When 'neutral' evidence still has probative value (with implications from the Barry George Case)

Norman Fenton[1], Daniel Berger, David Lagnado, Martin Neil, Anne Hsu

July 2013

Abstract

The likelihood ratio (LR) is a probabilistic method that has been championed as a 'simple rule' for evaluating the probative value of forensic evidence in court. Intuitively, if the LR is greater than one then the evidence supports the prosecution hypothesis; if the LR is less than one it supports the defence hypothesis, and if the LR is equal to one then the evidence favours neither (and so is considered 'neutral' - having no probative value). It can be shown by Bayes' theorem that this simple relationship only applies to pairs of hypotheses for which one is the negation of the other (i.e. to mutually exclusive and exhaustive hypotheses) and is not applicable otherwise. We show how easy it can be - even for evidence experts - to use pairs of hypotheses that they assume are mutually exclusive and exhaustive but are not, and hence to arrive at erroneous conclusions about the value of evidence using the LR. Furthermore, even when mutually exclusive and exhaustive hypotheses *are* used there are extreme restrictions as to what can be concluded about the probative value of evidence just from a LR. Most importantly, while the distinction between source-level hypotheses (such as defendant was/was not at the crime scene) and offence-level hypotheses (defendant is/is not guilty) is well known, it is not widely understood that a LR for evidence about the former generally has no bearing on the LR of the latter. We show for the first time (using Bayesian networks) the full impact of this problem, and conclude that it is only the LR of the offence level hypotheses that genuinely determine the probabitive value of the evidence. We investigate common scenarios in which evidence has a LR of one but still has significant probative value (i.e. is *not* neutral as is commonly assumed). As illustration we consider the ramifications of these points for the case of Barry George. The successful appeal against his conviction for the murder of Jill Dando was based primarily on the argument that the firearm discharge residue (FDR) evidence, assumed to support the prosecution hypothesis at the original trial, actually had an LR equal to one and hence was 'neutral'. However, our review of the appeal transcript shows numerous examples of the problems with the use of hypotheses identified above. We show that if one were to follow the arguments recorded in the Appeal judgment verbatim, then contrary to the Appeal conclusion, the probative value of the FDR evidence may not have been neutral as was concluded.

---

[1] Professor and Director of Risk and Information Management Research Group (Queen Mary University of London) and CEO Agena Ltd. Email: n.fenton@qmul.ac.uk

1. Introduction

One way to determine the probative value of any piece of evidence *E* (such as a footprint matching that of the defendant found at the crime scene) is to use the likelihood ratio (LR) [22] [3]. This is the probability of *E* given the prosecution hypothesis (e.g., 'defendant guilty') divided by the probability of *E* given the alternative, complementary defence hypothesis (e.g, 'defendant not guilty'). Increasingly, it is recommended as a 'simple rule' for evaluating forensic evidence in courts [13][10][25][31]. Broader questions about how well the LR can capture the legal concept of relevance are discussed in [27][28][29].

Because the LR involves probabilities – and ultimately some understanding of Bayes' theorem – its actual use in courts is often controversial, as can be seen from the RvT judgement [2], which seemed to suggest that it should only be applicable to evidence (such as DNA) where the relevant probabilities are based on extensive databases of evidence. Numerous papers have criticized the RvT judgement, highlighting its misunderstandings not just about the LR but about the about the role of probabilistic inference in the law generally [8][25][32][34]. It is not the intention of this paper to revisit these arguments. In fact, for simplicity, we will assume that there is no disagreement about the specific probability values used in a given LR (the potential for such disagreement was the focus of the RvT debate and does not need to be repeated). Rather, we focus on a much more fundamental concern about the LR, namely the circumstances under which it actually provides correct information about the probative value of the evidence. We believe this is the first paper to identify these concerns in full.

This paper argues that there are many circumstances in which the actual probative value of evidence may be very different from what can be concluded from the LR. This includes the fact that, contrary to received opinion, evidence with a LR equal to one can often still have significant probative value, i.e. is *not* neutral. Similarly, evidence with LR > 1, may actually have greater probative value on the prosecution hypothesis than on the defence hypothesis (and conversely an LR < 1 can be of greater probative value on the defence hypothesis than on the prosecution hypothesis). This is because there are several significant subtleties to consider when interpreting LRs. Consideration of these subtleties requires careful, precise definitions of the hypotheses and the evidence being evaluated. We will show that, to interpret the LR as a meaningful measure of probative value of evidence (as opposed to a comparison between hypotheses), requires consideration of only pairs of hypotheses that are both mutually exclusive and exhaustive, which means that exactly one of the hypotheses must be true. This point (together with the fact that we cannot sidestep the need to consider prior probabilities when considering the LR) has been considered by others in the research community (see [26][7][12][37][24]). However, in practice, these concerns do not seem to have been well understood, and we will show that even the most senior evidence experts have encountered difficulty in formulating relevant hypotheses that are mutually exclusive and exhaustive. Second, even when hypotheses *are* mutually exclusive and exhaustive, there remains the potential during a case to confuse what in [11] were referred to as *source-level* hypotheses (such as blood at the scene belonging to or not belonging to the defendant) and *offence-level*

hypotheses[2] (such as defendant being guilty or not guilty). Sometimes one may mutate into another through slight changes in the precision with which they are expressed. A LR for the source-level hypotheses will not in general be the same as for the offence-level hypotheses. Indeed, we will show it is possible that an LR that strongly favours one side for the source-level hypotheses can actually strongly favour the *other* side for the offence-level hypotheses even though both pairs of hypotheses seem very similar. Similarly, an LR that is neutral under the source-level hypotheses may actually be significantly non-neutral under the associated offence-level hypotheses.

To illustrate the issues we raise, we use the Barry George Appeal judgment [1] in which the use of LR gained widespread attention because of it central role. We believe there are examples of many of the above problems in the transcript. Barry George had previously been convicted of the murder of TV celebrity Jill Dando. In the Appeal it was argued that the Firearm Discharge Residue (FDR) evidence, that had formed a key component of the prosecution case at the original trial, actually had a LR equal to one. The defence argued that this meant that the evidence was 'neutral' i.e. it had no probative value. The Judge duly quashed the original conviction as unsafe. Our critique of the Barry George appeal case is aimed towards the judgment transcript and not the actual expert testimonies during the trial. We have good reason to believe that careful testimonies may have been inaccurately presented in the appeal judgment. The extent of the confusion and mistaken reasoning present in the judgment document shows that these issues regarding the interpretation of the LR remain widely misunderstood.

In Section 2 we provide an overview of the role of likelihoods and the definiton of LR. We explain exactly what is meant by *probative value* of evidence and why the LR may be used to evaluate this. We also explain precisely what is meant by 'neutral' evidence. Our presentation clears up a number of widely held misunderstandings. In particular, we show why Bayes' theorem is critical and that the use of prior probabilities for hypotheses cannot be side stepped (many texts assume that the LR can be understood without either Bayes' theorem or the consideration of priors). In Sections 3 and 4 we focus on the special case of evidence for which the LR is one. Withthe help of Bayesian networks we use scenarios to exemplify how, in many circumstances, a LR of one does not ensure neutral evidence. Specifically, in Section 3, we show examples where the hypotheses are not mutually exclusive and exhaustive. In Section 4 we show that, even when evidence has a LR of one for mutually exclusive and exhaustive hypotheses (thus, really is neutral with respect to those hypotheses), the evidence has probative value. This means it is not neutral with regard to other relevant hypotheses; this includes the offence-level hypotheses of whether or not the defendant is guilty. Section 5 provides a thorough analysis of the Barry George appeal case judgment and shows how this document contains many examples of hypotheses used for the FDR evidence that were potentially not mutually exclusive and were not properly linked to the offence-level hypotheses. We demonstrate that if one were to follow the arguments recorded in the Appeal judgment verbatim, the probative value of the FDR evidence may not have been neutral (contrary to the Appeal conclusion) bur rather still supported the prosecution.

---

[2] Also often referred to as 'ultimate' hypotheses

Some of what appears in Sections 2-4 is known to probability experts and a small number of forensic experts, but the ramifications do not appear to have been made explicit anywhere, nor have there been appropriate examples demonstrating the problems. This is the first paper to reveal the full extent of the problems. We use the formalism of Bayesian networks [17][36] both to model explicitly the causal relationships between hypotheses and evidence and also to automatically compute the necessary probability calculations. However, to ensure as wide a readership as possible most of the necessary calculations and detailed model descriptions appear only in the supplementary material [38]. The models themselves (which can be run in the free version of the sotware tool [3]) are all provided in supplementary material [39].

2. Likelihoods, the likelihood ratio and the probative value of evidence

Any legal trial seeks to determine whether one or more hypothesis is either true or false. In the simplest case the prosecution has a single hypothesis H_p (defendant guilty) and the defence has a single alternative hypothesis H_d (defendant innocent). In this simplest case we assume that H_d is the same as "not H_p" (formally this means that H_p and H_d are mutually exclusive and exhaustive events).

Belief in a hypothesis is expressed as a probability. The ***prior*** probability of a hypothesis H_p, written $P(H_p)$, is the probability of H_p before we observe any evidence. When there are two mutually exclusive and exhaustive hypotheses, H_p and H_d, the greater our belief in one, the less our belief in the other since $P(H_d) = 1-P(H_p)$ by a basic axiom of probability. When we observe evidence E we revise our belief in H_p (and similarly H_d). This revised probability is called the ***posterior probability*** of H_p and is written $P(H_p \mid E)$ which means the 'probability of H_p given E'. Bayes' theorem (see Appendix 1) provides a formula for computing this posterior probability. If the posterior probability is greater than the prior probability then it makes sense to say that the evidence E ***supports*** the hypothesis H_p, because our belief in H_p has increased after observing E. And if our belief in H_p has *increased* then our belief in H_d must have decreased since they are mutually exclusive explanations for the evidence, E. So, in such situations, it is both natural and correct to say that the evidence supports H_p over H_d. The bigger the increase the more the evidence E supports H_p over H_d.

Because many lawyers assume that prior probabilities are for jury members only (as they are 'personal and subjective') it is widely assumed that they should not be considered in court by forensic experts [17]). Instead, a comparison of the probability of evidence E being found under both of the hypotheses is used to capture the probative value of evidence. Specifically, we compare

- The probability of E assuming H_p is true - this is written $P(E \mid H_p)$ and is called the ***prosecution likelihood***

- The probability of E assuming H_d is true - this is written $P(E \mid H_d)$ and is called the **defence likelihood**[3]

and calculate the **likelihood ratio (LR)**[4], which is the prosecution likelihood divided by the defence likelihood.

$$\frac{P(E \mid H_P)}{P(E \mid H_d)}$$

A simple example of how the LR describes the impact of evidence on hypotheses is shown in Appendix 1. We also prove in Appendix 1 that when prosecution and defence hypotheses are mutually exclusive and exhaustive, a LR of greater than one supports the prosecution hypothesis and a LR of less than one supports the defence hypothesis. Hence, the LR has a simple interpretation for the probative value of the evidence under these assumptions.

The proof of the probative value of evidence in terms of the LR depends on Bayes' theorem. Typically textbooks 'prove' the simple LR rule by comparing the prior odds (of the prosecution hypothesis against the defence hypothesis) with the posterior odds. This 'odds' approach (which is also explained in Appendix 1) is considered a 'simple rule' because it demands only that we consider relative probabilities of alternative hypotheses rather than additionally focus on the prior probabilities of one or other hypothesis. However, we believe that this rule is confusing. Not only does it hide the assumption that the hypotheses need to be mutually exclusive for it to be correct, but it also fails to tell us clearly what we most need to know: namely, that for the evidence E to 'support' the hypothesis H_p it is necessary that the posterior probability of H_p, i.e. $P(H_p \mid E)$, is greater than the prior probability $P(H_p)$: in other words our belief in H_p being true increases after we observe E.

This also leads us to a natural and rigorous definition of 'neutral' evidence. Specifically, the evidence E is neutral for H_p if the posterior is unchanged from the prior after observing the evidence, i.e. $P(H_p \mid E) = P(H_p)$. Appendix 2 provides a mathematical proof that, when H_p and H_d are mutually exclusive and exhaustive and the LR equals one, then the evidence is neutral for H_p and must also be neutral for H_d and vice versa. However, Appendix 2 also proves that when H_p and H_d are **not** mutually exclusive and exhaustive, all we can actually conclude when the LR is equal to one is that the ratio of the posterior probabilities of H_p and H_d is equal to the ratio of the prior probabilities. In Section 3 we will show examples where the evidence in such cases is **not** neutral with respect to H_p and H_d. First, however, there are two fundamental, points that must be noted about the limitations of the use of the LR that are not widely understood:

---

[3] Other researchers have correctly pointed out the difficulties in defining the impact of evidence on a single defence hypothesis H_d because in actuality H_d is often made up of multiple hypotheses that are difficult to articulate and quantify [7]. This has been described in detail for DNA evidence [26]. However, for our purposes we will assume that $P(E \mid H_d)$ can be meaningfully computed.

[4] It is argued that one of the most important benefits of the LR is that it forces experts to consider both the prosecution and defence likelihoods. Hence it helps avoid well-known problems associated with focusing on a single likelihood, such as interpreting a low defence likelihood as synonymous with a low probability of H_d being true, or even worse committing the prosecutor's fallacy [17] .

5

The 'prior misconception': the LR is popular with forensic experts precisely because it can be calculated without having to consider any prior probabilities for the hypotheses [30]. But this is something of a misconception for two reasons. First, the LR actually tells us nothing about the probability that either hypothesis is true, no matter how high or low it is. We can only make conclusions about such (posterior) probabilities if we know the prior probabilities. Although this observation has been well documented [16][23] this issue continues to confound not just lawyers, but also forensic experts and statisticians. An indication of the extent of the confusion can be found in one of the many responses by the latter community to the RvT judgement. Specifically, in the otherwise excellent position statement [5] (signed by multiple experts) is the extraordinary point 9 that asserts:

> "It is regrettable that the judgment confuses the Bayesian approach with the use of Bayes' Theorem. The Bayesian approach does not necessarily involve the use of Bayes' Theorem."

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

The second reason why it is a misconception is because it is impossible to define $P(E \mid H_p)$ and $P(E \mid H_d)$ meaningfully without knowing something about the priors $P(H_p)$, $P(H_d)$ (in strict Bayes' terms[5] we say the likelihoods and the priors are all ***conditioned on some background knowledge K***). For example, 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 one, there is a problem with the defence likelihood, $P(E \mid H_d)$. For DNA evidence such as this, the defence likelihood is usually assumed to be the random match probability (RMP) of the DNA type. This can typically be as low as one in a 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 :

a) $P(H_p) = 0.5$, where the defendant is one of two people seen grappling with the victim before one of them killed the victim;
b) $P(H_p) = 1/40$ million.where nothing is known about the defendant other than he is one of 40 million adults in the UK who could have potentially committed the crime.

Whereas a value for $P(E \mid H_d) = $ RMP seems reasonable in case b), it is clearly not in case a). In case a) the defendant's DNA is very likely to be on the victim irrespective of whether or not he is guilty. 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.

---

[5] Specifically, the priors $P(H_p)$, $P(H_d)$, really refer to $P(H_p|K)$ and $P(H_d|K)$ 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 $P(E|H_p, K)$ and $P(E|H_d, K)$.

<u>The simple evidence misconception.</u> In many cases the 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. Consider the following example E: "tiny matching DNA trace found". Suppose that the DNA trace has a profile with a random match probability of 1/100 (such relatively 'high' match probabilities are common in low-template samples [6]). It would be typical to assume that $P(E \mid H_p) = 1$ and that $P(E \mid H_d) = 1/100$ 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 actually need the formalism of Bayesian networks [20][21][36] and its supporting tools [19], as explained in Appendix 3. In fact Appendix 3 describes the two components of the Bayesian network (BN) method that we use throughout this work:

(i)   A method for defining the nodes of the BN in such a way as to carefully distinguish between those hypotheses that are mutually exclusive and exhaustive and those that are not. This ensures that experts are only ever required to consider pairs of mutually exclusive hypotheses when specifying likelihoods.

(ii)   Modelling the causal story between hypotheses and different pieces of evidence.

Appendix 3 shows how to model the separate pieces of evidence $E1$ and $E2$ (in the above example) and their relationship to H_p and H_d in a BN. The resulting calculations confirm that, under a set of very reasonable assumptions, the LR is less than one. Hence, the evidence actually supports the defence hypothesis H_d rather than H_p.

3. The problems when hypotheses are not mutually exclusive and exhaustive

When the assumption of mutually exclusive and exhaustive hypotheses is either wittingly or unwittingly undermined, the relationship between the LR and the notion of 'probative value' of the evidence can change dramatically.

3.1   *Hypotheses that are not mutually exclusive*

If H_p and H_d are not mutually exclusive then H_p and H_d could *both* be true.

Example 1: Suppose a defendant is charged with murder and that:

> H_p is "defendant guilty" and
> H_d is "defendant not at the crime scene"

> Then H_p and H_d may **both** be true (as would be the case if the defendant paid a hired killer). It is also the case that neither may be true. Suppose the priors for H_p and H_d are both 0.5. Now suppose we get the following evidence E

> > E: Ten minutes before the crime took place the defendant – seen at a different location - was overheard on the phone saying 'go ahead and kill him'.

The evidence E clearly supports *both* hypotheses, and it is quite conceivable that both $P(E \mid H_p)$ and $P(E \mid H_d)$ are equal. In supplement [38] Section 1 we provide the details of a situation in which both are equal to 0.9 and so the LR equals one. The supplement [38] also shows that $P(H_p \mid E) = P(H_d \mid E) = 0.666$ in this case. So the ratio of the posterior probabilities has remained the same as the ratio of the prior probabilities but the probability of both hypotheses have increased from 0.5 to 0.666. Because of the nature of the hypotheses, the fact that the probability of guilt has increased, even though the defence hypothesis has increased by a similar amount, means that the evidence E has genuine probative value despite its LR being equal to one. In this case we are not only more confident of guilt, but we are also more confident that the suspect was not at the crime scene. In other words, the increase in the probability of the defence hypothesis does not necessarily increase belief in innocence, but rather provides an explanation for guilt.

In the case where there is separate evidence E, such as a murder motive, to support H_p the probative value of the supposedly 'neutral' evidence E can become even more dramatic as is shown in supplement [38] .

If readers feel that Example 1 is unsatisfactory because the likelihood values were 'plucked out of the air' then the following example demonstrates the same core point for likelihoods that are indisputable:

> **Example 2:** The defendant rolls two dice – a black die which he owns and a red die randomly selected by a member of the public from a batch provided by a reputable dice company. The evidence E against the defendant is that both dice rolls are 6s. However, in this case the prosecution hypothesis H_p concerns only the black die, while the defence hypothesis H_d concerns only the red die:

> H_p: "The black die is fixed with all sides being 6s"
> H_d: "The red die is fixed with all sides being 6s"

The reason H_d is the defence hypotheses is because it was subsequently discovered that the red die came from a batch in which 50% were faulty in the sense of having all sides sixes. Hence, the prior $P(H_d) = 1/2$. Suppose the prior $P(H_p) = 1/2$ because it is known that 50% of the defendant's black dice are fixed with all sides being 6s. In supplement [38] Section 2 we prove that $P(E \mid H_p) = P(E \mid H_d) = 7/12$, so the LR is 1. But the evidence is **not** neutral. We show that the posterior probability of H_p is 6/7

and so the prosecution hypothesis is now very likely. Just because the defence hypothesis has increased by the same amount as the prosecution hypothesis, is essentially irrelevant.

We also use a slight variation of the dice example to show in supplement [38] Section 3 that it is possible to have an LR of less than 0.5 even though the evidence favors the prosecution hypothesis rather than the defence. The variation is to assume that 10%, rather than 50%, of the red dice were faulty. In this case $P(E \mid H_d) = 7/12$ as before. However, $P(E \mid H_p) = ¼$ meaning the LR is 0.43. According to the guidelines on the LR in [30] we should be able to conclude that the evidence provides support for the defence hypothesis. But this is nonsense because what matters is that, as before, $P(Hp \mid E) = 6/7$.

A further problem introduced by non-mutual exclusivity is that likelihoods are apt to be incorrectly reported. This is because the following probabilities are no longer all the same:

$$P(E \mid H_p)$$
$$P(E \mid \text{not } H_d)$$
$$P(E \mid H_p \text{ and not } H_d)$$

If an expert fails to understand that H_p and H_d are not mutually exclusive it is likely that when asked to give the likelihoods for $P(E \mid H_p)$ and $P(E \mid H_d)$ they will provide instead $P(E \mid H_p \text{ and not } H_d)$ and $P(E \mid H_d \text{ and not } H_p)$, which are *not* equal to $P(E \mid H_p)$ and $P(E \mid H_d)$. In this circumstance, it is possible that the likelihoods provided by the expert are equal, i.e. $P(E \mid H_p \text{ and not } H_d) = P(E \mid H_d \text{ and not } H_p)$ when the true likelihoods, $P(E \mid H_p)$ and $P(E \mid H_d)$, are not equal at all. Supplement [38] Section 4 demonstrates exactly such a scenario using the same dice example. This shows how it is easy to be deceived into thinking evidence might have a LR of one when it does not.

3.2   *Hypotheses exclusive but non-exhaustive*

If H_p and H_d are exclusive but not exhaustive then although they cannot both be true, it is possible that *neither* may be true.

> **Example 3:** Fred and Bill attempt to rob a man. When the man resists he is struck on the head and dies. During the long police investigation Bill dies while on remand so Fred is tried on his own for murder. Fred's defence is that it was Bill and not him who struck the blow to the head. Hence we have:
>
> H_p is "Fred struck fatal blow"
> H_d is "Bill struck fatal blow"

Although these hypotheses are mutually exclusive they are not exhaustive since they fail to consider the possibility H_a that the man did not actually die from the blow to the head. Suppose we know that 10% of people who die while being physically assaulted actually die from a heart attack induced by the stress (hypothesis H_a). Then it is reasonable to assume that the priors for H_p and H_d are both 0.45 and the prior for H_a is 0.1. Now suppose we discover the following evidence E:

9

> *E*: The dead man is known to have had a heart attack six months before the assault.

It is clear in this case that $P(E \mid H_p) = P(E \mid H_d)$ and hence the LR is 1 with respect to the hypotheses for H_p and H_d. However, it is also clear that, whatever the value is for $P(E \mid H_p)$ and $P(E \mid H_d)$, the value for $P(E \mid H_a)$ is much higher. The effect of this is that, after observing *E*, the posterior probabilities of both H_p and H_d reduce (albeit by the same amount). In particular, $P(H_p)$ drops after observing *E*. So, even though its LR is equal to one the evidence has clear probative value.

Supplement [38] Section 5 shows that if $P(E \mid H_p)$ and $P(E \mid H_d)$ equal 0.01 and $P(E \mid H_a)$ is 0.5, then the posterior probability of H_p drops from 0.45 to 0.07627.

3.3   Ensuring hypotheses are mutually exclusive and exhaustive can be extremely difficult

It follows from the above that, in order to use the LR to determine the probative value of evidence, the need to select hypotheses that are mutually exclusive and exhaustive is paramount. In practice (we will show this more extensively in Section 5) it is easy to veer towards pairs of hypotheses that are not mutually exclusive and exhaustive. To give an indication of the extent of this problem we refer to the paper [28] which includes an extensive discussion between leading evidence experts about the LR.

On page 8 of [28] Allen introduces an example where, for a pair of prosecution/defence hypotheses he shows that a piece of evidence has LR equal to one, but he claims the evidence is clearly probative (favours the prosecution) and admissible. He finds this paradox puzzling and feels it exposes some problems with the LR and Bayesian reasoning. Various experts (including those who write extensively about Bayes' theorem) subsequently weigh into the debate with complex philosophical arguments and explanations, and reasons why Bayes and the LR cannot be used etc. But *not one of them* identifies that Allen's example is fundamentally flawed because it is based on a pair of hypotheses that are neither mutually exclusive nor exhaustive[6]. Allen's example is of a defendant accused of murder committed in a small town. The evidence *E* is that the defendant was seen driving to town shortly before the murder was committed. The 'prosecution' and 'defence' hypotheses are respectively:

- H_p: "Defendant drove to town to commit murder"
- H_d: "Defendant drove to town to visit his mother who lives there"

These are neither mutually exclusive nor exhaustive because it is possible that *both* H_p and H_d could be true and also that *both* could be false.

Hence, none of the evidence experts in [28] appears to have grasped the implications explained in Sections 3.1 and 3.2: That unless the hypotheses are mutually exclusive

---

[6] In [29]Picinali discusses the example further and implicitly addresses some of the concerns we raise here.

and exhaustive the LR may tell us nothing about the probative value of the evidence and that an LR of one certainly does not mean 'no probative value'. All that was needed was for someone to point out to Allen that - by the most simple application of Bayes theorem - an LR of one did not mean in this case that the evidence was not probative. End of discussion. We highlight this case to point out how easy it is to make the mistakes we describe above: If even the world's leading evidence experts can fail to notice non-mutually exclusive hypotheses, it is something the legal community must pay attention to. It cannot be dismissed as yet another 'obvious' or 'trivial' fact because clearly it is easily overlooked.

Allen's hypotheses mix up the evidence ("seen driving to town") with the ultimate hypothesis ("committed murder"). Careful use of the BN causal modelling method we describe in Appendix 3 would help to avoid this problem, much as it helped to disentangle separate pieces of evidence presented as a single piece of evidence. The only natural way to model it - and to show that Allen's example is not a paradox at all - is to use the causal model (this is an example of a Bayesian network) shown in Figure 1. Here we have a Boolean node "Committed murder" (whose true and false states represent the offence level hypotheses), a separate Boolean node "Visited mother" (whose true and false states represent the source level hypotheses introduced by the defence) and a separate evidence node E "seen driving to town".

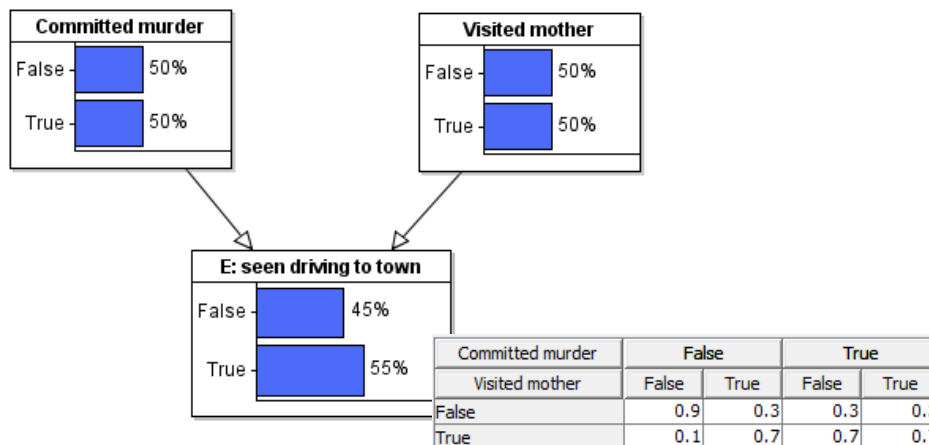| Committed murder	False		True	
Visited mother	False	True	False	True
False	0.9	0.3	0.3	0.3
True	0.1	0.7	0.7	0.7

Figure 1. Model and prior probabilities for Allen example. The prior conditional probablities for E are shown in the table, where for example the entry 0.1 is the probability that E is true given that both 'committed murder' is false and 'visited mother' is false

In the model, as priors we use Allen's assumption that:

P(E | Murder) = P(E | visited mother)

We set each of these to be equal to 0.7 although the exact values chosen do not affect the overall argument that follows) and for simplicity we assume 'ignorant prior' assumptions for the two pairs of hypotheses (this choice also does not affect the following argument.

Allen is wrong to conclude that the LR of the evidence with respect to the prosecution hypothesis is one. Instead of comparing P(E | Murder) to P(E | visited mother) as he does we have to compare it to P(E | not Murder).

11

In fact, while P(E | Murder) = 0.7, it turns out that P(E | not Murder) = 0.4 so the LR is not one but rather favours the prosecution hypothesis of 'Murder' being true.

When we perfom the calculations for this model after observing E (i.e. E is set to 'true') we get the result shown in Figure 2.
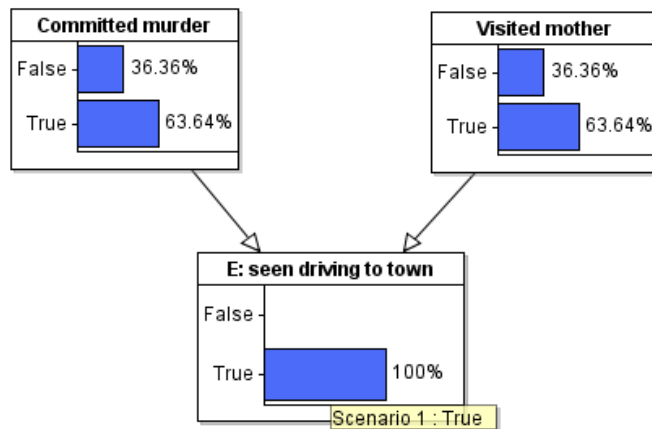

Figure 2 Updated probabilities after observing evidence E

So Allen *is* correct in his intuition that E supports the prosecution hypothesis because P(Murder | E) is greater than P(Murder). The fact that it ALSO supports the defence hypothesis (with the same increase in probability) simply confirms what is written in Section 3.1 above (i.e. it is a consequence of non mutually exclusive hypotheses).

The only way to really avoid Allen's supposed paradox without introducing more than a single pair of hypotheses is to have very explicit hypotheses that are genuinely mutually exclusive and exhaustive hypotheses, namely:

H_p: "(defendant committed murder) AND (defendant did not visit mother)"

not H_p: "(defendant did not commit murder) OR (defendant visited mother)"

By De Morgan's law those are negations of each other.

But the problem with such unnatural hypotheses is that it no longer seems either reasonable or feasible to assert that P(E | H) = P(E | not H) in this case (if it were then the evidence really would be neutral).

4. The problem with neglecting the impact of evidence on offence-level hypotheses

Even when the hypotheses *are* mutually exclusive and exhaustive, the LR of the evidence may tell us nothing about its probative value on other superficially similar hypotheses. This is especially concerning if the source-level hypotheses (for which the evidence is neutral) are confused with offence-level hypotheses (the defendant is

innocent/guilty). Such concerns have been raised by other researchers in previous work [23][12], but without any detailed elaboration.

Hence, we next consider the case where we have mutually exclusive and exhaustive source-level hypotheses, and evidence that has a LR of one, but which still has probative value for the (very closely related but different) offence-level hypotheses (which are also mutually exclusive and exhaustive).

> **Example 4***:* Suppose two men, Fred and Joe live at the same address. It is known that gun X is registered to that address, but the identity of the owner is not known. A man, Bob, is found murdered from a gun shot. Shortly after the murder the police find the following evidence *E*: there is a gun in Fred's house with firearm discharge residue (FDR) that matched that from the crime scene. Fred is charged with the murder of Bob. The offence level hypotheses are:
>
> H_p: Fred fired the shot that killed Bob
> H_d: Fred did not fire the shot that killed Bob
>
> However, at the trial, instead of focusing on the offence-level hypotheses, the lawyers and experts focus on the question of who owned the gun, which they assume is directly related to the question of guilt. Specifically, they consider:
>
> $H1_p$: Fred owned the gun that killed Bob
> $H1_d$: Fred did not own the gun that killed Bob
>
> These source-level hypotheses are mutually exclusive and exhaustive and so there is no possibility of making any of the errors highlighted in Section 3.

If we assume (as is reasonable) that the gun is equally likely to be owned by Fred or Joe, then it follows that the probability of the evidence *E* given $H1_p$ is the same as the probability of the evidence *E* given $H1_d$. Hence, the LR of the evidence is one. A defence lawyer would therefore be correct in concluding that the evidence is neutral with respect to the hypotheses $H1_d$ and $H1_p$. However, if he were to also conclude that "the evidence has no probative value as it supports neither the prosecution nor defence case" he would be wrong. The evidence is *not* neutral with respect to the offence-level hypotheses H_p and H_d.

Again we use the modelling method proposed in Appendix 2 to arrive at the causal BN model shown in Figure 3.
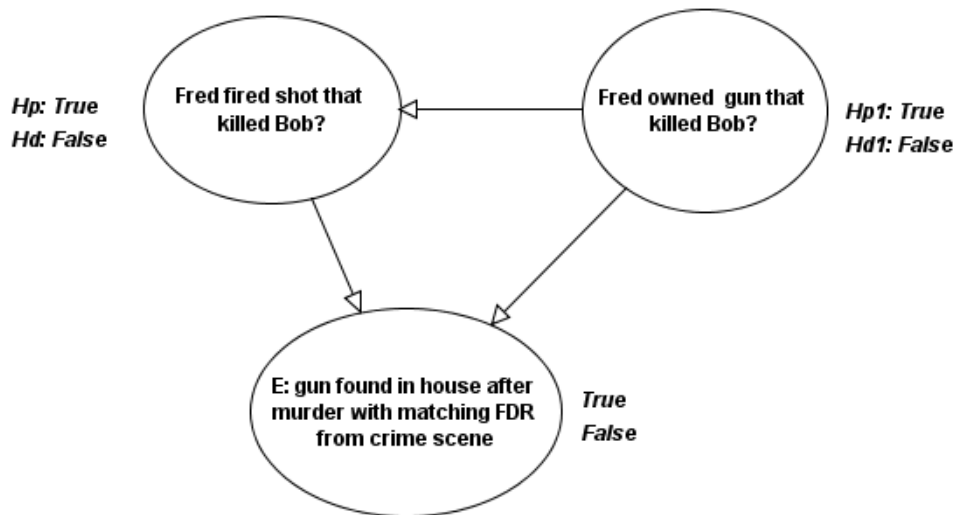
Figure 3 Simple BN for example 3

Using the following reasonable assumptions:

- The prior probability of $H1_p$ and $H1_d$ are equal (both 1/2)
- The probability of finding the evidence, E, is higher if Fred did fire the shot that killed Bob than if he did not fire the shot
- Fred is just as likely to have fired the shot that killed Bob whether he owns the gun or not. We assume the probability in each case is 1/100 (so the prior probability of H_p is 1/100).

We show in supplement [38] Section 6 that, while the posterior probability of the source-level prosecution hypothesis $H1_p$ remains unchanged after we enter the evidence E as true, the offence level prosecution hypothesis H_p changes from a prior of 1% to a posterior of 9.1%. Thus the evidence that was 'neutral' with respect to whether Fred owned the gun has real probative value towards the ultimate hypothesis of Fred's guilt.

It follows from the above discussion that the LR of any piece of evidence should be calculated against the offence level hypotheses. If (as is usual) the LR determined by forensic experts was calculated for source level hypotheses, then this will involve the kind of calculations that are automatically computed using an appropriate BN model and tool, but which are difficult and tiresome to calculate manually.

There is one additional problem that leads to LRs that can easily be misunderstood: in many situations even the offence level hypotheses "guilty or not guilty" may actually hide implict assumptions that suggest the hypotheses are not exhaustive (so that we are back in the situation described in Section 3.2). An example would be where there is an implict assumption that a crime has actually taken place (for example, a murder) even though there is a small probability (ignored by both sides) that the victim was not actually murdered. In such a situation the 'offence level' hypotheses H_p ("guilty") and H_d ("not guilty") may be wrongly assumed to be exhaustive because what they actually mean are, respectively, "murder was committed by the defendant" and "murder was committed by somebody other than the defendant". What is missing is the hypothesis H_n "no crime committed". Supplement [38] Section 7 provides a

14

comprehensive example of this in which we have the evidence *E*: "victim left a suicide note".

5. Implications for Barry George case

The scenarios above show that great care must be taken when using the LR. In particular, a LR of one can still have significant probative value depending on the choice of hypotheses and the precise meaning of the evidence. This has implications for the high profile case of Barry George, which centred around a piece of evidence that was later dismissed for having no probative value because it had a LR of one. We first provide a brief overview of the case. We then explain how the transcripts from the judgment of the Case Appeal [1] show how discussions of LRs can easily drift into examples where the simple interpretation of the LR was ill-formed and confused. We emphasize that the confusions highlighted are based on the Case Appeal judgment text alone, and we do not make any claims about how these corresponded to the actual arguments made during the appeal trial. Indeed, it is impossible to know how many of the confusions we describe below can be attributed to actual confusions made during the appeal versus confusions during the writing of the judgment document.[7]

Nevertheless, the muddled state of the Case Appeal judgment document itself is evidence that LR arguments are easily confused in such a way that simple interpretations of the LR are not warranted. In particular: the document shows hypotheses (for which likelihoods are reported) that are ill-defined, and appear to drift over the course of the case; the precise definition of the evidence associated with the likelihoods reported also appear to change over the course of the case; the relationship among the multiple hypotheses being considered, and their relationship to the ultimate hypotheses, are unclear; causal explanations are muddled with diagnostic reasoning, which may result in a confused interpretation of the statistics provided by expert witnesses.

5.1   Summary of the Case and the Appeal

In 2001 Barry George (BG) was convicted of the murder of TV celebrity Jill Dando (JD), who had been shot dead in 1999. An important part of the prosecution case centred on the following piece of evidence *E*:

> *E*: A single particle of firearm discharge residue (FDR) – which matched the constituent elements of FDR found at the crime scene – was found one year later in Barry George's coat pocket.

The prosecution expert witnesses (Mr Keeley and Dr Renshaw) asserted that it was not unusual for there to be just one particle found on the person firing the gun. The defence expert Dr Lloyd argued that the small size of the particle 'cast doubts on where it came from' — that it could be the result of contamination, including flawed police procedures. However, in [1] we see that

---

[7] The actual judgments made in the case were almost certainly more sound than those recorded. For example, the transcript contains an example of the fallacy of the transposed conditional [17]. Paragraph 18 of [1] claims that the testimony was made that: "It was no more likely to have come from the gun that killed Miss Dando than from some extraneous source". We understand from discussions with people involved in the case that this was certainly not stated in court.

> … the main part of his [Dr Lloyd's] evidence was directed to the places where innocent contamination of [the coat] could have taken place. Dr Lloyd was of the view that the police procedures had been flawed and contamination could have occurred at any stage, even before the events surrounding the victim's death …"

Consequently, analysis of the evidence had concentrated on the possibility that a particle would have got into the pocket as a result of indirect contamination on a number of identified occasions on which this might have occurred. The prosecution case had been that on each such occasion 'this was so unlikely that it could be discounted' [1]. Although no explicit probabilities were mentioned during the trial it is fair to conclude that this meant that there was a very low probability that the FDR in the coat pocket could have been caused by contamination.

Barry George's first appeal (on the basis of identification) in 2002 was unsuccessful. In the mean time Dr Ian Evett, who worked for the Forensic Science Service (FSS), became concerned about the way the FDR evidence was used. Evett is a pioneer of a technique called Case Assessment and Interpretation (CAI) that stresses the need to consider both the prosecution and defence likelihoods for any piece of forensic evidence to determine its probative value [24][25]. Evett was concerned that only the defence likelihood had been discussed at the trial. In his view the evidence of a very low probability that the FDR in the coat pocket could have been caused by contamination was essentially a statement about the defence likelihood. Specifically, that $P(E \mid H_\mathrm{d})$ was 'very low' where E is the FDR evidence and H_d is the defence hypothesis "BG did not fire the gun that shot JD".

Hence, Evett argued that $P(E \mid H_\mathrm{d})$ had been presented to the jury without also presenting $P(E \mid H_\mathrm{p})$, where H_p is the prosecution hypothesis "BG fired the gun that shot JD". A low value for $P(E \mid H_\mathrm{d})$, when presented in isolation, can easily be wrongly interpreted as implying a low value for $P(H_\mathrm{d})$ – a problem that is exacerbated by the prosecutor's fallacy [17]. Evett, therefore, asked Mr Keeley to provide probabilities for **both** $P(E \mid H_\mathrm{d})$ and $P(E \mid H_\mathrm{p})$. Keeley estimated that both $P(E \mid H_\mathrm{d})$ **and** $P(E \mid H_\mathrm{p})$ were equal to 1/100. Since H_p and H_d are mutually exclusive and exhaustive hypotheses it would follow correctly that the evidence had no probative value.

Evett took no action in relation to his conclusions, but in 2006 the FSS introduced new guidelines for reporting single particles and low level FDR [1]. Subsequently, in response to a request from the Criminal Cases Review Commission, the FSS reappraised the FDR evidence in the original trial and concluded that:

> "The significance of the FDR findings in this case can be put into context by considering two alternative propositions:
>
> Mr George is the man who shot Ms Dando
> Mr George had nothing to do with the incident.
>
> In our opinion the probability of finding a single particle of discharge residue in Mr George's coat pocket would have been the same, regardless of which of

the above propositions was true. The FDR evidence is thus inconclusive. In our opinion it provides no assistance to anyone asked to judge which proposition is true."

In 2007 Barry George was granted an appeal solely on the grounds that the FDR evidence, which was relied on by the prosecution at the trial as of great significance, was, in reality, 'neutral', i.e. of no probative value. The Appeal was successful with the judge concluding:

"It is impossible to know what weight, if any, the jury attached to the FDR evidence. It is equally impossible to know what verdict they would have reached had they been told as we were told, by the witnesses who gave evidence before us, that it was just as likely that the single particle of FDR came from some extraneous source as it was that it came from a gun fired by the appellant. The verdict is unsafe. The conviction will be quashed."

5.2   Confusions in the Appeal judgment transcript

The Appeal judgment document demonstrates the challenge of posing LR arguments correctly. There are several examples of forensic expert statements that appear to use inconsistent and/or ill-defined hypotheses and inconsistent and/or ill-defined statements of what the FDR evidence actually was. Indeed, throughout the text, there is a failure to state in clear and precise terms what the hypotheses were and what evidence was being considered. We re-emphasize that in the discussion below, we are showing what was stated in the Appeal Court judgment, and this does not necessarily reflect what was actually stated, i.e., statements such as "Keeley said" indicate what Keeley was reported to say according to the document *not* what Keeley necessarily actually said.

In what follows we shall assume, as Evett did, that:

H_p is the hypothesis: "BG was the man who shot JD"
E is the evidence: "A single particle of FDR matching that from the gun that killed JD is found in BG coat pocket"

5.2.1 Problems concerning the basic hypotheses

The first fundamental problem concerns both H_p and the alternative (i.e. defence) hypothesis H_d. Specifically:

- It is not clear that H_p stated above was really the same prosecution hypothesis considered by the experts
- Whereas Evett assumed, perfectly reasonably, that the defence hypothesis H_d was simply "not H_p", i.e. "BG was not the man who shot JD", the document suggests that the experts did not adhere to this assumption.

17

In fact, the judgment text contains an apparent statement by Keely which directly compromises Evett's assumptions. Specifically, Paragraph 26 of [1], describes Mr Keeley's testimony:

> "It was necessary to balance the likelihood that the particle came from a gun fired by the appellant and the likelihood that it came from some other source. Both were unlikely but both were possible. He did not and could not say that one was more likely than the other. In these circumstances the presence of the particle provided no support for the proposition that the wearer of the coat had fired a gun."

This statement suggests that Keely's prosecution hypothesis was not the H_p first stated above (that BG shot JD) but rather:

> $H1_p$: "The particle found in BG's pocket came from a gun fired by BG".

Moreover, the statement suggests that, rather than the original LR narrative presented, Keely was making a direct statement about the prior $P(H1_p)$ rather than about the likelihood $P(E \mid H_p)$. However, it would be inconsistent to conclude that $P(H1_p)$ and $P(\text{not } H1_p)$ are both 'unlikely', since these two probability values must sum to one. So it is unclear what Keely really meant by "the likelihood that the particle came from some other source".

Later, in Paragraph 38 we find:

> "A single particle of FDR had been found in the pocket of the appellant's coat. According to the evidence that Mr Keeley gave to us, this was an equally unlikely event, whether it had come from the cartridge that killed Miss Dando, or from some innocent source. There was an even chance that it had resulted from innocent contamination."

This statement suggests that Keely was assuming yet another different prosecution hypothesis:

> $H2_p$: "The particle found in BG's pocket came from the gun that killed JD".

But, since the evidence E forms a logical part of hypothesis $H2_p$, it is a tautological fact that $P(E \mid H2_p)$ must be equal to one. Hence, Keeley cannot have been thinking of $H2_p$ when he reported a prosecution likelihood of 1/100. However, it does give a very clear indication of what Keeley and the other experts and lawyers might have really meant as the defence hypothesis. Here (and in many other places) the suggestion is that the real alternative cause of the evidence E is a defence hypothesis $H2_d$ that refers to some form of "innocent contamination". By explicitly talking about contamination as an alternative cause of E, Paragraph 38 indicates that Keeley did not use "not H_p", i.e. "BG was not the man who shot JD", as the defence hypothesis. In fact it is highly unlikely that any expert could possibly have provided a realistic direct estimate for the probability $P(E \mid \text{not } H_p)$. And the indications are clear that the experts did not attempt to make such an estimate here. Instead, the experts seem to have considered a ***different*** but ill-defined defence hypothesis, namely that there were some circumstances that could have led to the FDR particle being inserted into BG's pocket

by contamination. So what *was* the defence hypothesis being considered, which we refer to as $H1_d$? It makes no sense to define $H1_d$ as "Particle inserted by contamination" because in that case $P(E \mid H1_d)$ is again, trivially equal to one and is not 1/100, as reported by Keeley. Instead, we can consider $H1_d$ as representing the set of conditions under which subsequent contamination may be possible. Paragraph 44 of [1] uses the notion of the "integrity of BG's coat being corrupted". We contend that Keeley must therefore have been using a defence hypothesis equivalent to:

$H1_d$: "Integrity of BG coat was corrupted"

When Keeley – and indeed others – referred to the equal likelihoods it seems reasonable to assume that they were referring to $P(E \mid H_p)$ and $P(E \mid H1_d)$. The problem is that H_p ("BG was the man who shot JD") and $H1_d$ "Integrity of BG coat was corrupted") **are not mutually exclusive**. It is possible that BG was the man who shot JD, but that the FDR particle in his pocket was unrelated to the gun, namely the result of contamination from the integrity of the coat being corrupted. In other words **both** H_p and $H1_d$ may be true (no matter how unlikely they may be jointly). Calculations in supplement [38] Section 8 using the BN structure of Figure 4 and the assumptions in the case - show that it is possible that $P(E \mid H_p) = P(E \mid H1_d)$ but the evidence E is not neutral as concluded by Keeley and accepted by the court. It favours H_p.
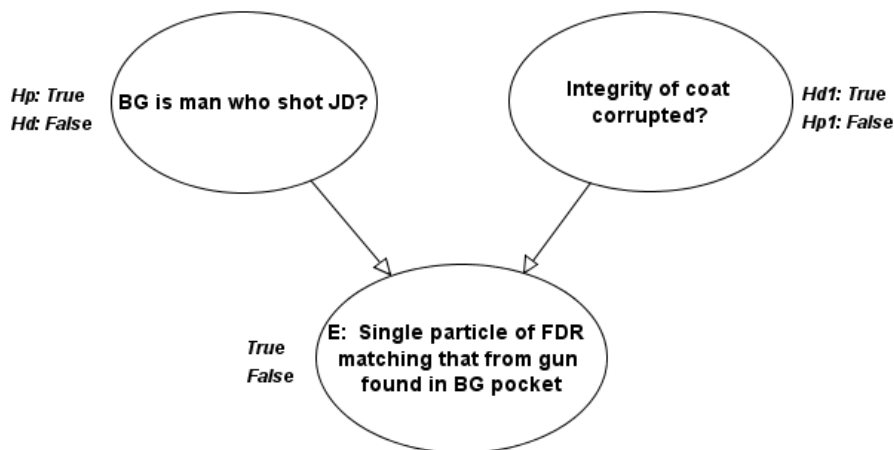

Figure 4 Simple BN model for Barry George case

5.2.2 Problems concerning the formulation of the evidence

A further problem with the Barry George case is in the formulation of the evidence. As in our example in Section 2, there are actually two **separate** pieces of evidence that make up E, namely:

E1: a single particle of FDR is found in BG coat pocket
E2: the single particle found matches that from the gun that killed JD

The failure especially to distinguish between *E*, *E1* and *E2* is a fundamental concern throughout the appeal judgment. Consider, for example the crucial FSS summarising statement (Paragraph 22):

19

> "In our opinion, it would be just as likely that a single particle of discharge residue would have been recovered from the pocket of BG's coat whether or not he was the person who shot Ms Dando…"

This statement is NOT an assertion that $P(E | H_p) = P(E | \text{not } H_p)$ as is clearly assumed. It is actually the assertion that $P(E1 | H_p) = P(E1 | \text{not } H_p)$.

Although it is possible to combine the two pieces of evidence $E1$ and $E2$ into the single statement E it is clear that doing so has created confusion for both the experts and the lawyers. For example, in Paragraph 17 of [1] we learn that

> Mr Keeley estimated the likelihood of his finding one or a few particles as 1 in 100 on either proposition.

Ignoring the additional complication of whether finding 'one particle' or 'a few particles' can really be considered the same, Keeley's assertion is interpreted as meaning $P(E | H_p) = P(E | H_d) = 0.01$. However, elsewhere the focus of E is not on the 'unusualness' of the single particle finding, but rather on the extent to which finding a particle that 'matches' the ammunition fired by a particular gun actually means the particle came from that ammunition. This is the 'random match probability' – the probability that you would find a particle matching some ammunition given that it comes from different ammunition, i.e.

$$P(E2 | \text{ not } H2_p)$$

where $H2_p$ is, as defined above, 'The particle found in BG's pocket is from the gun that killed JD'.

(note the key difference between $H2_p$ and the hypothesis $H1_p$ that Keeley may have unwittingly considered, namely that "The particle found in BG's pocket came from a gun fired by BG").

It is crucial to note that the likelihood $P(E2 | \text{ not } H2_p)$ was certainly non-zero, as confirmed by the FSS statement (Paragraph 22) which asserted

> "The particle is indistinguishable from some of those produced by the round of ammunition used to shoot Ms Dando, ***but a high proportion can produce such particles***.

It is a significant weakness of the arguments in the appeal judgment transcript that this point was not explored elsewhere and that nowhere is any estimate of $P(E2 | \text{not } H2_p)$ provided. However, the transcript suggests that, when Keeley provided a figure of 0.01 for $P(E | H_d)$ to Evett, what he actually meant was a completely different likelihood, namely $P(E2 | \text{not } H2_p)$. This possibility is supported by Paragraphs 26 and 38 of [1], already stated above.

If we were to properly incorporate the distinction between $E1$ and $E2$, along with $H2_p$, we would need the kind of causal model (represented as a Bayesian network) shown in Figure 5.
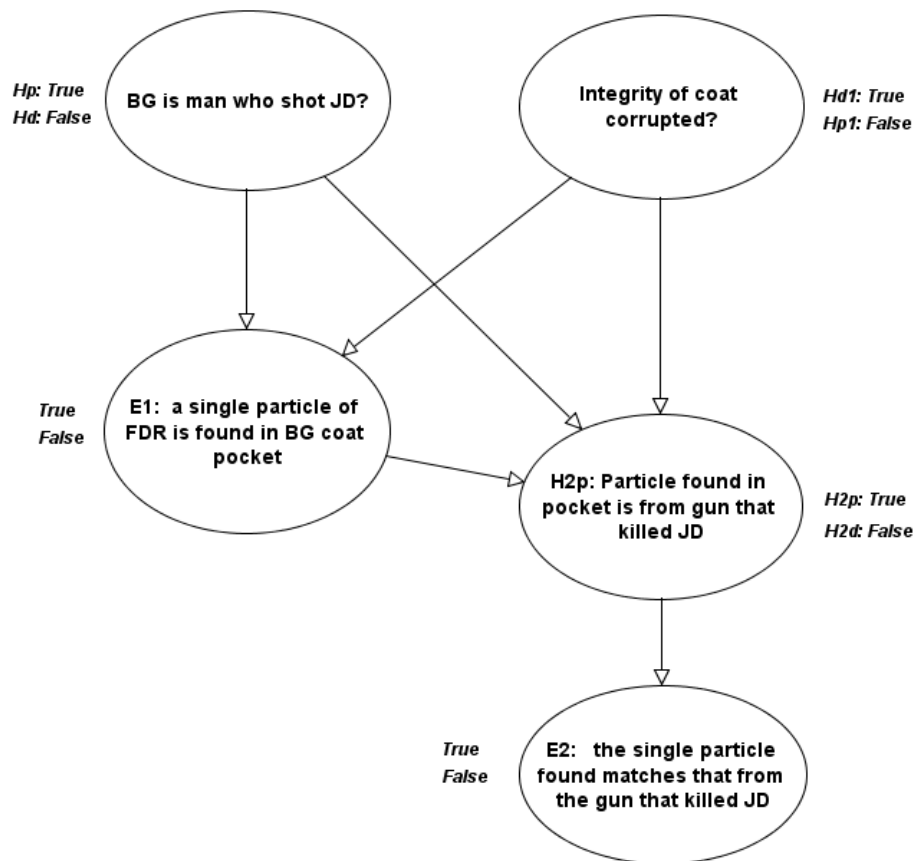
Figure 5 More comprehensive causal model

Completing the conditional probabilities (likelihoods) for this revised model (see supplement [38] Section 9) involves eliciting several more probabilities than were discussed at the trial. This would actually be helpful to all concerned – the required probabilities are not ambiguous like the original ones. Without knowing what these probabilities are, we have simply entered values that are very favourable to the defence case. Nevertheless the posterior probability of H_p given *E1* and *E2* still increases. So once again, even though we have preserved the 'equal likelihoods' agreed by the experts, the evidence is far from neutral. It again supports H_p.

What we have not explicitly considered, however, is the relevance of 'absence of evidence' in the case. Specifically, just as we showed for the example of the 'tiny trace' of DNA evidence in Section 2 and Appendix 3, it could be argued that the absence of anything other than a single FDR particle was more likely under the defence hypothesis than the prosecution. This is discussed in supplement [38] Section 10.

5.2.3 Failure to properly identify multiple different hypotheses and their relationships

Most of the above problems and confusions would have been avoided by a clearly stated set of hypotheses and evidence so that every explicit (or implicit) statement of

conditional probability could have been clearly stated as P(*A* | *B*) where *A* and *B* were unambiguous.

But the problems go much deeper because it is clear that there are many different hypotheses being considered in the evidence (and often these are being confused) and that it is inconceivable that the experts could have provided all of the appropriate likelihoods to come to the conclusions that they did. Here are examples of just some of the pairs of different prosecution and defence hypotheses that are mentioned in the Appeal ruling (although we accept that some of these are semantically equivalent) [1]:

1. Para 18:
Prosecution: FDR came from gun that killed victim
Defence: FDR came from some extraneous source

2. Para 22:
Prosecution: FDR recovered from BG's coat pocket, as he was the killer of the victim
Defence: FDR recovered from BG's coat pocket, but he was not the killer of the victim

3. Para 23:
Prosecution: BG is the man who shot victim
Defence: BG had nothing to do with the incident

4. Para 26 (i):
Prosecution: FDR came from a gun fired by BG
Defence: FDR came from some other source

5. Para 26 (ii):
Prosecution: The wearer of the coat fired the gun
Defence: The wearer of the coat did not fire the gun (the defence hypothesis is unspecified in the appeal report, but this is one possibility.)

6. Para 27:
Prosecution: FDR came from a gun fired at the time of the victim's murder
Defence: FDR came from some other source

7. Para 28:
Prosecution: FDR found as a result of BG firing a gun
Defence: FDR found as a result of secondary contamination

8. Para 32:
Prosecution: The particle is FDR
Defence: The particle is not FDR

9. Para 33:
Prosecution: FDR came from ammunition that killed the victim
Defence: FDR came from any other ammunition that had that kind of percussion primer

10. Para 37:
Prosecution: FDR did not come from secondary contamination
Defence: FDR came from secondary contamination

11. Para 38:
Prosecution: FDR came from the cartridge that killed the victim
Defence: FDR came from some innocent source

12. Para 50:
Prosecution: FDR was deposited on the coat other than innocently
Defence: FDR was deposited on the coat innocently

That there must be far more hypotheses involved is evident from the fact that numerous probabilistic and quasi-probabilistic statements mentioned in the trial are not encapsulated by the grossly simple original H_p, H_d and E alone [1]. A full set of is provided in supplement [38] Section 11.

Our final concern relates to the fact that the transcript suggests in places an unnatural diagnostic (as opposed to causal) view of evidence. This is discussed in supplement [38] Section 12.

6. Conclusions and recommendations

Justice is best served when the evidence and hypotheses under consideration are accessible and clear to all parties and are unambiguously defined. In particular, this is the only way to assign correct meanings to the likelihoods provided by expert witnesses. Furthermore, because the probative value of a piece of evidence on source-level hypotheses may be very different from its probative value on offence-level hypotheses, the relationships between source-level and offence-level hypotheses must be made clear. We have demonstrated that an an efficient way to achieve such clarity will require the construction of causal models through the aid of tools such as Bayesian networks. This approach helps to improve legal reasoning and by doing so demonstrates how hypotheses can be clarified, related and made precise enough for reliable quantificiation. Central to this approach is the distinction between hypotheses that are mutually exclusive and exhaustive and those that are not.

We contend that, in order to determine whether evidence has probative value – and therefore whether it should be excluded from proceedings or not – it should be evaluated against offence-level hypotheses. Any diversion from this key principle will carry the risk that evidence might be presented to the jury merely as a diversionary tactic, and persuade it to make decisions based on superfluous source-level hypotheses.

We have demonstrated serious concerns about the Barry George Appeal Court judgment. The case document suffered from oversimplification: what appeared as a superficially simple set of hypotheses were actually a set of ill-defined, but related, assumptions and vaguely defined hypotheses. Under these circumstances the 'simple LR rule' inadequately captured the probative value of the evidence. We have shown that evidence with an LR equal to one in this case was not necessarily 'neutral'.

23

The errors we highlight are taken from a judgment document, and although this may not always have accurately recorded what was said in court, the fact that it contains so much erroneous reasoning is cause for concern. Clearly, if a case judgment can be wraught with so many failings, similar problems are likely to occur in courtroom judgments too. Additionally, it is a concern that the careful and rigorous presentations of statistical evidence made by conscientious expert witnesses can become distorted in a case judgment.

Buried among the numerous muddled arguments present throughout the Barry George Appeal transcript, there is a lucid statement that captures a key point we have made in this paper: "It is often the case that a piece of evidence that proves nothing when viewed in isolation acquires probative value when considered in the context of other evidence." (Paragraph 33) [1]. Thus, the voice of reason was present, but it was outnumbered by the numerous imprecise and incorrect arguments scattered throughout the Appeal judgment document. In this paper we have provided formal and clear explanations in order to illuminate the above point. We present this work with the hope that in the future, the voice of reason will not be drowned out by erroneous arguments.

While we have focused on the technical issues of our approach to improve understanding about the probative value of evidence, it is clear that its success depends on incorporating not just semantic features of the argument, but also pragmatic and contextual features. These kinds of feature are extremely difficult to assess. They are influenced by the cultural background, principles of critical reasoning or conversational implicature in assessing the meaning of statements. Ultimately there is a need to train forensic scientists and lawyers not just in the use of BNs to interpret arguments, but a wider range of tools for analysing and understanding arguments.

7. Acknowledgements

8. References

[1]   (2007). R. v. George, EWCA Crim 2722
[2]   (2010). R v T. EWCA Crim 2439
      http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf
[3]   AgenaRisk software (2013). www.agenarisk.com
[4]   Aitken, C. G. G. and F. Taroni (2004). Statistics and the evaluation of evidence for forensic scientists (2nd Edition), John Wiley & Sons, Ltd.
[5]   Aitken, C. and many other signatories (2011). "Expressing evaluative opinions: A position statement." Science and Justice 51(1): 1-2.
[6]   Balding, D. J. (2005). Weight-of-Evidence for Forensic DNA Profiles, Wiley.
[7]   Balding, D., (2004) Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk 3 (1): 63-64.

[8]   Berger, C. E. H., J. Buckleton, C. Champod, I. Evett and G. Jackson (2011). "Evidence evaluation: A response to the court of appeal judgement in R v T." Science and Justice 51: 43-49.

[9]   Broeders, T. (2009). Decision-Making in the Forensic Arena. In "Legal Evidence and Proof: Statistics, Stories and Logic". (Eds H. Kaptein, H. Prakken and B. Verheij, Ashgate) 71-92.

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

[11]   Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA. (1998). A hierarchy of propositions: deciding which level to address in casework. Science and Justice, 38: 231-239.

[12]   Dawid, A.P., (2004) Which likelihood ratio (Comment on 'Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk 3(1):65–71.

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

[14]   Evett, I. W., L. A. Foreman, G. Jackson and J. A. Lambert (2000). "DNA profiling: a discussion of issues relating to the reporting of very small match probabilities." Criminal Law Review (May) 341-355.

[15]   Evett IW, Gill PD, Jackson G, Whitaker J, Champod C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. J Forensic Sci. May;47(3):520-30.

[16]   Fenton, N. E. (2011). "Science and law: Improve statistics in court." Nature 479: 36-37

[17]   Fenton, N.E. and Neil, M. (2011), 'Avoiding Legal Fallacies in Practice Using Bayesian Networks', Australian Journal of Legal Philosophy 36, 114-151

[18]   Fenton, N. E., M. Neil, and D. Lagnado (2012) "Modelling mutually exclusive causes in Bayesian networks", http://www.eecs.qmul.ac.uk/~norman/papers/mutual_IEEE_format_version.pdf

[19]   Fenton N., and Neil, M. Risk Assessment and Decision Analysis with Bayesian Networks, CRC Press, 2012.

[20]   Fenton, N., Neil, M., Lagnado, D.A. (2013). A general structure for legal arguments about evidence using Bayesian networks. Cognitive Science, 37, 61–102.

[21]   Korb K, Nicholson A. (2010), "Bayesian Artificial Intelligence", 2nd Edition, Chapman & Hall/CRC Computer Science & Data Analysis.

[22]   Lempert, R. O. (1977), " Modeling Relevance", Michigan Law Review Vol. 75, No. 5/6, pp. 1021-1057

[23]   Meester, R. and Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk 3 (1): 51-62.

[24]   Meester, R. and Sjerps, M. Response to Dawid, Balding, Triggs and Buckleton, Law, Probability and Risk (2004) 3(1): 83-86

[25]   Morrison, G. M. (2012). "The likelihood ratio framework and forensic evidence in court: a response to RvT." International Journal of Evidence and Proof 16(1).

[26]   Nordgaard, A. Hedell, R., and Ansell, R. (2012). Assessment of forensic findings when alternative explanations have different likelihoods-"Blame-the-brother"-syndrome. Science and Justice 52: 226-236.

[27]   Pardo, M. S. (2013). "The Nature and Purpose of Evidence Theory." Vanderbilt Law Review 66.

25

[28] Park, R. C., P. Tillers, et al. (2010). "Bayes Wars Redivivus — An Exchange." International Commentary on Evidence 8(1): 1-38.

[29] Picinali, F. (2012) "Structuring inferential reasoning in criminal fact finding: an analogical theory" Law, Probability and Risk 11, 197–223

[30] Puch-Solis, R., P. Roberts, S. Pope and C. Aitken (2012). PRACTITIONER GUIDE NO 2: Assessing the Probative Value of DNA Evidence, Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society
http://www.rss.org.uk/uploadedfiles/userfiles/files/Practitioner-Guide-2-WEB.pdf

[31] Redmayne, M. (2001). Expert Evidence and Criminal Justice, Oxford University Press.

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

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

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

[35] Saks, M. J. and J. J. Koehler (2007). "The Individualization Fallacy in Forensic Science Evidence." http://works.bepress.com/michael_saks/1

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

[37] Triggs C.M. and J.S. Buckleton, Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk (2004) 3(1): 73-82

[38] Supplementary Material 1: Supporting Calculations and Examples

[39] Supplementary Material 2: Bayesian network models

Appendix 1 Bayes' Theorem proof and the probative value of evidence for mutually exclusive hypotheses

The following is an example of how the likelihood ratio describes the impact of evidence on a pair of mutually exclusive hypotheses:

> **Appendix Example 1:** Suppose that a man is charged with a gambling offence, namely that he was using a 'fixed' die in which five of the six sides were 6's. Let H_p be the hypothesis that the die was fixed, and let H_d be the alternative hypothesis that the die was not fixed (i.e. it was a 'fair' die). The evidence E is the observation that the outcome of two consecutive rolls of the die were two 6s. Then

$$P(E \mid H_\mathrm{p}) = 25/36$$

Whereas

$$P(E \mid H_\mathrm{d}) = 1/36.$$

In this case the prosecution likelihood is 25 times greater than the defence likelihood, i.e. the LR is 25. In the above example, the fact that the prosecution likelihood is 25 times greater than the defence likelihood suggests intuitively that the evidence E supports the prosecution hypothesis. In fact, it can be proven that when prosecution and defence hypotheses are mutually exclusive, a LR of greater than one supports the prosecution hypothesis and a LR of less than one supports the defence hypothesis. Hence, the LR has a simple interpretation for the probative value of the evidence under these assumptions, and the proof is as follows:

In order to prove this important property of the LR, we need Bayes' theorem

Bayes' Theorem tells us that:

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

By applying Bayes' theorem to both H_p and H_d we get the equivalent form of Bayes (called the 'odds' version):

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)} \times \frac{P(H_P)}{P(H_D)}$$

In this version the term

$$\frac{P(E \mid H_P)}{P(E \mid H_D)}$$

is the likelihood ratio (LR) – it is simply the prosecution likelihood divided by the defence likelihood.

The term

$$\frac{P(H_P)}{P(H_D)}$$

represents the 'prior odds' – the relative prior belief in the prosecution hypothesis over the defence hypothesis.

The term

$$\frac{P(H_P \mid E)}{P(H_D \mid E)}$$

represents the revised 'posterior odds' – the relative (posterior) belief in the prosecution hypothesis over the defence hypothesis having observed the evidence E.

Most texts that attempt to explain the impact of the LR on the probative value of E use an argument based on the relative 'odds' of the hypotheses. The formula tells us that whatever our prior odds were in favour of the prosecution hypothesis, the posterior odds are the result of multiplying the prior odds by the LR. Hence, when the prosecution likelihood is greater than the defence likelihood the posterior odds in favour of the prosecution hypothesis must increase.

However, this argument it is unnecessarily confusing, because not only does it hide the assumption that the hypotheses need to be mutually exclusive for it to work, but it also fails to tell us clearly what we most need to know: namely that for the evidence E to 'support' the hypothesis H_p it is necessary that the posterior probability of H_p, i.e. $P(H_p \mid E)$ is greater than the prior probability $P(H_p)$ in other words our belief in H_p being true increases after we observe E.

What follows is a proof that $P(H_p \mid E) > P(H_p)$ when the LR is greater than 1:

From Bayes' Theorem:
$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)} \times \frac{P(H_P)}{P(H_D)}$$

But since the LR > 1 it follows that:

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} > \frac{P(H_P)}{P(H_D)}$$

But because $H_d = $ not H_p we know that

$$P(H_D) = 1 - P(H_P) \quad \text{and} \quad P(H_D \mid E) = 1 - P(H_P \mid E)$$

Hence, substituting these into the above inequality equation we get:

28

$$\frac{P(H_P \mid E)}{1 - P(H_P \mid E)} > \frac{P(H_P)}{1 - P(H_p)}$$

$$\Rightarrow P(H_P \mid E)(1 - P(H_p)) > P(H_P)(1 - P(H_P \mid E)$$

$$\Rightarrow P(H_P \mid E) - P(H_p)P(H_P \mid E) > P(H_P) - P(H_P)P(H_P \mid E)$$

$$\Rightarrow P(H_P \mid E) > P(H_P)$$

Appendix 2. Neutral evidence

First we prove that evidence E is neutral when the LR is 1 and when the prosecution and defence hypotheses are mutually exclusive.

Since the LR is 1 we know that $P(E \mid H_p) = P(E \mid H_d)$

$$P(H_P \mid E) = \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)P(H_P) + P(E \mid H_D)P(H_D)}$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)P(H_P) + P(E \mid H_P)P(H_D)} \text{ since } P(E \mid H_P) = P(E \mid H_D)$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)(P(H_P) + P(H_D))} \text{ since } P(E \mid H_P) = P(E \mid H_D)$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)} \text{ since } P(H_P) + P(H_D) = 1 \text{ as } P(H_P), P(H_D) \text{ mutually exclusive and exhaustive}$$

$$= P(H_P)$$

What happens when the LR =1 but H_p and H_d are not mutually exclusive? From the odds version of Bayes' we know that

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(H_P)}{P(H_D)}$$

So all we can *actually* conclude is that the ratio of the posterior probabilities of H_p and H_d is equal to the ratio of the prior probabilities.

Appendix 3: The Bayesian Network modelling approach

As Appendix 2 indicates, even the simplest application of Bayes' theorem (namely when we have a single pair of mutually exclusive hypotheses and a single piece of evidence) can involve quite complex calculations. When there are multiple hypotheses and different (possibly related) pieces of evidence it is impossible to do the calculations correctly by hand. Fortunately, there is a well established formalism – Bayesian networks [20][21][36] – supported by widely available tools [19], that enables all the Bayesian calculations to be performed automatically once the priors and likelihoods are specified by the experts.

In a Bayesian network (BN) the nodes represent variables (such as hypotheses and evidence) which have different state values (such as false and true) while the arcs drawn between nodes represent known dependencies between the variables. The set of states of a variable in a BN are, by definition, always mutually exclusive and exhaustive. Hence:

{'true', 'false'},
{"guilty", "not guilty"},
{"Fred shot the gun", "Joe shot the gun", "neither Fred nor Joe shot the gun"}
would all acceptable sets of states for a BN node, while
{"Fred shot the gun", "Fred or Joe shot the gun")
would not be acceptable.

While the calculations are automatically computed from a given model they are, of course, critically dependant on the choice of model. One of the confusions associated with using BNs for legal reasoning is that there are actually two fundamentally different ways to model hypotheses. If the prosecution and defence hypotheses are mutually exclusive and exhaustive then a simple and natural way to represent them is to model them as the states of a single hypothesis node, as in either of the versions of Figure 6(i).

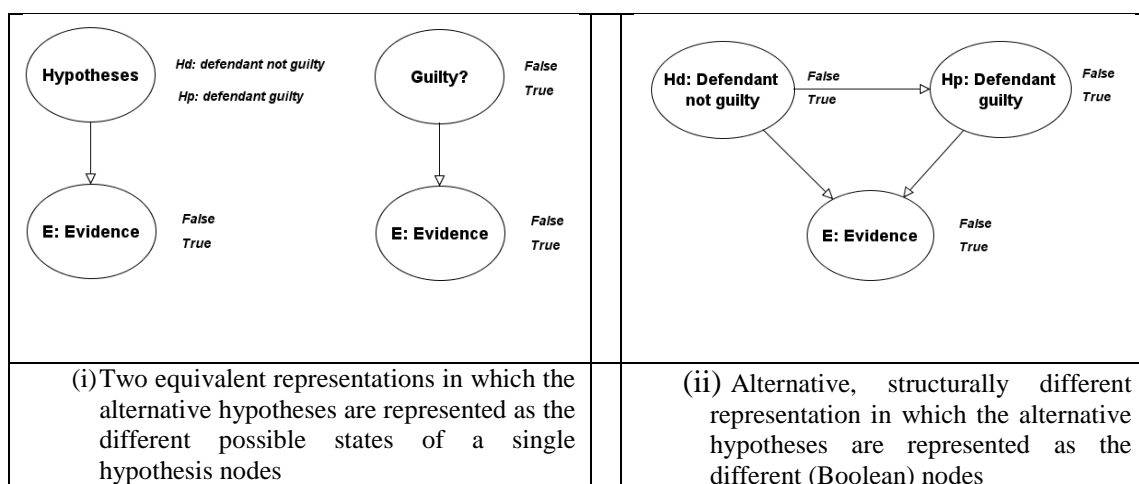| (i) Two equivalent representations in which the alternative hypotheses are represented as the different possible states of a single hypothesis nodes	(ii) Alternative, structurally different representation in which the alternative hypotheses are represented as the different (Boolean) nodes

Figure 6 Representing hypotheses as either (i) states of a single unknown variable or as (ii) separate unknown (Boolean) variables

However, for reasons explained in [18], it is sometimes preferable to use separate nodes for the prosecution and defence hypotheses. The fundamental problem with this

31

approach is that there is no general mechanism for ensuring that separate nodes in a BN are mutually exclusive and exhaustive. For the case of two hypotheses it is possible to 'force' mutual exclusivity by introducing a link as shown in Figure 6(ii). However, this does not extend to more than two mutually exclusive hypotheses [18].

When the defence and prosecution hypotheses are not mutually exclusive (for example, if Hp is "defendant guilty" and Hd is "defendant not at crime scene") then we can – and indeed must - represent the hypotheses as separate (Boolean) nodes as shown in Figure 7.

Figure 7 Representing hypotheses that are not mutually excusive and exhaustive

The method we propose here (and the one that is used in all examples) is:

- When the hypotheses are mutually exclusive use a single node whose states correspond to these hypotheses (so we adopt the approach in Figure 6(i) in this case)

- When the hypotheses are not mutually exclusive we use separate (Boolean) nodes to correspond to each hypotheseis (so we adopt the approach in Figure 7).

Once the hypotheses are carefully considered and modelled in this way we use the BN to model the causal story linking hypotheses and different pieces of evidence. In addition to its causal structure we have to specify, for each node in the BN having parents, a conditional probability table. Specifically this requires us to specify the probability of each state of the node conditional on each combination of parent states. For a node with one parent such as node E in Figure 6(i), this table would be one such as shown in Figure 10.

Hypotheses	Hd: defendant not guilty	Hp: defendant guilty
False	0.99	0.0
True	0.01	1.0

Figure 8 Conditional probability table for node with one parent

The value in the bottom left cell is precisely the defence likelihood , $P(E \mid H_d)$ while the value in the bottom right cell is precisely the prosecution likelihood , $P(E \mid H_p)$. The other cells are simply one minus these values respectively since they correspond to , $P(\text{not } E \mid H_d)$ and $P(\text{not } E \mid H_d)$ respectively. Hence specifiying the conditional probability table for E is nothing more and nothing less than specifying the likelihood values.

For a node with more than one parent, such as node E in Figure 7, the conditional probability table would be one such as shown in Figure 9.

Hd: Defence hypothesis	False		True	
Hp: Prosecution hypothesis	False	True	False	True
False	0.5	0.0	0.99	0.0
True	0.5	1.0	0.01	1.0

Figure 9 Conditional probability table for node with two parents

The cell entries in the bottom row again correspond to likelihood values. But note that, because H_p and H_d are not assumed to be mutually exclusive and exhaustive it is not sufficient to specify just $P(E \mid H_d)$ and $P(E \mid H_p)$. Instead we are forced to specify the likelihood values for the different mutually exclusive *state combinations*.

Many examples of completed BNs are provided in the complementary material [39] but we present one example here to solve the problem described in Section 2 of the main text where we have the evidence E: "tiny matching DNA trace found" that actually comprises two separate pieces of evidence:

- $E$1: tiny DNA trace found
- $E$2: DNA trace found matches defendant

The example assumes that H_p is 'defendant guilty' and H_p is "defendant not guilty" . Hence, we use a single node to model these mutually exclusive and exhaustive hypotheses. The oversimplistic model is therefore shown in Figure 10(i). The correct model is shown in Figure 10(ii).
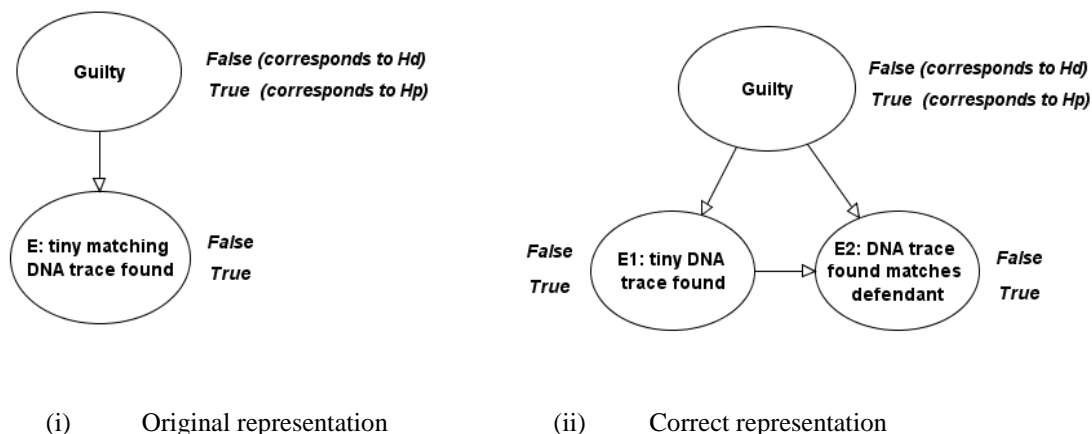

   (i)   Original representation   (ii)   Correct representation

Figure 10 Modelling complex evidence in a BN

Suppose that the DNA trace has a profile with a random match probability of 1/100. In the oversimplistic model the expert would typically provides the likelihoods:

$P(E| H_p) = 1$
$P(E| H_d) = 1/100$

In other words the conditional probabilty table for the node E is defined as shown in Figure 11.

Guilty	False	True
False	0.99	0.0
True	0.01	1.0

Figure 11 Conditional probability table for node E

The LR in this case is therefore 100. This can also be seen by executing the model, with the evidence E entered as true, as shown in Figure 12 (to ensure the LR is the same as the posterior odds of guilty we set the prior odds to be 50:50).

Guilty

False — 0.9901%

True — 99.0099%

E: tiny matching DNA

False —

True — 100%

Scenario 1 : True

Figure 12 Posterior odds of guilt.

However, what the oversimplistic model completely fails to capture is relevance of the fact that the trace was tiny. 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. A reasonable conditional probability table for $E1$ is shown in Figure 13.

34

Guilty	False	True
False	0.5	0.999
True	0.5	0.0010

Figure 13 Conditional probability table for E1

The conditional probability table for *E*2 shown in Figure 14 captures uses the same RMP information as was used in the oversimplified model.

Guilty	False		True	
E1: 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

Figure 14 Conditional probability table for E2

Calculating the overall LR manually in this case is much more complex, so we go directly to the result of running the model with *E*2 set as true (and the prior odds of guilt set at 50:50 again). This is shown in Figure 15. The LR is just the probability of guilty divided by the probability of not guilty, which is 0.2. So the evidence supports the defence hypothesis rather than the prosecution.
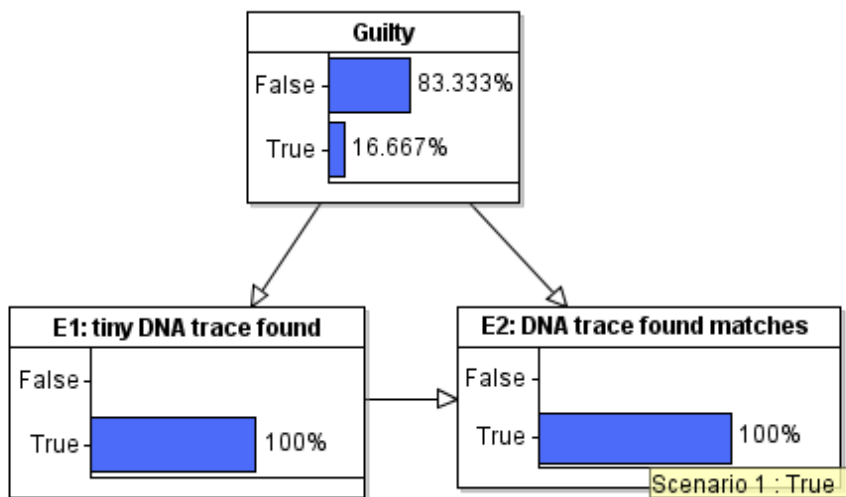

Figure 15 Posterior odds in correct model

This example also indicates the importance of taking account of absence of evidence.

35

Supplementary Material 1: Supporting Calculations and Examples

1. Calculations for Example 1

$P(H_P) = P(H_D) = 0.5$

$P(E \mid H_P, H_D) = 0.9$

$P(E \mid H_P, \text{not } H_D) = 0.9$

$P(E \mid \text{not } H_P, H_D) = 0.9$

$P(E \mid \text{not } H_P, \text{not } H_D) = 0$

Then

$P(E \mid H_P) = P(E \mid H_P, H_D)P(H_D) + P(E \mid H_P, \text{not } H_D)P(\text{not } H_D) = 0.9 \times 0.5 + 0.9 \times 0.5 = 0.9$

and

$P(E \mid H_D) = P(E \mid H_D, H_P)P(H_P) + P(E \mid H_D, \text{not } H_P)P(\text{not } H_P) = 0.9 \times 0.5 + 0.9 \times 0.5 = 0.9$

so

$P(E \mid H_P) = P(E \mid H_D)$ i.e. the LR is equal to 1

Now we can also use marginalisation to compute $P(E)$:

$P(E) = P(E \mid H_P, H_D)P(H_P)P(H_D) + P(E \mid H_P, \text{not } H_D)P(H_P)P(\text{not } H_D) +$

$+ P(E \mid \text{not } H_P, H_D)P(\text{not } H_D)P(H_D) + P(E \mid \text{not } H_P, \text{not } H_D)P(\text{not } H_D)P(\text{not } H_D)$

$= (0.9 \times 0.5 \times 0.5) + (0.9 \times 0.5 \times 0.5) + (0.9 \times 0.5 \times 0.5) + 0$

$= 0.675$

Hence by Bayes:

$$P(H_P \mid E) = \frac{P(E \mid H_P)P(H_P)}{P(E)} = \frac{0.9 \times 0.5}{0.675} = 0.666$$

Similarly:

$$P(H_D \mid E) = \frac{P(E \mid H_D)P(H_D)}{P(E)} = \frac{0.9 \times 0.5}{0.675} = 0.666$$

Next we add the motive evidence. Because this scenario involves four related variables the Bayesian calculations are rather too complex to demonstrate from scratch and so we use a Bayesian network software tool.
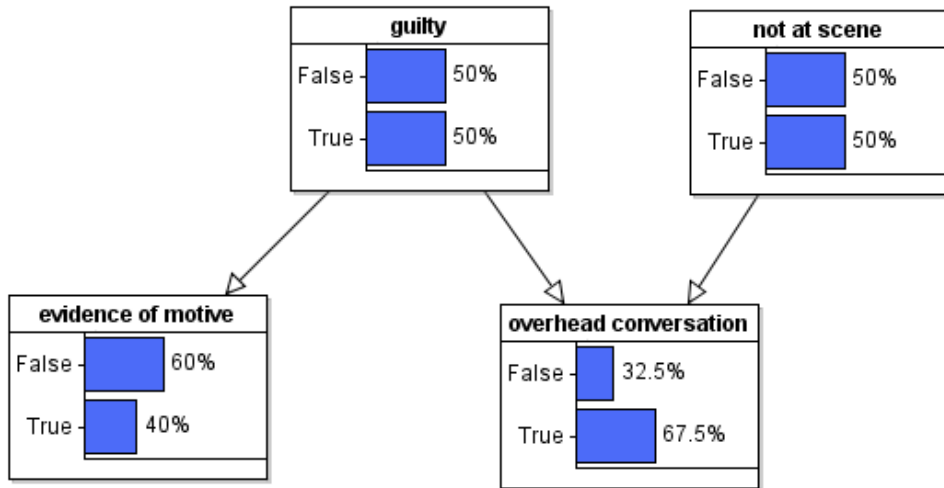
Figure 16 shows the prior probabilities.

Figure 16 Prior probabilities for murder model

Figure 17 shows the results of observing the motive evidence E'. In this case $P(H_p \mid E') = 0.875$, while $P(H_d \mid E')$ is unchanged at 0.5.
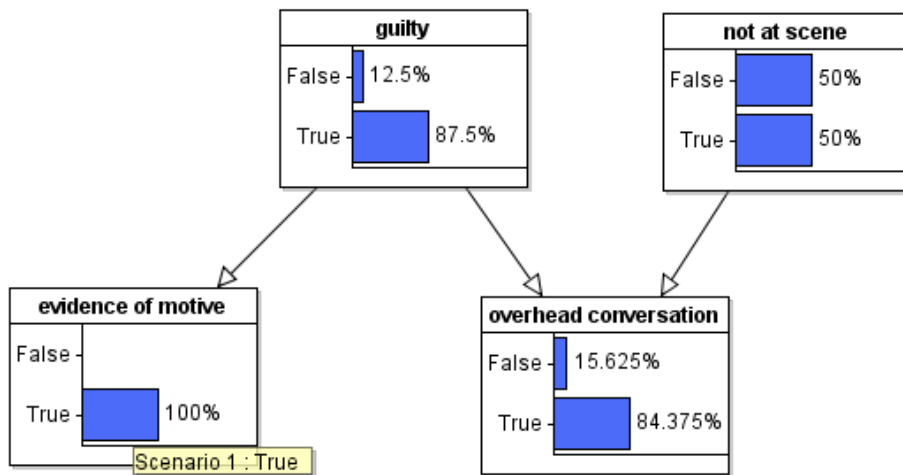

Figure 17 Posterior probabilities for murder model after observing evidence of motive

When we now observe E (Figure 18) we see that the probability of H_p, that is $P(H_p \mid E', E)$, jumps to 0.933. The evidence E therefore may be sufficient in this case to convince a jury to convict (if there were, say a threshold of 90% certainty required).
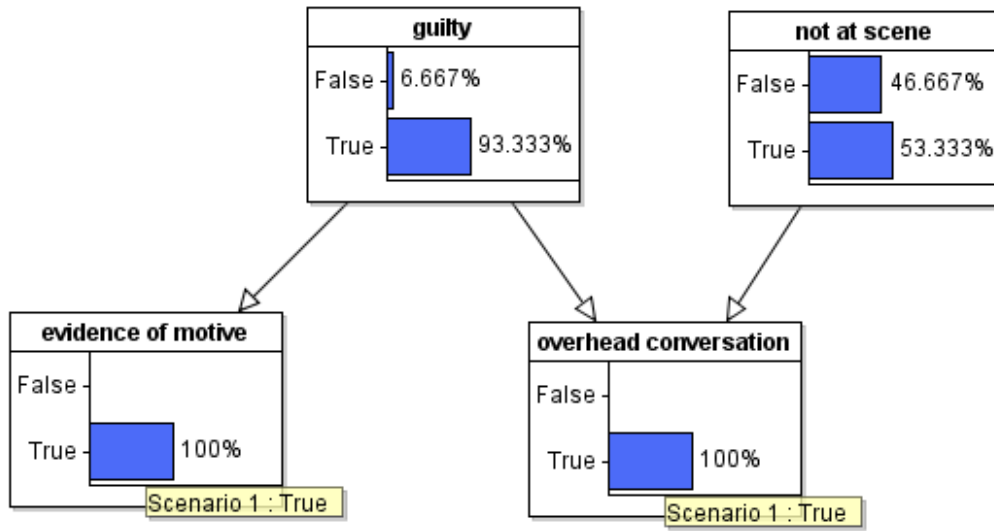
Figure 18 Posterior probabilities for murder model after observing evidence of motive and overheard conversation

2. Dice example (non-mutually exclusive hypotheses): LR=1 but evidence is not neutral

Example S1: The defendant rolls two dice – a black die which he owns and a red die randomly selected by a member of the public from a batch provided by a reputable dice company. The evidence E against the defendant is that both dice rolls are 6s. However, in this case the prosecution hypothesis H_p concerns only the black die, while the defence hypothesis H_d concerns only the red die:

H_p: "The black die is fixed with all sides being 6s"
H_d: "The red die is fixed with all sides being 6s"

The reason H_d is the defence hypotheses is because it was subsequently discovered that the red die came from a batch in which 50% were faulty in the sense of having all sides sixes. Hence, the prior $P(H_d)=1/2$. Suppose the prior $P(H_p)=1/2$ because it is known that 50% of the defendant's black dice are fixed with all sides being 6s.

Then we have:

$$P(H_p) = \frac{1}{2}$$

$$P(H_d) = \frac{1}{2}$$

$$P(E|H_p, H_d) = 1$$

$$P(E|H_p, \text{not } H_d) = 1/6$$
$$P(E|\text{not } H_p, H_d) = 1/6$$
$$P(E|\text{not } H_p, \text{not } H_d) = 1/36$$

It follows that $P(E|H_p)=P(E|H_d) = 7/12$ because:

$$P(E|H_p) = P(E|H_p, H_d)P(H_d) + P(E|H_p, \text{not } H_d)P(\text{not } H_d)$$
$$= \frac{1}{6} + \frac{1}{6} \times \frac{1}{2} = \frac{7}{12}$$

$$P(E|H_d) = P(E|H_d, H_p)P(H_p) + P(E|H_d, \text{not } H_p)P(\text{not } H_p)$$
$$= \frac{1}{6} + \frac{1}{6} \times \frac{1}{2} = \frac{7}{12}$$

So LR=1, but the evidence is ***not*** neutral as can be seen from the results of running the model in Figure 19.
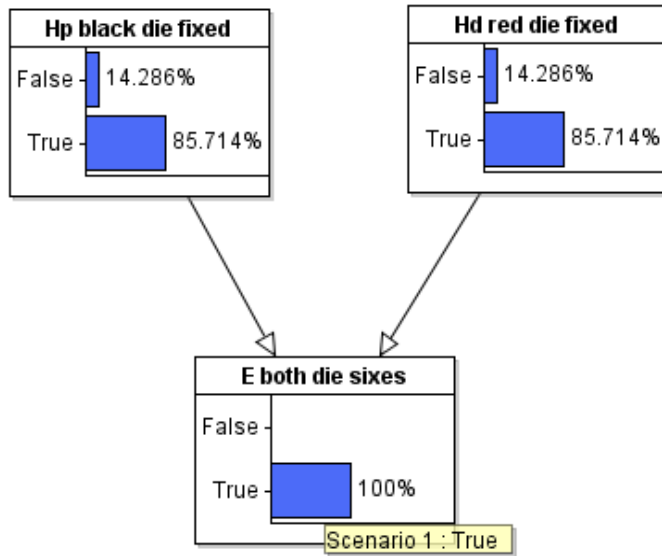
Figure 19 Evidence not neutral

The fact that $P(H_p|E)= 6/7 = 0.85714$ tells us that the prosecution hypothesis is now very likely. Just because the defence hypothesis has increased by the same amount is essentially irrelevant.

Formally, the calculations are based on Bayes' theorem and noting that the marginal $P(E)$ is

$$P(E) = P(E|H_p, H_d)P(H_p)P(H_d) + P(E|H_p, \text{not } H_d)P(H_p)P(\text{not } H_d)$$
$$+ P(E|\text{not } H_p, H_d)P(\text{not } H_p)P(H_d)$$
$$+ P(E|\text{not } H_d, \text{not } H_p)P(\text{not } H_p)P(\text{not } H_d)$$

$$= \left(1 \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{6} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{6} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{36} \times \frac{1}{2} \times \frac{1}{2}\right)$$

$$=49/144=0.34028$$

Hence by Bayes'

$$P(H_p) = \frac{P(E|H_p) \times P(H_p)}{P(E)} = \frac{\frac{7}{12} \times \frac{1}{2}}{\frac{49}{144}} = \frac{6}{7}$$

Thus, as in *Example 1* from the main text, the fact that the posterior for H_p and H_d increase in the same proportions from their priors is less important than the fact that the posterior for H_p is now more likely than unlikely.

40

3. Dice example (non-mutually exclusive hypotheses): LR<0.5 but evidence favours prosecution not defence.

Next we consider a slight variation of example S1 in which the LR is not 1 but actually favours the 'wrong' hypothesis.

Example S2: This is the same as Example S2, except in this case the red die came from a batch in which it was subsequently found that 10% were faulty ('all six') dice. Hence the prior $P(H\text{d})=0.1$

In this case $P(E|H_d) = 7/12 = 0.5833$ as before. However, $P(E|H_p) = ¼ = 0.25$

So the defence likelihood is more than twice the prosecution likelihood, i.e. LR <0.5. According to the FSS guidelines on the likelihood ratio we should be able to conclude that the evidence provides support for the defence hypothesis. But this is nonsense because what matters is that, as before, $P(H_p|E)= 0.86$ (see Figure 20).
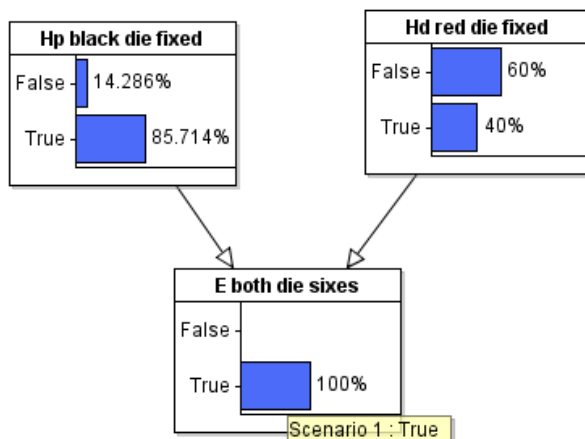

Figure 20 Evidence leads to strong belief in Hp

As before the evidence shows it is now very likely that the prosecution hypothesis is true. Just because the probability the defence hypothesis has increased by a bigger percentage is again essentially irrelevant.

So while it is true that the proportional increase in the probability of H_d (from 0.1 to 0.4) is greater than that of H_p (from 0.5 to 0.86), this is irrelevant from the perspective of the probative value of the evidence. If H_p is true it does not matter whether or not H_d is also true; the evidence E provides strong support for H_p being true.

In other words the 'probative value' of the evidence is exactly the opposite of what is normally assumed of a LR. But things can get much worse. If the evidence in Example 2 was three 6s rolled on each of the dice then the defence likelihood becomes about 50 times greater than the prosecution likelihood, but the impact of the evidence is such that the prosecution hypothesis is almost certain (0.9954).

41

4. Example of LR mistaken to be 1

Using exactly Example S2 we can illustrate how it is easy to be deceived into thinking evidence might have an LR of 1 when it does not. This is because both of the 'intuitive' likelihoods are the same:

$$P(E|H_p, \text{not } H_d) = 1/6$$
$$P(E|\text{not } H_p, H_d) = 1/6$$

However, the intuitive likelihoods fail to take account of the non-exclusivity of the hypotheses, i.e. the fact that **both** H_p and H_d might be true. Indeed as we already saw in Example 2, in this case $P(E|H_d) = 7/12 = 0.5833$, while $P(E|H_p) = ¼ = 0.25$

5. Calculations for Example 4 of main text (non exhautive hypotheses)

Figure 21(a) shows the prior marginals when $P(E \mid H_p) = P(E \mid H_d) = 0.01$ and $P(E \mid H_a) = 0.5$. Figure 21(b) shows the resulting posterior marginals when E is entered as true.

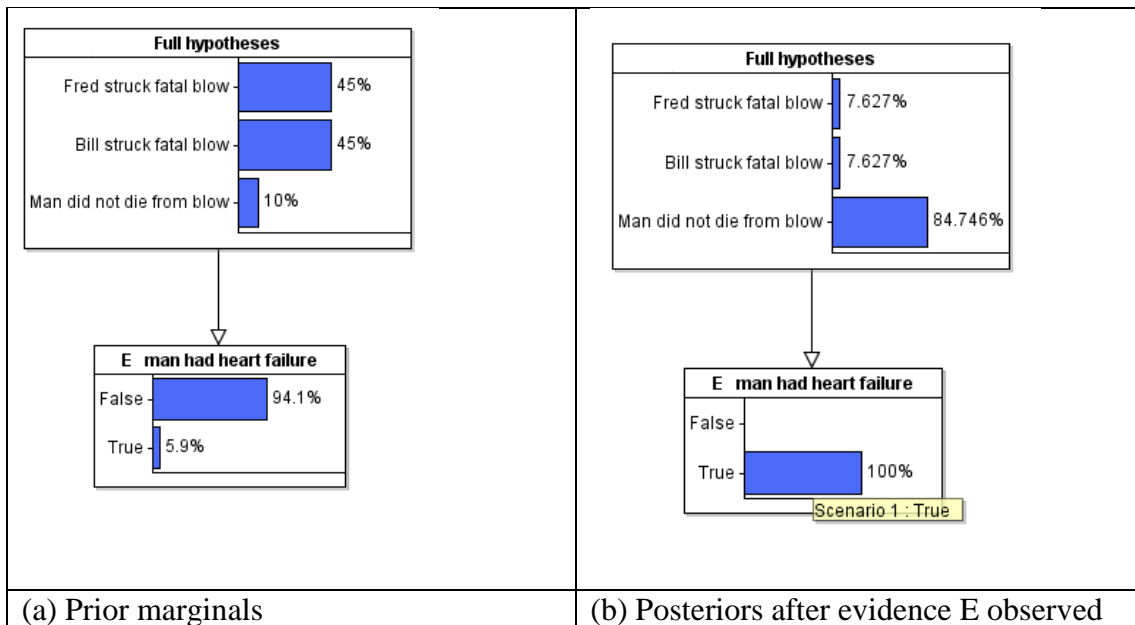| (a) Prior marginals	(b) Posteriors after evidence E observed

Figure 21 Prior and posterior probabilities for the problem in Example 2.

6. Calculations for Example 4 in main text

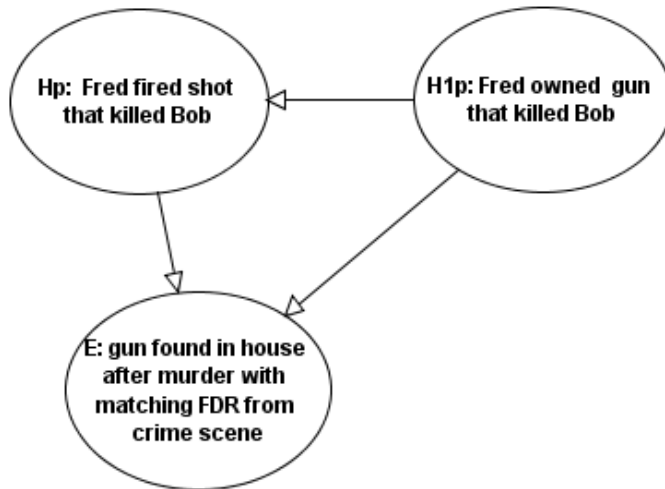The BN for Example 3 in the main text shown in Figure 22.

Figure 22 BN in Example 3

Here the secondary hypothesis $H1_p$ of whether Fred owned the gun is a parent of the ultimate hypothesis H_p. This is a classic example of an 'opportunity' node [26]. For our current analysis we assume that Fred will be equally likely to have fired the gun regardless of whether he owned it (i.e., chance that Fred fired the gun does not depend on whether he owned it). We assume that, before the evidence, the prior probability that Fred fired the gun is 1/00. This gives us the conditional probability table for P(H_p | H) shown in Figure 23:

H1p: Fred owned gun that killed Bob	False	True
False	0.99	0.99
True	0.01	0.01

Figure 23 Conditional probability table for Hp: Fred fired the gun that killed Bob

The table shows that Fred is just as likely to use the gun whether he owns it or not (thus the link between $H1_p$ and H_p is actually redundant). Now we need the conditional probability table for the evidence given the two sets of hypotheses, P(E | H_p, H1$_p$). This is shown in Figure 24

use

Hp: Fred fired shot that killed Bob	False		True	
H1p: Fred owned gun that killed Bob	False	True	False	True
False	0.99	0.99	0.9	0.9
True	0.01	0.01	0.1	0.1

Figure 24 Conditional probability table for evidence E: Gun found in Fred's house

Here we again made the reasonable assumption that if Fred fired the gun it is just as likely E will be true whether or not he owned it and this likelihood is 1/10. Similarly

44

if he did not fire the gun it is just as likely the evidence would be observed (E=true) whether or not he owned the gun, but in the case these equal probabilities are lower (1/100) than in the case if he did fire the gun.

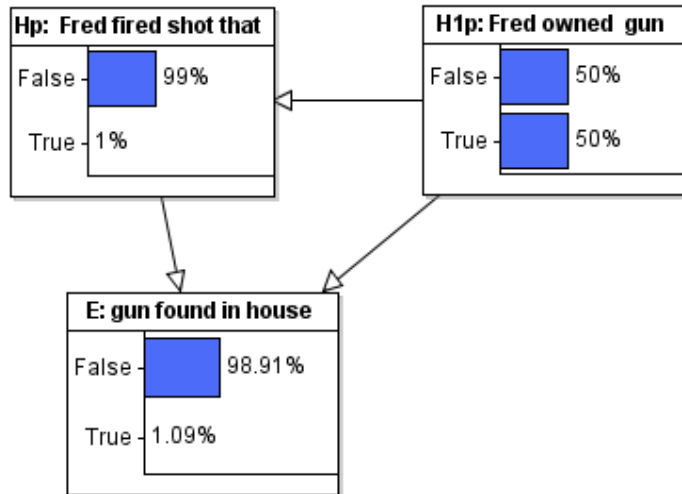The prior state of the model with the above conditional probability tables is shown in Figure 25

Figure 25 Prior state of the BN

Now we use the BN to calculate the probability of evidence E under the two values for $H1_p$, which means setting the node for 'Fred owned the gun' to True and False respectively. The screen shots from the BN calculation (Figure 26) show that the likelihoods $P(E|H1_p)$ and $P(E|\text{not } H1_p)$ are equal and hence the LR with respect to $H1_p$ is 1.

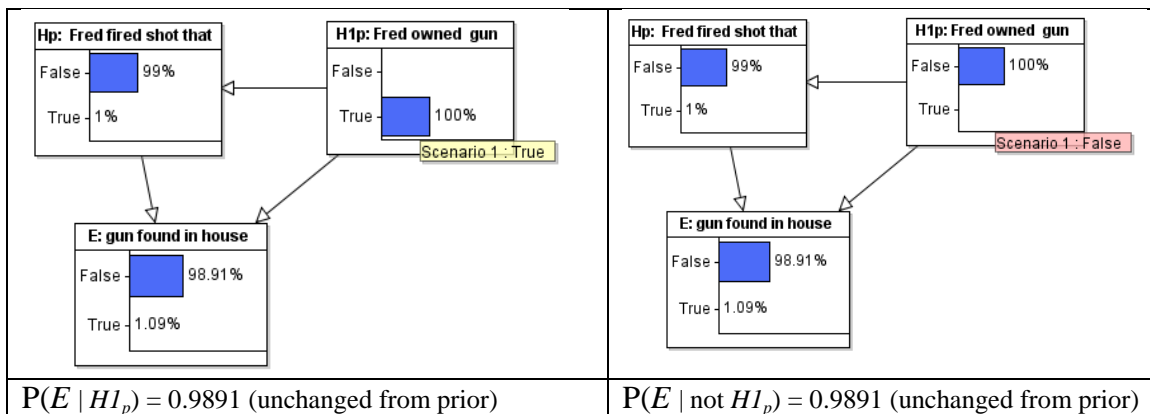| $P(E \mid H1_p) = 0.9891$ (unchanged from prior)	$P(E \mid \text{not } H1_p) = 0.9891$ (unchanged from prior)

Figure 26 Calculating the probability of evidence E under the two values for H1p (True on left hand side, False on right hand side)

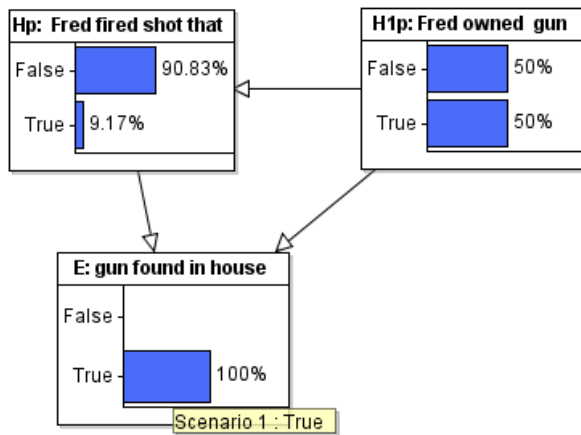When we observe the evidence (set E=true) we get the result shown in Figure 27.

Figure 27 Evidence is observed

We see that the evidence has not changed $H1_p$, i.s. $P(H1_p \mid E) = P(H1_p) = 50\%$, and thus the evidence is 'neutral' with respect to whether Fred owned the gun. However, $P(H_p \mid E)$ is *not* equal to H_p—the probability has increased from 1% to 9.17%. Thus, the evidence is *not* neutral with respect to the ultimate hypothesis: the evidence increases the chance of Fred being guilty of murder.

46

7.   Offence level hypotheses appear exhaustive but are not

In many situations even the offence level hypotheses "guilty or not guilty" may actually hide implict assumptions that suggest the hypotheses are not exhaustive (so that we are back in the situation described in Section 3.2). An example would be where there is an implict assumption that a crime has actually taken place (for example, a murder) even though there is a small probability (ignored by both sides) that the victim was not actually murdered. In such a situation the 'offence level' hypotheses H_p ("guilty") and H_d ("not guilty") are wrongly assumed to be exhaustive because what they *actually mean* are "murder was committed by the defendant" and "murder was committed by somebody other than the defendant" respectively. What is missing is the hypothesis H_n "no crime committed".

Suppose we have the evidence E: "victim left a suicide note".

Then clearly $P(E \mid H_p) = P(E \mid H_d)$ since the probability of finding a suicide note would not be changed by knowing the identity of the person charged with murder. It follows that the LR is equal to one for the 'offense level' hypotheses, and so this piece of evidence would be (erroneously) regarded as having no probative value. But while it has no impact on H_p and H_d when the probability of H_n is assumed to be zero (Figure 28, Case A), it certainly does have an impact when the probability H_n is not zero (Figure 28, Case B where we assume a 1% probability).
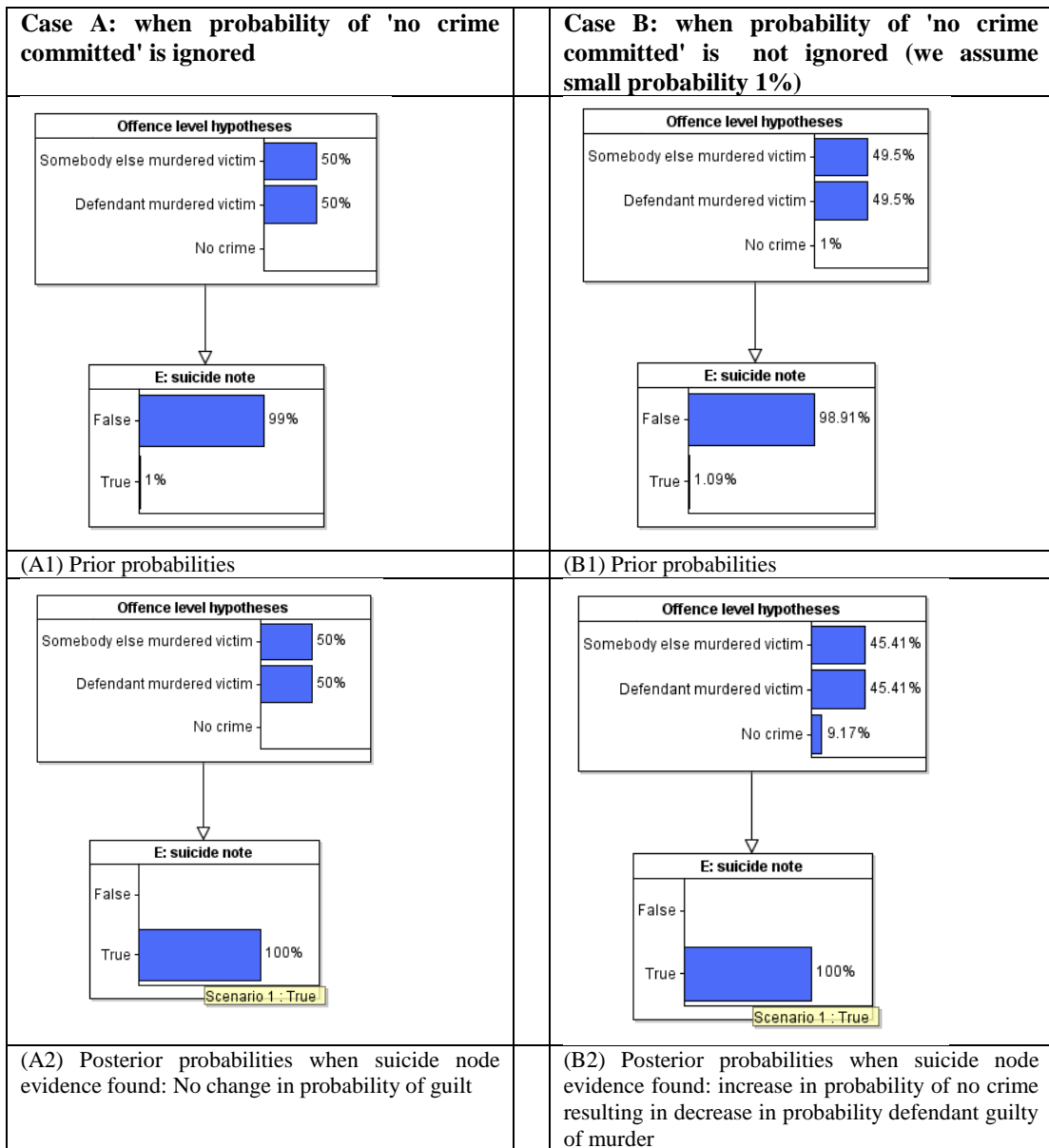
Case A: when probability of 'no crime committed' is ignored	Case B: when probability of 'no crime committed' is not ignored (we assume small probability 1%)
(A1) Prior probabilities	(B1) Prior probabilities
(A2) Posterior probabilities when suicide node evidence found: No change in probability of guilt	(B2) Posterior probabilities when suicide node evidence found: increase in probability of no crime resulting in decrease in probability defendant guilty of murder

Figure 28 The importance of considering the correct offence level hypotheses.

In both cases we make the same assumptions about $P(E \mid H_p)$, $P(E \mid H_d)$, $P(E \mid H_n)$. Specifically, that $P(E \mid H_p) = P(E \mid H_d) = 0.01$ and $P(E \mid H_n) = 0.1$. The actual values do not matter. All that matters is that $P(E \mid H_p) = P(E \mid H_d)$ are equal and that $P(E \mid H_n)$ is greater.

An alternative way to consider this problem is to separate out the pair of offence-level hypotheses "guilty/not guilty" from the ultimate hypotheses "crime committed/crime not committed". This is shown in Figure 29.
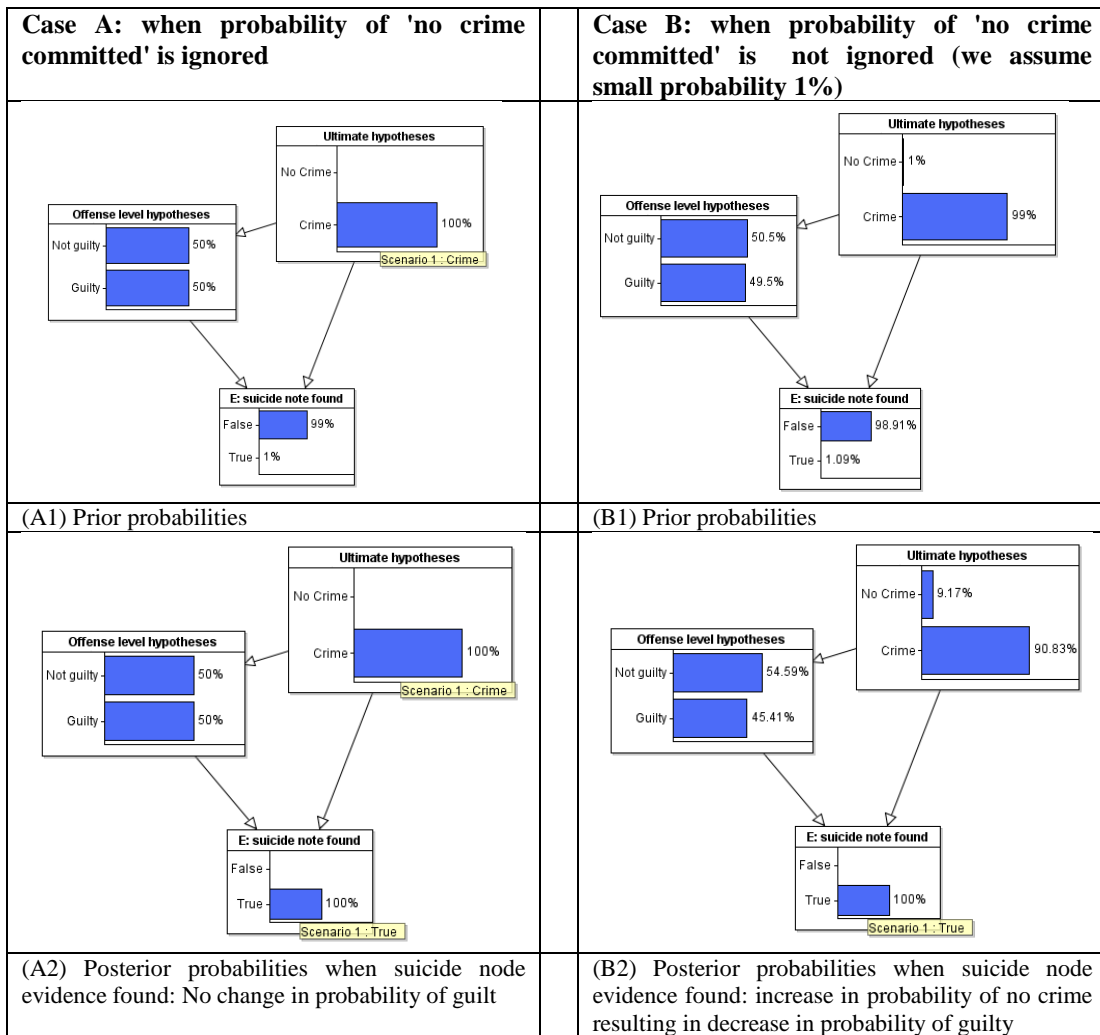
Case A: when probability of 'no crime committed' is ignored	Case B: when probability of 'no crime committed' is not ignored (we assume small probability 1%)
(A1) Prior probabilities	(B1) Prior probabilities
(A2) Posterior probabilities when suicide node evidence found: No change in probability of guilt	(B2) Posterior probabilities when suicide node evidence found: increase in probability of no crime resulting in decrease in probability of guilty

Figure 29 Separating out the pair of offence level hypotheses from the ultimate hypotheses.
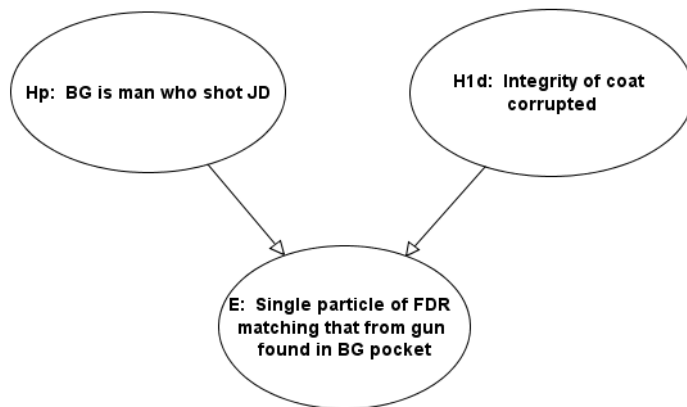
8. Basic Barry George FDR evidence

Figure 30 BN with separate pairs of hypotheses

We assume that the conditional probabilities are defined as:

$P(E| H_p$, not $H1_d) = 0.01$
$P(E| H1_d$, not $H_p) = 0.01$
- these are the most reasonable assumptions from Keeley's assertions to Evett.

However, since H_p and $H1_d$ can clearly both be true at the same time we also need to define $P(E| H_p$ and $H1_d)$ and it seems reasonable to define this as
$P(E| H_p$ and $H1_d) = 0.02$ (in fact the exact probability matters little as far as illustrating the problem, but it is certainly non-zero and it must be greater than 0.01).

Assuming $P(H_p)=P(H1_d)=0.5$ we get the following results:

$P(E| H_p) = P(E |H1_d) = 0.015$, as shown in Figure 31, which confirms that the likelihoods are equal as assumed by Keeley and Evett. However, this certainly does not mean the evidence E is neutral because, as shown Figure 31(b) $P(H_p | E) = 0.75$, which an ***increase*** from the prior 0.5 (shown in Figure 31(a)).
Even though $P(H1_d |E)$ is also increased by exactly the same amount (by virtue of the rather unrealistic equal priors for both hypotheses) the increase in $P(H_p |E)$ is what actually matters.
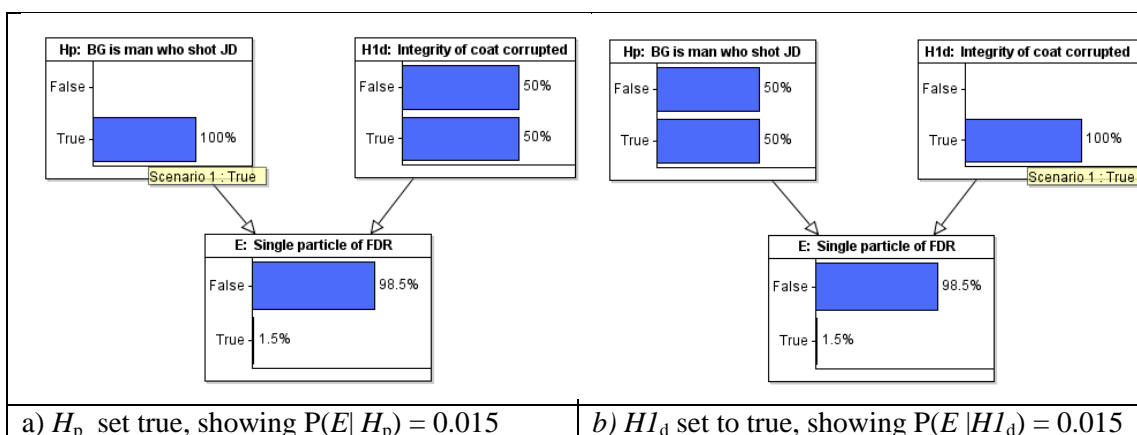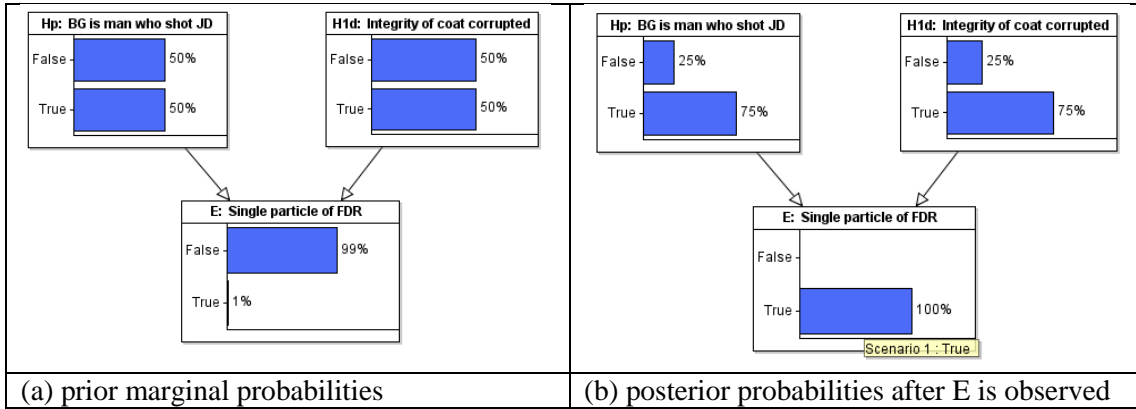

| a) H_p set true, showing $P(E| H_p) = 0.015$ | b) $H1_d$ set to true, showing $P(E |H1_d) = 0.015$ |

Figure 31 Calculating P(E|Hypothesis) for the two different hypotheses Hp and H1d respectively

Hp: BG is man who shot JD False — 50% True — 50%	**H1d: Integrity of coat corrupted** False — 50% True — 50%
E: Single particle of FDR False — 99% True — 1%	
(a) prior marginal probabilities	(b) posterior probabilities after E is observed

9. More realistic Barry George Model

The more realistic causal model incorporating appropriate hypotheses and evidence is shown in Figure 32.
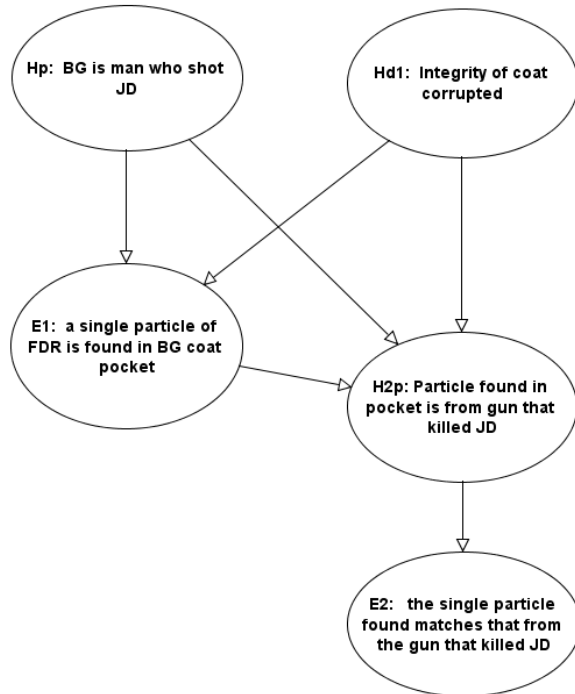

Figure 32 More realistic BN model

Completing the conditional probabilities (likelihoods) for this revised model involves eliciting several more probabilities than were discussed at the trial. This would actually be helpful to all concerned – the required probabilities are not ambiguous like the original ones. Without knowing what these probabilities are, we have simply entered values that are most favourable to the defence case. Nevertheless the posterior probability of H_p given *E1* and *E2* still increases as shown in

Appendix 3: The Bayesian Network modelling approach

As Appendix 2 indicates, even the simplest application of Bayes' theorem (namely when we have a single pair of mutually exclusive hypotheses and a single piece of evidence) can involve quite complex calculations. When there are multiple hypotheses and different (possibly related) pieces of evidence it is impossible to do the calculations correctly by hand. Fortunately, there is a well established formalism – Bayesian networks [20][21][36] – supported by widely available tools [19], that enables all the Bayesian calculations to be performed automatically once the priors and likelihoods are specified by the experts.

In a Bayesian network (BN) the nodes represent variables (such as hypotheses and evidence) which have different state values (such as false and true) while the arcs drawn between nodes represent known dependencies between the variables. The set of states of a variable in a BN are, by definition, always mutually exclusive and exhaustive. Hence:

{'true', 'false'},
{"guilty", "not guilty"},
{"Fred shot the gun", "Joe shot the gun", "neither Fred nor Joe shot the gun"}
would all acceptable sets of states for a BN node, while
{"Fred shot the gun", "Fred or Joe shot the gun")
would not be acceptable.

While the calculations are automatically computed from a given model they are, of course, critically dependant on the choice of model. One of the confusions associated with using BNs for legal reasoning is that there are actually two fundamentally different ways to model hypotheses. If the prosecution and defence hypotheses are mutually exclusive and exhaustive then a simple and natural way to represent them is to model them as the states of a single hypothesis node, as in either of the versions of Figure 6(i).

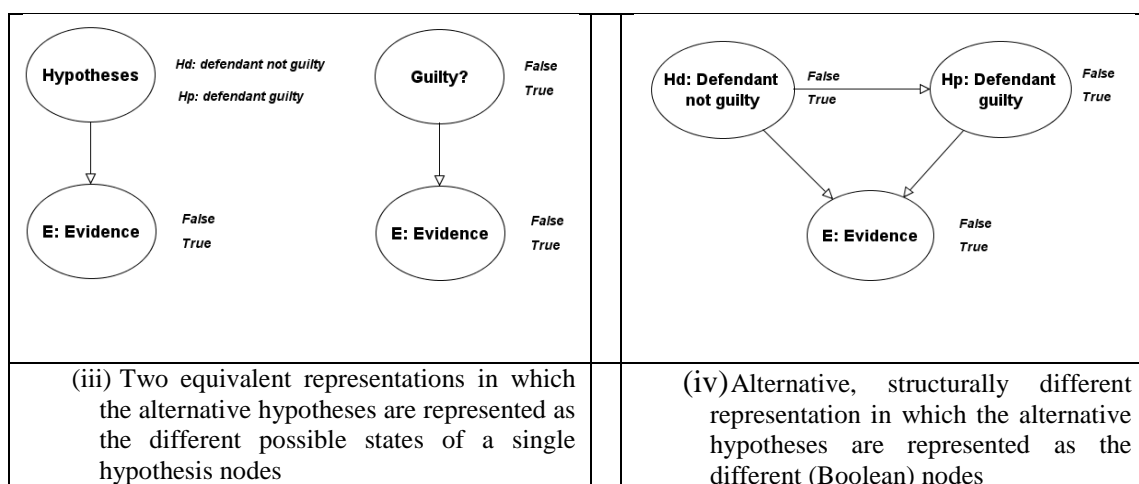| (iii) Two equivalent representations in which the alternative hypotheses are represented as the different possible states of a single hypothesis nodes	(iv) Alternative, structurally different representation in which the alternative hypotheses are represented as the different (Boolean) nodes

Figure 6 Representing hypotheses as either (i) states of a single unknown variable or as (ii) separate unknown (Boolean) variables

However, for reasons explained in [18], it is sometimes preferable to use separate nodes for the prosecution and defence hypotheses. The fundamental problem with this

approach is that there is no general mechanism for ensuring that separate nodes in a BN are mutually exclusive and exhaustive. For the case of two hypotheses it is possible to 'force' mutual exclusivity by introducing a link as shown in Figure 6(ii). However, this does not extend to more than two mutually exclusive hypotheses [18].

When the defence and prosecution hypotheses are not mutually exclusive (for example, if *H*p is "defendant guilty" and *H*d is "defendant not at crime scene") then we can − and indeed must - represent the hypotheses as separate (Boolean) nodes as shown in Figure 7.
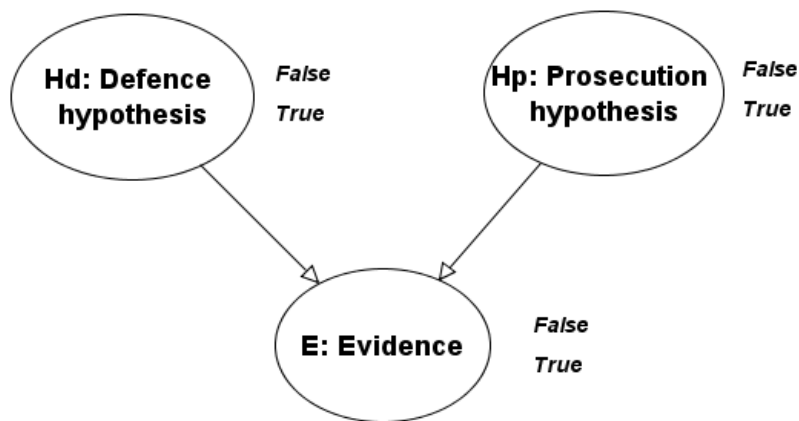

Figure 7 Representing hypotheses that are not mutually excusive and exhaustive

The method we propose here (and the one that is used in all examples) is:

- When the hypotheses are mutually exclusive use a single node whose states correspond to these hypotheses (so we adopt the approach in Figure 6(i) in this case)

- When the hypotheses are not mutually exclusive we use separate (Boolean) nodes to correspond to each hypotheseis (so we adopt the approach in Figure 7).

Once the hypotheses are carefully considered and modelled in this way we use the BN to model the causal story linking hypotheses and different pieces of evidence. In addition to its causal structure we have to specify, for each node in the BN having parents, a conditional probability table. Specifically this requires us to specify the probability of each state of the node conditional on each combination of parent states. For a node with one parent such as node E in Figure 6(i), this table would be one such as shown in Figure 10.

Hypotheses	Hd: defendant not guilty	Hp: defendant guilty
False	0.99	0.0
True	0.01	1.0

Figure 8 Conditional probability table for node with one parent

54

The value in the bottom left cell is precisely the defence likelihood , $P(E \mid H_\mathrm{d})$ while the value in the bottom right cell is precisely the prosecution likelihood , $P(E \mid H_\mathrm{p})$. The other cells are simply one minus these values respectively since they correspond to , $P(\text{not } E \mid H_\mathrm{d})$ and $P(\text{not } E \mid H_\mathrm{d})$ respectively. Hence specifiying the conditional probability table for E is nothing more and nothing less than specifying the likelihood values.

For a node with more than one parent, such as node E in Figure 7, the conditional probability table would be one such as shown in Figure 9.

Hd: Defence hypothesis	False		True	
Hp: Prosecution hypothesis	False	True	False	True
False	0.5	0.0	0.99	0.0
True	0.5	1.0	0.01	1.0

Figure 9 Conditional probability table for node with two parents

The cell entries in the bottom row again correspond to likelihood values. But note that, because H_p and H_d are not assumed to be mutually exclusive and exhaustive it is not sufficient to specify just $P(E \mid H_\mathrm{d})$ and $P(E \mid H_\mathrm{p})$. Instead we are forced to specify the likelihood values for the different mutually exclusive ***state combinations***.

Many examples of completed BNs are provided in the complementary material [39] but we present one example here to solve the problem described in Section 2 of the main text where we have the evidence E: "tiny matching DNA trace found" that actually comprises two separate pieces of evidence:

- $E$1: tiny DNA trace found
- $E$2: DNA trace found matches defendant

The example assumes that H_p is 'defendant guilty' and H_p is "defendant not guilty" . Hence, we use a single node to model these mutually exclusive and exhaustive hypotheses. The oversimplistic model is therefore shown in Figure 10(i). The correct model is shown in Figure 10(ii).
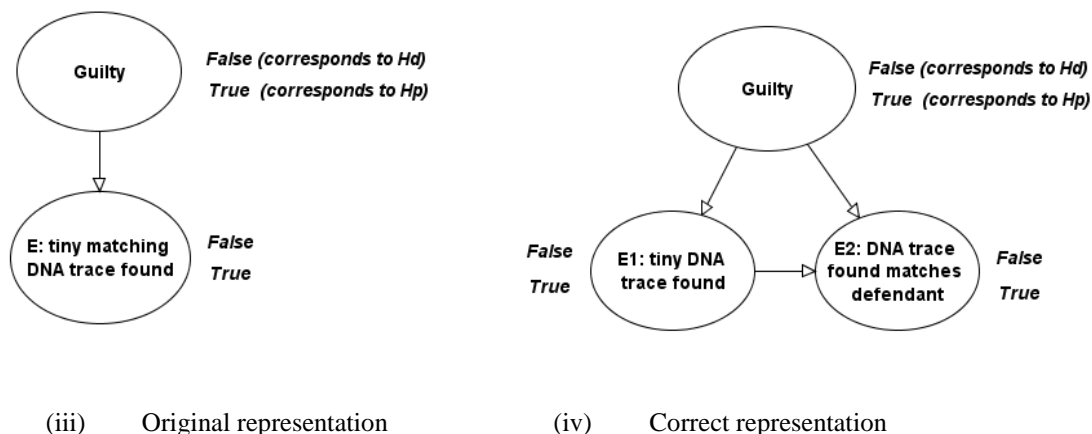

    (iii)   Original representation   (iv)   Correct representation

Figure 10 Modelling complex evidence in a BN

Suppose that the DNA trace has a profile with a random match probability of 1/100. In the oversimplistic model the expert would typically provides the likelihoods:

$P(E|H_p) = 1$
$P(E|H_d) = 1/100$

In other words the conditional probabilty table for the node E is defined as shown in Figure 11.

Guilty	False	True
False	0.99	0.0
True	0.01	1.0

Figure 11 Conditional probability table for node E

The LR in this case is therefore 100. This can also be seen by executing the model, with the evidence E entered as true, as shown in Figure 12 (to ensure the LR is the same as the posterior odds of guilty we set the prior odds to be 50:50).
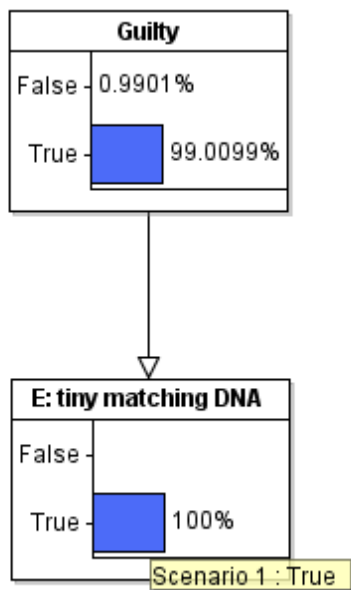

Figure 12 Posterior odds of guilt.

However, what the oversimplistic model completely fails to capture is relevance of the fact that the trace was tiny. 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. A reasonable conditional probability table for $E1$ is shown in Figure 13.

Guilty	False	True
False	0.5	0.999
True	0.5	0.0010

Figure 13 Conditional probability table for E1

The conditional probability table for *E*2 shown in Figure 14 captures uses the same RMP information as was used in the oversimplified model.

Guilty	False		True	
E1: 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

Figure 14 Conditional probability table for E2

Calculating the overall LR manually in this case is much more complex, so we go directly to the result of running the model with *E*2 set as true (and the prior odds of guilt set at 50:50 again). This is shown in Figure 15. The LR is just the probability of guilty divided by the probability of not guilty, which is 0.2. So the evidence supports the defence hypothesis rather than the prosecution.
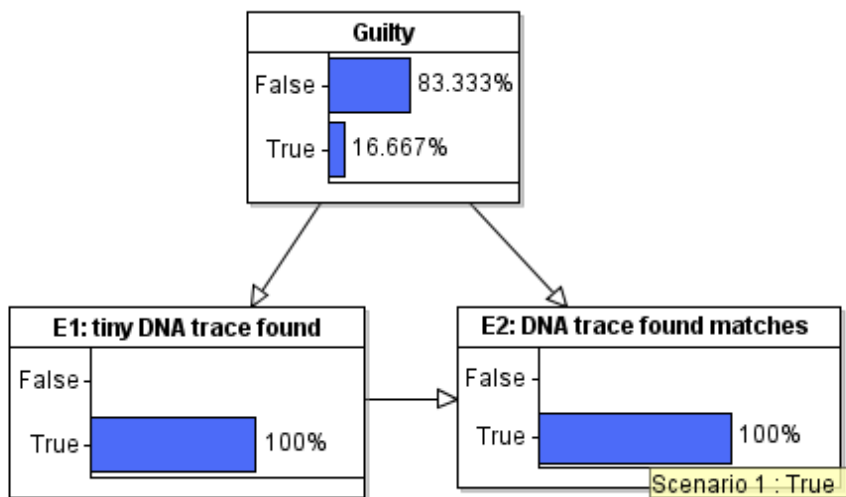

Figure 15 Posterior odds in correct model

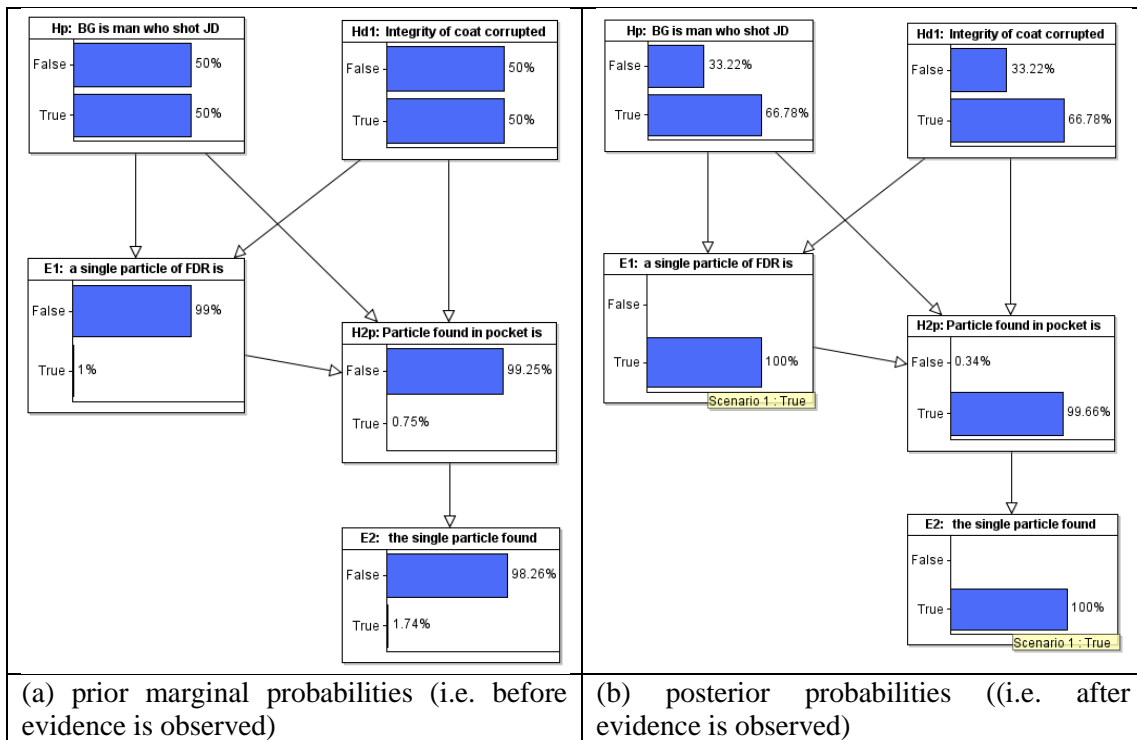This example also indicates the importance of taking account of absence of evidence..

| (a) prior marginal probabilities (i.e. before evidence is observed) | (b) posterior probabilities ((i.e. after evidence is observed) |

Figure 33 Impact of the evidence

10. Failure to properly incorporate the notion of 'absence of evidence'

There is great confusion throughout the Barry George appeal judgment transcript about the fact that a single particle was found rather than a large collection. There are suggestions in the transcript that finding only a 'small number of particles' - which can actually mean up to 20 - is more likely to result from contamination than the actual firing of a gun. There is no discussion of the different prosecution and defence likelihoods associated with different numbers of particles, presumably on the assumption that this would be redundant for any number other than one (one being the number in E). While the latter would certainly be unnecessary there is, however, one important additional piece of evidence (which surely favours the defence) that should have been considered in the likelihood values was:

E3: failure to find particles elsewhere in BG's apartment or possessions.

While this introduces significant extra complexity into the model (which would therefore be more complex than that in Appendix 9) the concern is that *E3* may actually have been consciously or unconsciously 'factored' in to the evidence E when the likelihoods were estimated. In other words the experts may have been considering likelihoods like P(*E1* and *E2* and *E3* | H_p) and if so it would have been difficult to do so consistently.

11. Probabilistic and quasi-probabilistic statements mentioned in the BG trial not encapsulated by the simple original hypotheses

1. Para 9:

"In Mr Keeley's experience FDR would more often than not be found on the firer of the gun, but would not be found on ordinary members of the public unless they had been associated with firearms."

2. Para 22:

"In our opinion, it would be just as likely that a single particle of discharge residue would have been recovered from the pocket of Mr George's coat whether or not he was the person who shot Ms Dando nearly a year previously."

3. Para 23:

"Mr George is the man who shot Ms Dando
Mr George had nothing to do with the incident."

"In our opinion the probability of finding a single particle of discharge residue in Mr George's coat pocket would have been the same, regardless of which of the above propositions was true."

4. Para 26:

"It was necessary to balance the likelihood that the particle came from a gun fired by the appellant and the likelihood that it came from some other source. Both were unlikely but both were possible. He did not and could not say that one was more likely than the other. In these circumstances the presence of the particle provided no support for the proposition that the wearer of the coat had fired a gun."

5. Para 27:

"Mr Keeley said that he had intended to convey to the jury that it was no more likely that the single particle of FDR came from a gun fired at the time of Miss Dando's murder than that it came from some other source."

6. Para 28:

"Dr Renshaw s (who had also given evidence at the trial) evidence accorded with that of Mr Keeley. While it was unlikely that the particle had resulted from secondary contamination of the coat it was equally unlikely that it was the result of the appellant firing a gun a year before."

7. Para 29:

"The evidence had concentrated on the possibility that a particle would have got into the pocket as a result of indirect contamination on a number of identified occasions on which this might have occurred. The prosecution case had been that on each such occasion this was so unlikely that it could be discounted."

8. Para 37:

"The remainder of Mr Keeley's evidence was devoted to consideration of the likelihood of secondary contamination. He was taken in detail through each occasion on which the defence suggested that there might have been a possibility of such contamination and gave his opinion in respect of each instance that contamination was 'most unlikely"

9. Para 38:

"A single particle of FDR had been found in the pocket of the appellant's coat. According to the evidence that Mr Keeley gave to us, this was an equally unlikely event, whether it had come from the cartridge that killed Miss Dando, or from some innocent source. There was even chance that it had resulted from innocent contamination."

10. Para 44:

"I am not going to attempt a statistical probability of this happening by reason of innocent contamination. We submit that it is so unlikely that you can safely ignore the possibility of innocent contamination."

12. Problem of replacing causal explanations with purely diagnostic reasoning

There appears to have been confusion about the entire notion of conditional probability that makes the use of the LR irrelevant. Specifically, some of the discussion is couched not in terms of a natural causal model (whereby the truth or otherwise of a hypothesis leads to certain types of evidence being more or less likely to be observed) but rather a purely diagnostic model in which the experts consider explicitly which of the alternative hypotheses most naturally explains the observed evidence. This difference is summarised in Figure 4.
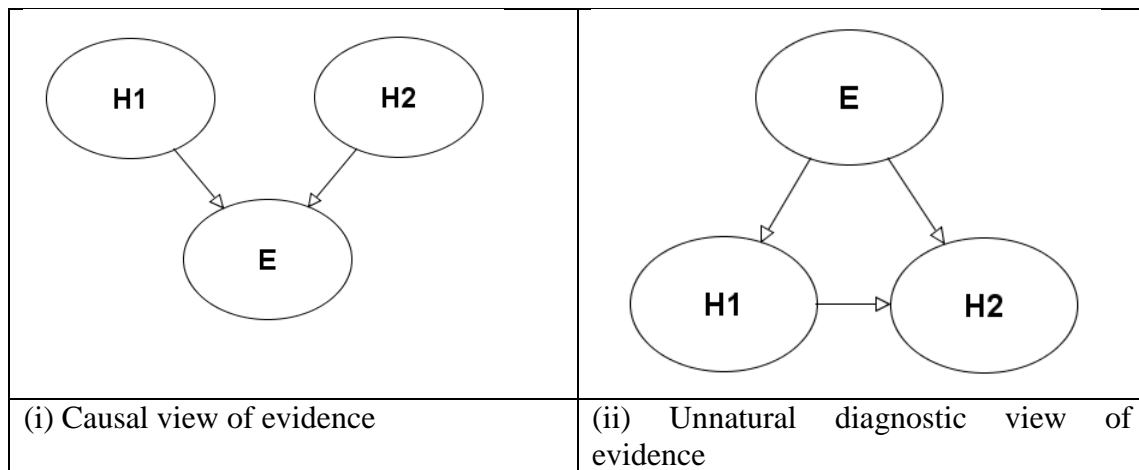

(i) Causal view of evidence	(ii) Unnatural diagnostic view of evidence

Figure 34 Causal versus diagnostic view of evidence

Specifically, some discussion focuses on whether the evidence E of the discovery of a single particle of FDR in the coat is more likely to imply $H2_p$ ("Particle found in pocket is from gun that killed JD") or *not $H2_p$*

That the unnatural diagnostic model seems to have been unwittingly used is evident from the fact that in this case it make no sense to elicit from experts $P(E \mid H2_p)$ and $P(E \mid \text{not } H2_p)$ because in both cases the probabilities are equal to one. Instead we have to consider $P(H2_p \mid E)$ and $P(\text{not } H2_p \mid E)$. Much of what is contained in the transcript indicates that at least some of the experts and lawyers involved were assuming that it was $P(H2_p \mid E) = P(\text{not } H2_p \mid E) = 0.01$. If we assume this then it is true that the evidence has no probative value on $H2_p$ but that does not mean the evidence has no probative value on Hp.