The Opportunity Prior: A Simple and Practical Solution to the Prior Probability Problem for Legal Cases

Norman Fenton,
School of Electronic Engineering and Computer Science, Queen Mary, University of London and Agena Ltd, UK
n.fenton@qmul.ac.uk

David Lagnado,
Division of Psychology and Language Science, University College London, UK
d.lagnado@ucl.ac.uk

Christian Dahlman,
Law Department, Lund University, Sweden
Christian.Dahlman@jur.lu.se

Martin Neil,
School of Electronic Engineering and Computer Science, Queen Mary, University of London and Agena Ltd, UK
m.neil@qmul.ac.uk

ABSTRACT

One of the greatest impediments to the use of probabilistic reasoning in legal arguments is the difficulty in agreeing on an appropriate prior probability for the ultimate hypothesis, (in criminal cases this is normally "Defendant is guilty of the crime for which he/she is accused"). Even strong supporters of a Bayesian approach prefer to ignore priors and focus instead on considering only the likelihood ratio (LR) of the evidence. But the LR still requires the decision maker (be it a judge or juror during trial, or anybody helping to determine beforehand whether a case should proceed to trial) to consider their own prior; without it the LR has limited value. We show that, in a large class of cases, it is possible to arrive at a realistic prior that is also as consistent as possible with the legal notion of 'innocent until proven guilty'. The approach can be considered as a formalisation of the 'island problem' whereby if it is known the crime took place on an island when *n* people were present, then each of the people on the island has an equal prior probability $1/n$ of having carried out the crime. Our prior is based on simple location and time parameters that determine both a) the crime scene/time (within which it is certain the crime took place) and b) the extended crime scene/time which is the 'smallest' within which it is certain the suspect was known to have been 'closest' in location/time to the crime scene. The method applies to cases where we assume a crime has taken place and that it was committed by one person against one other person (e.g. murder, assault, robbery). The paper considers both the practical and legal implications of the approach. We demonstrate how the opportunity prior probability is naturally incorporated into a generic Bayesian network model that allows us to integrate other evidence about the case.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

/ICAIL '17,/ June 12-16, 2017, London, United Kingdom

© 2017 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-4891-1/17/06…$15.00
http://dx.doi.org/10.1145/3086512.3086519

CCS CONCEPTS

• **Mathematics of computing** →Bayesian networks; • **Applied computing** → Law;

KEYWORDS

Opportunity Prior Probability; Island Problem; Prior Probability; Bayesian Networks; Crime Scene and Time.

1 INTRODUCTION

When the police apprehend someone suspected of a crime, one of the first questions they ask the suspect is where he was at the time of the crime. This is potentially a very diagnostic question: if the suspect can show that he was elsewhere, then he cannot (except in special cases that we ignore in this paper) have committed the crime. If, however, it can be shown that he was at the crime scene at the time of the crime, then he is ruled into a (relatively) small subset of possible perpetrators. In classic investigative terms this establishes ***opportunity*** which, along with *motive* and *means*, is often considered necessary for conviction. Thus, finding out about the suspect's whereabouts in relation to the crime scene and time is a critical starting point for most investigations.

In this paper we will argue that the same logic applies to later stages of the legal process: in particular, when the suspect is charged with the crime and how we evaluate the strength of evidence against him. Information about the suspect's whereabouts in relation to the crime scene is crucial – often as a starting point for establishing a case for or against him, before other evidence is presented.

One key point, typically neglected in formal analyses of evidential reasoning, is that case information often allows us to make reasonable judgments about the probative force of opportunity evidence. Indeed, such information, typically used to drive an investigation, is equally relevant in court when we are assessing the probability that a suspect is guilty.

Why is such information often neglected in formal analyses? We will argue that this neglect hinges on several flawed assumptions, including: (i) the belief that at best opportunity evidence 'fails to exclude' a suspect, but it does not have a positive confirmatory value on the hypothesis that he committed the crime; and (ii) that placing someone at the crime scene (close to time of crime) at most means he is one of N other possible perpetrators, and this set includes all people who could have committed the crime, so N is a very large number (which is also hard to estimate). Thus, information about opportunity at best gives imprecise and typically very small prior probabilities, e.g. 1/(some large population size).

As we shall see, both assumptions are wrong. We present a principled approach to quantifying information about opportunity evidence that corrects these misconceptions. This approach maps naturally onto typical investigative practices[1] and shows how opportunity evidence is often a key factor in determining a suspect's guilt.

Our objective is to provide a simple and realistic method for estimating what we call the "opportunity prior probability"; this is the probability that a suspect is guilty of the crime for which he is accused based ***only*** on evidence about his proximity (in space and time) to the crime scene location and time.

The paper is structured as follows. In Section 2 we explain how our approach can be viewed as a natural and practical extension of the classic 'Island problem' scenario for the use of Bayes to assess probability of guilt given evidence and a prior probability of guilt. In Section 3 we provide a formal definition of the Crime Scene (*CS*) and Crime Time (*CT*), while in Section 4 we define the formal notion of the ***extended*** *CS* and *CT*, which is necessary for establishing an opportunity prior in cases where the suspect has not been proven to be at the *CS* during the *CT*. This leads to an opportunity prior $1/N$ for the suspect. In Section 5 we explain how N can be reduced in certain pathological cases. In Section 6 we address the concern of convicting innocent bystanders when N is very low. Finally, in Section 7 we describe a generic Bayesian Network model that enables us to incorporate the opportunity prior along with all other relevant evidence in order to compute a posterior probability of guilt.

2 BAYES AND THE LAW: THE CLASSIC ISLAND PROBLEM

To illustrate the full potential of using Bayesian probabilistic reasoning in legal arguments it is common to consider the classic 'island problem' whereby a crime is assumed to have taken place on an island when it is known n people were present[2–4]. This set of people is the 'reference class' for the crime and the defendant is one of them. Before any evidence is considered each of the people on the island has an equal prior probability $1/n$ of having carried out the crime. The Bayesian approach for legal arguments, as described for example in[5–8] can be summarised as follows: The prior odds against guilt (n-1 to 1 in this case) are multiplied by the *likelihood ratio* (LR) to arrive at the posterior odds of guilt. The LR is the probability of the evidence under the prosecution hypothesis divided by the probability of the evidence under the defence hypothesis. Suppose, for example, we discover evidence linking the defendant to the crime such as that he/she has a DNA profile matching a DNA trace left by the person who committed the crime. Suppose the LR for this evidence is 10,000. Then if n = 100 there is a very strong posterior probability that the defendant is guilty (about 99.9%), whereas if n = 1,000,000 the posterior probability of guilt is only 1%. It is clear, therefore, that while the LR offers important information about the probative value of evidence, it is our prior probability of guilt that determines whether or not we believe the evidence is sufficient to convict.

We now provide a formalisation of the island problem whereby we seek to narrow down the crime scene and time as far as can possibly be agreed. If, for example, the crime definitely took place in a particular village in the island on a particular quiet evening then, although ***all*** the islanders ***could*** have been present, we should be able to provide a good estimate of those who actually ***were*** there during that evening. On an island where n = 1,000,000 this might reduce the relevant n to 100. If it is proven that the defendant was one of the people in the village then the prior probability of guilt is 1/100. If the defendant insists he/she was not present that evening, then the first task for the prosecution is to determine the 'closest' place/time where it is agreed the defendant was. If, for example, he/she was certainly at a garage one mile from the village two days earlier then we need to consider the area covering the village and a one mile circumference outside it, and the two days leading up the crime being committed. We consider this to be the 'extended crime scene/time'. Let N be an estimate of the number of people who were in this extended crime scene and time. Then we know that the crime must have been committed by one of these N people and that the defendant is one of these.

In the next two sections we define formally the notions of crime scene and crime time (section 3) and the extended versions of these (section 4) in order to arrive at objective values for the number n and N respectively.

Note that the parameter values we set in all of the examples presented could easily be replaced with distributions rather than point values, but we have used point values for ease of explanation, without any loss of generality.

3 DEFINING THE CRIME SCENE AND TIME

In what follows we assume a crime has taken place and that it was committed by one person against one other person (e.g. murder, assault, robbery).

The Crime Scene (CS): this is the smallest physical area within which it is certain the crime happened.

Example 3.1:

a) If a person was attacked in a 20-metre alleyway, but it is not certain which specific point then the *CS* is the entire alleyway.
b) If a person was mugged while standing next to a particular lamppost by Piccadilly Circus tube station then the *CS* would be an area about one metre around the lamppost.

c) If the victim was shot in a theatre, then the *CS* is the area of the theatre covering any point from which a shot could have been fired.

The Crime Time (CT): this is the smallest time interval (t, t') between which it is certain the crime took place. This interval could be as short as a millisecond (in cases where we have verified time recordings of the crime) or as long as months or even years for old cases.

The above examples are typical of scenarios in which our proposed method is most useful since the crime scene is reasonably specific and any disagreements and uncertainty about it are tightly bounded. However, in some scenarios there may be fundamental disagreement about the *CS*.

> **Example 3.2:** Suppose the police claim the crime (a murder) took place in the alleyway between 01.00 and 01.30 but the defence shows that the body could have been dumped there and that the time of death could have been any time after midnight (when the victim was last seen alive at a night club one mile away) and 01.30. Then, in such a case, the crime scene would be an area covering not just the night club and alleyway, but anywhere within which it would be possible for the victim to have got to, such that it would also be possible for the victim's body to have got back to the alleyway by 01.30. Even for such a relatively short period of time this could be a very large area since the victim could have been taken in a car and driven 30 miles away before being returned to the alleyway.

The number of people, n, at CS during CT: Although we generally do not know who was present at CS during CT it is possible to estimate the number of people *n* (other than the victim) who were. For example:

- If, in Example 3.1(a) *CT* is between 01.00 and 01.30 (i.e. early hours of the morning) *n* might be up to 5, whereas if CT is between 08.00 and 08.15 (a shorter, but much busier period) it could be 30.
- If, in Example 3.1(b) *CT* is 17.30-17.33 on a Thursday then *n* could be as high as 200, whereas if CT is 04.00-04.15 on a Tuesday morning *n* might be closer to 5.
- If, in Example 3.1(c) the victim was an actor on stage during a crowded performance then *n* would be the capacity of the theatre. If, however, the attack took place in the foyer while the performance was taking place then *n* would be a very low number (typically only a handful of people would be in the foyer at any time during the performance).
- In Example 3.2 the *CS* covers an area about 30 miles around the area of the alleyway and night club over a period of 90 minutes at night. If this was an urban area including a major city then *n* would be an estimate of the number of residents and visitors present during that period (a very large number).

Whoever committed the crime must, on the above assumptions, be among the *n* people who were present at *CS* during *CT*. In the absence of any other evidence each such person has a 1/*n* probability of being the criminal. **If it is proven that the defendant is one of these people, then the prior probability of guilt is 1/*n***.

We shall deal with the case where it is not proven that the defendant was in the *CS* during *CT* in the next section, but it is important first to clear up a very important and common misunderstanding that applies to all of the above examples. As mentioned earlier what we have done is formalise the reference class of possible suspects. Variations of this approach have been considered many times and criticised on the basis that more or less anybody in the world *could have been there* and that therefore none of these people can be ruled out as suspects. We can dismiss this concern by emphasizing that the number *n* is an estimate of the actual number of people who *were actually there*, *NOT* the number of people who *could have been there.* To hammer this difference home consider the following:

> **Example 3.3**: From CCTV footage, two men Fred and Bill are known to have been in a room when a third man was murdered. No other people were in the room at the time and so one of Fred or Bill committed the murder. Fred is charged with the crime. In the absence of any other evidence there is no doubt that ½ is a reasonable prior probability for Fred's guilt. Suppose, however, that the CCTV footage only shows that Fred plus a second man whose identity we do NOT know was in the room. In theory, any man in the world *could have been there*. But Fred's prior probability of guilt must be unchanged at ½. It is a fallacy to claim – as some have – that the probability of guilt is 1/*k* where *k* is the number of all people who could theoretically have been at the scene.

Hence, when *n* is small and it is proven that the defendant was at CS during CT the prior probability of guilt 1/*n* is relatively high.

> **Example 3.4**: In a Dutch murder case in Simonshaven in 2011[9] , the *CS* is known to be a very small area of a quiet forest and the *CT* is known to be a fairly short period of time (between 8.00pm and 8.30pm on a Saturday evening). The suspect *X* was found there close to the body of his wife. He claims they were both attacked by a man coming out of the bushes. There is no dispute that the suspect was present at *CS* during *CT*. The question is: how many other people were in the *CS* during the *CT*. Based on local knowledge, a generous estimate for the defence might be *n* = 5. So, the prior probability of guilt before considering any evidence may be set no lower than 1/5.

4 DEFINING THE EXTENDED CRIME SCENE AND TIME WHEN IT HAS NOT BEEN PROVEN THAT THE DEFENDANT WAS AT THE CRIME SCENE DURING THE CRIME TIME

In general, the defendant will dispute having been present at *CS* during *CT*. Our task is hence to determine the *extended crime scene and time* based on the suspect's 'closest' proven time and location to *CS* and *CT*. The notion of 'closeness' is derived from both *distance* from *CS* (i.e. location) and *time* from *CT* and defined by considering the most 'recent' known locations where *X* was either before or after *CT*. Specifically, we consider:

- *Case 1*: Let *L* be any location and t_1 the time where it is proven *X* certainly was *before* time t (this could include location *CS* of course) such that it was physically possible for *X* to get from *L* to *CS* before time t'. Consider the area whose centre is *CS* and whose perimeter is *d(L)* where *d(L)* is the distance of *L* from *CS*. Let *N* be the total number of people who were in this area between time $t_1$and time t'. Then exactly one of these people – which includes *X* - must have committed the crime.
- *Case 2*: Let *L* be any location and t_2 the time where it is proven *X* certainly was *after* time t (this could include location *CS* of course) such that it was physically possible for *X* to get from *CS* to *L* between time t and t_2. Consider the area whose centre is *CS* and whose perimeter is *d(L)* where *d(L)* is the distance of *L* from *CS*. Let *N* be the total number of people who were in this area between time *t* and *t2*. Then exactly one of these people – which includes *X* – must have committed the crime.

Note that, in general, there will be at least one instance of each of cases 1 and 2. Each instance results in a number $N_1, N_2, \dots, N_k$.

Let $N = \min\{N_1, N_2, \dots, N_k\}$

Then, by definition, *X* is one of exactly *N* people who could have committed the crime and, in the absence of other evidence, 1/*N* is a lower bound for the probability that *X* committed the crime. This probability is also a realistic and sensible prior for the probability of *X*'s guilt. We can also conclude that a reasonable prior probability for *X* being at the crime scene is *n*/*N*.

Example 4.1: Consider the case of a murder in the foyer of a theatre during a performance. The murder occurred sometime between 21.00 and 21.10. Suppose that it is proven that the suspect was in the theatre that evening but the suspect denies being in the foyer between 21.00 and 21.10. However, it is proven he was in the foyer when he entered the theatre at 19:00 and when he left at 21:50. So we calculate:

- N_1: the number of people who were in the foyer between 19:00 and 21:10. If the performance started at 19:30 then this will be approximately the number of people who watched the performance plus the staff in the foyer.
- N_2: the number of people who were in the foyer between 21:10 and 21:50. Assuming the performance finished at 21:45 this number will likely be slightly smaller than N_1 since some people will have left before 21:10 and some will not yet have left the theatre.

In this example it would be pointless considering any agreed time or location where *X* was before or after he entered and left the theatre since it would give rise to a number *N* larger than N_2. We conclude that $N = N_2$ in this case.

Example 4.2: In example 3.1(a) above suppose the CS is an alleyway in Barking (East London) and *CT* is 03.00 to 03.30, but the suspect claims he has never visited the *CS* and that he was at home *H* (20 miles from *CS*) during the entire *CT*. It is proven he was at home *H* at 01:00 (the last known time before *t*) and at work *W* (10 miles from *CS*) at 06:30 (the first known time after t').

In this example we have

- Location *H*: in this $N_1$is the number of people who were within an area 20 miles around the alleyway between 01:00 and 03:30.
- Location *W*: in this case N_2 is the number of people who were within an area 10 miles around the alleyway between between 01:00 and 06:30.

It is unlikely any other known location would lead to a smaller number than N_1or N_2 so *N* is the minimum of $N_1$and N_2. In other words: the set of possible perpetrators is the smallest extended crime scene/time (in terms of the number of people who were there) of which the defendant is a member.

Example 4.3 In example 3.1(a) above suppose the suspect lived 200 miles from *CS* and denies ever being close to *CS*. Suppose also that the crime took place several years before the suspect was arrested. In such a case the 'closest' known time and location could be a very long way from the CS and a long time before or after *CT*. In such a situation *N* could be very large - the number of people who lived and visited a very large area over a prolonged period of time. If, for example, the *CS* is somewhere in central London and if the 'closest location' we have for the suspect is a location 15 miles from the *CS* 6 months before the *CT*, then *N* would be in the order of 40 million – the number of people who lived in or visited an area including the whole of Greater London during that 6-month period.

5 REDUCING THE NUMBERS *n* AND *N* AND HANDLING PATHOLOGICAL CASES

While the above approach attempts to constrain the 'reference class' of potential suspects as much as possible in most cases, it should be possible to reduce *n* and *N* further. For example, it seems reasonable to always exclude from *n* and *N* an estimate of the number of people in the area during the period that could not physically have carried out the crime. Depending on the type of crime this could mean excluding all people under a certain age, all people of a certain sex, all people with certain types of physical disabilities etc. However, this touches on the critical notion of 'capability/means' which is distinct from opportunity and which would normally be considered as evidence during the case (unlike the pure location/time based opportunity evidence that we have argued should determine the prior probability). In this case the lawyers should make it clear that, if the capability evidence was explicitly used to determine the prior then it should not be counted again as evidence against the defendant. One of the benefits of the BN approach that we describe in Section 7 is that it explicitly avoids such double-counting.

There are also pathological cases whereby the number *N* (for the extended crime scene location) is massively inflated by the inclusion of people passing through the area during the extended time period who could not have been present at the actual *CS* during the *CT*. This is especially true for events that attract large numbers of people to an area for brief periods.

Example 5.1: Suppose the crime took place on 1 October in a small village with 100 residents. Suppose the 'closest' location the defendant is known to have been was 10 miles away on 5 October. Then *N* is the number of people who were within 10 miles of the village between 1 and 5 October. However, suppose that this area includes a football stadium and that on 3 October a match took place that attracted 20,000 fans from abroad. It is known – from airport and hotel records – that almost all the visitors arrived and left on the same day (3 October) and therefore could not physically have been at the *CS* during the *CT*. However, by our definition, *N* includes all 20,000 visitors. This is clearly inflated. Although we cannot rule out any of the 20,000 visitors from committing the crime, there is no reason why we cannot reduce the number based on an estimate of the number who came and left on 3 October. To not do so would be similar to the fallacy highlighted in Example 3.3 - treating the unknown 'other man' in the room of two suspects as requiring a different prior from the case where we know the identity of both men.

Another possible complication is where the defendant has – *a priori* – a lower probability of having been at the *CS* than others.

Example 5.2: Suppose the *CS* is a town in North East Scotland, but the closest known location for the suspect is his home town Bournemouth where he is known to have been both the day before and the day after the *CT* (the defendant claims to have been there for the whole period but only the day before and after have been independently confirmed). Because Bournemouth and North East Scotland are 500 miles apart the extended area includes the whole of the UK. Now, while it is certainly possible for the suspect to have got to North East Scotland in the time interval, in the absence of any other evidence it is surely reasonable to assume that the prior probability he did so is much less than people living closer to the town in North East Scotland. This suggests that we might need to consider a distance-weighted computation when calculating priors.

6 THE RISK OF CONVICTING INNOCENT BYSTANDERS

The solution presented above works well when it is known that the defendant was in the vicinity of the *CS* fairly close to the *CT*, but it may be problematic when it is proven that the defendant was very close to the CS and very close to the CT. In such cases *N* will be a very small number, and the prior probability, $1/N$, will, consequently, be high.

Example 6.1: Suppose that a man *Y*, living alone, has been murdered in his house. This is in a very quiet district with no previously reported crimes, and with just one neighbouring house 20 metres away. The neighbour *X*, also a man living alone, was known to have been at home on the night of the murder, although it is not known if he visited *Y*. The *CS* is *Y*'s living room and the *CT* is 9.00-11.00pm. By considering an extended CS to be a 20-metre perimeter around the *CS*, it is accepted that *X* was in the extended *CS* during *CT*. In such a situation *N* will be extremely low, typically $N = 2$, allowing for the possibility of a rare visitor to *Y* (invited or uninvited). As *X* is one of the *N*, his prior probability of guilt is 50%.

In Example 6.1, *X* may well become a suspect simply based on this 50% prior. The danger is that, with a prior probability of 50%, little additional evidence may be needed to meet the standard of proof. If the standard of proof requires a posterior probability of 95%, it is sufficient for conviction that the evidence presented by the prosecution has a likelihood ratio of 19 ($0.50/0.50 \times 19 = 0.95/0.05$). This is problematic since some of the cases that start with a high prior, because it is known that the defendant was very close to the *CS* very close to the *CT*, are cases where the defendant is innocent, and just happened to be nearby. The proposed solution for determining the prior probability makes it very easy for innocent bystanders to be wrongfully convicted.

One way to handle this problem is to add a strong requirement of robustness to the standard of proof, so that a defendant can only be convicted on evidence with a low likelihood ratio if other hypotheses have been investigated so thoroughly that it is highly unlikely that evidence could be produced against someone else.

7 INCORPORATING THE PRIORS INTO A BN MODEL THAT ALSO HANDLES UNCERTAINTY ABOUT *N* AND *n*

In order to properly incorporate the prior opportunity probability with other potentially complex and related evidence (as well as unknown hypotheses) to compute a rational posterior probability of guilt, it has been widely acknowledged that a Bayesian Network (BN) model is an ideal formalism[10–12].

A BN is a directed graph, together with an associated set of probability tables. The graph consists of nodes and arcs as shown in Figure 1.

The nodes represent variables – some of which are discrete and non-numeric, such as the Boolean variable 'Suspect committed the crime' which has two states "True" and "False"; and some of which, like N as described above, are numeric and may be discrete or continuous. The arcs represent causal or influential relationships between variables, and so enable us to represent dependencies between different pieces of evidence. Associated with each node is a Node Probability Table (NPT). For a discrete node with discrete parents the NPT captures the relationship between the node and it parents by specifying the probability of each of its states given every combination of parent states. For a numeric node with parents the NPT is generally specified as a conditional probability distribution. For a discrete non-numeric node without parents, the NPT simply specifies the prior probability associated with each state. For a numeric node without parents the NPT is normally specified as a probability distribution.

Once a BN has been constructed we can enter observations (evidence) on any node and perform Bayesian inference to update the probability of each unobserved node – here we are, of course, especially interested in the updated probability of the node 'Suspect committed the crime'. This process (called Bayesian propagation) is complex for all but the smallest models but widely available BN tools (that implement standard propagation algorithms) enable us to easily build and run the computations automatically[13].

Figure 1 presents a generic BN model for incorporating the values n and N described above in such a way that – before any other evidence is presented – the prior probability of guilt is n/N as demanded of our method. Some intermediate nodes primarily used to enable us to transform continuous probability values into Boolean nodes according to the 'Binomial trick'[14] are hidden (this trick simply inserts a hidden integer node of two values {0,1} as a child of the continuous node c and defines its probability as a *Binomial*(1, c) distribution); dotted edges signify that there is at least one such hidden node on the path. The full model is available for download[15] and may be run in the freely available version of AgenRisk[16]. In addition to the nodes n and N we have nodes:

- "Suspect at *CS*" is a Boolean node for which the probability of True is equal to n/N
- "Suspect committed the crime" is a Boolean node for which the probability of True is equal to:
    - $1/n$ when "Suspect at *CS*" is True;
    - 0 when "Suspect at *CS*" is False;
- "Other committed the crime" is a Boolean node which is True when "Suspect committed the crime" is False, and False when "Suspect committed the crime" is True
- Various Evidence nodes (shaded) that are defined according to the evidence accuracy idiom[17] and whose NPTs encapsulate the Likelihood Ratio of the evidence, and are dependent on the particular type of evidence.

When we enter exact values for n (e.g. n = 10) and N (e.g. N = 100) and execute the model we get the expected prior probability values for both suspect at CS and suspect committed the crime (see Figure 2).

We can also enter uncertain evidence about n and N in the form of probability distributions as shows in Figure 3. In Figure 4 we have entered some evidence that the suspect was at the *CS*.

This could be, for example, forensic evidence found at the scene that matches the suspect or a credible eye witness - the NPT for the latter is defined in Table 1, which takes accounts of reasonable errors in such identifications.

Table 1 NPT for node "Evidence suspect at CS"

Suspect at CS	False	True
False	0.999	0.1
True	0.001	0.9

It should be noted that the BN propagation also provides us with the likelihood ratio for any Boolean hypothesis node H (such as 'Suspect at crime scene' or 'Suspect committed the crime') with respect to any evidence observed E (such as the eye witness evidence). This is because the BN provides us with the values of P(H) and P(*not* H) before the evidence is entered and the values of P(H | E) and P(*not* H | E) after the evidence is entered. By Bayes theorem the likelihood ratio is:

$$\frac{P(H|E)}{P(not\ H|E)} \times \frac{P(not\ H)}{P(H)}$$

In this example, reading off the relevant probabilities from Figures 3 and 4 we get a likelihood ratio of approximately 898 for the eye witness evidence with respect to the 'Suspect at crime scene' hypothesis, but a lower likelihood ratio of approximately 8.5 for the same evidence with respect to the 'suspect committed crime' hypothesis. Note that the BN calculations take account of all of the other dependencies and prior information in these computations.

Finally, suppose we also have evidence supporting the hypothesis that the suspect committed the crime (where the evidence has a likelihood ratio of 50 that is encoded into its NPT) then the posterior probability that the suspect committed the crime increases to nearly 98% as shown in Figure 5. However, any contrary evidence supporting the hypothesis that somebody else committed the crime would, of course, reduce this probability.

8 CONCLUSIONS AND RECOMMENDATIONS

We have presented a novel approach to modelling opportunity information in a legal context. It clarifies several misconceptions about the question of prior probability, thus avoiding some of the key objections to using Bayesian approaches to evaluate evidence. It also unifies good inferential practices, used during police investigations to identify the whereabouts of a suspect at the time of the crime, with the corresponding application of opportunity information in the evidence evaluation phase. This has implications both pre-trial and in the courtroom.

As it allows for a systematic treatment of opportunity information our approach should be of use pre-trial to help investigators and prosecutors assess the evidential case against a suspect and thus inform subsequent decisions about whether there is sufficient evidence to prosecute. It should also be relevant to how prosecution or defence teams formulate their arguments, allowing them to incorporate opportunity evidence in a principled manner. We also believe that it has relevance for how evidence is presented in court. One common criticism levelled against Bayesian approaches is that prior probabilities are solely the province of the trier of fact (which in UK and USA might be a jury of laypeople). A critical problem here is that if the jury is presented with statistical evidence (such as DNA evidence) they are left with the difficult task of combining this quantitative information with their prior beliefs. By showing the jury how to factor in opportunity evidence along with other key evidence in the case our proposed approach could help alleviate some of these difficulties.

Another common objection to the use of prior probabilities is that it seems to conflict with the legal presumption of innocence. Our approach addresses this objection since our account interprets the presumption of innocence to say that the defendant should be treated no differently from any other person who also had the opportunity to commit the crime. In other words the defendant is as probable to be the perpetrator as anyone else with the same opportunity, absent other evidence in the case.

Finally, we acknowledge that our proposal only applies to cases where it is known that a crime has been committed, but there is uncertainty as to the identity of the perpetrator. But in some cases we are uncertain as to whether a crime has been committed at all, for example in cases where a mother is accused of killing her baby, or someone is charged with dangerous driving. In such cases identity is not an issue, and so the prior must be calculated differently. A solution for this problem is a question for future research.

ACKNOWLEDGMENTS

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Probability and Statistics in Forensic Science (FOS) where some of the work on this report was undertaken. This work was supported by EPSRC grant no EP/K032208/1 and the Simons Foundation. This work has also been supported by the European Research Council (ERC) through project, ERC-2013-AdG339182-BAYES_KNOWLEDGE.

REFERENCES

[1] Gardner, T. J. & Anderson, T. M. *Criminal Evidence: Principles and Cases*. (Cengage, 2014).

[2] Balding, D. J. & Donnelly, P. Inference in forensic identification. *J. R. Stat. Soc. Ser. A* **158,** 21–53 (1995).

[3] Dawid, A. P. & Mortera, J. Coherent Analysis of Forensic Identification Evidence. *J. R. Stat. Soc. Ser. B* **58,** 425–443 (1996).

[4] Lindley, D. V. The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence andExpert Systems. *Stat. Sci.* **2,** 17–24 (1987).

[5] Dawid, A. P. in *Bayes's Theorem: Proceedings of the British Academy* **113,** 71–90 (Oxford University Press, 2002).

[6] Fenton, N., Neil, M. & Berger, D. Bayes and the Law. *Annu. Rev. Stat. Its Appl.* **3,** (2016).

[7] Lindley, D. V. Probability and the Law. *Stat.* **26,** 203 (1977).

[8] Friedman, R. A Presumption of Innocence, Not of Even Odds. *Stanford Law Rev.* **52,** 873–887 (2000).

[9] The Hague Appellate Court, 18th February 2015, ECLI:NL:GHDHA:2015:282.

[10] Dawid, A. P., Mortera, J. & Vicard, P. Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Sci. Int.* **169,** 195–205 (2007).

[11] Fenton, N. E., Lagnado, D. A. & Neil, M. A General Structure for Legal Arguments Using Bayesian Networks. *Cogn. Sci.* **37,** (2013).

[12] Taroni, F., Aitken, C., Garbolino, P. & Biedermann, A. *Bayesian Networks and Probabilistic Inference in Forensic Science*. (John Wiley & Sons, 2014).

[13] Fenton, N. E. & Neil, M. *Risk Assessment and Decision Analysis with Bayesian Networks*. (CRC Press, 2012).

[14] Fenton, N. E. Transforming a continuous probability value into a Boolean. (2015). www.youtube.com/watch?v=5Yp9OsnK6Wk.

[15] Fenton, N. E. Bayesian network models for download. (2017). Available at: http://www.eecs.qmul.ac.uk/~norman/Models/.

[16] Agena Ltd. AgenaRisk. (2017).

[17] Lagnado, D. A., Fenton, N. E. & Neil, M. Legal idioms: a framework for evidential reasoning. *Argument Comput.* **4,** 46–63 (2013).

FIGURES

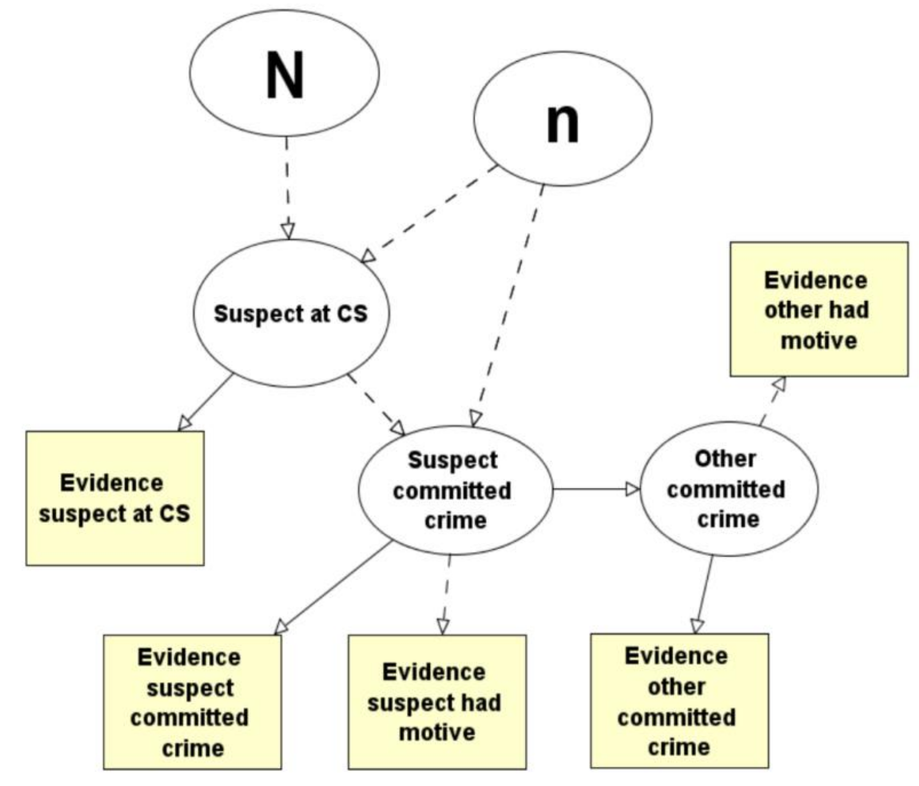

Figure 1 Generic BN model structure

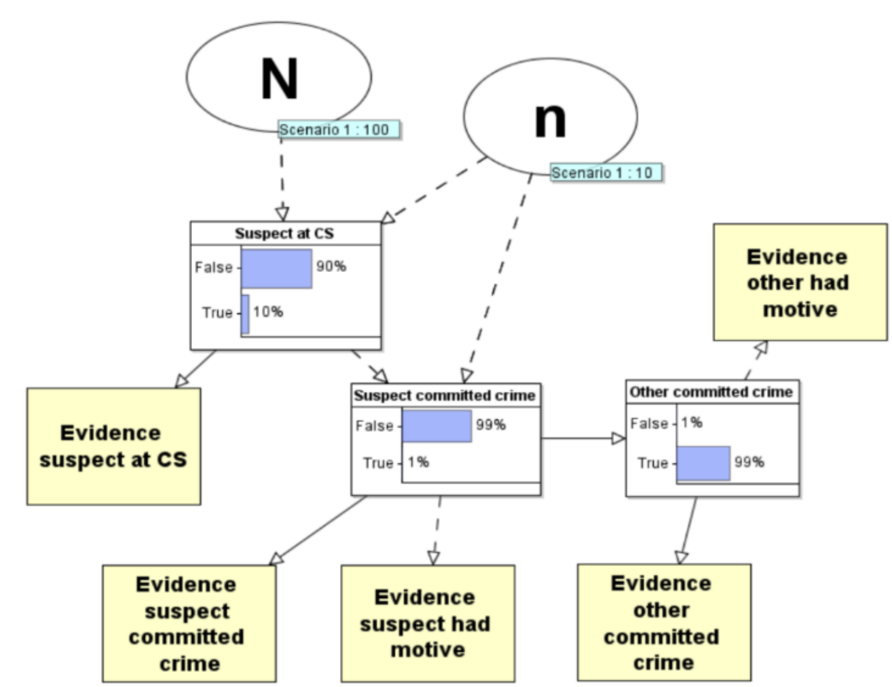

Figure 2 State of model when n = 10 and N = 100

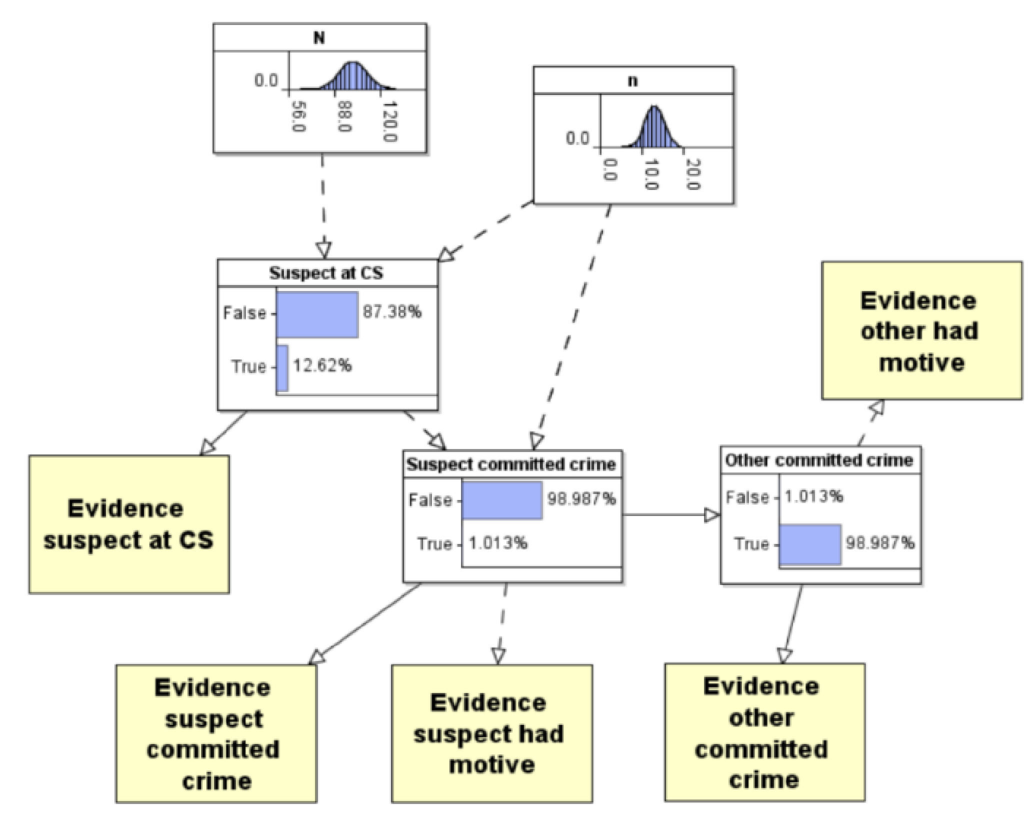

Figure 3 State of model when *n* and *N* are distributions rather than point values

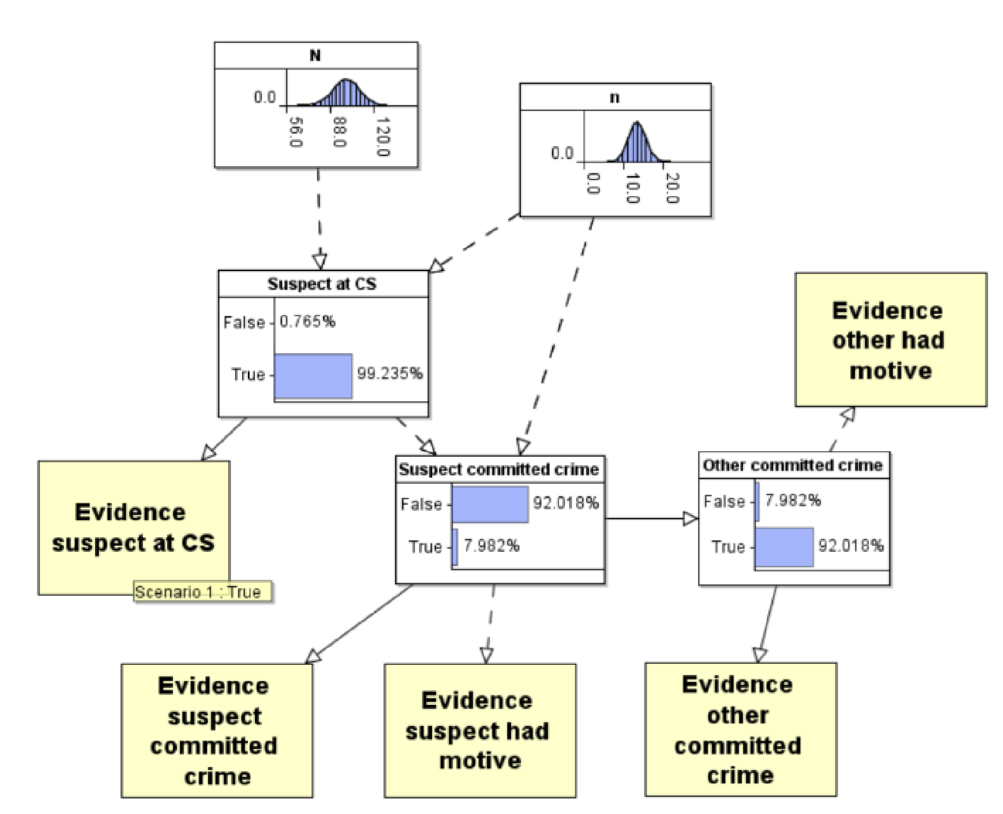

Figure 4 State of model after some evidence is entered

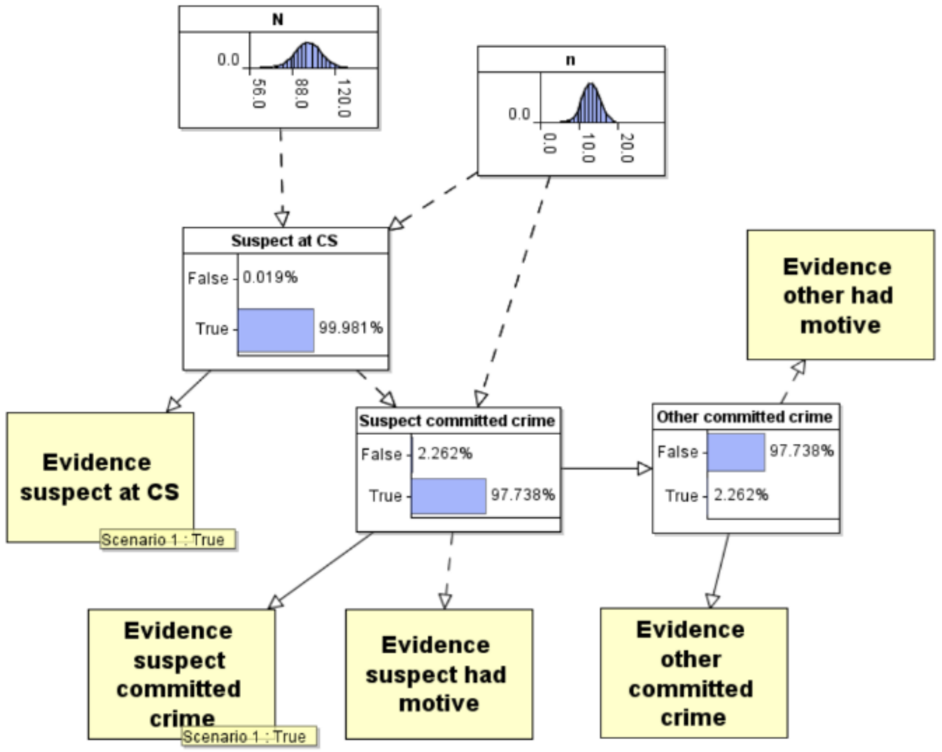

Figure 5 Evidence suspect committed crime (with LR = 50) entered