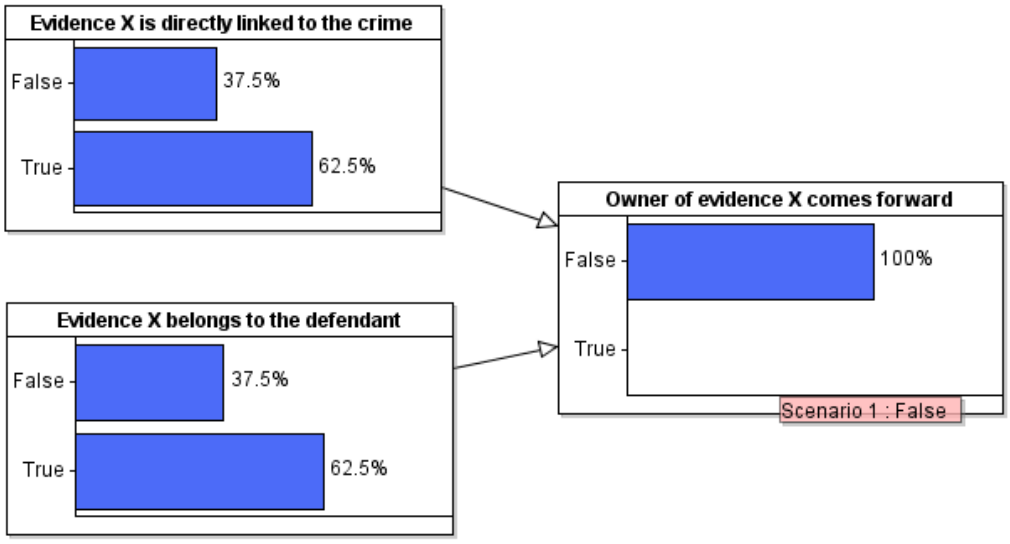


Figure 15 Evidence now makes a difference

5 Conclusions

Despite fairly extensive publicity and many dozens of papers and even books exposing them, probabilistic fallacies continue to proliferate legal reasoning. In this paper we have presented a wide range of fallacies (including one new one) and a new simple conceptual approach to their classification. While many members of the legal profession are aware of the fallacies, they struggle to understand and avoid them. This seems to be largely because they cannot follow Bayes Theorem in its formulaic representation. Instead of continuing the painful struggle to get non-mathematicians to understand mathematics we must recognise that there is an alternative approach that seems to work better.

In simple cases equivalent visual representations of Bayes, such as event trees, enable lawyers and maybe even jurors to fully understand the result of a Bayesian calculation without any of the mathematics or formulas (websites that promote public understanding of probability, such as [42] and [103] are now using such visual techniques extensively). This approach has already been used with considerable effect in real cases. However, it does not scale up. As more pieces of evidence and dependencies are added no first principles argument that the lawyers can fully understand is ever going to be possible. In such cases we have proposed to use

Bayesian networks (BNs). This proposal is not new; indeed, as long ago as 1991 Edwards [37] 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, he was grossly optimistic for two reasons.

1. Even within the community of statisticians interested in legal arguments there has been some lack of awareness that tool support for BNs now makes them easily usable;
2. Acceptance of BNs by members of the legal community requires first an understanding and acceptance of Bayes theorem. For reasons explained in this paper, there have been often insurmountable barriers to such acceptance.

What is new about our proposal is our strategy for addressing the latter.

We feel the strategy presented could feasibly work in both pre-trial evidence evaluation and in court. In pre-trial we envisage a scenario where any evidence could be evaluated independently to eliminate that which is irrelevant, irrational or even irresponsible. This could, for example, radically improve and simplify arguments of admissibility of evidence.

During trial, our proposed approach would mean that the jury and lawyers can focus on the genuinely relevant uncertain information, namely the prior assumptions.

Crucially, 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 calculator to compute a long division. Lay people do not need to understand how the calculator works in order to accept the results of the calculations as being correct to a sufficient level of accuracy. The same must eventually apply to the results of calculations from a BN tool.

We have demonstrated practical examples of our approach in real cases. We recognise that there are significant technical challenges we need to overcome to make the construction of BNs for legal reasoning easier, notably overcoming the constraints of existing BN algorithms and tools that force modellers to specify unnecessary prior probabilities (the work in [102] may provide solutions to some of these issues). We also need to extend this work to more relevant cases and to test our ideas on more lawyers. And there is the difficult issue of who, exactly, would be the most appropriate people to build and use the BN models. However, the greater challenges are cultural. There is clearly a general problem for the statistics community about how to get their voices heard within the legal community. We gain comfort from the fact that similar, seemingly impenetrable, cultural barriers have been breached in other domains. For example, the case of Duckworth-Lewis [36] demonstrated it was possible to gain acceptance of complex mathematical formalisms in the very conservative, non-mathematical environment of test cricket (when a new complex mathematical formalism was used to replace the previous unsatisfactory method for determining fair outcomes of one-day matches). If we can similarly break down the barriers between the mathematics and law communities, then there could be a transformative impact in the way legal evidence is analysed and presented

Because of this, we believe that in 50 years time professionals of all types involved in the legal system will look back in total disbelief that they could have ignored these available techniques of reasoning about evidence for so long.

6 Acknowledgements

This paper evolved from a presentation at 7th International Conference on Forensic Inference and Statistics (ICFIS) Lausanne, 21 August 2008. The trip was funded as part of the EPSRC project DIADEM. We would like to thank the organisers of the conference (especially Franco Tarroni) for their support, and we also acknowledge the valuable advice and insights provided by Colin Aitken, Richard Ashcroft, David Balding, Bill Boyce, Phil Dawid, Itiel Dror, Philip Evans, Ian Evett, Graham Jackson,

Martin Krzywinski, Jennifer Mnookin, Richard Nobles and David Schiff. Finally, we would like to thank one anonymous referee for their very helpful insights.

7 References

[1] (1968). People v. Collins, 438 P. 2d 33 (68 Cal. 2d 319 1968)

[2] (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

[3] (1997). R v Doheny and Adams 1 Crim App R 369.

[4] (2006). R v Richard Bates (Case No: 200503241C3). Mr Justice Fulford, Supreme Court of Judicature Court of Appeal (Criminal Division)

[5] (2007). R. v. George, EWCA Crim 2722.

[6] (2008) R. v Levi Bellfield, http://www.timesonline.co.uk/tol/news/uk/crime/article3397051.ece

[7] (2010). R v T. EWCA Crim 2439 http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf

[8] Aitken, S.G.G., Interpretation of evidence, and sample size determination, in Statistical SCience in the Courtroom, Gastwirth, J.L., Editor. 2000, Springer: New York. p. 1-24

[9] 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.

[10] Aitken, C.G.G. and Taroni, F., Statistics and the evaluation of evidence (2nd Edition). 2004 John Wiley & Sons, Ltd.

[11] Aitken, C. (2008). Interpreting the Value of Evidence. Royal Statistical Society, 9 Sept 2008. http://www.rss.org.uk/PDF/Colin%20Aitken%20-%209%20Sept%2008.pdf

[12] Aitken, C. G. G., et al. (1995). Bayesian belief networks with an application in specific case analysis. Computational Learning and Probabilistic Reasoning. A. Gammerman, John Wiley and Sons Ltd.

[13] Allen, R. J. and Miller, J. S. (1993). "The Common Law Theory of Experts: Deference or Education?" Northwestern Law Review 87: 1131-1147.

[14] Anderson, T., Schum, D. A. and Twining, W. (2005). Analysis of Evidence, Cambridge University Press.

[15] Ancker, J. S., Senathirajaha, Y., Kukafka, R. and Starren, J. B. (2006). "Design Features of Graphs in Health Risk Communication: A Systematic Review " J Am

Med Inform Assoc. 13: 608-618.

[16] Ask, K. and Granhag, P.A., Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure. Journal of Investigative Psychology and Offender Profiling, 2005. 2(1): p. 43-63. http://dx.doi.org/10.1002/jip.19

[17] Balding, D. J. (2005). Weight-of-Evidence for Forensic DNA Profiles, Wiley.

[18] 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.

[19] Balding, D.J. and Donnelly, P., Prosecutor's fallacy and DNA evidence, in Criminal Law Review, 711-722. 1994.

[20] 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.

[21] 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.

[22] 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.

[23] Buckleton, J., C. M. Triggs and S. J. Walsh (2005). Forensic DNA Evidence Interpretation, CRC Press.

[24] 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.

[25] 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

[26] Cowell, R. G. (2003). "Finex: a Probabilistic Expert System for forensic identification." Forensic Science International 134(2): 196-206.

[27] Cowell, R., Lauritzen, S. and Mortera, J. (2006). "Identification and separation of DNA mixtures using peak area information." Forensic Science International 166(1): 28-34.

[28] 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

[29] Darroch, J. (1987). "Probability and criminal trials: some comments prompted by the Splatt trial and the Royal Commission." Professional Statistican 6: 3-7.

[30] 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.

[31] Dawid, A. P. (2002). Bayes's theorem and weighing evidence by juries. In Bayes's Theorem: Proceedings of the British Academy. R. Swinburne. Oxford, Oxford University Press. 113: 71-90.

[32] Dawid, P., Statistics on Trial. Significance, 2005. 2(1): p. 6-8. 10.1111/j.1740-9713.2005.00075.x

[33] Donnelly, P. (2005). "Appealing Statistics." Significance 2(1): 46-48.

[34] Dror, I.E. and Charlton, D., Why experts make errors. Journal of Forensic Identification, 2006. 56(4): p. 600-616.

[35] Dror, I. E. and Rosenthal, R. (2008). "Meta-analytically quantifying the reliability and biasability of forensic experts." Journal of Forensic Sciences 53(4): 900-903.

[36] Duckworth, F. C. (2008). "The Duckworth/Lewis method: an exercise in communicating a mathematical concept to a non-mathematical audience." Mathematics Today (Bulletin of the IMA) 44(5): 212-214.

[37] Edwards, W. (1991). "Influence Diagrams, Bayesian Imperialism, and the Collins case: an appeal to reason." Cardozo Law Review 13: 1025-1079.

[38] Evett, I. W. (1995). "Avoiding the transposed conditional." Science and Justice 35(3): 127-131

[39] Evett, I. W. and Weir, B. S. (1998). Interpreting DNA evidence : statistical genetics for forensic scientists, Sinauer Associates.

[40] 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.

[41] 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

[42] Fenton, N.E. and Neil, M., Avoiding Legal Fallacies in Practice Using Bayesian Networks (Seventh International Conference on Forensic Inference and Statistics. 2008: Lausanne, Switzerland). www.eecs.qmul.ac.uk/~norman/papers/fenton_neil_prob_fallacies_3_July_09.pdf

[43] Fenton, N. E. (2008). Probabilistic and Logical Fallacies in the Prosecution Opening. Expert Report for the Criminal Court in the case of R v Levi Bellfield.

[44] Fenton, N. E. (2009). "Making Sense of Probability: Fallacies, Myths and Puzzles." https://www.dcs.qmul.ac.uk/~norman/papers/probability_puzzles/Making_sense_of_probability.html.

[45] 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.

[46] Fenton, N.E. and Neil, M., Managing Risk in the Modern World: Bayesian Networks and the Applications, 1. 2007, London Mathematical Society, Knowledge Transfer Report. http://www.lms.ac.uk/activities/comp_sci_com/KTR/apps_bayesian_networks.pdf

[47] Fenton, N. and Neil, M. (2010). "Comparing risks of alternative medical diagnosis using Bayesian arguments." Journal of Biomedical Informatics 43: 485-495.

[48] Fienberg, S. E. and Straf, M. L. (1991). "Statistical Evidence in the US Courts: An Appraisal." Journal of the Royal Statistical Society. Series A (Statistics in Society) 154(1): 49-59

[49] Finkelstein, M. O. and Levin, B. A. (2001). Statistics for Lawyers, Springer.

[50] Forrest, A. R. (2003). "Sally Clark - a lesson for us all." Science & Justice: 43: 63-64.

[51] Freckelton, I. and Selby, H. (2005). Expert Evidence: Law, Practice, Procedure and Advocacy. Australia, Lawbook Co.

[52] Giannelli, P. C. (2007). "Confirmation Bias." Criminal Justice 22(3): 60-61

[53] Gigerenzer, G. (2002). Reckoning with Risk: Learning to Live with Uncertainty. London, Penguin Books.

[54] Gilovich, T., Griffin, D., and Kahneman, D., Heuristics and Biases: The Psychology of Intuitive Judgment. . 2002, Cambridge: Cambridge University Press.

[55] Good, I.J., When batterer turns murderer. Nature, 1995. 375: p. 541.

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

[57] Goodwin, W., Linacre, A. and Hadi, S. (2007). An Introduction to Forensic Genetics, John Wiley and Sons.

[58] Green, P. (2002). Letter from the President to the Lord Chancellor regarding the use of statistical evidence in court cases, 23 January 2002. The Royal Statistical Society.

[59] Haigh, J. (2003). Taking Chances: Winning with Probability, Oxford University Press.

[60] Hill, R. (2005). "Reflections on the cot death case." Significance 2(13-15).

[61] Honess, T.M., Levi, M., and Charunan, E.A., Juror competence in process complex information: implications from a simulation of the Maxwell trial. 1998.

[62] Hunter, K., A new direction on DNA, in Criminal Law Review, 478-480, July. 1998.

[63] Huygen, P.E.M., Use of Bayesian Belief Networks in legal reasoning, in 17th BILETA Annual Conference. 2002: Free University, Amsterdam.

[64] 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

[65] Jowett, C., Lies, Damned Lies, and DNA Statistics: DNA Match testing. Bayes'

Theorem, and the Crminal Courts, in Med Sci. Law 41(3), 194-205. 2001.

[66] Jowett, C., Sittin' in the Dock with the Bayes, in New Law Journal Expert Witness Supplement of Feb 16. 2001.

[67] Joyce, H. (2002). "Beyond reasonable doubt." MathPlus 21, http://plus.maths.org/issue21/features/clark/.

[68] Kadane, J. B. and Schum, D. A. (1996). A Probabilistic Analysis of the Sacco and Vanzetti Evidence, John Wiley & Sons

[69] Kahneman, D., Slovic, P., and Tversky, A., Judgement Under Uncertainty: Heuristics and Biases. 1982: New York: Cambridge University Press.

[70] Kaye, D. H. (2001). "Two Fallacies About DNA Databanks for Law Enforcement." Brooklyn Law Review 67(1): 179-206.

[71] Kaye, D. H. (2007). "Revisiting 'Dreyfus': A More Complete Account of a Trial by Mathematics." Minnesota Law Review 91(3): 825-835.

[72] 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.

[73] Koehler, J. J. (1993). "The Base Rate Fallacy Myth." Psycoloquy 4(49).

[74] Koehler, J. J. (1993). "Error and Exaggeration in the Presentation of DNA Evidence at Trial." Jurimetrics 34: 21-39

[75] Krane, D. E., et al. (2008). "Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation." J Forensic Sci 53(4): 1006-1007

[76] Kwan, M., R. Overill, et al. (2011). Sensitivity Analysis of Digital Forensic Reasoning in Bayesian Network Models. Advances in Digital Forensics VII, Proc. 7th Annual IFIP WG 11.9 International Conference on Digital Forensics. Orlando, Florida, USA, 30 January - 2 February 2011, Springer

[77] 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. Statis. Soc. B, 50, No 2, pp 157-224.

[78] Leader, L. and Schofield, D., Madness in the Method? Potential Pitfalls in Handling Expert Evidence Journal of Personal Injury Law, 2006. 6(1): p. 68 - 86.

[79] Lucas, J.R., The Concept of Probability. 1970, Oxford: Clarendon Press.

[80] 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

[81] Matthews, R. (1995). "The Interrogator's Fallacy." Bulletin of the Institute of Mathematics and its Applications 31 3-5

[82] Matthews, R. (1997). "Tipping the scales of justice." New Scientist (2112): 18-19.

[83] 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.

[84] Murphy, P. (2003). Evidence, Proof, and Facts: A Book of Sources, Oxford University Press.

[85] Nickerson, R.S., Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 1998. 2: p. 175-220.

[86] Nobles, R. and Schiff, D. (2005). "Misleading statistics within criminal trials: the Sally Clark case." Significance 2(1): 17-19.

[87] Overill, R E and Silomon, J A M, (2011) Six Simple Schemata for Approximating Bayesian Belief Networks, in Proc. Cyberforensics-2011, Glasgow, UK, 27-28 June 2011

[88] Pearl, J. (1988). Probabilistic reasoning in intelligent systems. Palo Alto, CA, Morgan Kaufmann

[89] Percy, D. F. (2005). "The power of Bayes." Mathematics Today (Bulletin of the IMA) 41(4): 122-125.

[90] Redmeyne, M., DNA evidence, probability and the courts, in Criminal Law Review, 464. 1995.

[91] Reed, C. and Rowe, G. (2007). "A pluralist approach to argument diagramming." Law, Probability & Risk 6: 59-85.

[92] Redmayne, M., P. Roberts, C. Aitken and G. Jackson (2011). "Forensic Science Evidence in Question." Criminal Law Review (5): 347-356.

[93] Robertson, B., G. A. Vignaux and C. E. H. Berger (2011). "Extending the confusion about Bayes." The Modern Law Review 74(3): 444-455.

[94] Robertson, B. and Vignaux, T., Bayes' Theorem in the Court of Appeal. The Criminal Lawyer, 1997: p. 4-5.

[95] 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.

[96] Robertson, B. and Vignaux, T., Explaining Evidence Logically. New Law Journal, 1998. 148(6826): p. 159-162. http://www.mcs.vuw.ac.nz/

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

[98] Rossman, A. J. and Short, T. H. (1995). "Conditional Probability and Education Reform: Are They Compatible?" Journal of Statistics Education 3(2)

[99] Rossmo, D.K., Criminal investigative failures: avoiding the pitfalls. FBI Law Enforcement Bulletin, Oct 2006., 2006. http://findarticles.com/p/articles/mi_m2194/is_10_75/ai_n18765135

[100] 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.

[101] Sox, H. C., Blatt, M. A., Higgins, M. C. and Marton, K. I. (2007). Medical Decision Making, ACP Press.

[102] Smith, J. Q. and Anderson, P. E. (2008). "Conditional independence and chain event graphs." Artif. Intell. 172(1): 42-68.

[103] Spiegelhalter, D. (2009). "Understanding Uncertainty." http://understandinguncertainty.org/.

[104] Steventon, B., Statistical Evidence and The Courts -- Recent Developments. Journal of Criminal Law, 1998. 67: p. 176-184.

[105] Taroni, F. and Biedermann, A., Inadequacies of posterior probabilities for the assessment of scientific evidence, in Law Probablity and Risk 4: 89-114. 2005.

[106] Taroni, F., Aitken, C., Garbolino, P., and Biedermann, A., Bayesian Networks and Probabilistic Inference in Forensic Science 2006: John Wiley & Sons.

[107] Taroni, F., Bozza, S., and Biedermann, A., Two items of evidence, no putative source: an inference problem in forensic intelligence. J Forensic Sci, 2006. 51(6): p. 1350-61.

[108] 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

[109] Thompson, W.C., DNA Evidence in the O.J. Simpson Trial. Colorado Law Review, 1996. 67(4): p. 827-857.

[110] 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.

[111] 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.

[112] Tillers, P. (2007). "Introduction: visualizing evidence and inference in legal settings" Law, Probability and Risk 6: 1-4.

[113] Tribe, L. H. (1971). "Trial by Mathematics." Harvard Law Review 84: 1329.

[114] Triggs, C.M. and Buckleton, J.S., Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. Law, Probability and Risk, 2004. 3: p. 73-82.

[115] Verheij, B. (2007). "Argumentation Support Software: Boxes-and-Arrows and Beyond." Law, Probability & Risk 6: 187-208.

[116] Wigmore, J. H. (1913). "The problem of proof." Illinois Law Journal 8(2): 77-103.

[117] Woffinden B., Miscarriages of Justice, Hodder and Stoughton, 1988