The "Jury Observation Fallacy" and the use of Bayesian Networks to present Probabilistic Legal Arguments

Norman Fenton and Martin Neil

Computer Science Department
Faculty of Informatics and Mathematical Sciences
Queen Mary and Westfield College, London E1 4NS.

and

Agena Ltd
11 Main Street
Caldecote
Cambridge, CB3 7NU

email: norman@agena.co.uk

Version 5.0, 27 March 2000

Abstract

Probability theory, especially Bayesian probability, is widely misunderstood by the general public. Lawyers are no different from ordinary members of the public in falling victim to arguments that have been known to mathematicians for decades to be fallacies. The so-called prosecutor's fallacy and the defendant's fallacy are two well-known examples that arise from a basic misunderstanding of conditional probability and Bayes' Theorem. In this paper we introduce what we believe is a previously unreported fallacy, which we refer to as the *jury observation fallacy*. In this fallacy there is a basic misunderstanding about the belief in probability of guilt when a prior similar conviction by a defendant is revealed *after* the jury returns a not guilty verdict. Specifically, it is widely believed that the information about the prior conviction might suggest to external observers that the jury verdict was wrong (the belief is that probability of guilt increases). In fact, using very reasonable (and indeed conservative) assumptions it can be shown, using Bayesian reasoning, that such a response is irrational in many situations. To explain the Bayesian argument without exposing readers to any of the mathematical details we use Bayesian Networks (BNs) and a tool (Hugin) to execute them. Hence, a secondary objective of this paper is to show that there *is* a way of making all of the implications of Bayesian reasoning clear to lay people, without them having to understand any of the underlying mathematics. The implications of this in the legal profession are profound. Courts could eventually accept Bayesian arguments just as they accept forensic evidence without having to resort to explanations from first principles. Additionally, the results presented suggest that there may be reason for disquiet about the use of previous convictions as a basis for selecting suspects as is common Police practice.

1. Introduction

Most people, including recognised experts in all walks of life, make basic mistakes when reasoning about probability. This has been confirmed by numerous authors (see, for example, [Ayton 1994, Ayton and Pascoe 1995, Kahneman et al 1982]). The book [Kahneman et al 1982] provides a comprehensive listing of the most common fallacies arising from a basic misunderstanding of the rules of probability theory. Many of these fallacies arise, in particular, from misunderstandings about conditional probability. Examples of such fallacies are the so-called *prosecutor's fallacy* and the *defendant's fallacy* summarised by [Aitken 1996] as:

> "Suppose a crime has been committed. Blood is found at the scene for which there is no innocent explanation. It is of a type which is present in 1% of the population."
>
> The Prosecutor's fallacy is the assertion: "There is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus, there is a 99% chance that he is guilty."
>
> The Defendant's fallacy is the assertion: "This crime occurred in a city of 800,000 people. This blood type would be found in approximately 8,000 people. The evidence has provided a probability of 1 in 8,000 that the defendant is guilty and thus has no relevance."

The prosecutor's fallacy is to assume that $P(A|B)$ (the conditional probability of A given B) is the same as $P(B|A)$ where A represents the event "Defendant innocent" and B represents the event "Defendant has the matching blood type". The defendant's fallacy is to ignore the large change in the odds in favour of the defendant's guilt. [Aitken 1996] provides a mathematical explanation of both fallacies.

A less well known fallacy is the one described by Matthews in [Matthews 1995]. In this "interrogator's fallacy" Matthews provides a Bayesian explanation of why confessions secured under interrogation offer dubious evidence to support guilt.

The mathematical arguments used by Aitken and Matthews to explain the fallacies are straightforward for mathematicians and statisticians. However, when we have shown them to highly intelligent and experienced lawyers they remain flummoxed. It appears that no amount of careful explanation can make a non-mathematician understand arguments about conditional probability. It is therefore little surprise that mistakes about probability continue to be made by members of the public, including experts such as forensic scientists and lawyers. Aitken himself highlights a recent trial involving DNA evidence such that, if the defendant were not the source of the crime sample, then the probability of a match was 1 in 3 million. However, the forensic scientist mistakenly stated that:

> "The likelihood of this [the source of the sample] being any other man but [the defendant] is 1 in 3 million".

This and related cases have been discussed extensively in a number of papers (see, for example [Balding and Donnelly 1994], [Balding and Donnelly 1996], [Robertson and Vignaux 1995, 1998b]). In successive appeals against the original conviction the defence used an expert witness, Professor Donnelly, to explain to the jurors how to combine evidence using Bayes' theorem. Specifically, he explained how to combine the DNA evidence with other evidence (such as the failure of the victim to pick out the accused on an ID parade) and also how to combine each of the other items of evidence with each other. Applying Bayes' Theorem in this way reduces the probability of guilt significantly. Although all three judgements of the case assume that Bayes' Theorem is a technical matter appropriate for expert evidence, the Court noted that:

> "To introduce Bayes' Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task."

The Court of Appeal essentially rejected the admissibility of the use of Bayes' Theorem and the appeal was rejected. The reaction of the Court of Appeal is both alarming, yet also understandable. It is alarming because (as [Robertson and Vignaux 1998] argue superbly) it rejects the only logical approach to formalising and reasoning about uncertainty. It therefore implies that rigorous scientific proof should be rejected simply on the basis that lay people could not understand the science. Yet it is understandable because the defence's Bayesian arguments were presented from "first principles". There is a rich irony here. It was a very complicated scientific theory (DNA) which led to the original conviction; the prosecution would not have dreamed of trying to explain the underlying theories of DNA from first principles. The scientific community accepts the theories of DNA and forensic science and hence it is sufficient for an expert witness to state that the DNA evidence is valid and to explain the impact of the evidence. Whereas DNA and forensic science is relatively new, Bayesian probability is over 200 years old. To date, the statistical community has not mustered either the political or technical know-how to ensure that Bayesian arguments are as widely accepted (without having to be understood) within the legal profession as, say, DNA evidence. We hope that the excellent recent work of people such as Aitken, Balding and Donnelly, Robertson and Vignaux, Matthews, and Paulos (see for example [Paulos 1995]) will eventually lead to Bayesian reasoning being widely accepted in the law and elsewhere.

In this paper we address the important issue of presenting Bayesian arguments by way of a new example. We describe in Section 2 a fallacy, which we shall refer to as the *jury observation fallacy*. This fallacy does not appear to be have been discussed previously, even though it is related to those described above. As with the prosecutor's and defendant's fallacies, the jury observation fallacy results from a basic misunderstanding of probability theory. The fallacy involves more variables than the other fallacies and hence requires more complex Bayesian calculations and reasoning. The first objective of the paper is to explain that reasoning. However, this paper has an equally important second objective: We wish to show that it is not necessary to follow or even *see* any of the underlying Bayesian calculations in order to be convinced of the results. To do this we present in Section 3 the argument in terms of a visual model, namely a Bayesian Network (BN), together with a tool (Hugin) to perform all of the Bayesian calculations. In Section 4 we describe the assumptions underlying the model and summarise the results of various sensitivity analyses that identify the circumstances under which the fallacy may be 'valid'.

We believe that, with the BN approach, it is possible to make all of the implications of Bayesian reasoning clear to lay people, without them having to understand any of the underlying mathematics or perform any of the calculations. The implications of this in the legal profession could be profound. In future Bayesian arguments could be presented to juries (and accepted by them) in just the same way as forensic evidence is now. BNs provide the mechanism to do this.

2. The Jury observation fallacy

Consider the following situation that we shall refer to subsequently as ***the questionable verdict scenario***:

> "The jury, in a serious crime case, has found the defendant not guilty. It is subsequently revealed that the defendant had a previous conviction for a similar crime."

We now pose a question that we shall refer to subsequently as ***the external observer's question***:

> Does the subsequent evidence of a previous similar conviction (in the questionable verdict scenario) make you less confident that the jury were correct in their verdict?

By an external observer here we mean that the perspective is that of somebody simply observing that a trial has taken place, a not guilty verdict has been delivered and subsequent previous conviction information has been revealed. The external observer is assumed *not* to have been a party to the courtroom proceedings. The fallacy, which we shall refer to as the *jury observation fallacy,* is that most people answer ***yes*** to the *observer's question* irrespective of a range of underlying assumptions. In an extended version of this paper available on the web [Fenton and Neil 2000] we present examples of media reporting of real cases of the questionable verdict scenario. The way these cases are reported also implies that the writers at least assume that the answer should be yes. The implication of such media reporting is clear: 'we' (meaning the public observers) feel that the jury may have made a mistake. The evidence of a previous conviction (which cannot normally be revealed to the jury) 'must surely have increased the likelihood of the defendant being guilty. If only they had known about the previous similar conviction then they may have found the person guilty'. But this is, generally speaking, a fallacy.

In *most* cases that come to court (62% as cited in [Home Office 1998]) the defendant has a previous conviction – so it should hardly come as a surprise if evidence of a previous similar conviction is revealed after the trial. If you are already sure that the probability of guilt is fairly high, then evidence of a prior similar conviction does indeed increase the probability of guilt. However, it is reasonable to assume that the not guilty verdict was delivered because this is what the jury genuinely believed on the basis of the evidence presented. In such circumstances, using very reasonable (and indeed conservative) assumptions about the British legal system we will show that the external observer's response is irrational with respect to Bayesian reasoning. The knowledge that the person had a previous similar conviction should actually make them more convinced in the correctness of the jury's "not guilty" decision; the rational answer to the *observer's question* is NO. If anything, the public should feel more uncomfortable about the not guilty verdict if they subsequently discovered that the defendant had *no* previous similar conviction.

3. The Jury Fallacy explained using a Bayesian Network

Providing that readers are prepared to accept as valid:

1. Bayes' Theorem; and
2. software tools that calculate Bayes' Theorem are accurate

then it is possible to explain the Jury fallacy without exposing the mathematical details. The vehicle for doing this is a visual model called a Bayesian Network (BN) as shown in Figure 1.

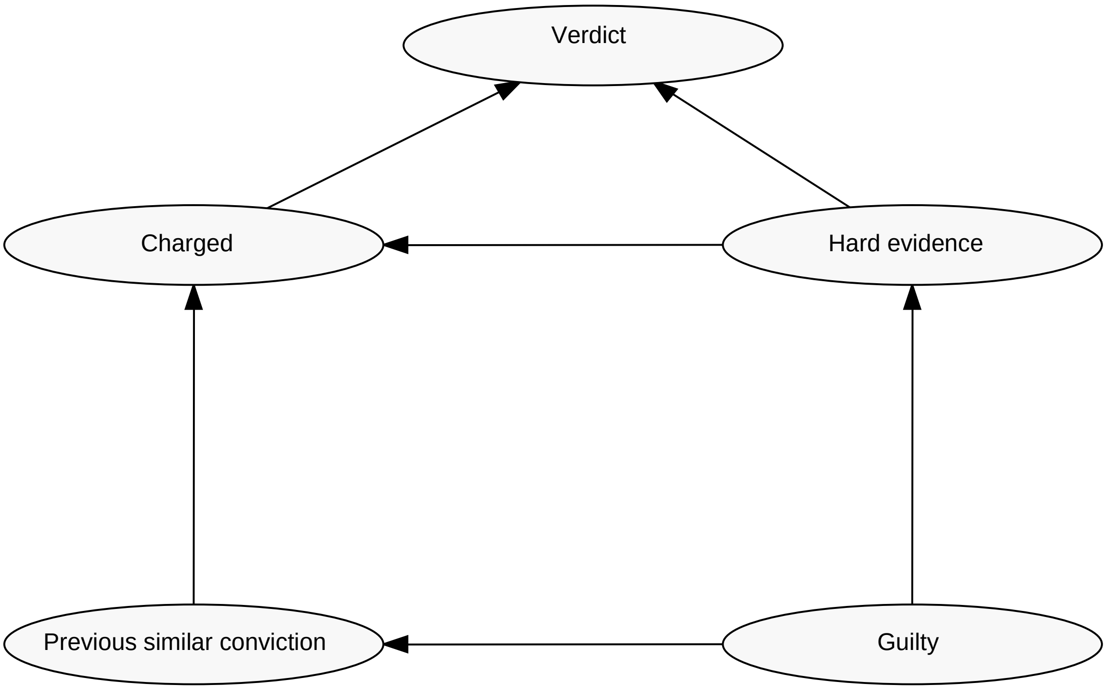

Figure 1: BBN showing causal structure

The nodes in a BN represent uncertain variables, while the arcs represent causal or influential relationships. In the BN in Figure 1 all nodes, except for the *Verdict* node, have two values: true or false. The *Verdict* node has values "Guilty", "Innocent", "No trial" (*Verdict* is synonymous with the outcome of the judicial process). Obviously a guilty or innocent verdict can only be delivered if the person is actually charged, hence the causal link from the node *Charged* to the node *Verdict* (*Charged* here is synonymous with *charged and tried*). Apart from this one trivial dependence, the British legal system decrees that the jury verdict must be influenced by only one thing: the evidence about the case. Only the existence of hard evidence should lead to a guilty verdict. Hence, there is a link from the node *Hard evidence* to *Verdict*.

Associated with each node in a BN is a probability table. Except for nodes that have no parents, this table provides the probability distribution for the node values conditional on each combination of parent values. For example, the probability table for the node *Verdict* (which is conditional on the nodes *Charged* and *Hard evidence*) might look like the one shown in Table 1.

Table 1 *Verdict* probability table

	Charged	Yes		No	
	Hard evidence	Yes	No	Yes	No
Guilty		0.99	0.01	0	0
Innocent		0.01	0.99	0	0
No trial		0	0	1	1

Here, the number 1 in the very last column and row is the probability that the verdict is “no trial” given that there is no charge and no hard evidence. Thus, the last two columns simply assert that “if the person is not charged then the probability that the verdict is guilty must be 0” (because a trial does not take place). The first two columns of figures give the probability of a guilty verdict given that the person has been charged. The assumption here is that the existence of hard evidence means there is a 0.99 probability of a guilty verdict (thus allowing for an error margin of 1%). Without hard evidence the probability of conviction is 0.01.

For nodes without parents (such as “Guilty”) the table represents the prior probability distribution. For example, if the crime were committed in a city of 10,000 people then the probability table would look like the one shown in Table 2 (all of the detailed probability assumptions in the BN are given in Section 4).

Table 2: *Guilty* probability table

Yes	0.0001
No	0.9999

It must be stressed that the BN shown in Figure 1 caused much debate and initial criticism when we presented it to learned colleagues. However, the key points to note are:

1. As with any Bayesian argument it must be clear whose perspective it is from and whose it is not from. For example, this model is definitely *not* from the perspective of the person charged, the police, or indeed the jury. A member of the jury who really wished to use a Bayesian argument to reveal more about the probability of guilt would certainly not need a node *Verdict*.
2. The particular perspective we have adopted is of the one relevant to the *observer’s question*. In other words it is relevant to somebody who is not party to the courtroom proceedings, but who might read about the results in a newspaper. This is the appropriate perspective for the problem that we are studying. Hence, for example, from this perspective it is perfectly reasonable to assume that the prior probability of guilt is simply the probability that a randomly selected person from this population is guilty.
3. The BN represents an idealised observer’s capacity to reason about evidence normally not available to the jury. Presentation of evidence about prior convictions is usually prohibited because it may lead the jury to convict the accused based on evidence about crimes other than the one under consideration. Likewise, the jury should not assume that, because a person has been charged, the police must have a good reason for doing so. Such arguments are essentially statistical and are unfair because they are not based on “hard” physical evidence. The BN therefore represents the unfair advantage the observer has over the jury and this unfairness is manifest in the following property of the BN model. The prior probability of guilt, when the only information known is that the person has been charged, is 88%. To the observer this represents a very high chance that the accused will be found guilty based on the historical knowledge of the conviction rate and the relationship between prior convictions and guilt (the figure comes directly from the 1998 official Home Office statistics [Home Office 1998]). From the jury’s perspective such a prior probability clearly verges on an assumption of guilt rather than of innocence - the very reverse of the tradition set in English law.
4. Like any model it is a massive simplification of reality, and hence it ignores variables that in specific circumstances might be crucial. It ignores all specific evidence, including circumstantial evidence. It simply classifies evidence as either ‘hard’ (meaning sufficiently convincing beyond reasonable doubt of guilt) or not. It ignores issues like whether criminals with previous convictions can learn how to ‘cheat the system’ by hiding evidence, and differences between men and women.

5. We are primarily interested in serious crimes since these are the ones most often reported in the press and for which the jury observation fallacy has been observed. The BN therefore is only relevant for serious criminal cases.
6. There was much discussion about whether the variable *Verdict* was really necessary since it is merely an (inaccurate) estimate of the variable *Guilty*, which is already there. By inaccuracy we mean the confidence the jury has in any probabilities assigned to p(hard evidence | guilty). This confidence will be influenced by the jury's ability to infer guilt from evidence in a realistic manner and any uncertainties inherent in the direct evidence itself. Clearly this is a much more complex situation and could be modelled by a BN, but would only serve to distract us from the central arguments.

It is important to note that, even when all of these assumptions are made clear when presenting the *external observer's question*, most lay persons still normally answer 'yes'.

BNs can be used to model complex chains of reasoning with conditional dependencies. When observations (that is, particular values) are entered all of the other probability values in the BN are updated by applying Bayes' theorem and propagating the results throughout. The notion of Bayesian propagation has been around for a long time. However, it is only in the last few years that efficient algorithms [Lauritzen and Spiegelhalter 1988, Pearl 1988], and tools to implement them [Hugin, Jensen 1996, SERENE 1999] have been developed. Hence it is only recently that it has been possible to perform propagation in networks with a reasonable number of variables. The recent explosion of interest in BNs is due to these developments which mean that for the first time realistic size problems can be solved.

The BN in Figure 1 is relatively small, but even in this case the Bayesian propagation computations would be difficult to do by hand. Using the Hugin tool we do not have to worry about the calculations. As we shall see below, they are performed instantly. In fact once a BN structure is agreed, explaining the Bayesian argument becomes a straightforward matter of running the tool. You enter observations and the tool automatically displays all of the revised probability values. Obviously, the results are only as good as the original model. If the underlying influences and probability tables are changed then so are the results. However, the key point is that discussion about the model is completely separated from showing the results of the assumptions. Discussions about the validity of particular probability tables can be localised and hidden from lay people to whom we wish to present the results. This is very important when we consider the real legal case discussed in Section 1 and others discussed by authors such as [Robertson and Vignaux 1998b]. In these cases much of the perceived confusion about the Bayesian argument was in the actual Bayesian propagation calculations rather than the underlying model assumptions.

Before we present the probability tables in Section 4 (which we stress again would be hidden from lay people, such as a jury, who need to understand the implications of applying Bayes' Theorem) we will actually demonstrate the fallacy by running the tool. What we have done is fixed throughout the 'size' of the population to be 10,000 for illustrative purposes (we could, in fact, have made the population size a variable in the BN, but this would have unnecessarily clouded the key points).

Figure 2 shows the result of running the tool without any observations (that is, without any evidence). In other words the tool is computing the prior marginal probability values of each variable (the probabilities are actually displayed as percentages rounded to two decimal figures). Thus, for example, before we have any evidence about the crime there is 0.01% chance (probability of 0.0001) that a randomly selected person from the population will be found guilty of the crime – about the same probability that the person really is guilty.

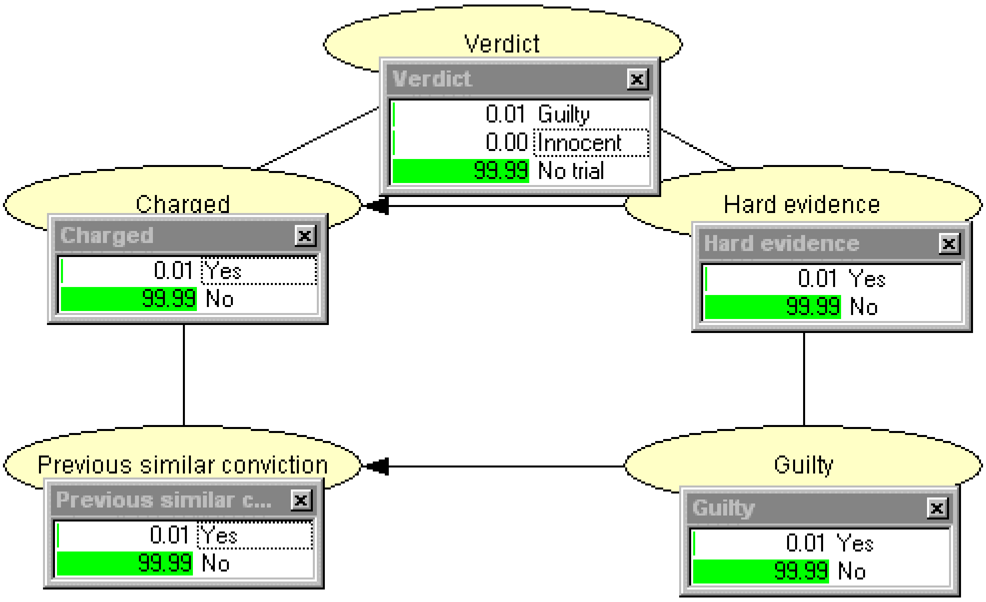

Figure 2: Initial probabilities

Next, suppose we know that a person has been charged and tried. Hence in Figure 3 we enter the observation "Yes" for the node *Charged*. All of the other probabilities in the BBN are now updated. Without knowing anything about the evidence there is 0.8802 probability that the jury will deliver a guilty verdict (in other words in 88% of all trials a guilty verdict is returned) and a 0.8788 probability that the person charged and tried really is guilty.

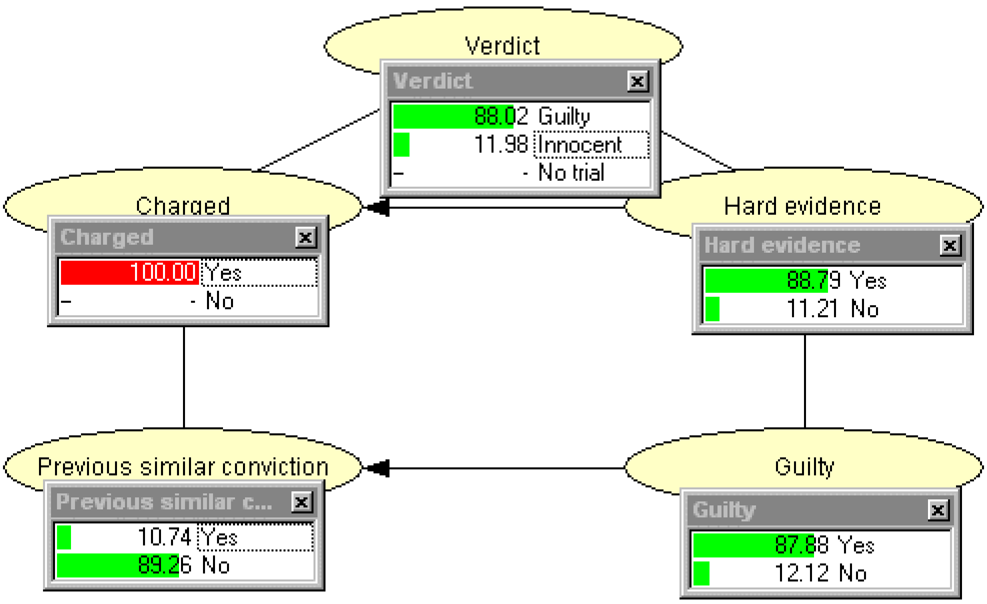

Figure 3: Person is charged and tried

Next, suppose the jury find the defendant not guilty. Figure 4 shows the result of entering this observation.

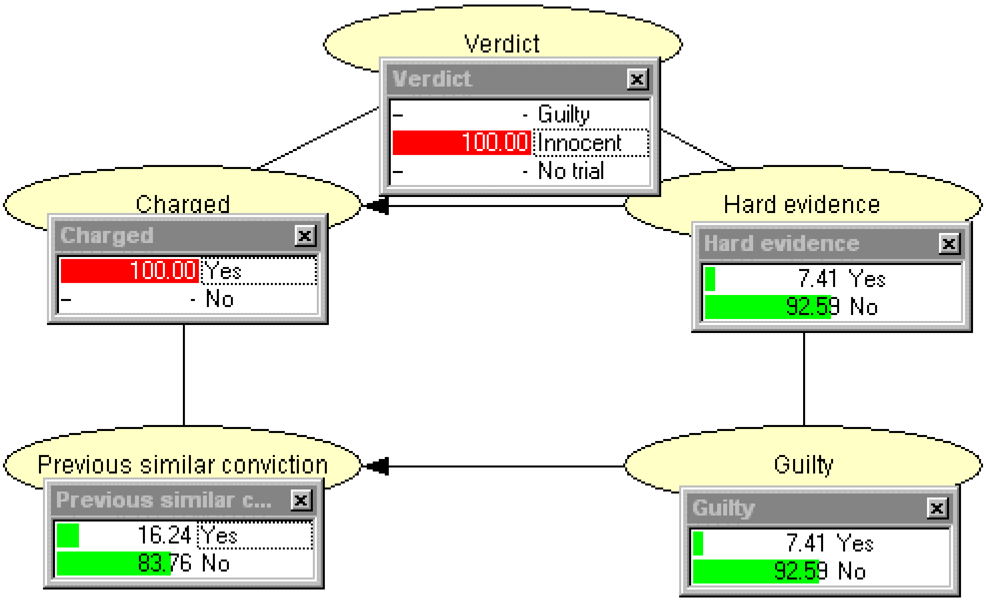

Figure 4: Person is found innocent

See how the observer's previous beliefs are drastically changed. There is now a 0.9259 probability that the person is really innocent, and this is explained primarily because we now have a strong belief that there was no hard evidence on which to make a conviction. Now we come to the precise scenario of the jury fallacy. Having found the defendant not guilty, we are informed that the defendant had a previous similar conviction. Figure 5 shows the results of entering this information. **Crucially, the probability that the defendant really is guilty has not increased (as assumed in the fallacy), but has decreased slightly** – from a previously low value of 0.0741 to an even lower one of 0.0503. Moreover, the BBN actually "explains away" this decrease. The probability that there was hard evidence has also decreased (from 0.0741 to 0.0456). In other words, it was previously almost certain that the not guilty verdict was the result of lack of hard evidence. Finding out that the defendant had a previous conviction actually explains away why, in the absence of hard evidence, the defendant was charged in the first place. Hence the probability of no hard evidence increases further explaining our increased belief in innocence.

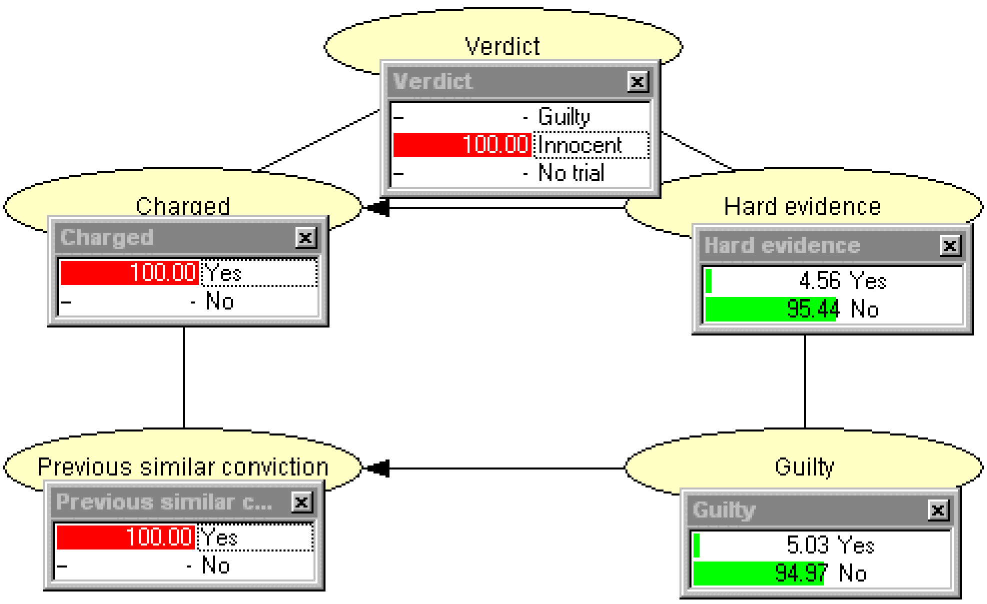

Figure 5: Previous conviction is revealed

Now let us compare what happens when, instead of finding out that the defendant had a previous conviction, we find that he had no previous conviction. This is shown in Figure 6.

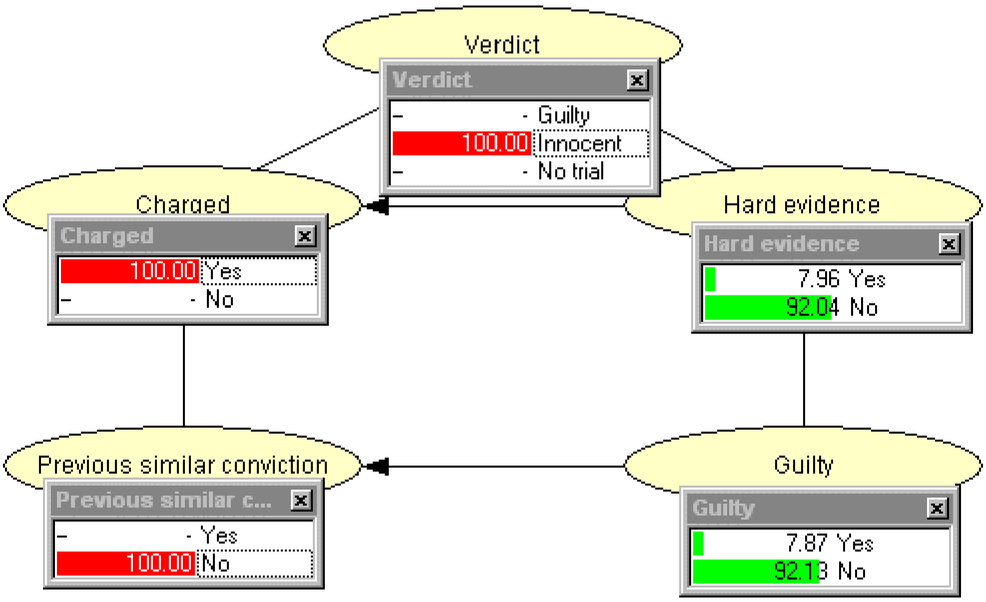

Figure 6: No previous conviction revealed

In this case we are still pretty sure that the verdict was correct because it is most likely there was no evidence. However, we are less sure than if there was a previous conviction (Figure 5) and we are even slightly less sure than before we knew anything about previous convictions (Figure 4).

As long as you believe the underlying model the above explanation shows that the jury fallacy is truly a fallacy. In the next section we describe exactly what the underlying assumptions of the model were.

4. The assumptions in the Bayesian Network

The crucial assumptions in the BN are the probability tables. In Section 3 we already showed the probability tables for the nodes *Verdict* and *Guilty*. The latter is uncontroversial being based purely on the population size. The former assumes that the jury is usually 99% "accurate" in the sense of their strict legal obligations; in other words 99% of the time a jury will deliver a guilty verdict if there is hard evidence and 99% of the time the jury will deliver a not guilty verdict is there is no hard evidence. The impact of various changes to this assumption (and indeed to all of the probability tables that we present below) is done in the Appendix and summarised at the end of this section.

Table 3 shows the probabilities for *Previous similar conviction.* Our assumption is that if a person is guilty then they are 1000 times more likely to have a previous similar conviction than if they are not guilty. **It is a basic misunderstanding of the influence of this information on Bayesian probability reasoning that results in the jury observation fallacy**.

Table 3: *Previous similar conviction* probability table

	Guilty	Yes	No
Yes		0.1	0.0001
No		0.9	0.9999

Next, we consider the node *Charged*. When a crime has been committed the police system works on the basis of collecting evidence and linking the evidence to a suspect (the whole issue of motive is important and can easily be modelled, but we have chosen not to do so to simplify the entire discussion). In the absence of hard evidence (and normally even before an attempt to gather hard evidence has been made) the police will identify “likely suspects”. There are a number of means by which this is done, including use of informants and circumstantial evidence, but by far the most common is to use a database of people convicted of previous similar offences to the one committed. Recent well-publicised cases have shown that the existence of a previous similar conviction, together with some circumstantial evidence is often sufficient for a suspect to be charged. A reasonable probability table for *Charged* is therefore the one shown in Table 4. The probability of being charged given previous similar convictions is 1 in 50 (0.02). In arriving at this figure the BN assumes a 10,000 population of whom 10 have previous similar conviction. Assume that in 20% of cases there is no hard evidence. In such cases (because it is a serious crime) the police will arrest someone on the basis of a previous conviction alone. Hence we arrive at the 0.02. The key point is that, although it is still very unlikely for a person to be charged on the basis of a previous similar conviction alone this small value is nevertheless significantly higher than the probability of a person being charged without any previous conviction or evidence (which we assume is very close to 0). Similarly, if there *is* hard evidence then it is almost certain (0.99 probability) that the person will be charged even without a previous conviction. However, this increases to 0.9999 (a factor of 100) if there is also a previous similar conviction.

Table 4: *Charged* probability table

	Hard evidence	Yes		No	
	Previous similar conviction	Yes	No	Yes	No
Yes		0.9999	0.99	0.02	0.00001
No		0.0001	0.01	0.98	0.99999

As with any of the tables in a BBN, Table 4 is typical inasmuch as the values could be obtained by statistical data if it is available and if not by subjective probabilities from relevant experts.

Table 5 shows the probabilities for *Hard evidence*. This simply assumes that in 95% of cases a guilty person will provide some hard evidence. We also allow for a very small non-zero probability that an error can be made (a not guilty person provides what is assumed to be hard evidence; this might happen, for example, if forensic evidence is analysed incorrectly).

Table 5: "Hard evidence" probability table

	Guilty	Yes	No
Yes		0.95	0.000001
No		0.05	0.999999

In [Fenton and Neil 2000] we performed a full sensitivity analysis that has been scrutinised by several experts. In this we looked at the impact on the fallacy of a wide range of adjustments to all of the probability tables. We found just two situations where the ‘correct’ answer to the external observer's question should be *yes* (and hence which support the jury fallacy). These are where:

1. ***The chance of a suspect being guilty given a conviction for a previous similar conviction was greater than 2000 times the chance of his being guilty given no such prior conviction***. In this case this would mean that, from a population of 10,000, the probability that a person, with a previous conviction, is guilty is 10% (approx.). Such a figure is clearly absurdly high given we had assumed only 10 people in this population (0.1%) has a previous conviction. For serious crimes the actual figure must be significantly lower than this because those who have a previous conviction for a serious crime tend to have served a long prison sentence (and so cannot re-offend during their sentence) and are statistically less likely to re-offend upon release. Unfortunately, the publicly available UK Home Office data does not shed any convincing light. It reveals that 67% of convicted offenders had a previous conviction [Home Office 1998], but this statistic is dominated by less serious offences because serious crimes are comparatively rarer. Also, the official figures do not reveal which proportion of previous convictions were for similar offences. When we restrict ourselves to serious crimes we cannot therefore assume a strong link between previous conviction and guilt.

2. ***The chance of being charged given a previous similar conviction and no hard evidence was less than 1 in 200.*** Again we can show that this case is unreasonable. We have already argued that, for less-serious offences, the connection between previous offending and future offending will be stronger than for serious offences. In the absence of hard evidence the number of potential suspects for less-serious offences is that much larger, making the chances of randomly selecting and charging the criminal lower. For serious crimes, where the population of suspects is much less, a 1 in 200 chance of being charged given a previous similar conviction and no hard evidence seems absurd. The actual figure must be much higher given the smaller population of suspects to choose from. In cases where the real uncertainty is not about mistaken identity, but about whether the crime really happened (as in the example cited from the Daily Telegraph in Section 2) there is normally just *one* suspect to choose from.

An implication of our model and assumptions is that there may be a serious problem with charging people who have been convicted with similar offences in the past. It appears that many people are being charged and convicted on the basis of prior convictions alone. The statistics on conviction rates appear to bear this out [Home Office 1998]:

- Over 60% of suspects arrested had a previous conviction
- 52% of suspects arrested are formally charged
- 62% of suspects charged had a previous conviction
- In Magistrates Court 92% of trials end in a guilty verdict and in a Crown Court 81% of trials end in a guilty verdict (remarkably, as high as it is this conviction rate is lower than in the US and Japan. For example, in Sept 1999 the San Francisco Chronicle reported a 91% felony conviction rate for cases that go to trial in the State of California, while in Japan 99.91% of all criminal trials result in conviction according to many sources on the Web).
- 67% of convicted offenders had a previous conviction

We have not had access to all of the various statistics available, nor have we built a BN to model all of the relevant factors. We feel, however, that based on the limited analysis performed there is justified reason for disquiet about the use of previous convictions as a basis for selecting suspects. Further study is clearly needed to investigate the matter properly.

5. Conclusions

The issue of using Bayesian reasoning in legal cases has taken some dramatic twists in recent years. This is especially true of cases involving DNA evidence [Balding and Donnelly 1994, 1996], [Hunter 1998], [Redmeyne 1995], [Robertson and Vignaux 1998], but is also relevant generally for presenting all types of evidence [Robertson and Vignaux 1995, 1998a]. Despite the wide acceptance of Bayesian reasoning as a logical means of formalising uncertainty, it has been regarded with great scepticism in courtrooms. Moreover, the third appeal court ruling of the case described in [Balding and Donnelly 1994] suggests that the prospects of Bayesian reasoning being more widely accepted in the near future are grim. This is an extremely regrettable situation, since it means that a valid scientific discipline is being excluded from an application where it can be especially useful.

Despite the recent pioneering work by the likes of Aitken, Balding and Donnelly, Matthews, Paulos, and Robertson and Vignaux, it is clear that the legal profession remains generally fairly ignorant of Bayesian reasoning. Considering that Bayes' theorem is over 200 years old this is a sad reflection on the ability of Bayesians to explain the broader relevance of their work outside academic circles. The public at large, including experts in all walks of life, are known to be very bad at even simple probabilistic reasoning. Hence, we feel strongly that the public at large (and crucially, this includes influential judges) will never really understand complex Bayesian arguments presented from first principles. But then the public at large would not be expected to understand the relevance of forensic evidence if it was presented from first principles either. Hence, we feel that the key to a broader acceptance of Bayesian reasoning is to be able show all of the implications and results of a complex Bayesian argument without requiring any understanding of the underlying theory or mathematics, and without having to perform any of the calculations. One of the objectives of this paper has been to show that there *is* such a way to present Bayesian arguments.

By way of example we have described a new fallacy, the jury observation fallacy, in which observers of a trial automatically lose some confidence in a not guilty verdict when evidence of a previous similar conviction by the defendant is subsequently revealed. We have shown that, with a range of different reasonable assumptions, finding out that the defendant had a previous similar conviction should actually make you more convinced of the correctness of the not guilty verdict. As with other well known fallacies in probability reasoning, this particular fallacy arises from a basic misunderstanding of conditional probability – people are aware that a guilty person is much more likely to have a previous conviction, but ignore the impact of a previous conviction on the probability of being charged. If a jury finds a defendant not guilty they do so because of a lack of hard evidence. In such circumstances a previous conviction "explains away" why a person innocent of this crime was charged and tried in the first place.

An unintended result of our work has also shown that there are reasons for disquiet about the use of information about previous convictions as a reason to charge suspects in serious criminal cases.

We have shown that, by using Bayesian nets and a tool such as Hugin, it is possible to show all of the implications and results of a complex Bayesian argument without requiring any understanding of the underlying theory or mathematics. Forensic (and indeed many other forms of evidence) is accepted because the scientists agree on what is a valid theory and argument. Hence, it is usually sufficient to present the results, without having to give an explanation of how the results were arrived at. The same should be true of Bayesian arguments, and BNs provide the mechanism to do this. In future, as with forensic science, it should be sufficient to provide the results of a Bayesian argument (possibly using a tool like Hugin) and rely only on the word of experts that the tool has been used correctly. Indeed, the results of [Honess et al 1998] confirm that juries are competent to process complex information providing there is some screening and supporting tools that are analogous to BNs.

We have already developed and applied BNs to predict the safety and reliability of software-intensive and traditional mechanical systems, with some success. Our projects [Fenton et al 1998, Agena 1999, Fenton et al 1999. Fenton and Neil 1999a] have demonstrated that engineers can benefit from BNs without any need for specialist education in statistics and probability; provided that they are shielded from the underlying mechanics of Bayesian updating. To help construct credible BNs we have developed methods to help reduce the practical problems faced by those wishing to use BNs in earnest. Specifically, our methods help build BN graph topologies and produce large conditional probability tables from smaller sets of probability distributions [Neil et al 1999]. We are also examining wider issues such as multi-criteria decision-making under uncertainty [Fenton et al 1999]. As well as applying BN technology to real problems our future research work will be directed at further improving information visualisation and looking at ways of investigating the sensitivity of results to changes in assumptions.

6. Acknowledgements

This work was supported in part by the EPSRC project IMPRESS and done mainly when the authors were still with the Centre for Software Reliability, City University. We are indebted to the following who provided valuable comments on earlier drafts: Simon Forey, Bev Littlewood, Robert Matthews, Martin Newby, Lorenzo Strigini and David Wright. We also acknowledge the valuable advice of an anonymous referee, which has led to important changes in the paper.

7. References

[Agena 1999] "Bayesian Belief Nets", Agena Ltd, Cambridge, UK. Article at http://www.agena.co.uk, 1999.

[Aitken 1996] Aitken, C, "Lies, damned lies and expert witnesses", Mathematics Today, Bulletin of the IMA, 32 (5/6), 76-80, 1996.

[Ayton 1994] Ayton P, On the Competence and Incompetence of Experts, in Expertise and Decision Support (Ed. G. Wright and F. Bolger), pp. 77-105, Plenum Press, 1994.

[Ayton and Pascoe 1995] Ayton P and Pascoe, Bias in Human Judgement under Uncertainty, Knowledge Engineering Review Vol 10.1, 21-41, 1995.

[Balding and Donnelly 1994] Balding DJ and Donnelly P, 'Prosecutor's fallacy and DNA evidence' , Criminal Law Review, 711-722, 1994.

[Balding and Donnelly 1996] Balding DJ and Donnelly P, 'Evaluating DNA profile evidence when the suspect is identified through a database search', Journal of Forensic Science 41, 599-603, 1996.

[Fenton et al 1998] Fenton NE, Littlewood B, Neil M, Strigini L, Sutcliffe A, Wright D, "Assessing Dependability of Safety Critical Systems using Diverse Evidence", IEE Proceedings Software Engineering, 145(1), 35-39, 1998.

[Fenton and Neil 1999a] N.E. Fenton and M. Neil "A Critique of Software Defect Prediction Research". IEEE Transactions on Software Engineering, 25(5) 675-689, 1999

[Fenton and Neil 1999b] Fenton NE and Neil M, "Making Decisions: Using Bayesian Nets and MCDA", to appear Knowledge-Based Systems, 2000.

[Fenton et al 1999] N.E. Fenton, M. Neil M and S. Forey. "TRACS (Transport Reliability And Calculation System) User Manual", CSR/TRACS/D12-v1.0, 30 March 1999. (Available from the Centre for Software Reliability, City University, Northampton Square, London EC1V 0HB, UK). www.agena.co.uk/tracs.

[Fenton and Neil 2000] Fenton NE and Neil M, "The Jury Observation Fallacy", www.agena.co.uk/papers

[Kahneman et al 1982] Kahneman D, Slovic P, Tversky A (eds), Judgement Under Uncertainty: Heuristics and Biases, Cambridge University Press, 1982.

[Home Office 1998] UK Home Office Research Study 185 Entry into the criminal justice system: a survey of police arrests and their outcomes, Crown Copyright 1998. (http://www.homeoffice.gov.uk/rds/publf.htm).

[Honess et al 1998] Honess TM, Levi M, Charunan EA, 'Juror competence in process complex information: implications from a simulation of the Maxwell trial', Criminal Law Review, 763-773, Nov, 1998.

[Hunter 1998] Hunter K, 'A new direction on DNA' , Criminal Law Review, 478-480, July, 1998.

[Lee 1997] Lee PM, "Bayesian Statistics: An Introduction, 2nd Edn", Arnold, London UK, 1997.

[Matthews 1995] Matthews R, "The Interrogator's Fallacy", Bulletin of the IMA, 31, 3-5, 1995.

[Matthews 1996] Matthews R, "Why is weather forecasting still under a cloud", Bulletin of the IMA, 32 (11/12), 168-170, 1996.

[Neil et al 1999] Neil M, Fenton NE, Nielsen L, "Building large-scale Bayesian Networks", submitted Knowledge Engineering Review, 1999.

[Paulos 1995] Paulos JA, A Mathematician Reads the Newspaper, Basic Books, 1995.

[Redmeyne 1995] Redmeyne M, 'DNA evidence, probability and the courts' , Criminal Law Review, 464, 1995.

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

[Robertson and Vignaux 1998] Robertson B and Vignaux T, Bayes' Theorem in the Court of Appeal, The Criminal Lawyer, January, pp 4-5 (also http://www.mcs.vuw.ac.nz/~vignaux/evidence/E96.017_Adams/Adams_NLJ.html), 1997.

[Robertson and Vignaux 1998a] Robertson B and Vignaux T, Don't teach Statistics to Lawyers!, Proceedings of the Fifth International Conference on Teaching of Statistics, Singapore, Vol 1, pp 541-547, 1998.

[Robertson and Vignaux 1998b] Robertson B and Vignaux T, Explaining Evidence Logically, New Law Journal, Vol. 148, no 6826, p.159-162, 1998.
(Also http://www.mcs.vuw.ac.nz/~vignaux/evidence/E98.003_LogicalJuries/logicalJuries.html).

8. Appendix: Sensitivity Analysis

In this section we present the results arising from various changes to the assumptions encoded in the probability tables. What we have done is to change the probability tables and then run the BN on precisely the same scenarios as shown in Figures 4-6. Essentially, we are only interested in the changes to the probability value "yes" in the node *Guilty*. Thus we look at this value for the three scenarios:

1. Innocent, no information of previous conviction
2. Innocent, previous conviction
3. Innocent, no previous conviction

For example, with the original assumptions we have probability guilty = "yes" as shown in Table 6.

Table 6: Probability of real guilt for original assumptions

Innocent, no information of previous conviction	0.07
Innocent, previous conviction	0.05
Innocent, no previous conviction	0.08

What we need to look for are examples where there are significant changes in 'direction'. If we find that, with certain assumptions, the value in row 2 is higher than the other values then these are the assumptions under which the jury fallacy is 'reasonable' especially if the actual value moves above 0.5.

8.1 Varying the jury accuracy

The original Table 1 assumed that juries are generally very accurate – 99% in the sense explained in Section 4. The "percentage inaccuracy" figure is shorthand for summarising the table for *Verdict*. Table 7 shows the results of changing the probability values to reflect varying degrees of accuracy of the jury. We show five different levels representing five different probability tables including the original (in each of the following examples we include the original probabilities for comparison in bold text).

Table 7: Probability of real guilt for varying jury accuracies

Observations	**1% inaccuracy**	5% inaccuracy	10% inaccuracy	20% inaccuracy	30% inaccuracy
Innocent, no information of previous conviction	**0.07**	0.29	0.46	0.66	0.76
Innocent, previous conviction	**0.05**	0.2	0.34	0.54	0.67
Innocent, no previous conviction	**0.08**	0.31	0.48	0.68	0.78

In each case the fallacy still exists. The probability of guilt decreases rather than increases as a result of knowing that the person has a previous conviction.

8.2 Varying the strength of link between guilt and previous convictions

The original Table 3 assumed that a guilty person was 1000 times more likely to have a previous similar conviction than an innocent person. Table 8 shows the results of varying this assumption.

Table 8: Probability of real guilt for varying strength of link between guilt and previous convictions

Observations	10 times more likely	100 times more likely	**1000 times more likely**	2000 times more likely	5000 times more likely
Innocent, no information of previous conviction	0.07	0.07	**0.07**	0.07	0.08
Innocent, previous conviction	0.0005	0.005	**0.05**	0.09	0.2

Innocent, no previous conviction	0.09	0.09	**0.08**	0.07	0.05

We deliberately set the original value to be very high. At 1000 times more likely, it meant that in a population of 10,000 we already assumed a relatively high probability (nearly 0.1) that a person with a previous conviction must be guilty without any other evidence. This is shown in Figure 7 using Hugin again. We felt that if the fallacy were exposed with this assumption then it would be even more dramatically exposed at more reasonable (lower) levels.

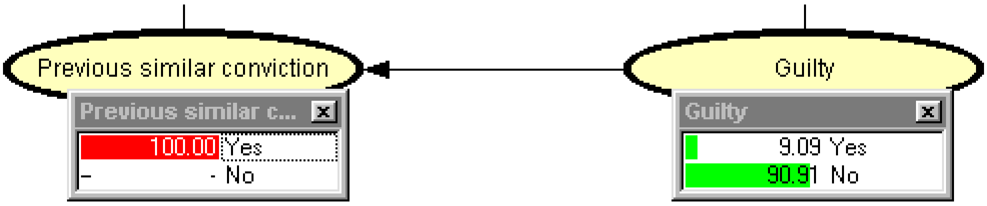

Figure 7 : Effect of previous conviction on assumption of guilt

Table 8 confirms this. If we assume that a guilty person is only ten times more likely to have a previous conviction then the already low probability of guilt drops to almost 0 when we discover that the person has a previous conviction. Only when we get to 2000 times more likely is there modest support *for* the fallacy. In the last column scenario (5000 times more likely) the probability of guilt does jump (from 0.08 to 0.2) when we discover a previous conviction. However, in this case we must already have assumed that there was 0.33 probability that the person was guilty before any evidence is known (as shown in Figure 8). What is remarkable is that even in this case the information about previous conviction, given the innocent verdict, still should not change our belief in the person's innocence. The probability the person is guilty is still only 0.2, which is also much lower than the probability (0.33) when we know nothing except the previous conviction.

Figure 8: Effect of previous conviction on assumption of guilt when guilty person is 5000 times more likely to have previous conviction

8.3 Varying the hard evidence 'inaccuracy'

The Table 5 for *hard evidence* assumed that there was only a 5% chance that a guilty person would not provide hard evidence. Table 9 shows the effects of changing this assumption. At 1% inaccuracy the fallacy is even more obviously exposed, while going up to 30% inaccuracy has almost no different effect at all.

Table 9: Probability of real guilt given varying inaccuracy of hard evidence

Observations	1% inaccuracy	**5% inaccuracy**	10% inaccuracy	20% inaccuracy	30% inaccuracy
Innocent, no information of previous conviction	0.07	**0.07**	0.07	0.07	0.06
Innocent, previous conviction	0.0005	**0.05**	0.05	0.06	0.06
Innocent, no previous conviction	0.09	**0.08**	0.07	0.07	0.06

8.4 Varying the strength of link between previous conviction, evidence and being charged

Table 4 (*Charged*) assumed that a person with a previous conviction was 100 times more likely to be charged than a person without. Although the police does not release such data this seems to be a reasonable assumption, especially given that the model also assumed that a person who was guilty was 1000 times more likely to have a previous conviction. Table 10 provides five alternative probability scenarios ranging from case 1 (where previous conviction has a larger impact on probability of being charged) to case 5 (where previous conviction has a much smaller impact on probability of being charged).

Table 10: Alternative probabilities for being charged given hard evidence and previous conviction

	Hard evidence	No	
	Previous similar conviction	Yes	No
Case 1		0.1	0.00001
Case 2		0.05	0.00001
Case 3		**0.02**	**0.00001**
Case 4		0.01	0.00001
Case 5		0.005	0.00001

Running these five cases through Hugin yields the probability values for guilt shown in Table 11. Only the last (case 5) provides an argument in favour of the fallacy, but makes the unlikely assumption that the probability of being charged given a previous similar conviction and no hard evidence is only 1 in 200 (0.005).

Moreover, even in this case the subsequent information about a previous conviction given the not guilty verdict still only increases the probability of guilt to 0.16 from 0.08.

Table 11: Probability of guilt given varying strength of link between previous conviction and being charge.

Observations	Case 1	Case 2	**Case 3**	Case 4	Case 5
Innocent, no information of previous conviction	0.04	0.06	**0.07**	0.09	0.08
Innocent, previous conviction	0.01	0.02	**0.05**	0.09	0.16
Innocent, no previous conviction	0.07	0.07	**0.08**	0.08	0.07