Bayes and the Law

Norman Fenton, Martin Neil, and Daniel Berger

School of Electronic Engineering and Computer Science, Queen Mary University London, London E1 4NS, United Kingdom; email: n.fenton@qmul.ac.uk

Annu. Rev. Stat. Appl. 2016. 3:51–77

First published online as a Review in Advance on March 9, 2016

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

This article's doi:
10.1146/annurev-statistics-041715-033428

Keywords

Bayes, Bayesian networks, statistics in court, legal arguments

Abstract

Although the use of statistics in legal proceedings has considerably grown in the last 40 years, primarily classical statistical methods rather than Bayesian methods 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 article 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 reasons include misconceptions by the legal community about Bayes' theorem, overreliance on the use of the likelihood ratio, and the lack of adoption of modern computational methods. We argue that Bayesian networks, which automatically produce the necessary Bayesian calculations, provide an opportunity to address most concerns about using Bayes in the law.

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 (Finkelstein 2009, Gastwirth 2000, Kadane 2008, Koehler 1992, Vosk & Emery 2014). The Howland case of 1867 is the earliest reported case in which detailed statistical analysis was presented as evidence (see Meier & Zabell 1980 for details). In this case Benjamin Peirce attempted to show that a contested signature on a will had been traced from the genuine signature, by arguing that their agreement in all 30 downstrokes was extremely improbable under a binomial model. Meier & Zabell (1980) highlight in Peirce's evidence the use and abuse of the product rule for multiplying probabilities of independent events. In any case, the court found a technical excuse not to use the evidence.

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, it turned out to be fundamentally flawed (Kaye 2007). Not until 1968 was there another well-documented case (*People v. Collins* 1968) 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 characterized by two errors: (*a*) 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 (*b*) it implied that the low probability from the calculation in (*a*) 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. Clark* (Forrest 2003, Hill 2005), *R v. George* (Fenton et al. 2013a), and *R v. de Berk* (Meester et al. 2007). Although the original bad use of statistics in each case (presented by forensic or medical expert witnesses without statistical training) was exposed through good use of statistics on appeal, it is the bad use of statistics that 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 use of statistics in legal proceedings has considerably grown in the last 40 years, its use in the courtroom has been restricted mostly to a small class of cases in which 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 (Royal Statistical Society 2015, Ziliak & McCloskey 2008), specifically in the context of legal and forensic evidence (Finkelstein 2009, Vosk & Emery 2014). In particular,

- The use of *p*-values can also lead to the prosecutor's fallacy because a *p*-value (which says something about the probability of observing the evidence given a hypothesis) is often wrongly interpreted as being the same as the probability of the hypothesis given the evidence (Gastwirth 2000).
- Confidence intervals are almost invariably misinterpreted because their proper definition is both complex and counterintuitive (indeed, it is not properly understood even by many trained statisticians) (Fenton & Neil 2012).

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, even though the Bayesian approach is especially well suited for a broad range of legal reasoning (Fienberg & Finkelstein 1996).

Although the natural resistance within the legal profession to a new statistical approach is one reason why Bayes has made only minimal impact to date, 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 (Faigman & Baglioni 1988, Fienberg 2011, Tillers & Green 1988, Tribe,

1971) and in more general policy decision-making (Fienberg & Finkelstein 1996). However, we contend that there is another rarely articulated but now dominant reason for its continued limited use: Most examples of the Bayesian approach have oversimplified the underlying legal arguments being modeled to ensure the computations can be carried out manually. Whereas such an approach may have been necessary 20 years ago, it is no longer necessary with the advent of easy-to-use and efficient Bayesian network (BN) algorithms and tools (Fenton & Neil 2012). This continuing oversimplification is an unnecessary and debilitating burden.

The article is structured as follows: Section 2 describes the basics of Bayes (and BNs) for legal reasoning, covering the core notion of the likelihood ratio (LR). Section 3 reviews the actual use of Bayes in the law. Section 4 describes why Bayes has made such a minimal impact, and Section 5 explains what can be done to ensure it gets the exploitation it deserves. We draw on primarily English-language sources and most cases cited are in English-speaking countries (normally with a jury system). Most are from the United Kingdom and the United States. Our review focuses on the use of Bayes and hence does not attempt to describe the broader use of statistics in the law, which is well covered elsewhere (Aitken & Taroni 2004, Fienberg 1989, Zeisel & Kaye 1997).

2. BASICS OF BAYES FOR LEGAL REASONING

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

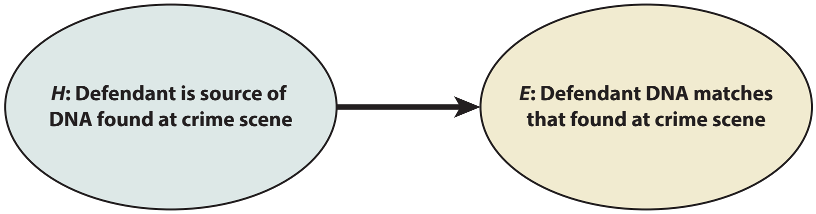
1. A hypothesis is a statement (typically Boolean) whose truth value we seek to determine but is generally unknown—and which may never be known with certainty. Examples include:
   a. "Defendant is innocent of the crime charged." (This is an example of an offense-level hypothesis, also called the ultimate hypothesis, because in many criminal cases it is ultimately the only hypothesis we are really interested in.)
   b. "Defendant was the source of DNA found at the crime scene." [This is an example of what is often referred to as a source-level hypothesis (Cook et al. 1998a).]
2. The alternative hypothesis is a statement that is the negation of a hypothesis.
3. 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 example in **Figure 1**, where we assume that (*a*) the evidence E is a DNA trace found at the scene of the crime (for simplicity we assume the crime was committed on an island with 10,000 people who therefore represent the entire set of possible suspects), and (*b*) the defendant was arrested and some of his DNA was sampled and analyzed.

The direction of the causal structure makes sense here because H being true (resp. false) can cause E to be true (resp. false), whereas 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. This latter type of inference is central to all legal reasoning because, 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



False	0.999
True	0.001

H: Defendant is source of DNA found at crime scene	False	True
False	0.999	0.0
True	0.001	1.0

Figure 1

Causal view of evidence, with prior probabilities shown in tables. This is a simple example of a Bayesian network. Abbreviations: *E*, evidence; *H*, hypothesis.

conditional probability of E given H, which we write as $P(E \mid H)$. Bayes' theorem provides the formula for updating our prior belief about H in light of observing E to arrive at a posterior belief about H, which we write as $P(H \mid E)$. In other words, Bayes calculates $P(H \mid E)$ in terms of $P(H)$ and $P(E \mid H)$. Specifically,

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

The probability table for H captures our knowledge that the defendant is one of 10,000 people who could have been the source of the DNA. The probability table for $E \mid H$ captures the following two assumptions. (*a*) The probability of correctly matching a DNA trace is 1 (so there is no chance of a false-negative DNA match). This probability $P(E \mid H)$ is called the prosecution likelihood for the evidence E. (*b*) The probability of a match in a person who did not leave their DNA at the scene (the random DNA match probability) is 1 in 1,000. This probability $P(E \mid \text{not } H)$ is called the defense 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 that the defendant is the source of the DNA at the crime scene moves from a prior of 1 in 10,000 to a posterior of 9%. Alternatively, our belief that the defendant is not the source of the DNA moves from a prior of 99.99% to a posterior of 91%.

Note that the posterior probability of the defendant not being the source of the DNA is very different from the random match probability (RMP) of 1 in 1,000. The incorrect assumption that the two probabilities $P(\text{not } H \mid E)$ and $P(E \mid \text{not } H)$ are the same characterizes 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 the source of this evidence is one in a thousand," when actually it is 91%. This simple fallacy of probabilistic reasoning has affected numerous cases (Balding & Donnelly 1994, Fenton & Neil 2011), 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 defense argues that because $P(\text{not } H \mid E)$ 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 counterintuitive (Casscells & Graboys 1978, Cosmides & Tooby 1996). 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' theorem tells us the posterior odds of H are the prior odds of H times the LR, where the LR is simply the prosecution likelihood of E divided by the defense likelihood of E, i.e.,

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

In the example in **Figure 1** the prosecution likelihood for the DNA match evidence is 1, while the defense likelihood is 1/1,000. So the LR is 1,000. This means that whatever the prior odds were in favor of the prosecution hypothesis, 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), whereas 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 because it leaves the posterior probability 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 1,000 meant the evidence was highly probative in favor 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 1,000 results in posterior odds of 1,000 to 1 in favor 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 (9,999 to 1 against H)—then the same LR of 1,000 results in posterior odds that still strongly favor the defense 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 case assessment and interpretation method (Cook et al. 1998b, Jackson et al. 2013) has the LR at its core. By forcing expert witnesses to consider both the prosecution and defense likelihood of their evidence—instead of just one or the other—it also avoids most common cases of the prosecutor's fallacy.

Although 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 three unknown hypotheses and two pieces of evidence, with the causal links shown in **Figure 2** (Dawid & Mortera 1998, Fenton et al. 2014), once we take account of the possibility of different types of DNA collection and testing errors (Foreman et al. 2003, Koehler 1993a, Thompson et al. 2003).

Moreover, further crucial hypotheses not shown in **Figure 2** (a full version of the model is provided in the supplemental material, follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**) include "Defendant was at the scene of the crime" and, the ultimate hypothesis, "Defendant committed the crime." These are omitted here only because, whereas the law might accept a statistical or forensic expert reasoning probabilistically about the source of the forensic evidence, it is presupposed that any probabilistic reasoning about the ultimate hypothesis is the province of the trier of fact, i.e., the judge and/or the jury.

Supplemental Material

With or without the additional hypotheses, **Figure 2** is an example of a 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

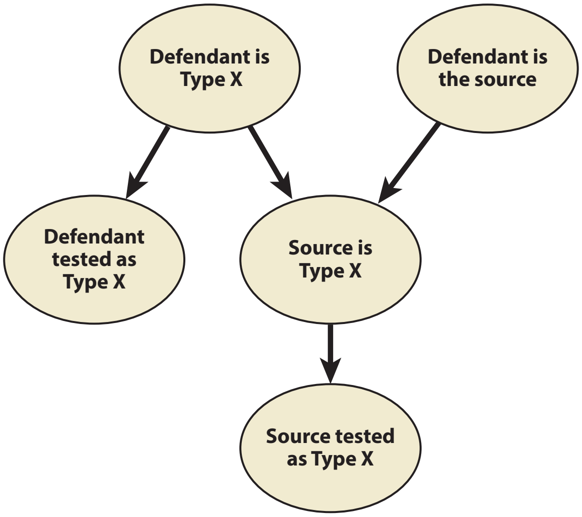


Figure 2

Bayesian network for DNA match evidence. Each node has states true or false. Figure adapted from Fenton et al. (2014) with permission.

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 1980s 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 BNs (Lauritzen & Spiegelhalter 1988, Pearl 1988). These algorithms have subsequently been incorporated into widely available graphical toolkits that enable users without any statistical knowledge to build and run BN models (Fenton & Neil 2012). Moreover, further algorithmic breakthroughs have enabled us to model an even broader class of BNs, namely those including numeric nodes with arbitrary statistical distributions (Neil et al. 2007). These breakthroughs are potentially crucial for modeling legal arguments.

To show the simple power of BNs, we consider an example of their ability to address claims made in several works, such as that of Cohen (1977), that probabilistic reasoning is inconsistent with legal reasoning because it leads to paradoxes. These claims have been challenged by Bayesians such as Dawid (1987), who paraphrased Cohen's classic conjunction paradox as follows (p. 92):

> Suppose that A and B are independent disputed issues of fact, and that the plaintiff produces two independent witnesses: W_1 attesting to A's occurrence, and W_2 to B's occurrence. The witnesses are both regarded as 70% reliable. We infer, on the witnesses' testimony, that $Pr(A) = Pr(B) = 0.7$ whence $Pr(C) = 0.49$. On Cohen's analysis the suit succeeds; on a 'Pascalian' analysis, it fails. But each witness clearly offered positive support for the plaintiff. How can it be that their combined support undermines his case? Does this paradox not show that Cohen's analysis is to be preferred?

Dawid provides a first-principles Bayesian argument to show that there is no paradox. The argument, which for nonmathematicians would be very hard to understand, is based on exposing

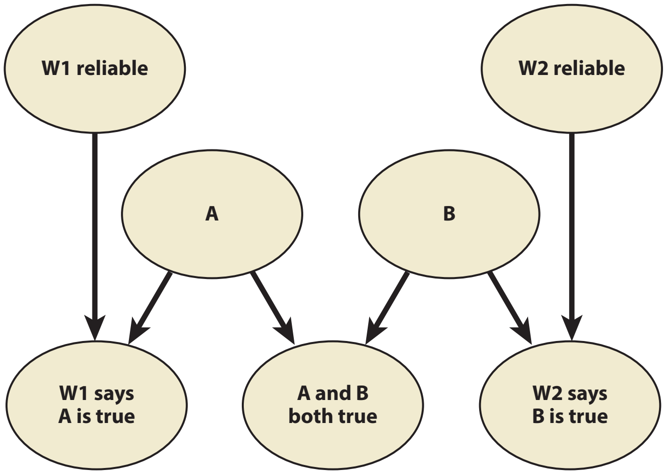


Figure 3

Bayesian network model for the conjunction paradox. A and B are independent disputed issues of fact. W1 is a witness attesting to fact A; W2 is a witness attesting to fact B.

hidden assumptions, notably that what the witness actually says about the fact is dependent on both whether the fact is true or not and whether the witness is reliable or not. No BN model is mentioned, but the underlying BN model is the one shown in **Figure 3**. The idea that the interpretation of any witness evidence must take account of the accuracy of the witness is a fundamental idiom (the evidence accuracy idiom) proposed in the general BN modeling approach of Fenton et al. (2013b).

Simply running this model with different prior conditional probability values [for example, using the free software AgenaRisk (Agena Ltd 2016) or GeNIe Modeler (BayesFusion 2016)] produces all the necessary results. For example, if we assume

- the priors for the nodes W1, W2 are 70:30 for True/False,
- the priors for nodes A, B are 50:50 True/False, and
- the conditional probability for the nodes representing what the witnesses say are as shown in **Table 1**,

then the result of running the model pre- and postobservations of the witness evidence is shown in **Figure 4*a*,*b***, respectively. This replicates all the complex calculations and formulas by Dawid (1987) and makes the argument much simpler and easier to understand.

Moreover, in contrast to the argument by Dawid (1987), the BN approach is also scalable, as explained in the supplemental material, in the sense that it can easily incorporate further assumptions that would make the Bayesian updating impossible to do manually. The supplemental material provides further detailed examples of the use of BNs.

Supplemental Material

Table 1 Conditional probability table for the nodes representing what the witnesses say

Witness reliable	True		False	
Fact	True	False	True	False
Witness says face is True	1	0	0.5	0.5
Witness says face is False	0	1	0.5	0.5

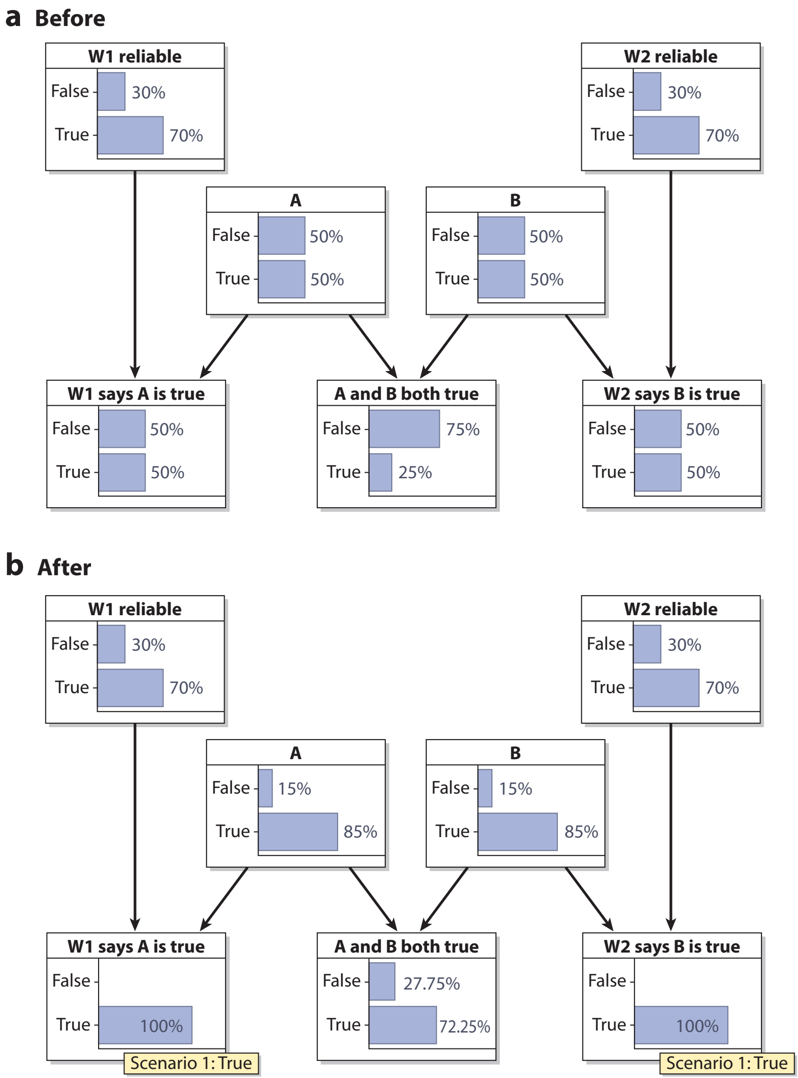


Figure 4

Running the model before and after observing the witness evidence. A and B are independent disputed issues of fact. W1 is a witness attesting to fact A; W2 is a witness attesting to fact B. (*a*) Marginal probabilities before evidence is observed. Note that the probability "A and B both true" is 25%. (*b*) Posterior probabilities after entering the witnesses' evidence. Note that the probability "A and B both true" is over 70%, hence avoiding Cohen's paradox.

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 behavior (Section 3.1), determining the extent to which trait evidence helps with identification (Section 3.2), using forensic evidence to infer the cause of effects (Section 3.3), combining multiple pieces of statistical evidence (Section 3.4), and combining multiple pieces of diverse evidence (Section 3.5). Additionally, Section 3.6 considers cases in which Bayesian solutions have been used indirectly or retrospectively.

3.1. Hypothesis Testing of Specific Dubious Behavior

The vast majority of reported cases of explicit use of statistics in legal proceedings fall under this classification, which in the supplemental material we further subclassify into (*a*) any form of class discrimination and bias, (*b*) any form of fraud/cheating, and (*c*) possession of illegal materials/substances.

The review of such cases in the supplemental material shows that depressingly few cases 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 Fienberg & Finkelstein 1996) 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 & 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.

Supplemental Material

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 behavior can be rejected at an appropriate level of significance. This use of classical statistical hypothesis testing occurs despite the known problems of interpreting the resulting *p*-values and confidence intervals, which the use of the Bayesian approach to hypothesis testing avoids (Fenton & Neil 2012, Press 2002). The potential for misinterpretation is enormous (Vosk & Emery 2014). For example, in *U.S. ex. Rel. DiGiacomo v. Franzen* (1982) a forensic expert quoting a *p*-value from a study published in the literature interpreted it as the probability that each of the head hairs found on the victim were not the defendant's. Sadly, there is a lack of awareness among statisticians that modern tools (such as those discussed in Section 2) make it possible to easily perform the necessary analysis for Bayesian hypothesis testing.

3.2. Determining the 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 Fenton et al. 2014) is used. Traits range from forensic physical features such as DNA (coming from different parts of the body), fingerprints, or footprints, to more basic features such as skin color, height, hair color, or even name. But traits can also refer to nonhuman artifacts (and their features) related to a crime or crime scene, such as clothing and other possessions, cars, weapons, glass, and soil.

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 rigor and validity of these estimates—with the errors made in the case of *People v. Collins* (1968) still prominent in many lawyers' minds.

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

Its use in presenting trait evidence (as recommended, for example, in Aitken & Taroni 2004; Balding 2004; Collins & Morton 1994; Dawid & Mortera 2008; Evett & Weir 1998; Evett et al. 2000; Jackson et al. 2006, 2013) has been most extensive in respect to DNA evidence for determining paternity (Aitken & Taroni 2004, Collins & Morton 1994, Fung 2003, Marshall et al. 1998). For example (see Univ. North Carol. 2009), 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, nongenetic probability of paternity, citing cases *Cole v. Cole* (1985), *State v. Jackson* (1987), and *Brown v. Smith* (2000). In contrast to the 50% prior recommended by North Carolina, in *Plemel v. Walter* (1987; discussed in Fienberg & Finkelstein 1996) the need to present results against a range of priors was recommended; indeed, this was also recommended in the criminal case *State of New Jersey v. J.M. Spann* (1993).

Outside of paternity testing the primary use of the LR in presenting DNA evidence has been to expose the prosecutor's fallacy in the original statistical presentation of an RMP. Notable examples of this scenario in the United Kingdom are the rape cases of *R v. Deen* (1994) and *R v. Alan James Doheny and Gary Adams* (1996) (both of which are discussed in Robertson & Vignaux 1997), as well as *R v. Adams* (1996, 1998) (discussed in Donnelly 2005). 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 (Thompson et al. 2001, see also the supplemental material).

Supplemental Material

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 in which the match arises from a database search (such as *People v. Puckett*, 2009; discussed in Roth 2010). In such cases Bayesian reasoning can again avoid errors (Balding & Donnelly 1996, Meester & Sjerps 2003).

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, fibers, and soil matches, and articles now promote its use in fingerprint evidence (Alberink et al. 2014, Nuemann et al. 2011). The most impressive well-publicized success concerns its use in relation to firearm discharge residue (FDR) in *R v. George* (2007). 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. Fenton et al. (2013a) discuss in detail the appeal court ruling in this case (along with our concerns about the oversimplistic use of the LR, which we return to in Section 4).

However, although 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 *R v. T* (2010), 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 it was recognized that Bayes was used for this purpose in the 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 LRs should not be used.

Given its potential to change the way forensic experts analyze and present evidence in court, the ruling has been criticized in numerous papers (Aitken et al. 2011, Berger et al. 2011, Morrison

2012, Nordgaard et al. 2012, Redmayne et al. 2011, Robertson et al. 2011, Sjerps & Berger 2012). These papers recognize that there were weaknesses in the way the expert presented the probabilistic evidence (in particular not making clear that LRs for different aspects of the evidence were multiplied together to arrive at a composite LR), but nevertheless express deep concern about the implications for future presentations of forensic evidence by experts. The papers also recognize positive features in the ruling (notably that experts should provide full transparency in their reports and calculations), but they compellingly argue why the main recommendations stated above are problematic. Unfortunately, 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.

Although most reported cases of the prosecutor's fallacy relate to poor presentation of trait evidence, the prosecutor's fallacy occurs in other ways. In *R v. Clark*, Sally Clark was convicted of the murder of her two young children who had died one year apart (Forrest 2003, Nobles & Schiff 2005). 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 these experts also used Bayes to explain the impact of failing to compare the prior probability of SIDS with the (also small) probability of double murder (Dawid 2005, Hill 2005). Clark was freed on appeal in 2003.

3.3. Forensic Evidence to Infer the Cause of Effects

While most uses of Bayesian reasoning with forensic evidence have been in the context of matching/identification as explained in Section 3.2, it is often the case that forensic evidence is used to infer the causes of effects (which, as is made clear in Dawid et al. 2015, is treated differently from the effects of causes). For example, items such as blood spatter are used to infer time of death and the cause of death (James et al. 2005), whereas a range of medical evidence is used in cases such as SIDS deaths to determine cause of death. Dawid et al. (2015) explain why Bayesian methods are crucial for interpreting this type of evidence.

3.4. Combining Multiple Pieces of Statistical Evidence

When multiple pieces of statistical evidence (such as trait evidence) are involved in a case, there is a need for careful probabilistic analysis to take account of potential dependencies between different pieces of evidence. Bayes is ideally suited to such an analysis, but sadly the failure to recognize this has led to many well-publicized 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 Fienberg & Finkelstein (1996) refer 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 (Meester et al. 2007) 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 at any given period of time. The extremely low probability P(evidence | 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 precisely on these concerns) is that of Benjamin Geen, who was also a nurse convicted of multiple murders and attempted murder (Gill 2014).

Fienberg & Kaye (1991) describe a number of earlier cases and note that, although Bayes is especially well suited because $P(H \mid E)$ increases as more similar events are included in the evidence, in none of them was a Bayesian perspective used.

3.5. Combining Multiple Pieces of Diverse Evidence

The idea that different pieces of (possibly competing) evidence about a hypothesis H are combined 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. The first is 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 reason (not totally unrelated to the first) is due to the spectacular failure in one well-publicized case for which Bayes was indeed used to combine diverse competing evidence.

The case was *R v. Adams* (1996, 1998), referred to above in connection to the misleading presentation of DNA evidence. This was a rape case (discussed in detail in Donnelly 2005) in which the only prosecution evidence was that the defendant's DNA matched that of a swab sample taken from the victim. The defense 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 emphasized the very low RMP (1 in 200 million) of the DNA evidence. The defense argued that if statistical evidence was to be used in connection with the DNA evidence, it should also be used in combination with the defense evidence and that Bayes' theorem was the only rational method for doing this. The defense called a Bayesian expert, Professor Peter Donnelly, who explained how, with Bayes, the posterior probability of guilt was much lower when the defense evidence was incorporated. The appeal rested on whether the judge misdirected the jury about the evidence in relation to the use of Bayes and left the jury unguided as to how that theorem could be used to properly assess the statistical and nonstatistical 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 that

> 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. (*R v. Adams 1996*)

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 was 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.

Although the subsequent *R v. T* ruling in 2010 dealt a devastating blow to 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. 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.

3.6. Cases In Which Bayesian Solutions Have Been Used Indirectly or Retrospectively

While there is a shortage of well-publicized 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) (Aitken & Taroni 2004, Balding 2004, Foreman et al. 2003, Good 1995, Kadane 2008, Lucy & Schaffer 2007, Redmayne 1995, Schneps & Colmez 2013, Taroni et al. 2014). Certain cases have, unsurprisingly, attracted numerous retrospective Bayesian analyses, notably the case of *People v. Collins* (1968; discussed in Dawid 2005, Edwards 1991, Finkelstein & Fairley 1970), the O.J. Simpson case (discussed in Good 1995, Thompson 1996), and the Nicola Sacco and Bartolomeo Vanzetti case (discussed in Fenton et al. 2013b, Hepler et al. 2007, Kadane & Schum 1996).

Bayes has also played an indirect (normally unreported) role in many cases. In our own experience (and 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 publicized. Rare exceptions include Kadane (1990), who describes his proposed Bayesian testimony in an age discrimination case (settled before trial); some of the works we have been able to make public retrospectively are *R v. Bellfield* (2007–2008) (Fenton & Neil 2011) and *B. v. National Health Service* (2005) (Fenton & Neil 2010).

4. WHY BAYES HAS HAD MINIMAL IMPACT ON THE LAW

The review in Section 3 leaves little doubt that the impact of Bayes on the law has been minimal. To understand how minimal, it is illustrative to consider the 350-page National Research Council report "Strengthening Forensic Science in the United States: A Path Forward" (National Research Council 2009). This important committee report has had a major impact in the United States with regard to the lack of data and statistical support for many forms of forensic evidence. However, it does not contain a single mention of Bayes, and it contains just one footnote reference to LRs. In contrast, in the United Kingdom a number of published guidelines recommend the extensive use of LRs for forensic evidence presentation (Jackson et al. 2013, Puch-Solis et al. 2012). However, the use of such methods is actually likely to decrease following rulings such as *R v. T*.

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 in the courtroom. Indeed, the role of probability—and Bayes in particular—was dealt another devastating and surprising blow in a 2013 UK appeal court case ruling (*Nulty & Ors v.*

Milton Keynes Borough Council 2013; discussed in Spiegelhalter 2013). The case was a civil dispute about the cause of a fire. The appeal court rejected the approach whereby the least unlikely of several possible causes of a fire (namely a discarded cigarette) was assumed to be the cause. In making this ruling, the Court effectively argued 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 overformulaic 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, Fenton & Neil 2012), supporters of it often point to Tribe's (1971) highly influential paper, which was written as a criticism of the prosecutor's presentation in *People v. Collins* (1968). While Tribe's paper did not stoop to the level of the "no such thing as probability" argument, it is especially skeptical of the potential use of Bayes' theorem because it identifies the following concerns (Berger 2014, Fienberg & Finkelstein 1996) that are especially pertinent for Bayes:

- An accurate and/or nonoverpowering prior cannot be devised.
- In using statistical evidence to formulate priors, jurors might use it twice in reaching a posterior.
- Not all evidence can be considered or valued in probabilistic terms.
- No probability value can ever be reconciled with "beyond a reasonable doubt."
- Owing to the complexity of cases and nonsequential nature of evidence presentation, any application of Bayes would be too cumbersome for a jury to use effectively and efficiently.
- Probabilistic reasoning is not compatible with the law, for policy reasons. In particular, 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 by Edwards (1991) and Koehler (1992), and more recently by Berger (2014) and Tillers & Gottfried (2007), the arguments against are far less well known among legal professionals than those in favor.

4.2. Likelihood Ratio Models Are Inevitably Oversimplified

We believe that much of the recent legal resistance to Bayes is due to confusion, misunderstanding, oversimplification, and overemphasis on the role of the LR. The issues (dealt with in depth in the recent papers Fenton 2014; Fenton et al. 2013a, 2014) are summarized below.

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, as explained in Section 2 (where **Figure 2** rather than **Figure 1** is the correct model). Even if we completely ignore much of the context (including issues of quality of trace sample, reliability of trace sample collection/storage, possible contamination, and potential testing errors), the LR may still be difficult or even impossible to elicit because somehow we have to factor into the hypothesis H_d (defendant is not the source of the DNA trace) every person other than the defendant who could have been the source (potentially every other person in the world) (Balding 2004, Nordgaard et al. 2012). For example,

$P(E \mid H_r)$ is much higher than $P(E \mid H_u)$, where H_r is the hypothesis "a close relative of the defendant is the source of the trace" and H_u is the hypothesis "a totally unrelated person is the source."

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 criticized; see Balding 2004) 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.

It is not just the hypotheses that may need to be decomposed. In practice, even an apparently single piece of evidence E actually comprises multiple, separate pieces of evidence, and it is only when the likelihoods of these separate pieces of evidence are considered that correct conclusions about probative value of the evidence can be made. This is illustrated in Example 1. In light of these observations it is perhaps not surprising that Wagenaar (1988) had difficulty getting the judge to interpret his LR descriptions.

Example 1: Consider the evidence E: "tiny matching DNA trace found." Suppose that the DNA trace has a profile with an RMP of 1/100 (such relatively "high" match probabilities are common in low-template samples). Assuming H_p and H_d are the prosecution and defense hypotheses, respectively, it would be typical to assume that

$$\mathrm{P}(E \mid H_p) = 1$$

and that

$$\mathrm{P}(E \mid H_d) = 1/100,$$

leading to a LR of 100, thus indicating quite strong support for the prosecution hypothesis. However, the evidence E actually comprises two separate pieces of evidence:

- $E1$: tiny DNA trace found
- $E2$: DNA trace found matches defendant

In particular, this makes clear the relevance of finding only a tiny trace of DNA when larger amounts would be expected to have been left by the person who committed the crime. So actually $P(E \mid H_p)$ will be much smaller than 1, because we would expect substantial amounts of DNA to be found rather than just a tiny trace. To elicit all the necessary individual likelihood values, and to carry out the correct Bayesian calculations needed for the overall LR in situations such as this, we again need to turn to BNs as shown in **Figure 5**.

The oversimplistic model **Figure 5*a*** fails to capture the relevance of finding only a tiny trace of DNA. If the defendant were guilty, it is expected that the investigator would have found significant traces of DNA. The significance of the tiny trace is properly captured by separating out $E1$ in the second model **Figure 5*b***. A reasonable conditional probability table for $E1$ is shown in **Table 2**. The conditional probability table for $E2$ shown in **Table 3** uses the same RMP information as was used in the oversimplified model.

Calculating the overall LR manually in this case is much more complex, so we go directly to the result of running the model in a BN tool with $E2$ set as true (and the prior odds of guilt set at 50:50 again). This is shown in **Figure 6**. The LR is just the probability of guilty divided by the probability of not guilty,

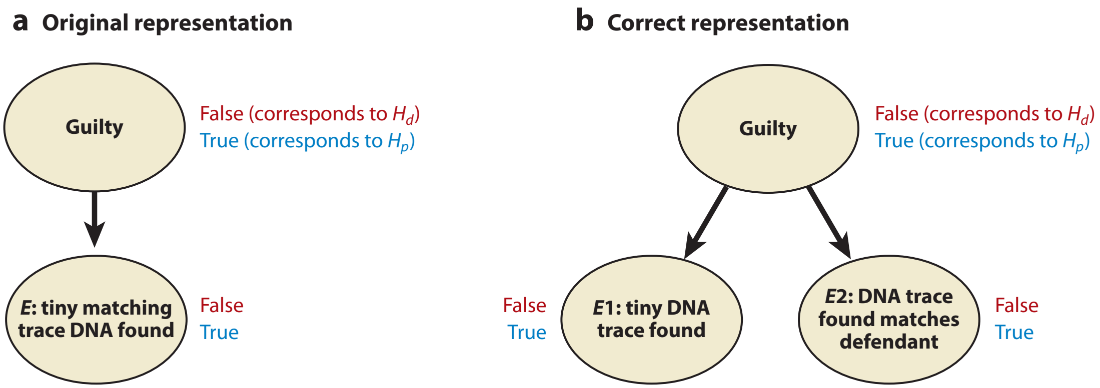


Figure 5

Modeling complex evidence in a Bayesian network. (*a*) Original representation. (*b*) Improved representation. Abbreviations: E, evidence; H_d, hypothesis of defense; H_p, hypothesis of prosecution.

which is 0.2. So the evidence supports the defense hypothesis rather than the prosecution hypothesis. This example also indicates the importance of taking account of absence of evidence.

4.3. The Exhaustiveness and Mutual Exclusivity of Hypotheses Is Not Respected in All Uses of the Likelihood Ratio

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 (defense hypothesis). But when we say the evidence E "supports the hypothesis H_p," we mean 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 on 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") (Fenton et al. 2013a). If H_p and H_d are mutually exclusive (as is generally assumed) but not exhaustive (as is often the case), then the LR tells us absolutely nothing about the relationship between the posterior and the priors of the individual hypotheses, as explained in Example 2.

Example 2: In the game Clue there are six people at the scene of a crime (three men and three women). One woman, Mrs. Peacock, is charged with the crime (so H_p is "Mrs. Peacock guilty"). The prosecution provides evidence E that the crime must have been committed by a woman. Because $P(E \mid H_p) = 1$ and $P(E \mid \text{not } H_p) = 2/5$, the LR is 2.5 and so the evidence is probative in favor of the prosecution. However, if the defense is allowed to focus on an alternative hypothesis H_d that is not "not H_p," say, "Miss Scarlet guilty," then $P(E \mid H_d) = 1$. In this case the defense can argue that the LR is 1 and so has

Table 2 Conditional probability table for *E*1

Guilty	False	True
False	0.5	0.999
True	0.5	0.0010

Table 3 Conditional probability table for *E*2

Guilty	False		True	
*E*1: tiny DNA trace found	False	True	False	True
False	1.0	0.99	1.0	0.0
True	0.0	0.01	0.0	1.0

no probative value. But while the evidence does not distinguish between the (nonexhaustive) H_p and H_d, it certainly is probative for H_p.

We already noted that it may be difficult in practice to elicit $P(E \mid H_d)$ 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 that seems a reasonable alternative to H_p for which $P(E \mid H_d)$ can be elicited. Example 2 and the arguments in Fenton et al. (2013a) show the dangers of 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. George* appeal may have been compromised by precisely this problem. Similarly, the failure to choose appropriate (mutually exclusive and exhaustive) hypotheses in the arguments used by Bayesians to expose the flaws in the *R v. Clark* case may have exaggerated the scale of the problem (Fenton 2014).

It is especially confusing (and even disturbing) that the guidelines (Jackson et al. 2013, Roberts et al. 2013) 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.

An additional danger of allowing H_d to be something different from "not H_p" is that, in practice, forensic experts may come up with a 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. George* appeal (Fenton et al. 2013a). Two hypotheses that are not mutually exclusive may be true and in such circumstances the LR is meaningless.

4.4. 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 (Press 2002). But this is something of a misconception for two reasons. First, the LR, no matter how high or low, actually tells us nothing

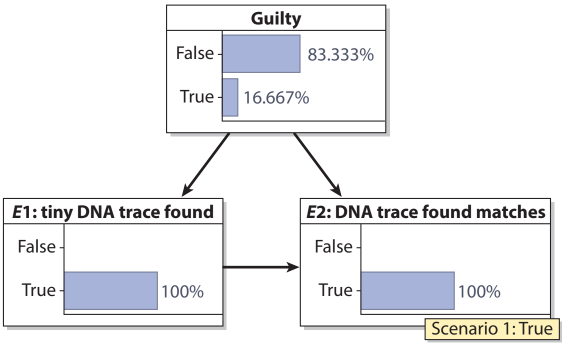


Figure 6

Posterior odds in the improved model shown in **Figure 5*b***. Abbreviation: *E*, evidence.

about the probability that either hypothesis is true. We can make conclusions about such (posterior) probabilities only if we know the prior probabilities. Although this issue has been well documented (Fenton 2011, Meester & Sjerps 2004, Triggs & Buckleton 2004), it 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 *R v. T* judgment. Specifically, in the otherwise excellent position statement (Aitken et al. 2011) is the extraordinary Point 9 (p. 2), which asserts that "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 refer specifically to the use of the LR, thereby implying that the use of the LR is appropriate, whereas the use of Bayes' theorem may not be.

Second, it is impossible to specify the likelihoods $P(E \mid H_p)$ and $P(E \mid H_d)$ meaningfully without knowing something about the priors $P(H_p)$, $P(H_d)$. This is because (in strict Bayesian terms[1]) we say the likelihoods and 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. This is illustrated in Example 3.

Example 3: Suppose the evidence E in a murder case is "DNA matching the defendant is found on victim." While the prosecution likelihood $P(E \mid H_p)$ might be agreed to be close to 1, there is a problem with the defense likelihood $P(E \mid H_d)$. For DNA evidence such as this, the defense likelihood is usually assumed to be the RMP of the DNA type. This can typically be as low as 1 in 1 billion. But consider two extreme values that may be considered appropriate for the prior $P(H_p)$, derived from different scenarios used to determine K:

1. $P(H_\mathrm{p}) = 0.5$, where the defendant is one of two people seen grappling with the victim before one of them killed the victim;
2. $P(H_\mathrm{p}) = 1/40$ million, where nothing is known about the defendant other than he is one of 40 million adults in the United Kingdom who could have potentially committed the crime.

Whereas a value for $P(E \mid H_d) = $ RMP seems reasonable in Scenario 2, it is clearly not in Scenario 1. In Scenario 1 the defendant's DNA is very likely to be on the victim whether or not he is the one who killed the victim. This suggests a value of $P(E \mid H_d)$ close to 1. It follows that, without an understanding about the priors and the background knowledge, we can end up with vastly different LRs associated with the same hypotheses and evidence.

4.5. The Likelihood Ratio 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 source-level and offense-level hypotheses (Cook et al. 1998a). Sometimes one may mutate into the other 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 for the offense-level hypotheses.

[1] Specifically, the priors $P(H_p)$ and $P(H_d)$ really refer to $P(H_p \mid K)$ and $P(H_d \mid K)$, respectively. The likelihoods must take account of the same background knowledge K that is implicit in these priors. So the real likelihoods we need are $P(E \mid H_p, K)$ and $P(E \mid H_d, K)$.

Using BNs that model dependencies between different hypotheses and evidence, Fenton et al. (2013a) show that a LR for a piece of evidence E that strongly favors one side for the source-level hypotheses can actually strongly favor the other side for the offense-level hypotheses, even though both pairs of hypotheses seem similar. Similarly, a LR that is neutral under the source-level hypotheses may actually be significantly nonneutral under the associated offense-level hypotheses. Once again these issues were shown to be a real problem in the transcript of the *R v. George* appeal (Fenton et al. 2013a).

4.6. Confusion About Likelihood Ratios Expressed on a Verbal Scale

The UK Forensic Science Service (Puch-Solis et al. 2012) recommends that a LR should be presented on an equivalent verbal scale to help lawyers and jurors understand its significance. This recommendation (criticized in Mullen et al. 2014) contrasts with US courts that have advised against verbal scales and instead recommended that posterior probabilities should be provided based on a range of priors for the given LR. We believe that the approach of US courts is correct, but recognize that in the United Kingdom explicit use of numerical LRs is increasingly snubbed following the *R v. T* 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 only DNA, but also handprints, fiber matching, footwear matching, soil and particle matching, matching specific articles of clothing, and matching cars and their license plates. In all cases there was some kind of database or expert judgment on which to estimate frequencies and RMPs, 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 a direct result of the *R v. T* ruling, which has effectively pushed explicit use of numerical LRs underground. Indeed, we have seen expert reports that contained the explicit data formally withdrawn as a result of *R v. T*. This is one of the key negative impacts of *R v. T*, and 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, many statisticians (including even some Bayesian statisticians) either are 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 nontrivial 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 the supplemental material we show other examples where statisticians provide unnecessarily complex arguments.

Supplemental Material

The manual approach is also not scalable; otherwise, one of the BN inference algorithms would have to be explained and computed, 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 in private at a recent legal conference (to the first author of this article[2]) by one of the United Kingdom'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 lack of understanding exists 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 informal evidence from cases in which we have been involved shows that 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 such as "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 frequently used by lawyers. Our own experiences as expert witnesses lead us to 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. And this issue has been extensively documented by numerous authors, including Anderson et al. (2005), Balding & Donnelly (1994), Edwards (1991), Evett (1995), Fenton & Neil (2000), Freckelton & Selby (2005), Jowett (2001), Kaye (2001), Koehler (1993b), Murphy (2003), Redmayne (1995), and Thompson & Schumann (1987), and has even been dealt with in populist books by Gigerenzer (2002) and Haigh (2003). There is near 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 by Evett (1995), Finkelstein & Levin (2001), Foreman et al. (2003), Good (2001), Jackson et al. (2006), Redmayne (1995), Saks & Thompson (2003), and Robertson & Vignaux (1995), although there is less of a consensus on whether experts are needed in court to present the results of all but the most basic Bayesian arguments (Robertson & Vignaux 1998).

We argue that the way forward is to use BNs to present probabilistic legal arguments, because this approach avoids much of the confusion surrounding both the oversimplistic 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. Edwards (1991, p. 1068) 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."

Unfortunately, Edwards was grossly optimistic. In addition to the reasons already identified above, there is the major challenge of 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 because it is easy to change the priors and do sensitivity analysis (Fenton & Neil 2012).

[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!

A basic strategy for presenting BNs to legal professionals is described in detail by Fenton & Neil (2011) 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. Although 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 enables us to do these calculations instantly.

The results from a BN tool can be presented using a range of assumptions including different priors. Legal professionals (and perhaps even jurors if presented in court) should never have to think about 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 (*a*) whether the correct nodes and causal links have been modeled, (*b*) what suitable values (or range of values) are required for the priors and conditional priors, and (*c*) whether the hidden value *K* for context has been considered and communicated throughout the model. But these assumptions are precisely what have to be considered in weighing any legal argument. The BN simply makes this all explicit rather than hidden, which is another clear benefit of the approach.

We recognize that significant technical challenges must be overcome to make the construction of BNs for legal reasoning easier. Recent work addresses the lack of a systematic, repeatable method for modeling legal arguments as BNs by using common idioms and an approach for building complex arguments from these idioms (Fenton et al. 2013b, Hepler et al. 2007). We also have to overcome various technical constraints of existing BN algorithms and tools, such as forcing modelers to specify unnecessary prior probabilities [this is being addressed in work, for example, by the European Research Council (2015)]. However, the greater challenges are cultural and presentational and these issues are being addressed in projects (European Research Council 2015, Isaac Newton Institute 2016, Royal Statistical Society 2015) that involve close interaction with legal professionals and empirical studies on how to best present arguments.

BNs offer a great opportunity 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 Langley 2012). 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 *R v. T* 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 (Balding 2004, Evett et al. 2002) is especially critical because a DNA match in such cases may have low probative value (especially if the potential for secondary and tertiary transfer is considered) but still have a powerful impact on lawyers and jurors; we are currently involved in appeal cases in which some low-template DNA evidence was assumed to trump all the opposing evidence in the case.

Supplemental Material

BN models demonstrate that the impact of DNA evidence has been exaggerated (examples are provided in the supplemental material).

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 prosecutors 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 of Bayes in the law in practice is rather depressing. It confirms that outside of paternity cases its impact on legal practice has been minimal. Although 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 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 *R v. T* 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 *R v. T* judgment, we have seen the devastating appeal judgments in *R v. Adams* (effectively banning the use of Bayes for combining nonstatistical evidence) and in *Nulty & Ors v. Milton Keynes Borough Council* (2013) (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 favorable to Bayes (*R v. George* 2007), subsequent research has discovered weaknesses in the LR reasoning (Fenton et al. 2013a). Similar weaknesses have been recently exposed in other successful Bayesian arguments, such as those used in the successful Sally Clark appeal (Fenton 2014).

The good news is that the difficulties encountered in presenting Bayesian arguments can be avoided once it is accepted that oversimplistic LR models and inference calculations from first principles are generally unsuitable. We have explained that BNs are the appropriate vehicle for overcoming these problems, because 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 in court complex Bayesian arguments from first principles. 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 & Vignaux (1998) 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 instead 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 the massive cultural barriers between the fields of science and law 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 sponsored by the European Research Council (2015), the Isaac Newton Institute (2016), and the Royal Statistical Society (2015).

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) that 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.

DISCLOSURE STATEMENT

N. Fenton and M. Neil are directors of a company, Agena Ltd., that specializes in Bayesian network tools.

ACKNOWLEDGMENTS

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

LITERATURE CITED

Agena Ltd. 2016. AgenaRisk. **http://www.agenarisk.com/**

Aitken C, Berger CEH, Buckleton JS, Champod C, Curran J, et al. 2011. Expressing evaluative opinions: a position statement. *Sci. Justice* 51:1–2

Aitken CGG, Taroni F. 2004. *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester, UK: Wiley. 2nd ed.

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

Anderson T, Schum DA, Twining W. 2005. *Analysis of Evidence*. Cambridge, UK: Cambridge Univ. Press

Balding DJ. 2004. *Weight-of-Evidence for Forensic DNA Profiles*. Chichester, UK: Wiley

Balding DJ, Donnelly P. 1994. Prosecutor's fallacy and DNA evidence. *Crim. Law Rev.* 1994:711–22

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

BayesFusion. 2016. GeNIe Modeler: Complete Modeling Freedom. **http://www.bayesfusion.com/#!genie-modeler/lf73d**

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

Berger D. 2014. *Improving legal reasoning using Bayesian probability methods*. PhD Thesis, Queen Mary Univ. London

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

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

Cohen LJ. 1977. *The Probable and the Provable*. Oxford, UK: Clarendon

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)

Collins A, Morton NE. 1994. Likelihood ratios for DNA identification. *PNAS* 91:6007–11

Cook R, Evett IW, Jackson G, Jones P, Lambert JA. 1998a. A hierarchy of propositions: deciding which level to address in casework. *Sci. Justice* 38:231–39

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. 1998b. A model for case assessment and interpretation. *Sci. Justice* 38:151–56

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

Dawid AP. 1987. The difficulty about conjunction. *J. R. Stat. Soc. D* 36:91–97

Dawid AP. 2005. Statistics on trial. *Significance* 2:6–8

Dawid AP, Mortera J. 1998. Forensic identification with imperfect evidence. *Biometrika* 85:835–49

Dawid AP, Mortera J. 2008. Probability and evidence. In *Handbook of Probability: Theory and Applications*, ed. T Rudas, pp. 403–22. Thousand Oaks, CA: Sage

Dawid AP, Musio M, Fienberg SE. 2015. From statistical evidence to evidence of causality. *Bayesian Anal.* doi:10.1214/15-BA968

Donnelly P. 2005. Appealing statistics. *Significance* 2:46–48

Edwards W. 1991. Influence diagrams, Bayesian imperialism, and the Collins case: an appeal to reason. *Cardozo Law Rev.* 13:1025–79

European Research Council. 2015. Effective Bayesian modelling with knowledge before data. **http://bayes-knowledge.com/**

Evett IW. 1995. Avoiding the transposed conditional. *Sci. Justice* 35:127–31

Evett IW, Foreman LA, Jackson G, Lambert JA. 2000. DNA profiling: a discussion of issues relating to the reporting of very small match probabilities. *Crim. Law Rev.* 341–55

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.* 47:520–30

Evett IW, Weir BS. 1998. *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sunderland, MA: Sinauer Assoc.

Faigman DL, Baglioni AJ. 1988. Bayes' theorem in the trial process. *Law Hum. Behav.* 12:1–17

Fenton NE. 2011. Science and law: Improve statistics in court. *Nature* 479:36–37

Fenton NE. 2014. Assessing evidence and testing appropriate hypotheses. *Sci. Justice* 54:502–4

Fenton NE, Berger D, Lagnado DA, Neil M, Hsu A. 2013a. When "neutral" evidence still has probative value (with implications from the Barry George Case). *Sci. Justice* 54:274–87

Fenton NE, Lagnado DA, Neil M. 2013b. A general structure for legal arguments using Bayesian networks. *Cogn. Sci.* 37:61–102

Fenton NE, Neil M. 2000. The Jury Fallacy and the use of Bayesian nets to simplify probabilistic legal arguments. *Math. Today* 36:180–87

Fenton NE, Neil M. 2010. Comparing risks of alternative medical diagnosis using Bayesian arguments. *J. Biomed. Inform.* 43:485–95

Fenton NE, Neil M. 2011. Avoiding legal fallacies in practice using Bayesian networks. *Aust. J. Leg. Philos.* 36:114–50

Fenton NE, Neil M. 2012. *Risk Assessment and Decision Analysis with Bayesian Networks*. Boca Raton, FL: CRC Press

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

Fienberg SE. 1989. *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer-Verlag

Fienberg SE. 2011. Bayesian models and methods in public policy and government settings. *Stat. Sci.* 26:212–26

Fienberg SE, Finkelstein MO. 1996. Bayesian statistics and the law. In *Bayesian Statistics 5*, ed. JM Bernardo, JO Berge, AP Dawid, AFM Smith, pp. 129–46. Oxford, UK: Oxford Univ. Press

Fienberg SE, Kaye DH. 1991. Legal and statistical aspects of some mysterious clusters. *J. R. Stat. Soc. A* 154:61–74

Finkelstein MO. 2009. *Basic Concepts of Probability and Statistics in the Law*. New York: Springer

Finkelstein MO, Fairley WB. 1970. A Bayesian approach to identification evidence. *Harv. Law Rev.* 83:489–517

Finkelstein MO, Levin BA. 2001. *Statistics for Lawyers*. New York: Springer

Foreman LA, Champod C, Evett IW, Lambert JA, Pope S. 2003. Interpreting DNA evidence: a review. *Int. Stat. Rev.* 71:473–95

Forrest AR. 2003. Sally Clark—a lesson for us all. *Sci. Justice* 43:63–64

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

Fung WK. 2003. User-friendly programs for easy calculations in paternity testing and kinship determinations. *Forensic Sci. Int.* 136:22–34

Gastwirth JL. 2000. *Statistical Science in the Courtroom*. New York: Springer
Gigerenzer G. 2002. *Reckoning with Risk: Learning to Live with Uncertainty*. London: Penguin
Gill R. 2014. How to become a convicted serial killer (without killing anyone). *Justice Gap*. Aug. 1. **http://thejusticegap.com/2014/08/become-convicted-serial-killer-without-killing-anyone/**
Good IJ. 1995. When batterer turns murderer. *Nature* 375:541
Good PI. 2001. *Applying Statistics in the Courtroom: A New Approach for Attorneys and Expert Witnesses*. Boca Raton, FL: CRC Press
Haigh J. 2003. *Taking Chances: Winning with Probability*. Oxford, UK: Oxford Univ. Press. 2nd ed.
Hepler AB, Dawid AP, Leucari V. 2007. Object-oriented graphical representations of complex patterns of evidence. *Law Probab. Risk* 6:275–93
Hill R. 2005. Reflections on the cot death cases. *Significance* 2:13–16
Isaac Newton Institute. 2016. *Probability and Statistics in Forensic Science.* **https://www.newton.ac.uk/event/fos**
Jackson G, Aitken C, Roberts P. 2013. *Practitioner guide no. 4. Case assessment and interpretation of expert evidence: guidance for judges, lawyers, forensic scientists and expert witnesses.* London: R. Stat. Soc. **http://www.maths.ed.ac.uk/~cgga/Guide-4-WEB.pdf**
Jackson G, Jones S, Booth G, Champod C, Evett IW. 2006. The nature of forensic science opinion—a possible framework to guide thinking and practice in investigations and in court proceedings. *Sci. Justice* 46:33–44
James SH, Kish PE, Sutton TP. 2005. *Principles of Bloodstain Pattern Analysis: Theory and Practice*. Boca Raton, FL: CRC Press
Jowett C. 2001. Lies, damned lies, and DNA statistics: DNA match testing Bayes' theorem, and the criminal courts. *Med. Sci. Law* 41:194–205
Kadane JB. 1990. A statistical analysis of adverse impact of employer decisions. *J. Am. Stat. Assoc.* 85:607–14
Kadane JB. 2008. *Statistics in the Law: A Practitioner's Guide, Cases, and Materials*. Oxford, UK: Oxford Univ. Press
Kadane JB, Schum DA. 1996. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. Hoboken, NJ: Wiley
Kaye DH. 2001. Two fallacies about DNA databanks for law enforcement. *Brooklyn Law Rev.* 67:179–206
Kaye DH. 2007. Revisiting "Dreyfus": a more complete account of a trial by mathematics. *Minn. Law Rev.* 91:825–35
Koehler JJ. 1992. Probabilities in the courtroom: an evaluation of the objections and policies. In *Handbook of Psychology and Law*, ed. DK Kagehiro, WS Laufer, pp. 167–84. New York: Springer-Verlag
Koehler JJ. 1993a. Error and exaggeration in the presentation of DNA evidence at trial. *Jurimetrics* 34:21–39
Koehler JJ. 1993b. The base rate fallacy myth. *Psycoloquy* 4(49)
Langley W. 2012. The case against DNA. *Telegraph*, Mar. 6. **http://www.telegraph.co.uk/news/science/9115916/The-case-against-DNA.html**
Lauritzen SL, Spiegelhalter DJ. 1988. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. R. Stat. Soc. B* 50(2):157–224
Lucy D, Schaffer B. 2007. *Nature of Legal Evidence*. Univ. Edinburgh, Scotland. **http://www.maths.ed.ac.uk/~cgga/nature-of-evidence.pdf**
Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7:639–55
Meester R, Collins M, Gill R, van Lambalgen M. 2007. On the (ab)use of statistics in the legal case against the nurse Lucia de B. *Law Probab. Risk* 5:233–50
Meester R, Sjerps M. 2003. The evidential value in the DNA database search controversy and the two-stain problem. *Biometrics* 59:727–32
Meester R, Sjerps M. 2004. Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. *Law Probab. Risk* 3:51–62
Meier P, Zabell S. 1980. Benjamin Peirce and the Howland Will. *J. Am. Stat. Assoc.* 75:497–506
Morrison GM. 2012. The likelihood ratio framework and forensic evidence in court: a response to RvT. *Int. J. Evid. Proof* 16(1)
Mullen C, Spence D, Moxey L, Jamieson A. 2014. Perception problems of the verbal scale. *Sci. Justice* 54:154–58

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

National Research Council. 2009. *Strengthening forensic science in the United States: a path forward*. Washington, DC: Natl. Acad. Press

Neil M, Tailor M, Marquez D. 2007. Inference in hybrid Bayesian networks using dynamic discretization. *Stat. Comput.* 17:219–33

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

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

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

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

Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Palo Alto, CA: Kaufmann

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

People v. Puckett. A121368 (Cal. Ct. App. 2009)

Press SJ, ed. 2002. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Hoboken, NJ: Wiley

Puch-Solis R, Roberts P, Pope S, Aitken C. 2012. *Practitioner guide no. 2. Assessing the probative value of DNA evidence: guidance for judges, lawyers, forensic scientists and expert witnesses*. London: R. Stat. Soc. **http://www.rss.org.uk/Images/PDF/influencing-change/rss-assessing-probative-value.pdf**

R v. Adams, [1996] EWCA Crim 222

R v. Adams, [1998] 1 Cr App R 377

R v. Alan James Doheny and Gary Adams, [1996] EWCA Crim 728

R v. Deen, [1994] EWCA Crim Case No: 90/1523/X3

R v. George, [2007] EWCA Crim 2722

R v. T, [2010] EWCA Crim 2439

Redmayne M. 1995. DNA evidence, probability and the courts. *Crim. Law Rev.* 1995:464–82

Redmayne M, Roberts P, Aitken C, Jackson G. 2011. Forensic science evidence in question. *Crim. Law Rev.* 2011:347–56

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*. London: R. Stat. Soc. **http://www.maths.ed.ac.uk/~cgga/Guide-3-WEB.pdf**

Robertson B, Vignaux GA, Berger CEH. 2011. Extending the confusion about Bayes. *Mod. Law Rev.* 74:444–55

Robertson B, Vignaux T. 1995. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester, UK: Wiley

Robertson B, Vignaux T. 1997. Bayes' theorem in the Court of Appeal. *Crim. Lawyer* January:4–5

Robertson B, Vignaux T. 1998. Don't teach statistics to lawyers! *Proc. Fifth Int. Conf. Teach. Stat.* pp. 541–47. Singapore

Roth A. 2010. Safety in numbers?: deciding when DNA alone is enough to convict. *New York Univ. Law Rev.* 85:101–53

Royal Statistical Society. 2015. *Statistics and the Law*. **http://www.rss.org.uk/RSS/Influencing_Change/Statistics_and_the_law/RSS/Influencing_Change/Current_projects_sub/Statistics_and_the_law.aspx**

Saks MJ, Thompson WC. 2003. Assessing evidence: proving facts. In *Handbook of Psychology in Legal Contexts*, ed. D Carson, R Bull, pp. 329–45. Chichester, UK: Wiley. 2nd ed.

Schneps L, Colmez C. 2013. *Math on Trial*. New York: Basic Books

Sjerps M, Berger C. 2012. How clear is transparent? Reporting expert reasoning in legal cases. *Law Probab. Risk* 11:317–29

Spiegelhalter DJ. 2013. Court of Appeal bans Bayesian probability (and Sherlock Holmes). *Understanding Uncertainty*, Feb. 25. **http://understandinguncertainty.org/court-appeal-bans-bayesian-probability-and-sherlock-holmes**

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

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

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

Thompson J, Danielson S, Hartmann E, Bucheger D, Koplitz J. 2001. *Trial by Probability: Bayes' Theorem in Court*. Work. Pap., Univ. Wis. Stout. **https://www.uwstout.edu/mscs/upload/Trial-by-Probability.pdf**

Thompson WC. 1996. DNA evidence in the O.J. Simpson trial. *Color. Law Rev.* 67:827–57

Thompson WC, Schumann EL. 1987. Interpretation of statistical evidence in criminal trials. *Law Hum. Behav.* 11:167–87

Thompson WC, Taroni F, Aitken CGG. 2003. How the probability of a false positive affects the value of DNA evidence. *J. Forensic Sci.* 48:47–54

Tillers P, Gottfried J. 2007. Case comment—*United States v. Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): a collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable? *Law Probab. Risk* 5(5):135–57

Tillers P, Green E, eds. 1988. *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism*. Dordrecht, the Neth.: Kluwer Academic

Tribe LH. 1971. Trial by mathematics. *Harv. Law Rev.* 84:1329–93

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

Univ. North Carol. 2009. *North Carolina Trial Judges' Bench Book District Court Edition*, Vol. 1, *Family Law*, Ch. 10: Paternity, ed. CD Howell, JS Simmons. Chapel Hill: Univ. North Carolina Sch. Gov.

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

Vosk T, Emery AF. 2014. *Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists*. Boca Raton, FL: CRC Press

Wagenaar WA. 1988. The proper seat: a Bayesian discussion of the position of expert witnesses. *Law Hum. Behav.* 12:499–510

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

Ziliak ST, McCloskey DN. 2008. *The Cult of Statistical Significance*. Ann Arbor: Univ. Michigan Press