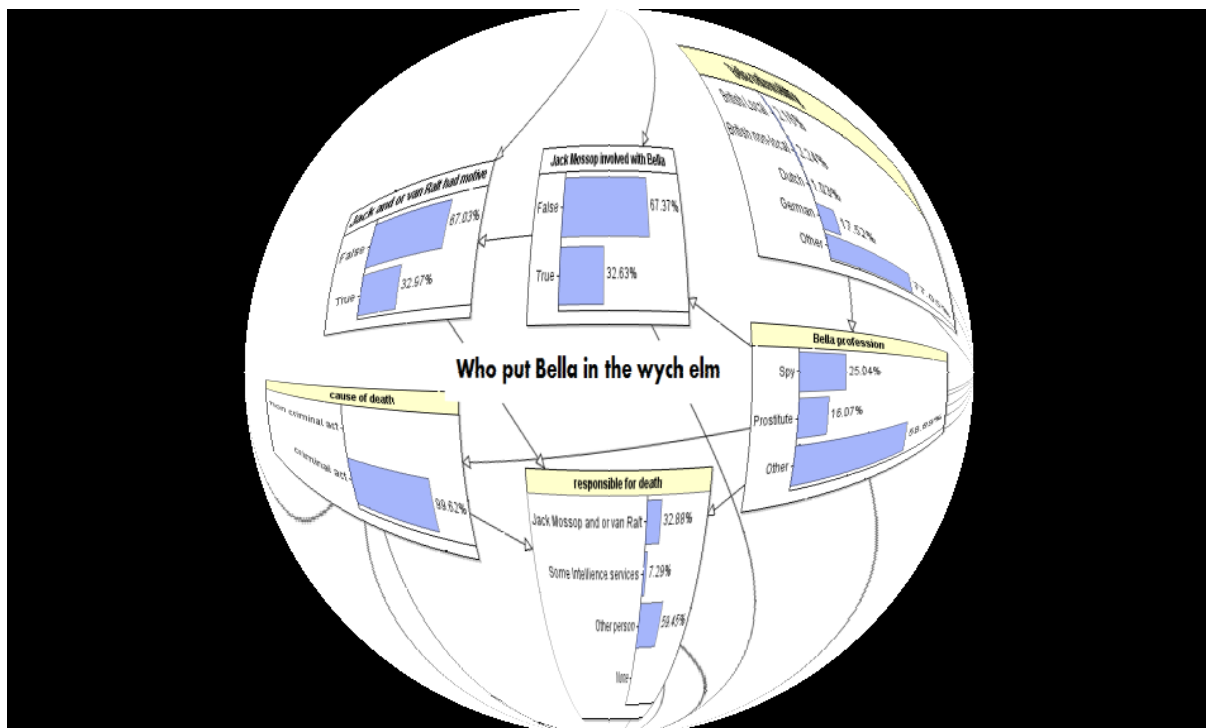


## Who put Bella in the wych-elm? A Bayesian analysis of a 70 year-old mystery



Norman Fenton and Martin Neil

Risk and Information Management Research Group  
School of Electronic Engineering and Computer Science

Queen Mary University of London

London E1 4NS

n.fenton@qmul.ac.uk and m.neil@qmul.ac.uk

2 August 2014

## 1. Background and Summary

The body of a female (who came to be known as Bella) found stuffed in a wych-elm tree in Worcester in 1943 was never identified. Nor was there ever any reliable witness of the events that led to her death. Based on forensic analysis of the body, as well as hearsay evidence and folklore, a number of hypotheses about how and why she died and who was responsible were proposed in subsequent years. One man, Jack Mossop, who died in 1942 (the year before Bella's body was discovered), was named after his death by two different relatives (his wife and a cousin) as having been involved in Bella's death. However, in the absence of hard evidence and the lack of promising leads the case was eventually closed as unsolved in 2005.

In June 2014 the BBC asked us whether statistical/probabilistic reasoning could help solve the mystery of who put Bella in the wych-elm, as this was to be the subject of one of the programmes in the Punt PI Series on BBC Radio 4 (<http://www.bbc.co.uk/programmes/b04c9dfn> first transmitted on 2 August 2014 and available on BBC iPlayer). Some of the background and theories about the case can be found in [7] and [8], although not all of the information provided there was consistent with the more detailed information given to us by the BBC team.

The case involves multiple connected unknown hypotheses, and multiple related pieces of uncertain information and is therefore suitable for analysis using a probabilistic technique called a Bayesian Networks [3]. A Bayesian Network (BN) is a model that combines all of the different types of relevant information together in order to calculate the probabilities of all of the unknown hypotheses. In the Bella case a BN model was developed using a method recently proposed for legal cases that is based on applying a small number of repeatable patterns of reasoning (idioms) [5]. Once all of the available knowledge available about the case was entered into the model, in the form of 'prior' probability assumptions, the following 'posterior' probabilities were determined:

- 99% that the cause of death was criminal
- 97% that Bella was not British
- 93% that Bella was still alive when put in the tree
- 33% that Jack Mossop was involved in her death (and 7% that it was some intelligence service)
- 25% that Bella was a spy, 16% that she was a prostitute.

While there is no possible way to 'formally validate' the model it is easy to use the model to see how the results vary under a wide range of different assumptions and to show the sensitivity of the results to these different assumptions. For example, the results are highly sensitive to the assumption that the Police really did exhaust all missing person leads. Moreover, the case confirms that the BN method for modelling complex legal cases provides important insights (like finding which pieces of evidence are most and least important) that are not easily available by other methods. The BN model is available for readers to experiment with and can be downloaded here:

<http://www.eecs.qmul.ac.uk/~norman/projects/bella/bella.html>

Execution of the model requires free software, AgenaRisk [1], which can be downloaded from [www.AgenaRisk.com](http://www.AgenaRisk.com).

## 2. Introduction

On 18 April 1943 four boys, trespassing while hunting for birds' nests in Hagley Woods Worcestershire, stumbled upon the body of what became known as Bella. The task of examining the body fell to Prof James Webster and Dr John Lund of the West Midlands Forensic Science Laboratory - they concluded that she had been asphyxiated and helped the police design the photofit believing she was around 35 years old, just under five feet tall, with mousy brown hair, irregular incisors, and had probably given birth. A few people came forward looking for loved ones, but none matched the body. Despite quite a trawl of missing persons Bella was never identified.

Around Christmas 1943, graffiti began to appear on the walls across the West Midlands "Who put Luebella down the wych-elm" which then became "Who put Bella down the wych-elm" but the police could not find out who wrote it. As it was copycat graffiti they eventually thought it was a hoax.

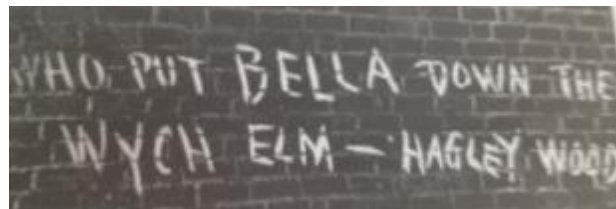


Figure 1 Graffiti found in late 1943

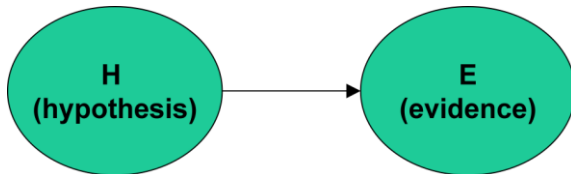
The rumour that she may have had her hand buried separately lead to the idea that it might have been witchcraft; the idea of her being involved in spying took off when journalists and authors went back over the story.

The Wolverhampton Express and Star reran the story in November 1953 and Anna of Calverley (whose real name was Una Mossop) came forward claiming to know that the murder victim, was a Dutch woman who had arrived in England illegally about 1941 and the perpetrator was dead. When the police tracked her down she said she had been married to Jack Mossop who had confessed to being involved in the demise of the woman. She pointed the finger at a "Dutchman called Van Ralt" and believed they may have meant to punish, but not kill Bella. She thought they were involved in spying. The police tried, but failed, to track down a person called Van Ralt. Jack Mossop died in Stafford Mental Hospital in 1942 and there is little evidence in the file that the Police did much on this case. Over the years various other people came forward claiming either to have seen or heard information relevant to the case (some of which is included in our analysis below), or to have been told information by family members. The most relevant was from a lady called Judith O'Donovan who was Jack Mossop's cousin. She claimed that it was "known in the family" that Jack had been involved in the death of a woman and also that Jack was a 'traitor'.

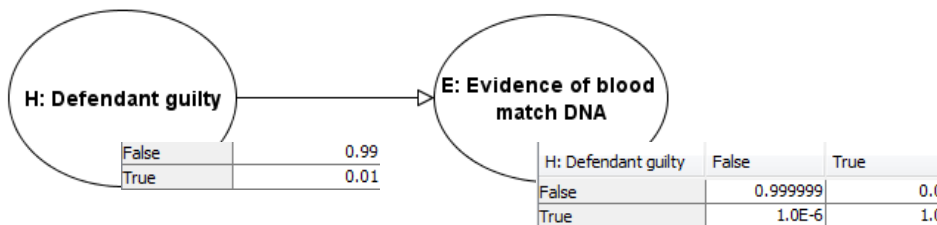
The case was finally closed in 2005.

### 3. Bayes and legal Reasoning

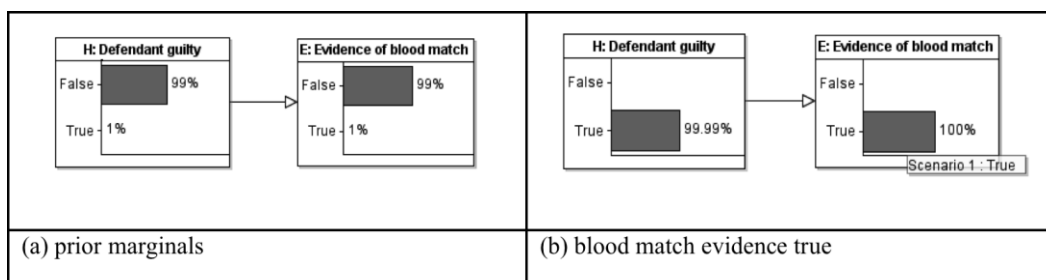
Probabilistic reasoning of legal evidence often boils down to the simple causal scenario shown in Figure 1a (which is a very simple BN): we start with some hypothesis  $H$  (normally the ultimate hypothesis that the defendant is or is not guilty) and observe some evidence  $E$  (such as an expert witness testimony that the defendant's blood does or does not match that found at the scene of the crime).



a.



b.



c.

**Figure 2 (a) Causal view of evidence. (b) BN for blood match DNA evidence with NPTs shown. (c) Running the simple model. Note: In this and all subsequent screenshots of the BN outputs all probabilities are expressed as percentages rather than values between 0 and 1. Hence the marginal probability for the defendant being guilty here is  $P(\text{Guilty})=0.01$  and  $P(\text{not Guilty}) = 0.99$**

The direction of the causal structure makes sense here because the defendant's guilt (innocence) increases (decreases) the probability of finding incriminating evidence. Conversely, such evidence cannot 'cause' guilt. Although lawyers and jurors do not formally use Bayes' Theorem (and the ramifications of this, for example in the continued proliferation of probabilistic reasoning fallacies are explained in depth in [3], they would normally use the following widely accepted intuitive legal procedure for reasoning about evidence:

- We start with some (unconditional) prior assumption about guilt (for example, the ‘innocent until proven guilty’ assumption equates to the defendant no more likely to be guilty than any other member of the population).
- We update our prior belief about  $H$  once we observe evidence  $E$ . This updating takes account of the *likelihood* of the evidence, which is the chance of seeing the evidence  $E$  if  $H$  is true.

This turns out to be a perfect match for Bayesian inference. Formally, we start with a prior probability  $P(H)$  for the hypothesis  $H$ ; the likelihood, for which we also have prior knowledge, is formally the conditional probability of  $E$  given  $H$ , which we write as  $P(E|H)$ . Bayes theorem provides the formula for updating our prior belief about  $H$  in the light of observing  $E$ . In other words Bayes’ calculates  $P(H|E)$  in terms of  $P(H)$  and  $P(E|H)$ . Specifically:

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

As an example, assume for simplicity that a blood trace found at the scene of a crime must have come from the person who committed the crime. The blood is tested against the DNA of the defendant and the result (whether true or false) is presented. This is certainly an important piece of evidence that will adjust our prior belief about the defendant’s guilt. Using the approach described above we could model this using the BN shown in Figure 1b where the tables displayed are the Node Probability Tables (NPTs) that are specified as prior beliefs.

Here we have assumed that the ‘random DNA match probability’ is one in a million, which explains the entry in the NPT for  $P(E | \text{not } H)$  (the probability of a match in an innocent person). We also assume, for simplicity, that we will definitely establish a match if the defendant is guilty, i.e.  $P(E | H)=1$ , and that the DNA analysis procedures are perfect (see [6] for a full discussion of the implications when these assumptions do not hold). Finally, we have assumed that the prior probability of guilt is 0.01 (which would be the case if, for example, the defendant was one of 100 people at the scene of the crime). With these assumptions the marginal distributions (i.e. the probabilities before any evidence is known) are shown in Figure 1c (left hand panel).

If we discover a match then, as shown in Figure 1c (right hand panel), when we enter this evidence, the revised probability for guilt jumps to 99.99%, i.e. 0.9999, so the probability of innocence is now one in 10,000. Note that, although this is a small probability, it is still significantly greater than the random match probability; confusing these two is a classic example of the **prosecutor’s fallacy** [3].

Despite its elegant simplicity and natural match to intuitive reasoning about evidence, practical legal arguments normally involve multiple pieces of evidence (and other issues) with complex causal dependencies. To handle the necessary complexities involved we use BNs [4] which can be regarded as extensions to the graphical model in Figure 2(a) with arbitrary number of nodes and links.

To develop the BN model for the Bella case we used the approach described in [5] that enables us to build relevant BN models, no matter how large, that are still conceptually simple because they are based on a very small number of repeated ‘idioms’ (where an idiom is a generic BN structure).

The most important idioms are the *cause-consequence idiom*, the *opportunity-motive idiom*, and the *evidence accuracy idiom*. The generic case of the evidence accuracy idiom is shown in Figure 3. It turns out that this is a very important idiom in the Bella case. This idiom makes clear that what a piece of evidence **really** tells us about the hypothesis is dependent on the accuracy of the evidence. If, for example, the stated evidence  $E$  is that there is a DNA match but it is known that the person making this statement was using contaminated then the accuracy  $A$  is low and our posterior belief in  $H$  once  $E$  is presented does not change as much as if  $A$  was high.

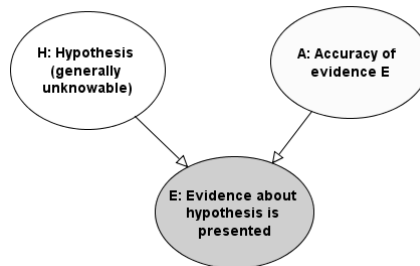


Figure 3 Evidence Accuracy idiom – generic form

## 4. The BN Model for the Bella case

The full BN model is shown in Fig 6 (Appendix). Here we explain the key parts of the model and its rationale. The approach described in [5] involves identifying variables (represented by nodes in the BN) that correspond to the core unknown hypotheses whose probabilities we wish to determine. The most important is what can be considered as the ultimate hypothesis, namely:

**“responsible for death”**: In the absence of any named suspects other than Mossop (with or without accomplice van Ralt) this variable has four distinct states of interest, namely:

- Jack Mossop and/or van Ralt
- Some Intelligence services
- Other person
- None (this state is included because we cannot rule out the possibility that Bella’s death was not a criminal act)

The other core hypotheses (and their associated mutually exclusive and exhaustive states) are shown in Table 1.

**Table 1 Core hypotheses and their possible states**

<b><i>Core Hypotheses</i></b>	<b><i>Possible states of the hypotheses</i></b>
Cause of death	<i>Criminal, Non-criminal</i> (Non-criminal covers: suicide, accident, or natural causes. Criminal covers both murder and the specific punishment hypothesis relating to Jack Mossop).
Bella still alive when put/got in tree	<i>True, False</i>
Bella nationality	<i>British local; British non-local; Dutch; German; Other</i>
Bella profession	<i>Spy; Prostitute; Other</i>

For some of the above hypotheses there is no or little direct evidence, but there is evidence for related intermediate hypotheses. For example, consider the “cause of death” hypothesis (Fig 2):

- If we know for certain that her “body was still warm when put in the tree”, then this would be strong evidence to support the “cause of death” hypothesis being “criminal”. However, we do not know for certain that this ‘evidence’ is true and so it is also treated as an (intermediate) unknown hypothesis and modelled using the evidence accuracy idiom where the reliability of the expert (Webster) has to be determined.
- If we knew for certain that Bella was asphyxiated (as claimed by the forensic expert) then this would be extremely strong evidence to support the “cause of death” hypothesis being “criminal”. However, we do not know for certain that “Bella was asphyxiated”. Instead, there is a piece of evidence “material embedded in teeth” that the forensic expert found that supports this hypothesis. But this evidence could be partly explained by the evidence that the scouts forced the skull into the tree (see Fig 3 which shows the changing probabilities for this part of the model as this evidence – and only this evidence – is entered)

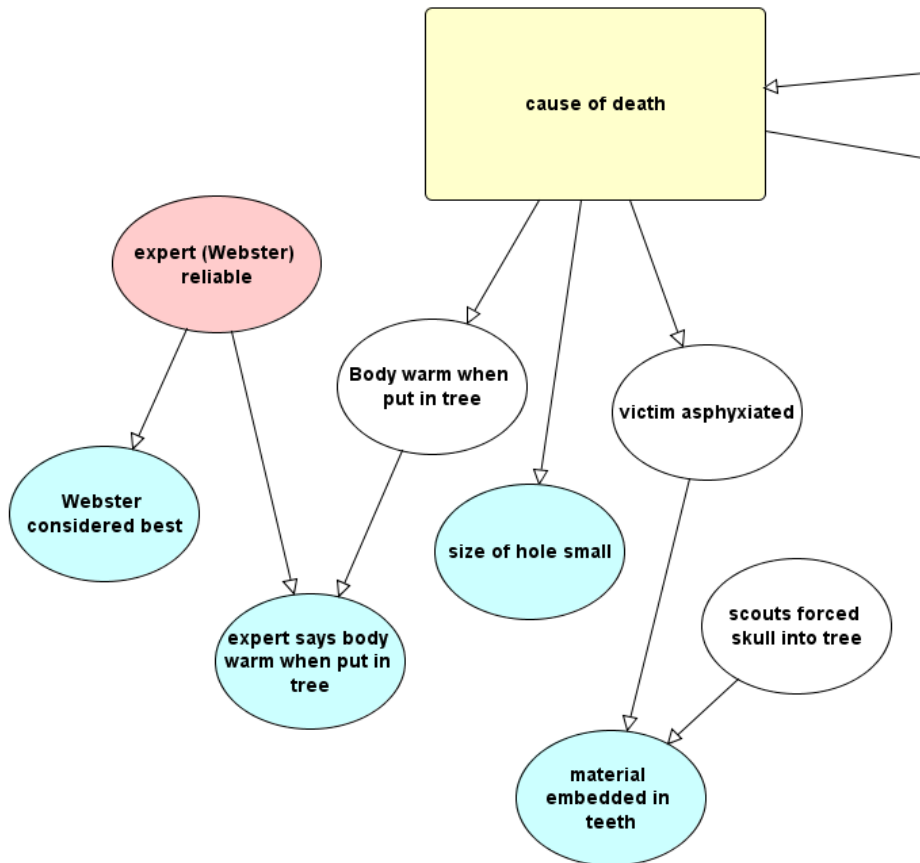


Figure 4 Part of BN model dealing with "cause of death"

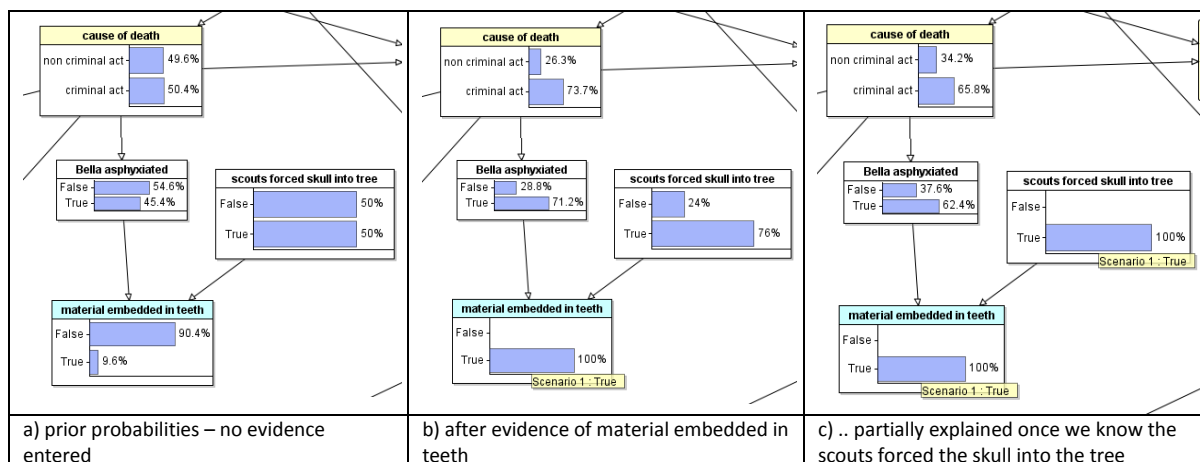


Figure 5 Was Bella asphyxiated?

For the “Bella nationality” hypothesis a key intermediate hypothesis is the police ‘evidence’ that there was “no valid matching report of person missing”. But we do not know if this is true or not – our belief depends on the accuracy/reliability of the police and their claims to have undertaken extensive appeals and investigations (which, given that this was in the middle of World War 2,



may have been extremely difficult). Hence, again we apply the evidence accuracy idiom as shown in Fig 4.

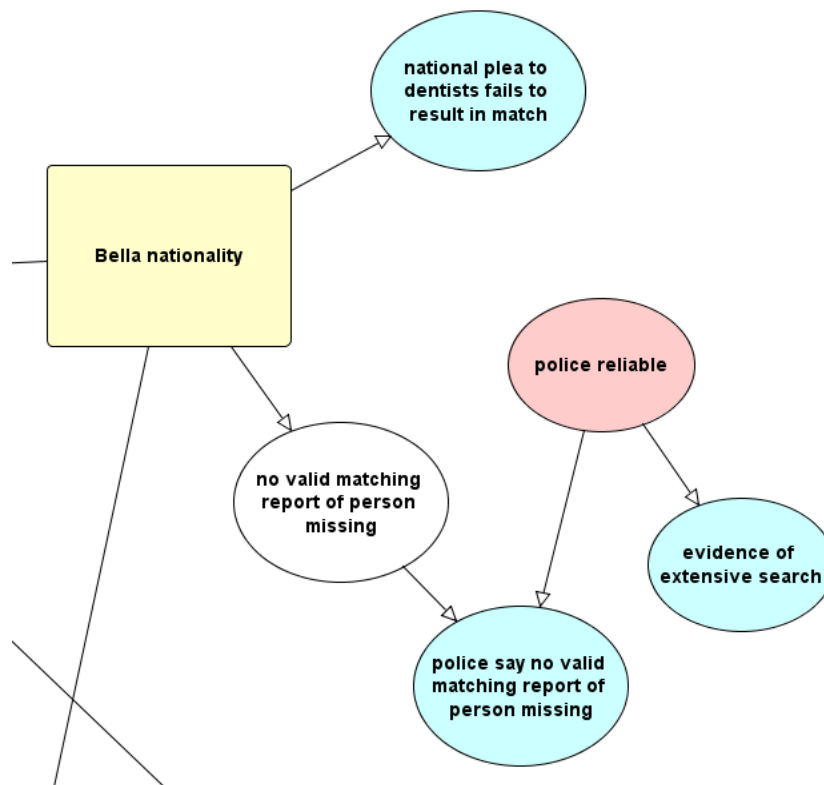


Figure 6 Part of BN relating to Bella nationality

The results – on the Bella nationality node - of incrementally entering relevant evidence is shown in Fig 5. The prior for Bella nationality is based on actual 1941 census data for the Worcester area (total 76448 women aged 21-40, of whom 6135 were non-British including 520 Germans and 56 Dutch. We assume 20% of British women come from outside the area).

<b>Bella nationality</b> British Local 72.4329% British non-local 18.1063% Dutch 0.0727% German 0.6751% Other 8.713%	<b>Bella nationality</b> British Local 63.7583% British non-local 16.223% Dutch 0.1529% German 1.5382% Other 18.3275%	<b>Bella nationality</b> British Local 6.7759% British non-local 3.8522% Dutch 0.6799% German 7.208% Other 81.484%
a) Prior probability (no evidence)	b) After police say no valid matching report of missing person	c) After evidence of extensive search

Figure 7 Bella nationality updated when evidence is entered

In line with [5] we use the ‘motive’ idiom, whereby an intermediate ‘motive’ hypothesis (“Jack and/or van Ralt had motive to harm Bella”) is a parent of the “responsible for death” node. This node also has, as a parent, an intermediate hypothesis “Jack Mossop involved with Bella”; this is important (and acts as an ‘opportunity node’ in the sense of [5]) because Una Mossop provides evidence of such a relationship.

The most important evidence to support the theory that Jack and/or van Ralt were involved in Bella's death comes from the statements of Una Mossop and Jack's cousin Judith O'Donovan. Once again the evidence accuracy idiom plays a key role here. If either of these witnesses is perfectly reliable then it would follow that their 'story' were true. But there is no other real evidence to corroborate their stories and the common features of their stories could be explained by being 'part of family folk-lore'<sup>1</sup>. In fact, in the absence of other real evidence, when we first enter as an observation the fact that "Una claims Jack Mossop and van Ralt were responsible for the death of Bella", this observation has two effects:

- 1) It does slightly increase the probability that Jack and/or van Ralt were responsible; but
- 2) It significantly decreases our belief in Una's reliability (starting with a 50:50 prior it drops immediately to less than 2%).

As evidence is entered (most of which does not contradict Una) our belief in her reliability creeps up. However, after all the evidence is entered our belief in Una's reliability only gets to 33%.

---

<sup>1</sup> It is also possible that, even if their story is true – i.e. that Jack Mossop had been involved in a gruesome event with a lady – it may have completely different woman to Bella. However, the model assumes that IF Jack was involved with a "Dutch woman" as claimed by Una and Judith then this woman was Bella – the model gets too complex otherwise.

## 5. Overall results and Conclusions from the model

Once all the evidence is entered and the model is run (using the BN tool available free from [1]) the revised probabilities for the core hypotheses are:

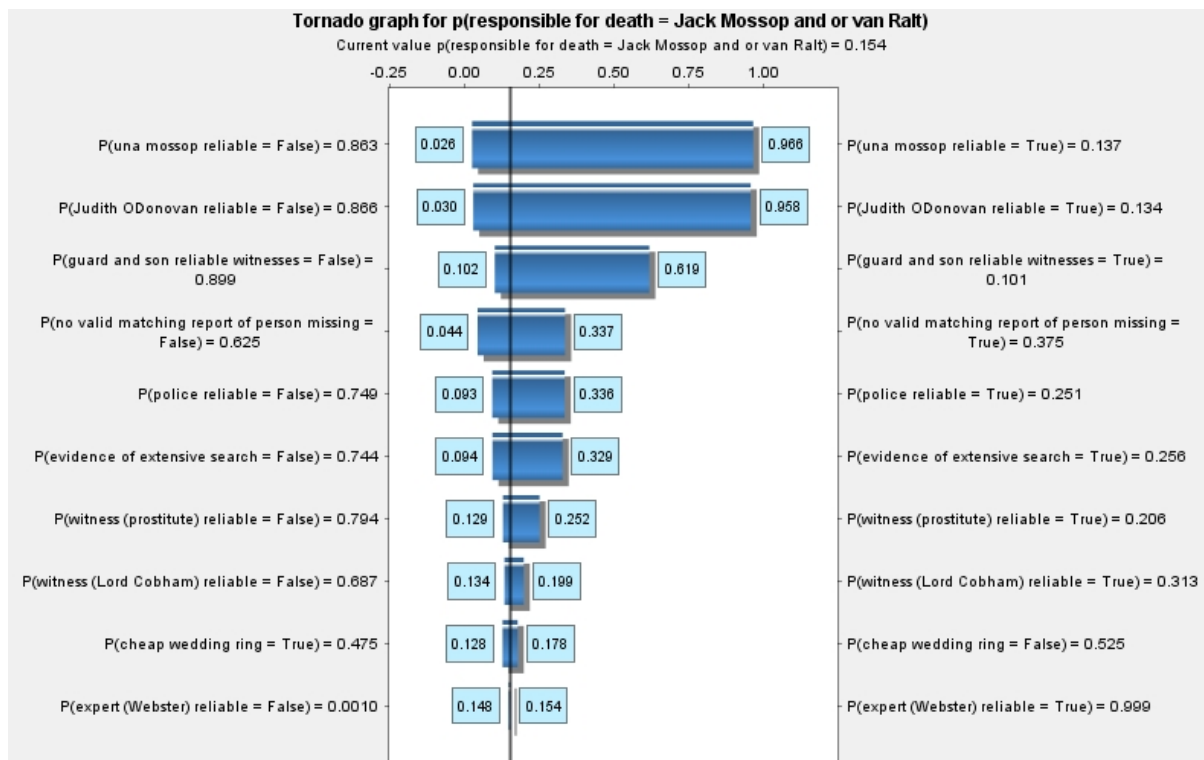
- 99% that the cause of death was criminal
- 97% that Bella was not British (however, there is less than a 2% chance she was Dutch – she was much more likely to be German at 18%)
- 93% that Bella was still alive when put in the tree
- 33% that Jack Mossop was involved in her death (and 7% that it was some intelligence service)
- 25% that Bella was a spy, 16% that she was a prostitute.

In order to get to at least a 95% probability that Jack Mossop was responsible we would need to find FOUR additional witnesses similar to Una and Judith. We know this because we are able ‘virtually’ to add to the model additional nodes that are equivalent to adding a witness like Una; each time we add such a witness ‘corroborating’ the story the belief in the reliability of the witnesses increases as does the belief in Jack’s involvement.

While there is no possible way to ‘formally validate’ the model it is easy to use the model to see how the results vary under a wide range of different assumptions and to show the sensitivity of the results to these different assumptions. For example, the results are highly sensitive to the assumption that the Police really did exhaust all missing person leads. If we change the observation on the node “evidence of extensive search” from true to false, then some of the key posterior probabilities change significantly. In particular, we find:

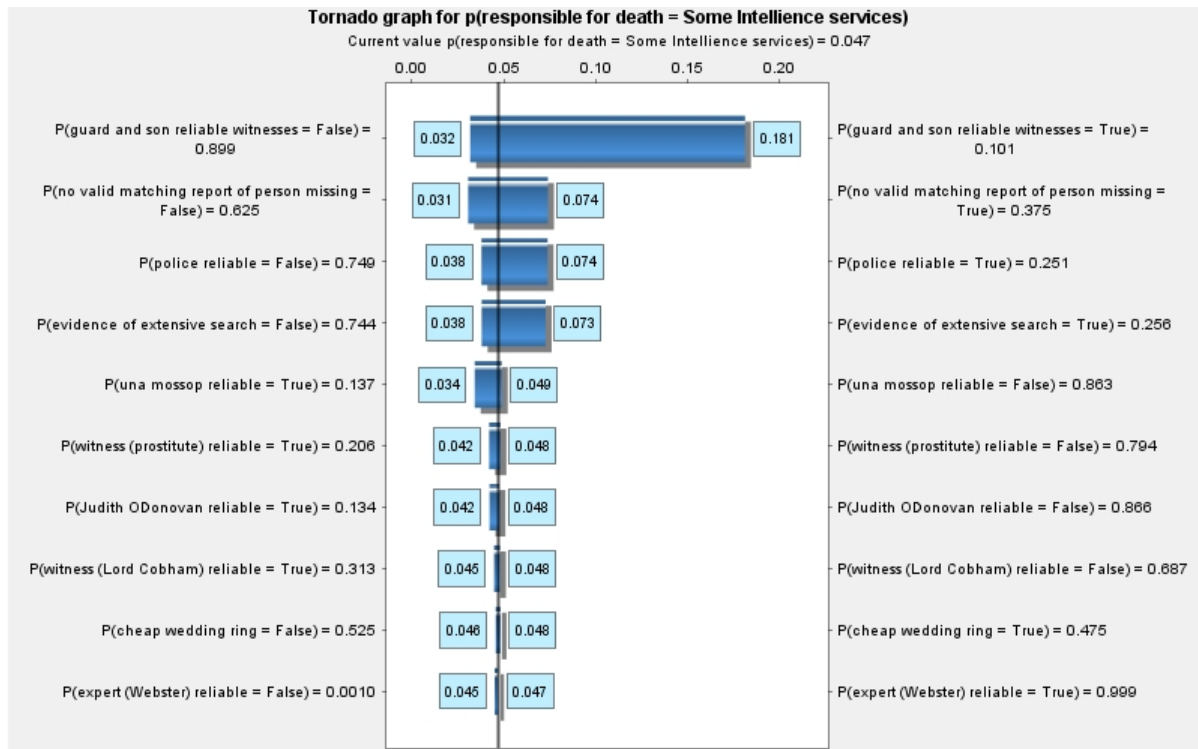
- 24% that Bella was not British (down from 97%)
- 9% that Jack Mossop was involved in her death (down from 33%)
- 6% that Bella was a spy (down from 25%) and 5% that she was a prostitute (down from 16%).

Figure 8 shows (using a Tornado graph produced by the AgenaRisk BN tool that runs the model) the unknown nodes that most impact the probability that Jack and/or van Ralt were responsible for Bella’s death. The sensitivity analysis takes account of nodes where evidence is entered, except in this case we removed the evidence of extensive search. Clearly, as expected, the reliability of the key witnesses Una and Judith have greatest impact, but the figure also shows the impact of the police reliability.



**Figure 8 Sensitivity analysis tornado graph for the probability that Jack and/or van Ralt were responsible for Bella's death**

Figure 9 shows a similar sensitivity analysis, but in this case with respect to some intelligence services being responsible for Bella's death. The impact here is, not surprisingly, dominated by the reliability and the guard and his son who essentially provided the only clear support for this hypothesis.



**Figure 9 Sensitivity analysis tornado graph for the probability that some intelligence services were responsible for Bella's death**

Readers are invited to experiment with different assumptions themselves in the BN model, which can be downloaded here:

<http://www.eecs.qmul.ac.uk/~norman/projects/bella/bella.html>

Execution of the model requires free software, AgenaRisk [1], which can be downloaded from [www.AgenaRisk.com](http://www.AgenaRisk.com).

We believe that using Bayesian methods to model complex legal cases can provides important insights (like finding which pieces of evidence are most and least important) that are not easily available by other methods. The use of Bayes also promotes the clear elucidation of legal arguments, highlights the quantitative underpinnings provided by expert evidence and hence can help detect biases and fallacies in legal reasoning.

## Acknowledgments

We are grateful to Sarah Bowen and Steve Punt of the BBC for inviting us to work on this case. Our work was partly funded by the ERC under Advanced Grant ERC-2013-AdG339182-BAYES\_KNOWLEDGE [2].

## References

- [1] Agena (2014). AgenaRisk software, [www.agenarisk.com](http://www.agenarisk.com)
- [2] ERC (2014) "Effective Bayesian Modelling with Knowledge Before Data (Short Name: BAYES-KNOWLEDGE)", [https://www.eecs.qmul.ac.uk/~norman/projects/B\\_Knowledge.html](https://www.eecs.qmul.ac.uk/~norman/projects/B_Knowledge.html)
- [3] Fenton, N. & Neil, M. (2011). Avoiding Legal Fallacies in Practice Using Bayesian Networks. *Australian Journal of Legal Philosophy*, 36: 114-150
- [4] Fenton, N. E., & Neil, M. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks*. Boca Raton: CRC Press.
- [5] Fenton, N. E., Lagnado, D. A., & Neil, M. (2013). A General Structure for Legal Arguments Using Bayesian Networks. *Cognitive Science*, 37(61-102).
- [6] Fenton, N. E., Neil, M., & Hsu, A. (2014). "Calculating and understanding the value of any type of match evidence when there are potential testing errors". *Artificial Intelligence and Law*, 22. 1-28
- [7] Haughton, B "Bella in the wych elm", [http://brian-haughton.com/articles/bella\\_in\\_the\\_wych-elm/](http://brian-haughton.com/articles/bella_in_the_wych-elm/)
- [8] Vale, A (2013) "Is this the Bella in the wych elm? Unravelling the mystery of the skull found in a tree trunk" *The Independent*, 18 April 2013  
<http://www.independent.co.uk/news/uk/home-news/is-this-the-bella-in-the-wych-elm-unravelling-the-mystery-of-the-skull-found-in-a-tree-trunk-8546497.html>

## Appendix: The Full BN Model

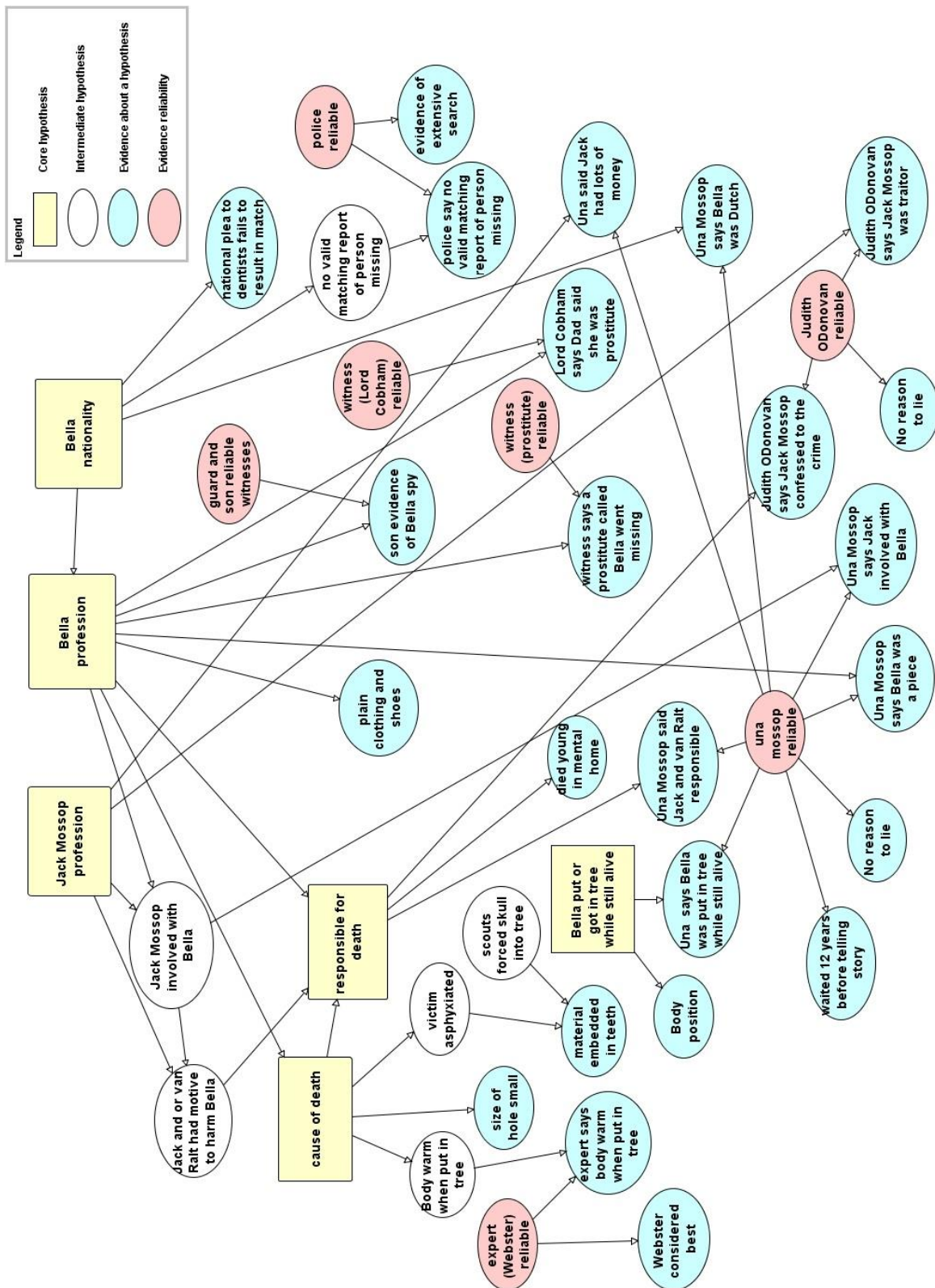


Figure 10 The full BN Model