View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Queen Mary Research Online

The Opportunity Prior:
A proof-based prior for criminal 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, Department of Experimental Psychology, 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 challenges to the use of probabilistic reasoning in the assessment of criminal evidence is the “problem of the prior”, i.e. the difficulty in establishing an acceptable prior probability of guilt. Even strong supporters of a Bayesian approach have often preferred to ignore priors and focus on the likelihood ratio (LR) of the evidence. But to calculate if the probability of guilt given the evidence reaches the probability required for conviction (the standard of proof), the LR has to be combined with a prior. In this paper, we propose a solution to the “problem of the prior”: the defendant shall be treated as a member of the set of “possible perpetrators” defined as the people who had the same or better opportunity as the defendant to commit the crime. For this purpose, we introduce the concept of an “extended crime scene” (ECS). The number of people who had the same or better opportunity as the defendant is the number of people who were just as close or closer to the crime scene, in time and space. We demonstrate how the opportunity prior is incorporated into a generic Bayesian network model that allows us to integrate other evidence about the case.

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 they were at the time of the crime. This is potentially a very diagnostic question: if the suspect can show that they were elsewhere, then they cannot (except in special cases that we ignore in this paper) have committed the crime. If, however, it can be shown that they were at the crime scene at the time of the crime, then they are ruled into a (relatively) small subset of possible perpetrators. In classic investigative terms this establishes ***opportunity*** which 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 we must evaluate the strength of evidence against them. Information about the suspect's whereabouts in relation to the crime scene is crucial – often as a starting point for establishing a case 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, frequently 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 neglected in formal analyses? We argue that this neglect hinges on several flawed assumptions, including: (i) the belief that at best opportunity evidence 'fails to exclude' a suspect, but does not have a positive confirmatory value on the hypothesis that the suspect committed the crime; and (ii) that placing someone at the crime scene (close to time of crime) at most means they are 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 divided by some large population size.

We will argue that both assumptions are wrong. We present a principled approach to quantifying information about opportunity evidence that corrects these misconceptions. This approach maps onto standard investigative practices (Gardner & Anderson 2014) and shows how opportunity evidence is often a key factor in determining a suspect's guilt or innocence. 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 they are accused based ***only*** on evidence about their proximity (in space and time) to the crime scene location and time. We are not assuming that the defendant is a suspect based only on having been 'present' at the (extended) crime scene. In most legal jurisdictions this would never be considered even a reason for arrest. In fact, the key point is that our approach is most relevant to cases where 'being present at the crime scene' is ***not*** a major component of the evidence. On the contrary, our approach simply provides a realistic bound on the prior probability of guilty ***before*** we consider ***any of the evidence*** that is independent of simple location.

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 theorem 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 present two real case study examples to demonstrate the computation of the opportunity prior. In Section 7 we address whether the opportunity prior is justified as a general rule for determining the prior probability in criminal cases, and discuss some problems with such a rule. Finally, in Section 8 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 ”PROBLEM OF THE PRIOR”

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 (Balding & Donnelly 1995; Dawid & Mortera 1996; Lindley 1987). 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 (Dawid 2002; Fenton et al. 2016; Lindley 1977; Friedman 2000), 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, the prior probability of guilt is decisive for whether the evidence is sufficient to convict.

However, one problem in real criminal cases, is that they rarely take place on islands completely isolated from the mainland. This makes it difficult to determine the number of possible perpetrators (*n*). Consider a murder in central Blackpool. Should *n* be the population of Blackpool (approx. 150 000), the population of Lancashire (approx. 1.5 million), the population of the UK (approx. 65 million) or the entire population of the world (approx. 7.6 billion)? The “problem of the prior” has been discussed by evidence scholars, and different solutions have been proposed, but each of these proposals are problematic for one reason or other. For example, Dennis Lindley has proposed that for a crime committed in the UK, *n* = the population of the UK (Lindley 1987). The obvious objection to this delimitation is that it is completely arbitrary. Why include Scots but exclude Belgians? A non-arbitrary solution is, of course, to let *n* = the entire world population. An obvious problem with such a solution is that it would set the prior so low that it would be extremely difficult to get people convicted. In many cases, not even a DNA match would be sufficient. There would be too many wrongful acquittals. In (Walsh et al. 1994) the authors propose that the ‘prior odds considering geography’ should be the following (where L is the ‘location’ where the crime C took place)

$$\frac{P(\text{suspect would be resident of L} \mid C)}{P(\text{suspect would be resident of L})}$$

where the numerator is derived from survey information about crimes of type C in L and the denominator is the population size of L divided by the world population. However, the practicality of this solution crucially depends on the availability of statistically sufficient and accurate data on the residence of perpetrators in previous crimes of the same kind committed in the same location. Needless to say, such data is rarely available. In (Wixted et al. 2019) and the associated three commentaries there is much discussion about the prior, but this focuses on the special case where DNA from the crime scene is found to match the DNA of a suspect from a DNA database search. (Wixted et al. 2019) consider estimates of the 'active criminal population' rather than location information.

(Dahlman 2018) has proposed a "normative prior", i.e. a legal rule that stipulates a number for *n* to be applied in every case no matter where the crime was committed (e.g. *n* = 1000). An objection to the "normative prior" is that it is not "proof-based". According to this objection, the prior should be based on the proven circumstances of the case, which the "normative prior" is clearly not.

The solution presented in this paper – the "opportunity prior" – is proof-based since it is based on the ***known*** whereabouts of the suspect. As we shall see, the "opportunity prior" offers a solution based on the idea of an "extended crime scene" (ECS) – which enables us to take account of the fact that the exact crime scene may not be known.

We provide a formalisation of the island problem whereby we seek to narrow down the crime scene and time as far as possible. 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 proven that 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 to 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 numbers *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 (and indeed we deal with this extension in Section 8), but we have used point values for ease of explanation, without any loss of generality.

3 DEFINING THE CRIME SCENE and TIME (CST)

In what follows we assume a crime has taken place and that it was committed by one person – typically against one other person (e.g. murder, assault, robbery). We define the ***crime scene and time*** (which we will write as ***CST***) as comprising the two components

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

2. ***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.

Example 3.1: In the following examples the CS (and hence CT) is reasonably specific and any disagreements and uncertainty about it are tightly bounded

a) A person was attacked in a 20-metre alleyway while walking home from work. It is not certain which specific point in the alleyway, so the *CS* is the entire alleyway, while, based on work hours the CT is known to be between 18.15 and 18.20.
b) A person was mugged while standing next to a particular lamppost by Piccadilly Circus tube station, so the *CS* would be an area about one metre around the lamppost.
c) A person was shot in a theatre. The *CS* is the area of the theatre covering any point from which a shot could have been fired.

The above examples are typical of scenarios in which our proposed method is most easily applied. However, in some scenarios there may be fundamental disagreement about the *CS and CT*.

Example 3.2: Suppose the police claim the crime (a murder) took place in an alleyway like that in Example 3.1(a) between 01.00 and 01.30, but the defence shows that a) the body could have been dumped there and b) 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 up to 30 miles away before being returned to the alleyway.

Next we define ***the number of people, n, at CS during CT (or simply the number of people, n, in CST)***: 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 18.00 and 18.15 (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 in CST. In the absence of any other evidence each such a 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 must emphasize that this is the prior probability of guilt before any other evidence is considered. For example, it explicitly ignores evidence such as whether the suspect knew the victim or was at the crime scene helping the victim etc.

We shall deal with the case where it is not proven that the defendant was in CST in the next section, but it is important first to clear up a very important and common misunderstanding that applies to all 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 must have 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. However, suppose, 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 in CST the prior probability of guilt 1/*n* is relatively high.

Example 3.4: In a Dutch murder case in Simonshaven in 2011 (Fenton et al. 2019), 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 the victim (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 in CST. The question is: how many other people were in CST. 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 (ECST) WHEN IT HAS NOT BEEN PROVEN THAT THE DEFENDANT WAS AT THE CRIME SCENE DURING THE CRIME TIME

Typically, the defendant will dispute having been present at *CS* during *CT*. Our task is hence to determine the *extended crime scene and time* (***ECST***) based on the suspect's 'closest' proven time and location to *CS* and *CT.* Recall that CT is defined as the interval (t, t'). 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 t_2. 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. So *t*=21.00 and *t'*=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 21:00 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 03:00 and 06:30.

It is unlikely any other known location would lead to a smaller number than N_1 or 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.

The numbers *n* (and *N* where necessary) can be reduced significantly in cases where the crime took place in an open area (as opposed to inside a closed space such as a house or building). This is because, if it is accepted that the defendant was in an open area during ECST, then we can restrict *n* (or *N*) to an estimate of the total number of people who were in an open area during ECST.

Example 5.2: In the case of Example 3.1(b) suppose the mugging took place in a public outside square late at night and the nearest confirmed location for the defendant is "walking towards a pub 100 metres away 5 minutes later". Although the extended CS includes several pubs and restaurants where hundreds of people were ***inside*** we can restrict *N* to an estimate of the number

of people who were outside of any building within the ECST (a much smaller number in this case) since the defendant is one of these and the offender must be included in that number (whoever the real offender is he/she must have been outside at the time of the offence even if they also moved inside during that period). However, suppose that, instead of the nearest confirmed location for the defendant being "walking towards a pub 100 metres away 5 minutes later", it is "inside a pub 100 metres away 5 minutes later". Then we can no longer assume the defendant was one of the (relatively small number of) people who were outside any building within the ECST. However, we do know that the offender is certainly one of the people who were either outside or inside a building within the ECST and the defendant is one of those. But *N* therefore has to include an estimate of all the people inside buildings in ECST plus those outside.

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

Example 5.3: 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. Specifically, while the 1/N prior is a valid upper bound on guilt it is unfairly high to the defendant. In such cases we might need to consider a distance-weighted computation when calculating priors.

6. REAL CASE EXAMPLES

We next present two real case examples which provide clear illustrations of the Opportunity Prior estimation. The first is a very well known and high profile murder, while the second is a case in which the first author was involved as an expert witness.

Case 1 The murder of Olof Palme

On Friday 28 February 1986, Swedish Prime Minister Olof Palme was assassinated in central Stockholm. Palme was walking home from the cinema with his wife, unaccompanied by bodyguards. The perpetrator came up from behind and fired a single fatal gunshot from close range into Palme's back. The perpetrator fired a second shot at Mrs Palme that just missed the target, and then fled the scene. Three years later, Christer Pettersson, a small time criminal, was identified in a police line-up by Mrs. Palme, and prosecuted for the murder.

In this case there is no dispute about the actual crime scene location and time: we know that the murder took place at 23:20 on the corner of Sveavägen and Tunnelgatan. Although the defendant did not admit to having been at the crime scene we DO have a very specific and well defined ***extended crime scene***, because there is a specific location and time close to the crime scene where the defendant was present. Specifically, we know that Pettersson left a gambling hall called "The Ox" at 22:30 on the night of the murder, approximately 250 meters from the crime scene and 50 minutes earlier. He claims that he went directly home on the subway. This sets a time-frame and a geographical perimeter for the opportunity prior. The number of possible perpetrators, that had the same opportunity as Pettersson to commit the crime, is the number of people who moved around within 250 meters of the crime scene within 50 minutes of the crime (note that, as explained in Example 5.2, we do not need to include the potentially large number of people who were ***inside*** buildings within the ECS). Considering the location in central Stockholm, but also considering that it was a very cold February night, the number can be estimated around 500. This sets the prior probability at 0.2%. It is important to note that the suspect's criminal record and the fact that he was identified in a police line-up are not considered at all in this prior – these are pieces of evidence that could be used subsequently.

Case 2 Croydon riots

Between 6-10 August 2011, England was struck by a series of extremely violent riots, with one of the worst occurring in Croydon on 8 August. Many shops were attacked and looted and some were burnt down. The first author was involved as an expert witness in a case of one defendant accused of causing damage to, and looting from, one of the shops - a sports shop - at 22.50. The main evidence against the defendant was a low-resolution image taken at the scene of the crime which the prosecution claimed was the defendant. The image did NOT show the offender's face and it was not even clear if the offender was a man or woman. But it did show clearly that the offender was wearing a particular type of hooded tracksuit top. The defendant denied being at the scene, but accepted that he was wearing that type of tracksuit top in Croydon that evening.

What is especially relevant about the case for our paper is that, while the defendant denies being at the crime scene we DO have a very specific and well defined extended crime scene. There was a clear CCTV image of the defendant (wearing the particular type of tracksuit top) 43 minutes earlier (at 22:07) at a location 200 metres from the sports shop. What is also pertinent in this case is that, from witnesses and CCTV images, we have a good estimate of the number of people who physically were in that area (comprising a circle of radius 200 metres centred on the sports shop) between 22:07 and 22:50 (again note that, as explained in Example 5.2, we do not need to include the potentially large number of people who were ***inside*** buildings within the ECS). The number was estimated at 300. By definition the defendant was one of these 300 people and so, ignoring all other evidence, the prior probability of him being the offender in the sports shop was 0.333%, i.e. a prior probability of 99.666% innocence.

Our own specialist skills were used to refine this prior probability by taking account of the likely number of other people among the 300 (excluding a number who had been eliminated as

possible offenders) who would have been wearing the same type of tracksuit top. The particular track suit top was actually very popular (we used data on actual sales from different sports shops in the area around Croydon) and many of the 300 in the extended crime scene were known to have been wearing hooded tracksuit tops. We produced (ranges) of revised prior probabilities of the defendant's guilt based on different assumptions. For example, the 'most likely' proportion of youths in the area wearing that particular tracksuit top was estimated at 0.45%. Using the Binomial Theorem and assuming 200 of 300 people in the extended crime scene had not been eliminated as suspects, this results in a 59% probability that at least one person other than the defendant was wearing the same tracksuit top and hence a revised prior probability of 59% that the defendant is innocent.

7 POSSIBLE PROBLEMS WITH THE OPPORTUNITY PRIOR

In this article, the opportunity prior is proposed as a general rule for determining the prior probability of guilt for all criminal cases where it is clear that the offense was committed by a single perpetrator and there is evidence of the defendant's whereabouts. While the method is mathematically correct and practically feasible, that does not necessarily mean it is legally acceptable. In this section we will discuss some possible legal and other objections.

As we have seen, the opportunity prior defines the number of possible perpetrators (N) as the number of people who had just as much of an opportunity as the defendant to commit the crime, i.e. the number of people who were as close or closer to the crime scene as the defendant within the established time frame. A person accused of a crime could object that determining the prior probability of his guilt in such a way is legally forbidden, since it violates the presumption of innocence. According to this objection, the presumption of innocence requires that everyone who could have committed the crime should be counted as a possible perpetrator, and N should not be limited to people who had just as much of an opportunity as the defendant. This interpretation of the presumption of innocence is hard to sustain, since it implies that a fact-finder is legally obligated to disregard proven facts about the defendant's whereabouts. Furthermore, it leads us back to a way of determining the prior where N = the entire world population (minus some people who could not possibly have committed the crime, e.g. infants). As we have explained in section 2 this approach is flawed as itleads to priors that are artificially low, and would produce too many wrongful acquittals.

There are also potential legal objections in cases where N is very low and hence the opportunity prior is very high. In the most extreme scenario we could even have N=1: for example, suppose video evidence shows the victim and Fred entering a room known to be empty and one minute later Fred leaving the room where the victim's body is found with a knife in the back. Using the opportunity prior, the probability of Fred's guilt is 1 and so, logically, no further evidence should be needed to ensure conviction. However, lawyers would certainly be uncomfortable about a conviction without some direct evidence (e.g. forensic evidence) linking Fred to the murder.

It is not just such an extreme case where there may be legal concerns. As we have seen with the murder of Olof Palme and the Croydon riots (section 6), the opportunity prior produces a neat prior in cases where the crime took place in a densely populated area and the defendant can be placed a couple of hundred meters from the crime scene at around an hour before the crime, but the

opportunity prior produces a high prior in cases where the crime took place a in a low-populated area and/or the defendant can be placed very close to the crime scene. In such cases, the opportunity prior will produce a high prior probability of guilt, and it may therefore take little additional evidence (a modest likelihood ratio) to reach the standard of proof, although the defendant might just be an innocent bystander. Using the opportunity prior in such cases may, therefore, produce too many wrongful convictions. This objection to the opportunity prior has been explored by Dahlman.[12] Consider the following case.

Example 7.1: 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 proven 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%.

Suppose now that, in Example 7.1, a shoe mark is found that is determined to be from a person of size 12 – an attribute shared by 1 in 19 people - and that X is size 12. If the standard of proof requires a posterior probability of 95%, then the matching shoe size evidence is sufficient for conviction. This is because the likelihood ratio (LR) of the evidence is 19 and so, with prior odds of guilt being even (50%), the posterior odds of guilt are 19 to 1, i.e. 95%.

This is problematic since there is generally a tendency to search for confirmatory evidence once there is a suspect, and coincidental evidence with this kind of likelihood ratio is not difficult to find. In such situations there is a clear danger of wrongly convicting a person who is a defendant just because they happened to be nearby.

One way to handle this objection to the opportunity prior 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. Also. it should be emphasized that, although X is one of only two people who could have been at the CS in Example 7.1, the fact that there WAS another unknown person there - who could have been any other person in the world – means that ideally X should not be considered as a suspect unless there is other actual evidence linking him to the crime. To avoid confirmation bias in such cases it is necessary to consider evidence that is probative on innocence as well as guilt. For example, because of confirmation bias it might be easy to find evidence like "defendant had a dispute with victim" if the defendant was a neighbour. However, if that argument was 10 years ago, then the fact that as neighbors they had no argument for 10 years is actually probative in favour of innocence. Similarly, if no incriminating forensic traces from the defendant were found then that is probative in favour of innocence and should easily counterbalance (from an LR perspective) any circumstantial evidence used as a result of confirmation bias.

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

While we have so far assumed that reasonable estimates for *n* and *N* can be agreed, in this section we show how uncertainties with respect to the assignments of *n* and *N* can be handled in a principled and rigorous way. We also illustrate what can be done with the opportunity prior, when no other input to the prior is available.

In general, in order to combine a prior probability of guilt with potentially complex and related evidence about the case to compute a rational posterior probability of guilt, it has been widely acknowledged that a Bayesian Network (BN) model is an ideal formalism (Dawid et al. 2007; Fenton et al. 2013; Taroni et al. 2014). 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.

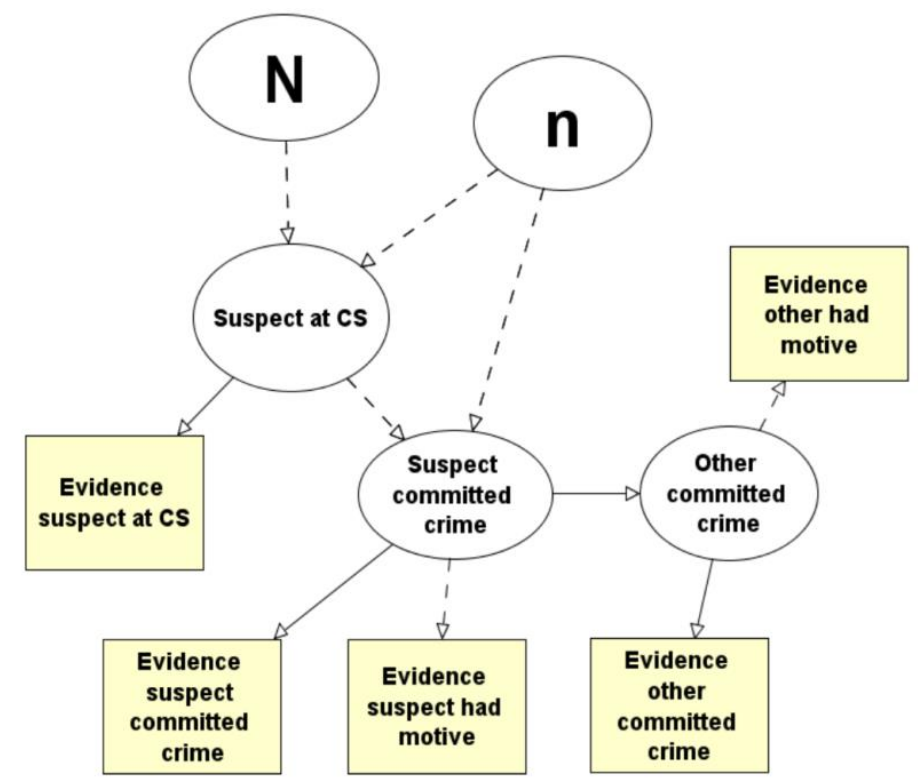

Figure 1 Generic BN model structure

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 its 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 (Fenton & Neil 2018).

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' (Fenton 2015) 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 (Fenton 2019) and may be run in the trial version of AgenaRisk (Agena Ltd 2019). 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(Lagnado et al. 2013) 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).

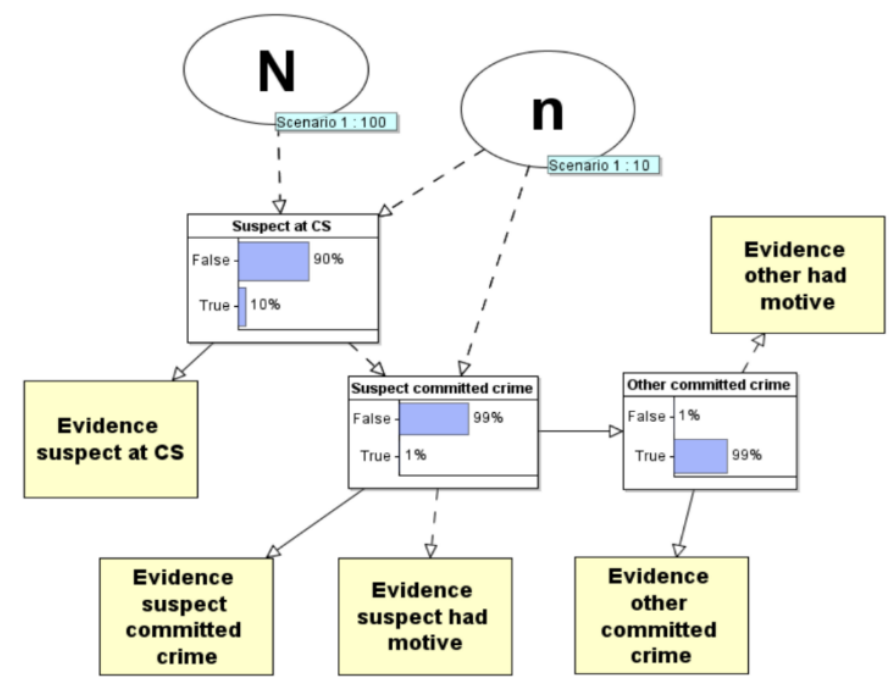

Figure 2 State of model when n = 10 and N = 100

Crucially, we can also enter uncertain evidence about *n* and *N* in the form of probability distributions as shows in Figure 3.

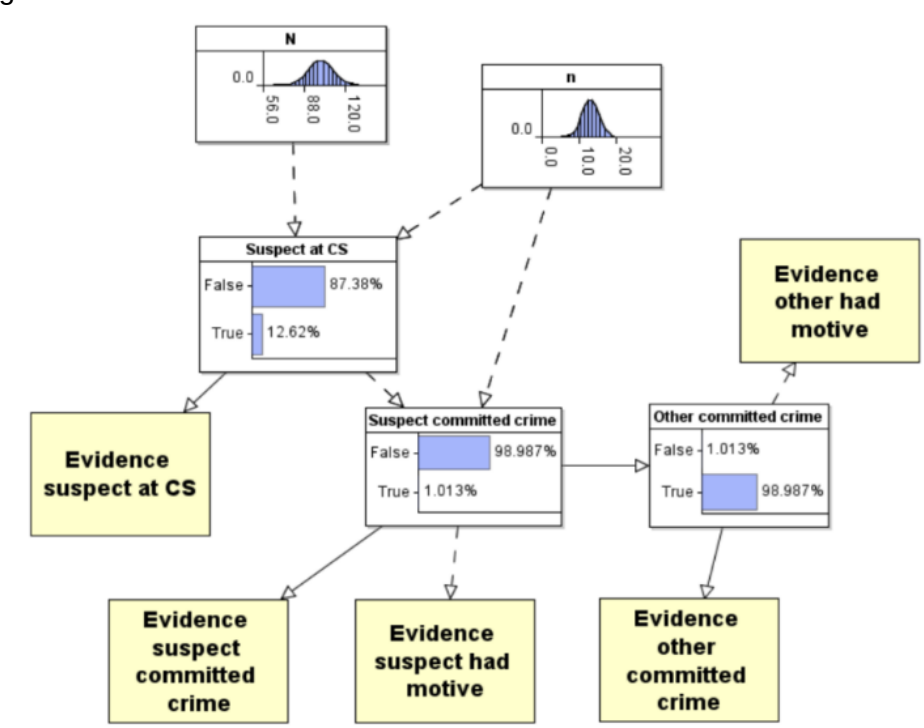

Figure 3 State of model when *n* and *N* are distributions rather than point values

In Figure 4 we have entered some evidence that the suspect was at the *CS*.

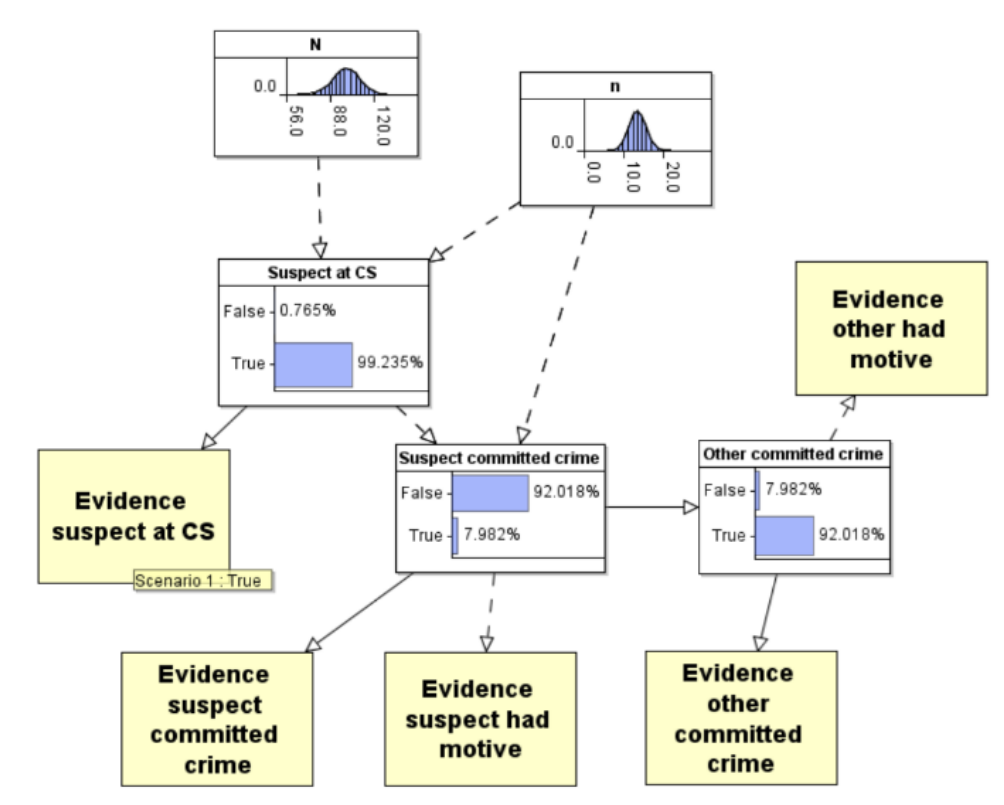

Figure 4 State of model after some evidence is entered

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.

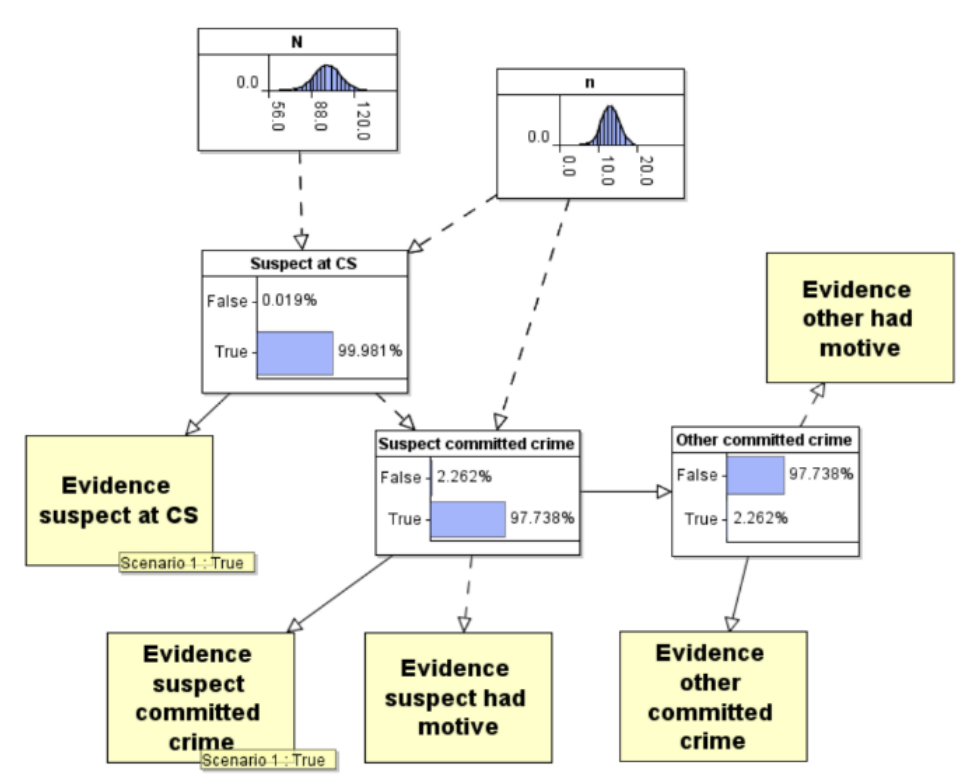

Figure 5 Evidence suspect committed crime (with LR = 50) entered

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. We also propose that it should be used by legal fact-finders (judges and jurors) to determine the prior probability of guilt. The opportunity prior provides the fact-finder with a proof-based method that is easy to use.

A 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 had the same opportunity or a better opportunity to commit the crime. In other words the defendant is as probable to be the perpetrator as anyone else who was equally close or closer to the crime scene, in time and space.

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. The work has also been improved as a result of the comments of the anonymous referees and Editor.

REFERENCES

Agena Ltd, 2019. AgenaRisk. Available at: http://www.agenarisk.com.

Balding, D.J. & Donnelly, P., 1995. Inference in forensic identification. *Journal of the Royal Statistical Society, Series A*, 158, pp.21–53.

Dahlman, C., 2018. Determining the Base Rate for Guilt. *Law, Probability and Risk*, 17(1), pp.15–28.

Dawid, A.P., 2002. Bayes's theorem and weighing evidence by juries. In . In *Bayes's Theorem: Proceedings of the British Academy*. Oxford: Oxford University Press, pp. 71–90.

Dawid, A.P. & Mortera, J., 1996. Coherent Analysis of Forensic Identification Evidence. *Journal of the Royal Statistical Society. Series B*, 58(2), pp.425–443.

Dawid, A.P., Mortera, J. & Vicard, P., 2007. Object-oriented Bayesian networks for complex forensic DNA

profiling problems. *Forensic Science International*, 169, pp.195–205.

Fenton, N.E. et al., 2019. Analyzing the Simonshaven Case using Bayesian Networks. *Topics in Cognitive Science*, to appear.

Fenton, N.E., 2019. Bayesian network models for download. Available at: http://www.eecs.qmul.ac.uk/~norman/Models/.

Fenton, N.E., 2015. Transforming a continuous probability value into a Boolean. *https://www.youtube.com/watch?v=5Yp9OsnK6Wk*. Available at: https://www.youtube.com/watch?v=5Yp9OsnK6Wk.

Fenton, N.E., Lagnado, D.A. & Neil, M., 2013. A general structure for legal arguments using Bayesian networks. *Cognitive Science*, 37(61–102).

Fenton, N.E. & Neil, M., 2018. *Risk Assessment and Decision Analysis with Bayesian Networks* 2nd ed., CRC Press, Boca Raton.

Fenton, N.E., Neil, M. & Berger, D., 2016. Bayes and the law. *Annual Review of Statistics and Its Application*, 3(1), pp.51–77. Available at: http://www.annualreviews.org/doi/pdf/10.1146/annurev-statistics-041715-033428 [Accessed April 4, 2016].

Friedman, R., 2000. A Presumption of Innocence, Not of Even Odds. *Stanford Law Review*, 52(4), pp.873–887. Available at: http://repository.law.umich.edu/articles/161 [Accessed January 26, 2017].

Gardner, T.J. & Anderson, T.M., 2014. *Criminal Evidence: Principles and Cases* 9th ed., Boston MA, USA: Cengage.

Lagnado, D.A., Fenton, N.E. & Neil, M., 2013. Legal idioms: a framework for evidential reasoning. *Argument and Computation*, 4(1), pp.46–63.

Lindley, D. V., 1977. Probability and the Law. *The Statistician*, 26(3), p.203. Available at: http://www.jstor.org/stable/2987898?origin=crossref [Accessed January 26, 2017].

Lindley, D. V., 1987. The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence andExpert Systems. *Statistical Science*, 2(1), pp.17–24.

Taroni, F. et al., 2014. *Bayesian Networks and Probabilistic Inference in Forensic Science* 2nd ed., Chichester, UK: John Wiley & Sons.

Walsh, K.A.J., Buckleton, J.S. & Trigg, C.M., 1994. Assessing prior probabilities considering geography. *Journal of the Forensic Science Society*, 34, pp.47–51.

Wixted, J.T., Christenfield, N.J.S. & Rouder, J.N., 2019. Calculating the posterior odds from a single-match DNA database search. *Law, Probability and Risk*, 18(1–23).