

RESEARCH

Open Access



Using Bayesian networks to guide the assessment of new evidence in an appeal case

Nadine M. Smit^{1,2*} , David A. Lagnado³, Ruth M. Morgan^{1,2} and Norman E. Fenton⁴

Abstract

When new forensic evidence becomes available after a conviction there is no systematic framework to help lawyers to determine whether it raises sufficient questions about the verdict in order to launch an appeal. This paper presents such a framework driven by a recent case, in which a defendant was convicted primarily on the basis of audio evidence, but where subsequent analysis of the evidence revealed additional sounds that were not considered during the trial. The framework is intended to overcome the gap between what is generally known from scientific analyses and what is hypothesized in a legal setting. It is based on Bayesian networks (BNs) which have the potential to be a structured and understandable way to evaluate the evidence in a specific case context. However, BN methods suffered a setback with regards to the use in court due to the confusing way they have been used in some legal cases in the past. To address this concern, we show the extent to which the reasoning and decisions within the particular case can be made explicit and transparent. The BN approach enables us to clearly define the relevant propositions and evidence, and uses sensitivity analysis to assess the impact of the evidence under different assumptions. The results show that such a framework is suitable to identify information that is currently missing, yet clearly crucial for a valid and complete reasoning process. Furthermore, a method is provided whereby BNs can serve as a guide to not only reason with incomplete evidence in forensic cases, but also identify very specific research questions that should be addressed to extend the evidence base and solve similar issues in the future.

Keywords: Bayesian networks, Bayesian inference, Audio evidence, Forensic reasoning, Appeal

Background

Extensive research is being undertaken in various scientific disciplines to extend the evidence base that can help the detection, collection, and analysis of forensic evidence (INTERPOL 2013). Such empirically based knowledge, together with experiential insights, can begin to address questions regarding the way in which evidence behaves [e.g. (French and Morgan 2015)] to assess the probability of observations under proposed defence and prosecution hypotheses. In practice, the personal opinion of the expert, based on their experience, is generally accepted by the court as valid evidence (Morgan and Bull 2007; R v. Weller 2010; The Law Commission 2011). Even though this kind of expert judgment can be informative, especially in the absence of an empirical evidence base,

these subjective arguments raise questions about issues such as the validation of the arguments put forward by the expert (National Research Council 2009) and susceptibility to cognitive influences on the interpretation of evidence under uncertainty (Kassin et al. 2013; Nakhaei-zadeh et al. 2014). Similarly, interpretation issues can be faced by the court when experts do provide empirical evidence, due to the lack of agreement and consistency in the application of current interpretation approaches (National Research Council 2009), and the difficulties in presenting the reasoning processes of the scientists transparently in court (Sjerps and Berger 2012).

Approaches to interpretation

To offer a scientific methodological approach to the forensic interpretation and decision-making process, the *case assessment and interpretation (CAI)* model was developed in the late nineties by Cook et al. (1998b). This framework requires one to make a structured and

*Correspondence: nadine.smit.12@ucl.ac.uk

² Centre for the Forensic Sciences, University College London, 35 Tavistock Square, London WC1H 9EZ, UK

Full list of author information is available at the end of the article

systematic assessment of the requirements, questions, and outcomes of the analysis to optimise the contribution of the evidence to the case (Jackson et al. 2013). Because the formulation of the hypotheses is of high significance in the interpretation of the evidence (Fenton et al. 2014), it is important to consider the hypotheses that the different stakeholders realistically can address, using their expertise, to come to a defensible conclusion. Therefore, Cook et al. developed the so-called *hierarchy of propositions* as part of the CAI (1998a) model, in which relevant hypotheses are defined at the offence, activity and source levels. While the scientist is able to make inferences on the source and activity level, the ultimate question of guilt at the offence level lies with a judge or jury.

To make inferences and guide decision-making, logical reasoning processes such as multivariate analytical methods can be used (Dawid and Evett 1997). One of these methods is the use of Bayesian networks (BNs) (Fenton and Neil 2012; Taroni et al. 2014), which has been applied to the interpretation of various kinds of evidence, including DNA (Evett et al. 2002), fire incidents (Biedermann et al. 2005), gunshot residue (Biedermann et al. 2009; Fenton et al. 2014), glass (Zadora 2009), and document examinations (Biedermann et al. 2011). In BNs, one can graphically formalise the probabilistic (uncertain) relations between the hypotheses in question and the available evidence. These are represented as nodes, and the relations are formalised in node probability tables (NPTs). New evidence is incorporated by using Bayes' theorem to update the previous (prior) belief in a hypothesis to a new (posterior) probability in the light of the strength of this evidence. Equivalently, it is also possible to compute the overall likelihood ratio (LR) of pieces of evidence (either individually or in combination) with respect to any particular hypothesis; the benefit of the LR approach is that it avoids the need to define prior probabilities for hypotheses that are not themselves conditioned on other hypotheses or evidence.

Despite the research successes, the Bayesian approach in a legal setting suffered a setback as a result of the case of *R v. T* (2010), which questioned the use of probabilistic approaches in court with regards to shoeprint evidence. This led to significant debates about the role of Bayesian approaches in presenting evidence types that do not have an established statistical base akin to DNA and more generally, how forensic evidence should be presented in court (Berger et al. 2011; Hamer 2012; Morrison 2012; Redmayne et al. 2011; Robertson et al. 2011). Thompson (2012) concluded that, in order to avoid such problems in the future, the forensic scientist should only receive necessary and specific information, valid scientific bases should accompany their presented evidence, their reasoning processes should be transparent and lastly, the

results should be presented to legal professionals in an understandable way. This could all be better regulated by the use of LRs (Thompson 2012), although [as was demonstrated in (Fenton et al. 2014)], there are several important limitations and caveats that need to be considered when using the likelihood ratio to summarise the impact of evidence in a Bayesian argument. Ultimately, if strategic and cultural challenges (such as support for and understanding by the criminal justice system) can be overcome, the Bayesian approach could become the central method for the evaluation of evidence (Fenton et al. 2013a).

Objectives

This paper aims to contribute to overcoming the aforementioned challenges by showing how BNs can be used to structure and represent the set of information and questions in the context of a criminal case, as a structural and transparent tool to evaluate the value of new evidence that is acquired in the stages prior to an appeal hearing. It presents a method to develop the BN that captures the relation between the acquired forensic evidence and the proposed hypotheses. The BN is subsequently used to help define the experimental studies necessary to evaluate the new evidence, to infer the evidentiary value of the newly acquired evidence, and to evaluate the evidence that is of most value to the hypothesis of interest.

Case study

In the case we consider here, the defendant was convicted of attempted murder. In summary (the full details are provided in "[Requirements and case pre-assessment](#)" section), the young child of the defendant was taken to hospital with serious injuries after an alleged incident took place while the baby and the defendant were alone in a room. The defendant allegedly hurt the baby, resulting in blood stains belonging to the baby being found on the wall in the room. One of the key items of evidence presented by the prosecution was an audio recording of an emergency services call that the defendant made immediately after the incident. On the recording there were distinct sounds that the prosecution claimed to be linked to the criminal act, namely the defendant scraping blood off a wall. During the trial, the jury was asked to decide whether or not they could be sure of either of the alternative counts on the indictment: attempted murder or causing grievous bodily harm with intent. After the defendant was convicted of murder, questions were raised by the defence team that led to an appeal process being set in place, where the audio evidence was re-analysed by an audio expert. Following this, the most important sections in the grounds of appeal were:

- the prosecution theory concerning the recorded sound was presented during the trial as if it was an established fact without any evidence from an expert, and
- new reports of the defence expert undermine the contention that the sound in question must have been made by scraping blood off the wall.

Content

This paper firstly provides an overview of the methods that are used to define the relevant propositions, create the network and evaluate the evidence in this case (“**Methods**” section). The subsequent sections present the requirements and case pre-assessment applied to the case study (“**Requirements and case pre-assessment**” section), the BN created using this information (“**The BN models for the case**” section), and the assessment of the evidentiary value of the new evidence (“**Evaluation of the evidential values**” section).

Methods

The propositions and the Bayesian network (BN)

All the original case materials were made available and were examined to identify the propositions, at different levels, that were of importance in the trial and subsequently addressed in the preparation for the appeal. Once the relevant propositions and required outcomes were set, a BN was created using the basic connective structures and building methods previously discussed in the literature (Fenton et al. 2013b; Taroni et al. 2004), where the presented propositions represent the arguments raised by the defence and prosecution in the original case. The BN was then extended by incorporating the newly gained information in preparation for the appeal and created using the freely available software package AgenaRisk (AgenaRisk 2015).

Conditional probabilities and inference methods

When revising the probabilities for a pair of prosecution (H_p) and defence (H_d) hypotheses in the light of n items of evidence, the posterior probabilities $P(H_p|E1, \dots, En)$ and $P(H_d|E1, \dots, En)$ can be found using Bayes’ theorem, which can be derived from the general laws of conditional probability: the prior probability is multiplied with the support that the new evidence ($E1, \dots, En$) provides for the hypotheses:

$$P(H_p|E1, \dots, En) = \frac{P(E1, \dots, En|H_p)}{P(E1, \dots, En)} \cdot P(H_p)$$

$$P(H_d|E1, \dots, En) = \frac{P(E1, \dots, En|H_d)}{P(E1, \dots, En)} \cdot P(H_d)$$

By dividing these probabilities, we get the ‘odds’ form of Bayes theorem which shows that the posterior odds are the product of the prior odds and the likelihood ratio.

$$\frac{P(H_p|E1, \dots, En)}{P(H_d|E1, \dots, En)} = \frac{P(E1, \dots, En|H_p)}{P(E1, \dots, En|H_d)} \cdot \frac{P(H_p)}{P(H_d)}$$

= likelihood ratio · prior odds

Providing that the pair of prosecution and defence hypotheses are mutually exclusive and exhaustive (meaning that H_d is equivalent to “not H_p ”) the likelihood ratio *on its own is a valid* measure of probative value of the evidence ($E1, \dots, En$) on the pair of hypotheses (Fenton et al. 2014). Specifically, if the LR > 1 the evidence supports H_p (with higher values offering greater support) and if the LR < 1, the evidence supports H_d (with lower values offering greater support). This is very attractive for forensic scientists since it means they can assess the probative value of the evidence without having to provide any prior probability for hypothesis H_p . However, while computing the LR is simple for a single piece of evidence [the expert only has to provide two likelihoods $P(E|H_p)$ and $P(E|H_d)$], when there are multiple pieces of (possibly) dependent evidence along with other unknown but related hypotheses, the computation of the probabilities $P(E1, \dots, En|H_p)$ and $P(E1, \dots, En|H_d)$ needed for the LR requires a full inference over the entire BN model that captures all of these dependencies.

In the BN model, we only have to provide ‘local’ likelihoods of individual evidence nodes given the parents—these can be extracted from the case information where available, and stored in the NPTs. Once the BN structure is created (showing the conditional dependent relations) and these ‘local’ likelihood values are provided for the NPTs, the over-all likelihood ratio for any set of evidence with respect to any hypothesis node can be obtained by running the model in BN software (e.g., AgenaRisk, Hugin). Specifically for evidence $E = \{E1, \dots, En\}$ when we run the model with observations for that evidence, the model computes (for any hypothesis node H), the value $P(H|E)$ [and $P(\text{not}H|E)$, which must be $1 - P(H|E)$]. But, rearranging the formula for the odds version of Bayes, we get:

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

where $P(H)$ is the marginal prior probability value for H when the model is running without any observation for the evidence nodes. So, the BN computation provides the correct LR value for E with respect to H.

Representation of case information

To translate the verbal scales used in the case documentation to numerical probabilities, the translation table presented by Morgan (2009) was used. Even though it is recognized that this translation is still subjective, it is

merely used to show how BNs can be used in cases where likelihoods are or can actually be reliably assigned. However, if there is insufficient information regarding such probabilities (such as a lack of databases or experimental studies), the probabilities are represented as a variable which can be studied across a range of possible values.

Sensitivity analysis

In order to investigate the evidential value of each item of evidence to the hypothesis of interest a sensitivity analysis was performed. Such an analysis allows one to identify those with the strongest evidential value, by calculating and comparing the probabilities of the target node with and without assumed states for (a combination of) the evidence nodes. For example, by comparing $P(H_p)$, $P(H_p|E1 = e1)$, and $P(H_p|E2 = e2)$. The results of such changes will be visualised in a so-called *tornado graph*.

Additionally, the effect when the missing conditional probabilities are actually obtained was assessed. This method was designed to result in a transparent reasoning process which is understandable by legal professionals and jurors, and addresses some of the issues raised by Thompson (2012) which are summarised in “[Approaches to interpretation](#)” section.

Requirements and case pre-assessment

Case information

From the case documentation, it follows that, in relation to the audio evidence, the following information can be observed:

- A baby was injured during an incident on the top floor of a house
- Blood from the baby was found on the wall in one of the rooms upstairs
- On an audio recording of the emergency telephone call made by the suspect, a scraping sound (allegedly indicating scraping blood off a wall) can be heard (hereafter referred to as sound 1 at time t_1)
- The suspect was charged with attempted murder.

The audio evidence played a significant role in the trial. However, the assessment of observing the evidence that would be expected if the suspect did indeed scrape blood off the wall was not supported by any empirical studies or past experiences. During the appeal preparation process, the recorded emergency call was re-analysed by an audio expert on behalf of the defence, and four other sounds were identified on the same recording that, according to the expert, *showed similarities* to sound 1. In particular, one of these sounds (referred to as *sound 2*) was of interest because of background noise that could be heard

simultaneously at this time t_2 . This expert statement provides evidence for both the activity causing sound 1 and the activity causing sound 2. The background noise was presumed to be the television, which was located in a different room to where the prosecution argued the scraping of the blood took place. In summary, the following additional information was added in the preparation for the appeal:

- The proposition that the suspect was rubbing his face was given as an alternative explanation by the defence for the explanation of sound 1 (in what follows we assume that these two hypotheses are mutually exclusive)
- A second sound (sound 2) was noted on the emergency recording, which, according to the audio expert, *showed similarities* to sound 1
- During this second sound, the TV (located downstairs) could be heard simultaneously on the emergency recording
- A statement by the police reads that the suspect was frequently rubbing his face in their presence

Questions

Following the available information, the questions of interest (from an offence to a source level) to the case can be defined as:

- Did the suspect attempt to murder the victim?
- Did the suspect inflict any injuries on the victim?
- Did the suspect scrape blood off the wall during the emergency call causing sound 1 (time t_1)?
- Are sound 1 and 2 similar?
- Can the TV be heard during sound 2 (time t_2) if the suspect was positioned in the room upstairs?

Clearly, the formulation of the questions in the case assessment stage will impact the structure of the BN. For example, asking (as was the case in the trial)

“were the sounds *similar*”

is different from asking:

“were the sounds produced by the same activity?”

unless a similarity is defined as such. This highlights the importance of precisely defining the meaning of the terms used in the assessment.

Propositions

The ultimate question is whether the suspect attempted to murder the victim—as was the final statement in the trial. The other counts that were presented to the jury were innocence or causing GBH. Factors of relevance (and discriminatory power) to the three counts are the

degree of preparation, motive, opportunity, intent, and whether or not the criminal action actually took place. It is argued by Fenton et al. (2013b) that because opportunity, motive and intent are causes or pre-conditions for the offender to commit the crime, they should be parental nodes in a BN model. The same can be said about the degree of preparation (or malicious aforethought), which has a significant role in distinguishing between the counts of attempted murder and causing GBH. However, keeping in mind the objective of the appeal—to provide evidence which might influence the initial decision—only the propositions that are of relevance to the new evidence needs to be reassessed and is of relevance to the defence. The main forensic evidence available in this case relates to the question of whether the suspect scraped blood off the wall (which is subsequently said to be linked to the act of inflicting injuries on the victim). Hence, our BN model focuses only on source level hypotheses, ignoring the offence level hypotheses and the associated difficulties of modelling factors like opportunity, motive and intent.

The evidence presented in the trial to support the scraping/rubbing propositions was that of the sound, which is *consistent with the suspect scraping blood*. Ideally, of course, these hypotheses should be accompanied by an analysis that discusses the validity and accuracy of the evidence presented. However, as this was not done in the trial, it will therefore not be taken into consideration in this model. Statements about the presence of the other evidence types are defined similarly. Thus, the propositions presented in this case are:

- *The suspect scraped blood off the wall at time t_1 , and the alternative the suspect rubbed his face at time t_1 .*
- *A sound that is consistent with the suspect scraping blood at time t_1 is observed on the recording, and a sound that is consistent with the suspect scraping blood at time t_1 is not observed on the recording,*
- *The suspect scraped blood off the wall at time t_2 , and the suspect rubbed his face at time t_2 .*
- *A sound that is consistent with the suspect scraping blood at time t_2 is observed on the recording, and a sound that is consistent with the suspect scraping blood at time t_2 is not observed on the recording,*
- *Sound 1 and sound 2 are similar, and sound 1 and sound 2 are not similar.*
- *The TV can be observed on the recording at time t_2 , and the TV cannot be observed on the recording at time t_2 ,*
- *The police observed a frequent rubbing activity by the suspect, and the police did not observe a frequent rubbing activity by the suspect.*

These propositions highlight some important considerations in the inference process when one wants to assess LR's applied to real cases. The prosecution (scraping activity) and defence (rubbing activity) hypotheses are assumed to be mutually exclusive but not exhaustive, since the proposition that it is caused by another activity is not considered (Fenton et al. 2014). In theory, such a third hypotheses of 'other activity' could have been included in the activity nodes, if it were not for the fact that questions such as '*what is the probability of observing a similarity between a sound caused by scraping and one caused by any other activity*' are not possible to address using empirical data. Because of this, for example, a LR of 1 does not indicate that the evidence does not have any probative value: even though it might reduce the posterior probabilities of the hypotheses by equal amounts, it could increase the posterior probability of the third 'other' hypothesis (Fenton et al. 2014). Therefore, the LR will merely be used to determine whether the evidence provides more or less (or equal) support for one of the hypotheses over the other, rather than that it represents the aforementioned *probative* value of the evidence. This emphasizes the importance of formulating and thinking about the case information and desired and feasible outcomes before defining questions and hypotheses by using, for example, the Case Assessment model.

The BN models for the case

The relationship between the propositions defined in "Propositions" section and the evidence that can support or refute these hypotheses are now made explicit in the BN models.

The trial

In the trial, the prosecution hypothesis that the suspect scraped blood off the wall causing a sound at time t_1 (a_{1p}) was used to support the proposition that the suspect might have inflicted injuries to the victim. At this stage, the defence did not have an alternative state other than that 'the suspect did not scrape blood off the wall at time t_1 ' (a_{1d}). To support their hypothesis, the prosecution presented an audio recording (E1) on which sound 1 could allegedly be heard. The dependency between these is causal in such a way that the scraping activity causes the sound to be created, and therefore recordable and observable. This item of evidence can *be* heard (e_1), or simply *not be* heard (\bar{e}_1) to support or oppose the scraping proposition, and does not depend on the standpoint of either of the parties. The network that captures these nodes and the relation between them can be found in Fig. 1.

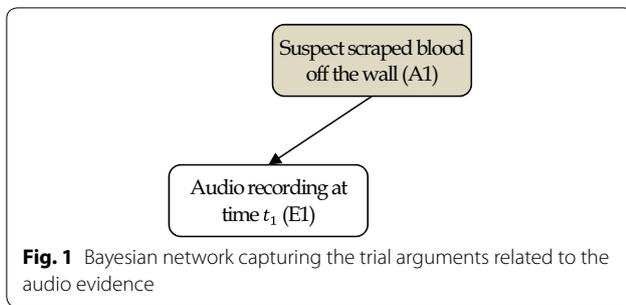


Fig. 1 Bayesian network capturing the trial arguments related to the audio evidence

The appeal

In the preparation for the appeal, the state for the defence proposition $a1_d$ changed when they provided the alternative explanation that the suspect was rubbing his face at t_1 . The second sound ($A2$) can now be added to the network with the accompanying evidence ($E2$). The assumption is made that the activities causing both sounds are dependent in such a way that if the suspect is acting in a certain way at time t_1 , it is reasonable to assume this is related to the activity at time t_2 . Following the case information, there is a link between whether or not the television could be heard (TV) and the activity causing the second sound, due to the distance between the location of the TV and the specific wall. Additionally, the evidence on the similarity (S) between the sounds is conditionally dependent on, logically, the activities that caused both sound 1 ($A1$) and sound 2 ($A2$). The last piece of evidence relevant to these propositions is a statement given by the police (P) that a police officer *observed a frequent rubbing activity by the suspect*. The new network is shown in Fig. 2, highlighting the node ($A1$) that is of ultimate relevance to the appeal.

The conditional probabilities

In order to perform the necessary Bayesian inference to determine the impact of the evidence we must provide all of the NPTs for the model. Typically, these are provided

either from data, experimentation or expert judgment. In the absence of substantive data, in what follows we consider ranges of values where relevant and perform sensitivity analysis over them, relying on the wording used in the case documentation and the verbal scale translation table shown in Table 1. As discussed before, these probabilities are merely used to how BNs can be used in cases where probabilities can be assigned based on more robust empirical grounds.

The activity nodes A1 and A2

The prior probability of activity 1 being scraping is unknown, since this may depend on background information that is not available in this case. Additionally, it was assumed that there is a dependency between the two activities where it is a bit more likely that $A2$ (with states $a2_p$ and $a2_d$) has a certain state given that this was also the case during the first activity:

$$P(A1 = a1_p) = x$$

$$P(A2 = a2_p|A1 = a1_p) = P(A2 = a2_d|A1 = a1_d) = 0.6$$

Note that in the remainder of this paper only the (unique) state names will be used to refer to the state of the nodes in the equations to maintain conciseness [e.g. $P(A2 = a2_p|A1 = a1_p) = P(a2_p|a1_p)$].

The audio evidence E1 and E2

The first evidence to assess are the sounds that are audible on the recording, which will be modelled in the same way and are claimed to be scraping sounds by the prosecution. Therefore, if the suspect was scraping the wall at the particular times (node states $a1_p$ and $a2_p$), it is expected that the audible sound appears to be scraping: $E1$ and $E2$ are *very likely* to be present (node states $e1$ and $e2$):

$$P(e1|a1_p) = P(e2|a2_p) = 0.95$$

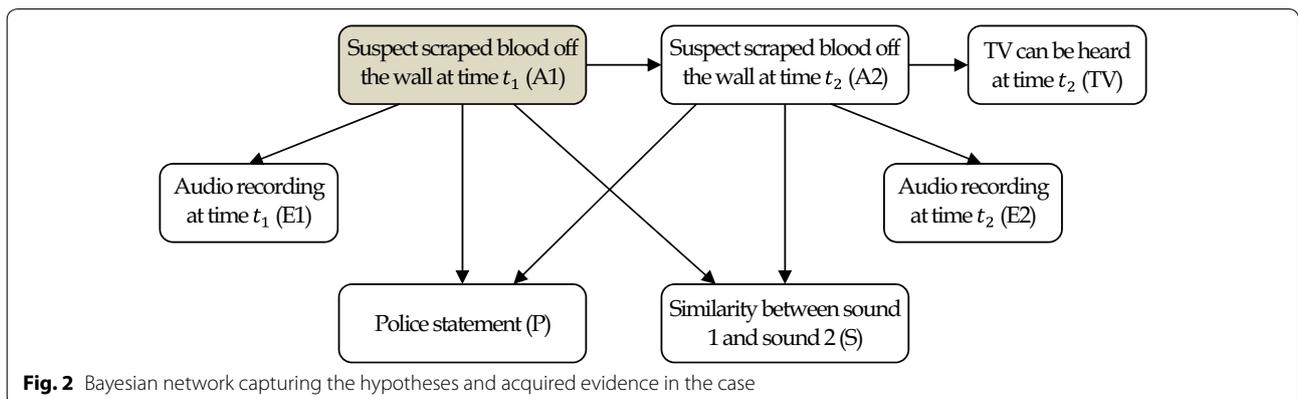


Fig. 2 Bayesian network capturing the hypotheses and acquired evidence in the case

Table 1 Probability ranges for verbal descriptions, taken from Morgan (2009)

Verbal term	Probability range
Virtually certain	>0.99
Very likely	0.9–0.99
Likely	0.66–0.9
Medium likelihood	0.33–0.66
Unlikely	0.1–0.33
Very unlikely	0.01–0.1
Exceptionally unlikely	<0.01

If, however, the suspect was rubbing his face at the particular times ($a1_d$ and $a2_d$), not much can be said about observing a scraping sound since the required information for the conditional probability is unavailable:

$$P(e1|a1_d) = P(e2|a2_d) = y$$

Evidence of similarity *S*

Following the expert statement of the audio expert, a certain level of *similarity* is observed between sound 1 and 2 (with states $S = s$ and $S = \bar{s}$). The probability of observing a similarity ($S = s$) between these sounds if they are produced by the same activity is assumed to be *almost certain*:

$$P(s|a1_d, a2_d) = P(s|a1_p, a2_p) = 0.99$$

The probability of observing a similarity between the two sounds given that each of the sounds are caused by a *different* activity is not assessed. This information is however crucial for the likelihood ratio of the ultimate hypothesis in this case (whether the suspect scraped blood off the wall) and therefore assessing the support of evidence of the expert witness statement on the degree of similarity:

$$P(s|a1_d, a2_p) = P(s|a1_p, a2_d) = z$$

In order to obtain these probabilities, experimental studies should have been performed in which the case situation is simulated and the different sounds resulting compared.

Police statement *P*

The probability of the police observing a rubbing activity and writing this into a statement (states $P = p$ and $P = \bar{p}$) is higher if at both times the suspect rubbed his face than if he only did this at one point (e.g. an increase by 0.05), and less likely if the suspect was scratching the wall both times (e.g. a decrease by 0.05). Additionally, knowing that the suspect was scraping the wall at both points in time does not give an indication as to the likelihood of observing *P*. Since the relative values in the NPT are more important rather than the exact value of the

conditional probability, the following assumptions are made:

$$\begin{aligned} P(p|a1_p, a2_p) &< P(p|a1_d, a2_p) \\ &= P(p|a1_p, a2_d) < P(p|a1_d, a2_d) \end{aligned}$$

$$P(p|a1_p, a2_p) = 0.5$$

Evidence regarding the background noise from the TV

Because the television was, according to the case records, positioned in a downstairs room relatively far away from the upstairs bedroom in which the scraping allegedly took place, it is considered to be *a lot more likely* that the television could not be heard ($TV = \bar{tv}$) than that it could ($TV = tv$), given that the suspect was scraping the wall at time $t2$:

$$P(tv|a2_p) = 0.1$$

$$P(\bar{tv}|a2_p) = 0.9$$

The full set of (conditional) probabilities for the BN model is presented in Fig. 3.

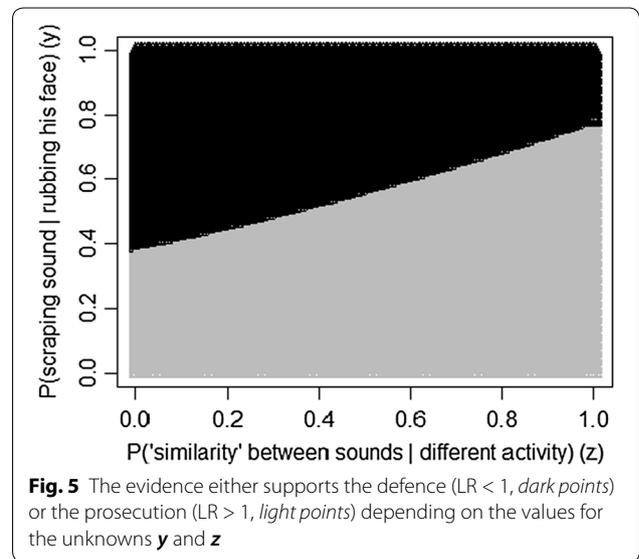
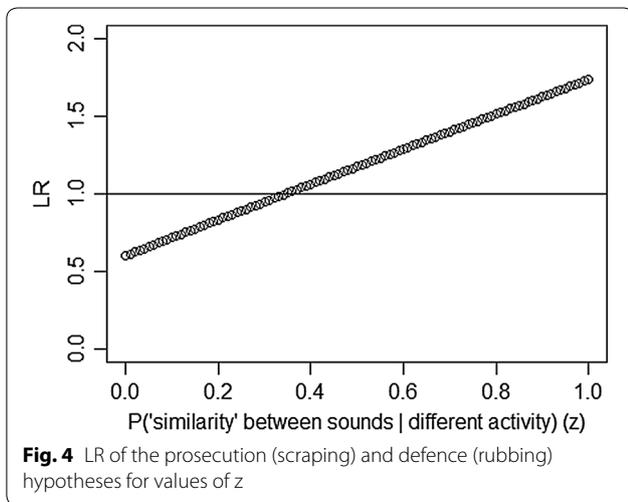
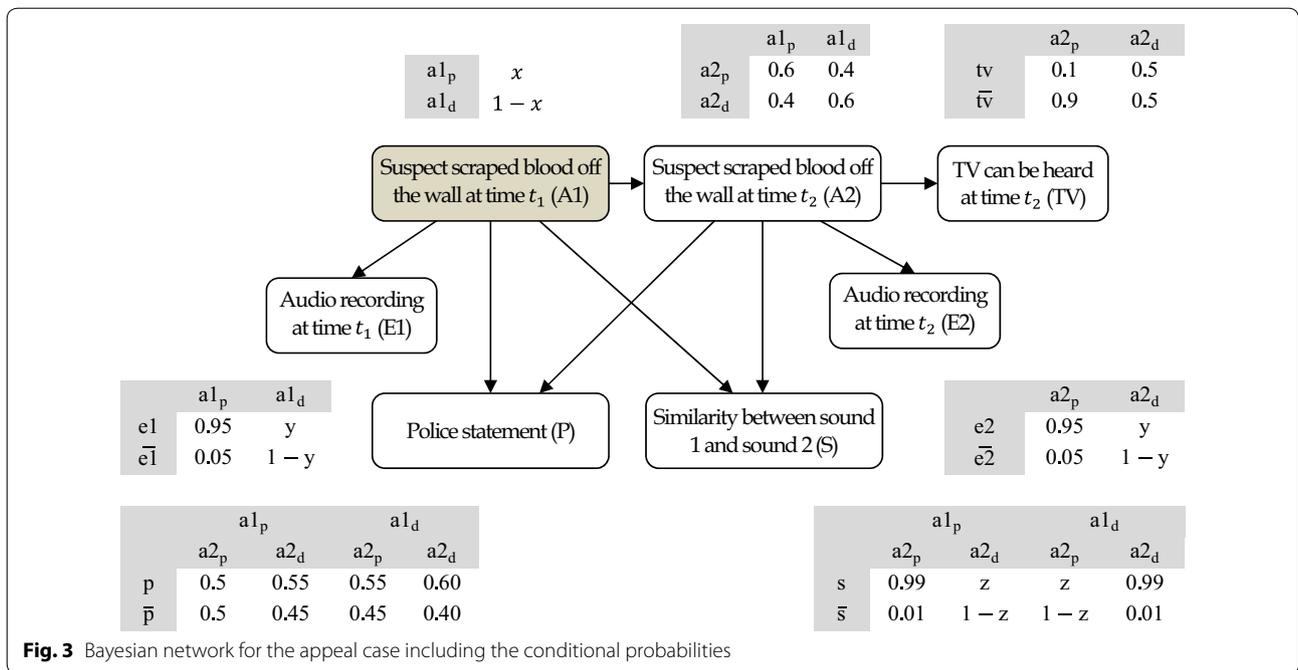
Evaluation of the evidential values

It follows from the BN in Fig. 3 that the probability of *A1* is conditional and depends on different factors that need to be assessed. Even though there is scope for disagreement regarding the probabilities and assumptions, it is possible to make inferences from the BN that can aid the interpretations of evidence within the case.

Evidential value and experimental studies

The question that is of interest in a legal setting is: ‘*does the new appeal evidence change the belief in $a1_p$* ’? And if so, given the aim of the defence for the appeal: ‘*does the presence of the new appeal evidence lower the probability of $a1_p$* ’? The variables x , y and z discussed in the previous section are different in such a way that x (the prior probability of $a1_p$) is based on human considerations while attempts can be made to approach a more precise value of y and z through experimental studies. However, because the expert statement regarding the audio evidence does not provide any information about the likelihood of observing a similarity between two sounds that are caused by a different activity (z), key evidence is missing to come to any conclusion. Figure 4 shows how, for $y = 0.5$, a difference in the z -value changes the LR to such an extent that the evidence supports different hypotheses depending on this value:

- For $0 \leq z < 0.35$, $LR < 1$, indicating that the evidence supports the hypothesis that the suspect was undertaking a normal behaviour, which is beneficial for the defence.

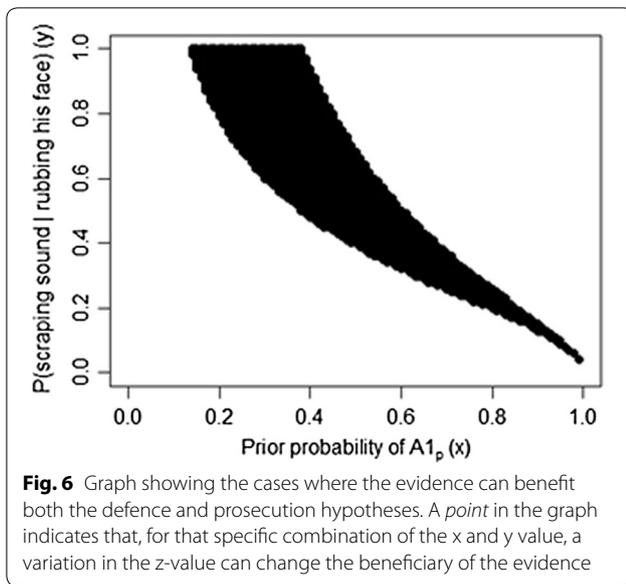


- For $0.35 < z < 1.0$, $LR > 1$, indicating that the evidence supports the hypothesis that the suspect was undertaking a ‘suspicious’ behaviour, which is beneficial for the prosecution.

Following this, Fig. 5 shows that the evidence either supports the defence ($LR < 1$, dark points) or the prosecution ($LR > 1$, light points) depending on the values for the unknowns y and z . This is especially the case for higher values of y .

As noted before, the LR only provides information about to what extent the evidence supports one of the

hypotheses over the other. However, the court is interested in addressing which of the two hypotheses is more likely based on the evidence (confusing these two is labelled the *prosecutors fallacy*). In order to do this, we need to incorporate the prior probabilities of the prosecutor and defence hypotheses, a value which can vary depending on the background information that one assumes to be present ($P(H_p = x)$). Figure 6 shows the results of varying the prior assumption x and the (currently) unknown value of y . A point in the graph indicates



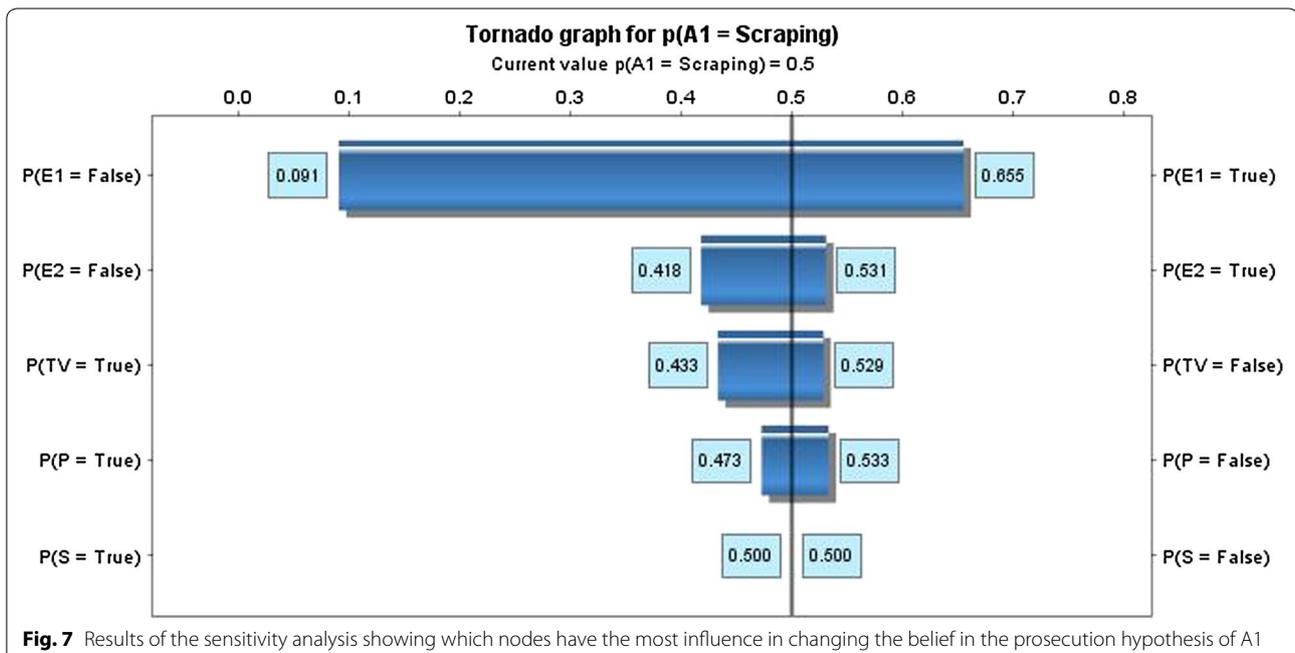
that for that specific combination of x and y , varying the value of the other unknown, z , can cause a switch in which sides' hypothesis is most likely based on the evidence presented at the trial. This is especially the case for lower x values (which would favour the defence) and higher y values (also in favour of the defence). Because of this, the defence should be very careful in presenting the new evidence in its favour when insufficient information is given by the expert with regards to the other unknown variables. However, if experimental studies were to show

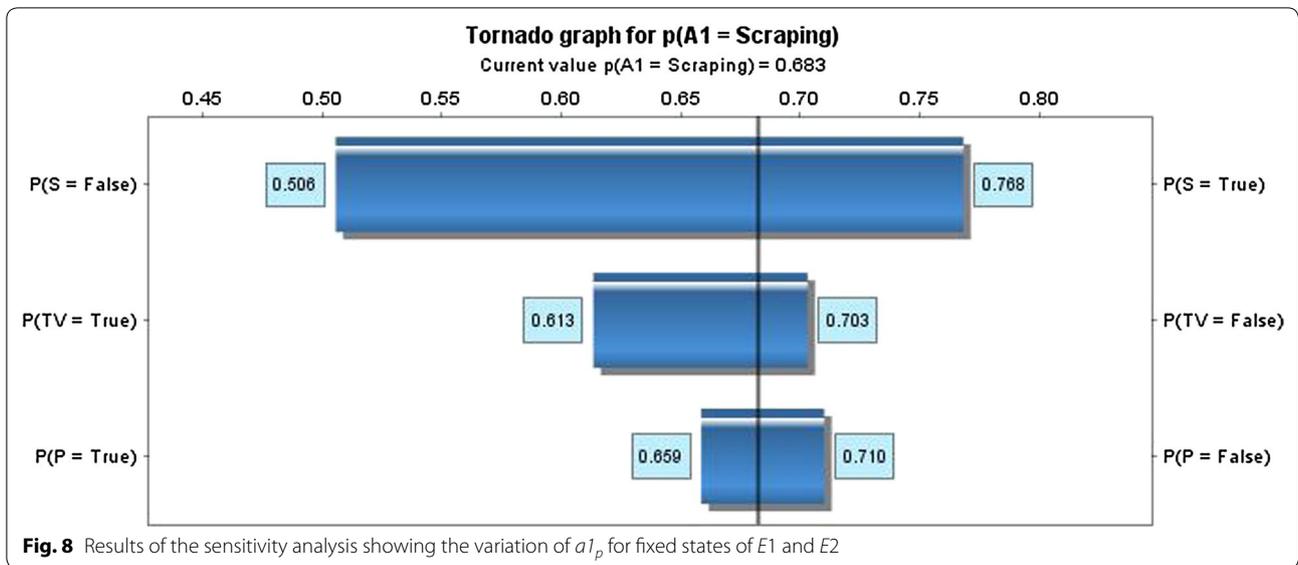
that the two possible activities do produce dissimilar sounds (and the value of z tends towards zero), the expert evidence reduces the posterior probability of the defendant scraping the wall greatly, which was the aim of the defence. Experimental studies are therefore needed to offer the empirical evidence base to establish the missing parameters in a Bayesian network, leaving less room for speculations.

Most influential items of evidence

A sensitivity analysis was performed to identify the items of evidence that have the most impact on the hypothesis A1 (whether the activity that caused the first sound was related or unrelated to the crime). Because y cannot be determined using the current case information, it is assumed that, given that the activity was rubbing, no distinction can be made between the two possible states of the evidence node (a y -value of 0.5). The prior probability x of ($A1 = a1_p$) is also set to 0.5 since no background evidence is assumed to be present before the evidence was presented at the trial. Since the defence assumed a low value for z , a value of 0.1 is used. Note that these values are only selected in order to show the effect and interaction of the items of evidence and are not used to show the evidential value.

The results of the sensitivity analysis are presented in the form of a tornado graph, see Fig. 7. The vertical line shows the prior probability where no evidence is assumed to be present yet ($P(a1_p) = 0.5$). For each of the other nodes in the network, the bars represent the extent an assumption about their state impacts the belief

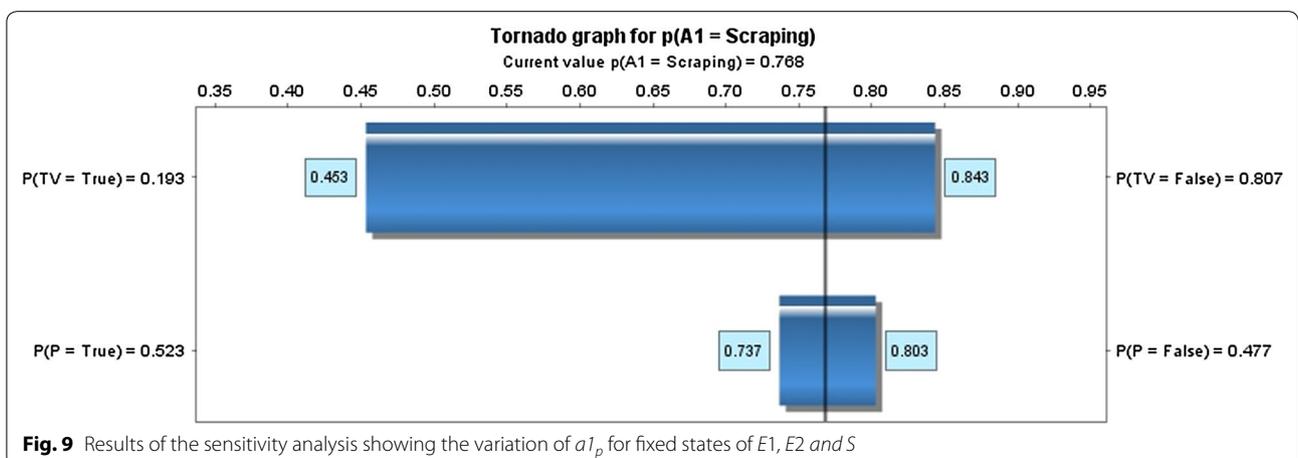




that the suspect was scraping blood off the wall at time t_1 . It follows from Fig. 7 that the probability of the suspect scraping the wall is most impacted by the presence of the two sounds, followed by the observations regarding the TV and the police statement. In this case, the evidence of similarity has negligible impact, since no assumptions are made regarding the presence of the two sounds.

The subsequent assumption is made that the items of evidence $E1$ and $E2$ are observed to draw further inferences. Figure 8 shows the results of the sensitivity analysis. Because of the chosen state of $E1$ and $E2$, the evidence now favours the prosecution without considering any of the other evidence: $P(a1_p|e1, e2) = 0.68$. It follows that the evidence of similarity between the sounds is most important, which can vary the probability of $a1_p$ between 0.51 and 0.77. Subsequently, the evidence related to the television and the observation of the police statement are

less significant in changing the belief in $a1_p$. However, because the value of $a2_p$ is not known, the related items of evidence are dependent where a change in the belief of S influences the state of the television evidence. To further assess this issue, a similar analysis is performed where the evidence of similarity is assumed to be present (Fig. 9). The posterior probability of $a1_p$ increased to 0.77 without considering the states of the police statement and TV evidence. It follows that the television evidence can change the probability of $a1_p$ to lower (0.45) or higher (0.84) than its prior (0.5). This sensitivity analysis thus shows that the inclusion or exclusion of items of evidence in the interpretation process can have a great impact on the significance of other items. Without the presence of the television evidence, the evidence on the similarities brought forward by the defence actually increases the probability of the prosecution hypothesis.



Conclusion

This study has demonstrated how the Bayesian framework of reasoning can be used to explore how items of evidence are related to each other and to the relevant hypotheses, illustrated by a real case. It highlighted the importance of assessing the information and defining the questions and propositions at different levels before creating the BN model. When one makes reasonable and justifiable decisions regarding aspects such as conditional independence, integrity of expert evidence and assumptions that are made in favour of the defence or prosecution, the BN can be a useful tool to provide transparent and workable results regarding the relevance of new items of evidence.

The changes in the key posterior probability was shown to be dependent on the different assumptions, ultimately showing that the incomplete defence expert evidence could also work to weaken the defence case. More specifically, the evidence about similarity of the sounds is the most valuable when aiming to address the activity level proposition, followed by the observation of the TV evidence. Because of this, it is important that experimental studies are undertaken in order to define the probabilistic relationships (Morgan et al. 2009), a process which can be guided using the presented methods. Additionally, it was shown that the importance of certain items of evidence is influenced by the true state of others, illustrating the importance of evaluating the dependencies and incorporating all the relevant available evidence.

This research therefore addresses a critical issue within the forensic sciences. The importance of presenting—in a holistic, transparent and reproducible manner—all the relevant evidence and hypotheses is, we believe, paramount for optimising the use of valuable forensic evidence in court. This approach enables judges and juries to evaluate the weight of not only individual items of evidence in a case but also identify the interaction of different items of evidence within that case to be able to draw conclusions. Such an approach offers a significant step that enables forensic science to address a number of the critiques that have been made in recent years, and to ensure that the value of physical evidence continues to be realised in forensic investigations and the pursuit of justice.

Abbreviations

BN: Bayesian network; CAI: case assessment and interpretation; NPT: node probability table; LR: likelihood ratio.

Authors' contributions

NS, DL and RM conceived the study. NS developed the research design, analysed the data, interpreted the results, and wrote the manuscript. RM, DL and NF participated in the interpretation of the results and provided guidance throughout the research process. All authors read and approved the final manuscript.

Author details

¹ Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK. ² Centre for the Forensic Sciences, University College London, 35 Tavistock Square, London WC1H 9EZ, UK.

³ Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK. ⁴ School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Engineering and Physical Sciences Research Council of the UK through the Security Science Doctoral Research Training Centre (UCL SECRt) based at University College London (EP/G037264/1), and the European Research Council (ERC-2013-AdG339182-BAYES_KNOWLEDGE).

Received: 15 March 2016 Accepted: 18 May 2016

Published online: 25 May 2016

References

- AgenaRisk. (2015). <http://www.agenarisk.com>. Accessed 2 Mar 2016.
- Berger, C. E. H., Buckleton, J. S., Champod, C., Evett, I. W., & Jackson, G. (2011). Evidence evaluation: a response to the court of appeal judgment in R v T. *Science and Justice*, 51(2), 43–49. doi:10.1016/j.scijus.2011.03.005.
- Biedermann, A., Bozza, S., & Taroni, F. (2009). Probabilistic evidential assessment of gunshot residue particle evidence (part I): likelihood ratio calculation and case pre-assessment using Bayesian networks. *Forensic Science International*, 191(1–3), 24–35. doi:10.1016/j.forsciint.2009.06.004.
- Biedermann, A., Taroni, F., Bozza, S., & Mazzella, W. D. (2011). Implementing statistical learning methods through Bayesian networks (part 2): Bayesian evaluations for results of black toner analyses in forensic document examination. *Forensic Science International*, 204(1–3), 58–66. doi:10.1016/j.forsciint.2010.05.001.
- Biedermann, A., Taroni, F., Delémont, O., Semadeni, C., & Davison, A. C. (2005). The evaluation of evidence in the forensic investigation of fire incidents. Part II. Practical examples of the use of Bayesian networks. *Forensic Science International*, 147(1), 59–69. doi:10.1016/j.forsciint.2004.04.015.
- Cook, R., Evett, I. W., Jackson, G., Jones, P. J., & Lambert, J. A. (1998a). A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*, 38(4), 231–239. doi:10.1016/S1355-0306(98)72117-3.
- Cook, R., Evett, I. W., Jackson, G., Jones, P. J., & Lambert, J. A. (1998b). A model for case assessment and interpretation. *Science and Justice*, 38(3), 151–156. doi:10.1016/S1355-0306(98)72099-4.
- Dawid, A. P., & Evett, I. W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences*, 42(2), 226–231.
- Evett, I. W., Gill, P. D., & Jackson, G. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*, 47(3), 520–530.
- Fenton, N. E., Berger, D., Lagnado, D. A., Neil, M., & Hsu, A. (2014). When “neutral” evidence still has probative value (with implications from the Barry George Case). *Science and Justice*, 54(4), 274–287. doi:10.1016/j.scijus.2013.07.002.
- Fenton, N. E., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks*. Boca Raton: CRC Press.
- Fenton, N. E., Neil, M., & Hsu, A. (2013a). Calculating and understanding the value of any type of match evidence when there are potential testing errors. *Artificial Intelligence and Law*, 22(1), 1–28. doi:10.1007/s10506-013-9147-x.
- Fenton, N. E., Neil, M., & Lagnado, D. A. (2013b). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science*, 37(1), 61–102. doi:10.1111/cogs.12004.
- French, J. C., & Morgan, R. M. (2015). An experimental investigation of the indirect transfer and deposition of gunshot residue: further studies carried out with SEM-EDX analysis. *Forensic Science International*, 247, 14–17. doi:10.1016/j.forsciint.2014.10.023.

- Hamer, D. (2012). Discussion paper: the R v T controversy: forensic evidence, law and logic. *Law, Probability and Risk*, 11(4), 331–345. doi:10.1093/lpr/mgs021.
- INTERPOL. (2013). *17th INTERPOL International Forensic Science Managers Symposium*, Review Papers (edited by Prof. Niamh Nic Daeid) (Lyon, 8–10 Oct 2013).
- Jackson, G., Aitken, C. G. G., & Roberts, P. (2013). *Case assessment and interpretation of expert evidence: guidance for judges, lawyers, forensic scientists and expert witnesses*. London: Royal Statistical Society.
- Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, 2(1), 42–52. doi:10.1016/j.jarmac.2013.01.001.
- Morgan, M. G. (2009). *Best practice approaches for characterizing, communicating and incorporating scientific uncertainty in climate decision making*. Collingdale: DIANE publishing.
- Morgan, R. M., & Bull, P. A. (2007). The philosophy, nature and practice of forensic sediment analysis. *Progress in Physical Geography*, 31(1), 43–58. doi:10.1177/0309133307073881.
- Morgan, R. M., Cohen, J., McGookin, I., Murly-Gotto, J., O'Connor, R., Mures, S., & Bull, P. A. (2009). The relevance of the evolution of experimental studies for the interpretation and evaluation of some trace physical evidence. *Science and Justice*, 49(4), 277–285. doi:10.1016/j.scijus.2009.02.004.
- Morrison, G. S. (2012). The likelihood-ratio framework and forensic evidence in court: a response to R v T. *The International Journal of Evidence and Proof*, 16(1), 1–29. doi:10.1350/ijep.2012.16.1.390.
- Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. *Science and Justice*, 54(3), 208–214. doi:10.1016/j.scijus.2013.11.003.
- National Research Council. (2009). *Strengthening forensic science in the United States: a path forward*. Washington D.C.: The National Academies.
- Redmayne, M., Roberts, P., Aitken, C. G. G., & Jackson, G. (2011). Forensic science evidence in question. *Criminal Law Review*, 5, 347–356.
- R v. Weller. (2010). EWCA Crim 1085.
- Regina v. T. (2010). EWCA Crim 2439.
- Robertson, B., Vignaux, G. A., & Berger, C. E. H. (2011). Extending the confusion about Bayes. *The Modern Law Review*, 74(3), 444–455. doi:10.1111/j.1468-2230.2011.00857.x.
- Sjerps, M. J., & Berger, C. E. H. (2012). How clear is transparent? Reporting expert reasoning in legal cases. *Law, Probability and Risk*, 11(4), 317–329. doi:10.1093/lpr/mgs017.
- Taroni, F., Biedermann, A., & Bozza, S. (2014). *Bayesian networks for probabilistic inference and decision analysis in forensic science*. Chichester: Wiley.
- Taroni, F., Biedermann, A., Garbolino, P., & Aitken, C. G. G. (2004). A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International*, 139(1), 5–16. doi:10.1016/j.forsciint.2003.08.004.
- The Law Commission. (2011). *Expert evidence in criminal proceedings in England and Wales*. London: HMSO.
- Thompson, W. C. (2012). Discussion paper: hard cases make bad law—reactions to R v T. *Law, Probability and Risk*, 11(4), 347–359. doi:10.1093/lpr/mgs020.
- Zadora, G. (2009). Evaluation of evidence value of glass fragments by likelihood ratio and Bayesian Network approaches. *Analytica Chimica Acta*, 642(1–2), 279–290. doi:10.1016/j.aca.2008.10.005.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
