Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2004/13

Handling Manipulated Evidence

Gianluca Baio, Fabio Corradi

Università degli Studi
di Firenze

Statistics

HANDLING MANIPULATED EVIDENCE

GIANLUCA BAIO AND FABIO CORRADI

University of Florence
Department of Statistics “G. Parenti”
59, Viale Morgagni, Florence 50134 - Italy
{*baio,corradi*}*@ds.unifi.it*

Abstract. Bayesian Networks have been advocated as a useful tool to describe the relations of dependence/independence among random variables and relevant hypotheses in a crime case. Moreover, they have been applied to help the investigator structure the problem and weight the observed evidence, typically with respect to the hypothesis of guilt of a suspect. In this paper we describe a model to handle the possibility that one or more pieces of evidence have been manipulated in order to mislead the investigations. This method is based on causal inference models, although it is developed in a different, specific framework.

Keywords: Bayesian Networks, Influence Diagrams, Forensic identification, Manipulated Evidence, Causal Inference.

1. Introduction

Bayesian Networks (BNs) have recently been advocated in forensic science as a powerful tool to establish the overall dependence between relevant hypotheses and observable random variables, and to evaluate the probabilistic effects of the latter on the former during an investigation, or in a trial (Dawid & Evett 1997, Garbolino & Taroni 2002). A further interesting feature of a BN is that it can be easily increased including relations with previously not considered variables, whenever required and as usually happens in real practice.

A possible subtle drawback in the use of BN-assisted investigations consists in overconfidence in the results obtained. The most treacherous possibility occurs if manipulated evidence is introduced, i.e. if observations not genuinely arisen from the context are produced to mislead the investigator. Examples of cases where police or a Court is induced to focus towards a person different from the culprit include false testimonies, blood traces left intentionally, and many others.

The aim of this work is to build a model that can help the investigator handle some possibly manipulated variables, in order to produce an updating of the probability that the evidence under suspicion is in fact genuine or not, as well as the posterior distribution of the hypotheses of interest.

The structure of the paper is the following: in section 2, we describe the very basics of BNs theory. Then, in section 3, we show how BNs can be used to formalise an investigation case, following its development. We presume that an expert, the investigator, supervises the construction of the models based on genuine evidence.

This paper was published as a Working Paper in the Department of Statistics ‘G. Parenti’ series: Baio, G. & Corradi, F. (2004), Handling Manipulated Evidence. *University of Florence Press, Florence, Italy, Working Paper* no. 2004/13.

In section 4 we derive the model that takes into account the possibility of manipulated evidence, modifying the graphical structure obtained under natural conditions, according to the intervention model originally introduced into the statistical literature by Pearl (1993) and Spirtes et al. (1993) in the causal inference framework.

In section 5, the model is extended to situations where more than one piece of evidence whose origin is unknown is possibly manipulated, to help the investigator detect a criminal plan aimed at misleading the inference. Finally, in section 6, we discuss the most relevant implications of this work.

2. Bayesian Networks

This section briefly reviews the basic features of BN theory. For a more thoroughly description, see for example Cowell et al. (1999), or Jensen (2001).

Formally, a BN is represented by $\mathcal{B} = \{\mathcal{G}, \mathbf{\Pi}\}$, where $\mathcal{G}$ is a Directed Acyclic Graph (DAG), and $\mathbf{\Pi}$ includes the conditional probability tables (CPTs) for the nodes in $\mathcal{G}$.

A DAG is a graphical structure $\mathcal{G} = (\mathbf{X}, \mathbf{E})$, where $\mathbf{X} = \{X_1, \ldots, X_n\}$ is the set of relevant nodes, each of which is associated with one of the random variables involved, and $\mathbf{E}$ is the set of edges connecting the nodes. Many examples of DAGs and CPTs are given in the rest of the paper (see for instance Figure 2 and Table 2).

The structure of the DAG is essentially provided by the expert knowledge. Conversely, the CPTs can be specified in different ways, depending on the context. First, epistemic probabilities elicited by experts can be used to formalise their opinions (Spiegelhalter et al. 1993). Otherwise, in case an appropriate sample is available, CPTs can be learned from empirical data, or specific experiments can be performed and probabilities can be derived from the results (Aitken & Taroni 2004).

The set $\mathbf{X}$ includes both *unobservable* (such as working hypotheses) and *observable* variables, which become pieces of evidence, once actually observed.

The set $\mathbf{E}$ specifies the alleged relations among the variables in $\mathcal{G}$. A node that 'points' another is said to be a *parent*, whereas the node that is reached by the arrow is a *child*. The set of the parents of a node X is indicated by $\text{pa}(X)$, and the set of its children is $\text{ch}(X)$. The set of the nodes in the directed path leaving X are named *descendants*, and is indicated by $\text{de}(X)$.

A direct arrow drawn from the node X_1 towards the node X_2 does not imply any causal effect *per se*, but only means that the probability distribution of X_2 is modified according to the value assumed by X_1.

More specifically, this circumstance expresses the fact that the expert is willing to: *a)* establish an explicit association between X_1 and X_2, and *b)* declare a preference in providing the joint distribution $\Pr(X_1, X_2)$ through the factorisation $\Pr(X_1) \times \Pr(X_2 | X_1)$, over any other alternative specifications.

On the contrary, the absence of a direct link from X_1 to X_2 encodes the assumption that the expert does not reckon that the conditional distribution of X_1 is directly dependent on the possible values that X_2 can take on. Nevertheless, observing X_1 can produce an undirect change in the probability distribution of X_2, through an open path connecting X_1 to X_2.

Independence among nodes is indicated by the symbol $\perp\!\!\!\perp$, commonly used in statistical literature (Dawid 1979), whereas the symbol $\not\!\perp\!\!\!\perp$ indicates dependence.

3. Modelling genuine evidence

In this section we show how an investigation can be translated into a BN framework, according to the information that successively becomes available to the investigator.

Unlike other works, such as those of Dawid & Evett (1997) and Garbolino & Taroni (2002), our focus is not in defining a collection of formulæ to be used in the calculation of the posterior probabilities of the relevant hypotheses and/or the associated weight of evidence.

In fact, despite the practice to highlight the role of some epistemic and population probabilities is quite common and formally attractive in forensic science, this approach proves of limited help, when the practitioner has to face the solution of his/her own case, which is invariably different from the examples provided.

On the contrary, following the suggestions of Lindley (2000), we rather aim at providing some indications to translate a real investigative case into a BN, and give less importance to the computational aspects, since efficient algorithms are freely available, i.e. `Hugin Lite` (`www.hugin.com`), and the `Matlab` package `BNets` by Kevin Murphy (`www.ai.mit.edu/~murphyk/Software/BNT/bnt.html`) that we customised to perform all the following calculations.

In this work, we focus on a single binary hypothesis H, such as '*is the suspect guilty?*', which takes on the values h_1 = guilty or h_2 = not guilty. Nevertheless, it is possible that more hypotheses are hierarchically related, as described by Garbolino & Taroni (2002).

In general, a hypothesis represents a state of nature, which is not observable, but influences probabilistically some of the other relevant variables, and is usually the main object of the inference. For this reason, in a BN a hypothesis node is a *root* of the graph, i.e. $\text{pa}(H) = \emptyset$.

3.1. **One single piece of evidence.** A crime is committed: a witness testifies to have seen an individual shooting a policeman during a robbery. Next, a suspect is individuated. A possible BN representation is depicted in Figure 1.

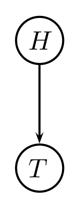

Figure 1: A testimony and the hypothesis of guilt

The variable H expresses the hypothesis of guilt. The variable T represents the witness testimony and its possible values are: $T = t_1$ in case the witness declares to recognise the suspect as the individual he has seen shooting; and: $T = t_2$ in case he does not. From now on, for the sake of simplicity, all the evidence is assumed binary, with the first state increasing the chance of guilt.

The investigator may define the CPT for the variable H, depending on the circumstances that led to the identification of the suspect. Alternatively, one can instrumentally set $\Pr(H = h_1) = \Pr(H = h_2) = 0.5$, so that the prior odds equal 1, and the weight of evidence is simply given by the posterior odds, directly available as a result from the evidence propagation algorithm in the BN.

As for the testimony, the 'flawless' witness should be associated with the CPT of Table 1a, i.e. his answer should be deterministically dependent on the true state of nature regarding H, without the possibility of mis-recognition.

More effectively, the witness may be tested by means of a psychological procedure to assess his capability to recognise a person, and the result can be used to build the CPT actually used for the node T, as in Table 1b.

	$H = h_1$	$H = h_2$
$T = t_1$	1	0
$T = t_2$	0	1

(a)

	$H = h_1$	$H = h_2$
$T = t_1$	0.9	0.3
$T = t_2$	0.1	0.7

(b)

Table 1: The CPT for the testimony T, given the hypothesis H. In case (a) the possibility of mis-recognition is not considered, whereas it is in case (b)

Notice that the figures in Table 1b are defined with specific reference to the witness of this case. In particular, when defining the CPT, the investigator should take into account any other information about the witness (i.e. his visual capacity, his criminal records, and so on).

Given the evidence $\mathbf{E}_1 = \{T = t_1\}$, i.e. that the witness claims he recognizes the suspect, it is straightforward to update the hypothesis of guilt as $\Pr(H = h_1 \,|\, \mathbf{E}_1) = 0.75$.

3.2. **More pieces of conditionally independent evidence.** Usually, the investigator is not satisfied with just one evidence, and is likely to look for other observable variables that can confirm (or disprove) its suggestions.

The most natural choice is to look for other variables directly influenced by H, but conditionally independent on the other evidence. A classical choice could be to question the suspect alibi.

Suppose, for instance, that the suspect declares that he was home watching TV with his wife, who is then interrogated.

The variable W in Figure 2 represents the woman testimony, and takes on the values w_1 in case she declares that her husband was not home by the beginning of the 6 o'clock news, exactly half an hour after that the crime was committed, and w_2 in case she declares that he was home by that time.

The graphical structure of Figure 2, known as *diverging connection*, encodes the relation $W \perp\!\!\!\perp T \,|\, H$, i.e. the distribution of the wife's statement is independent on the testimony, given that the value of the hypothesis H is actually known.

The CPT provided for the variable W from the investigator is shown in Table 2. Several experiments are performed to simulate the time needed to go from the crime scene to the suspect's house, and the possibility of doing so in less than half an hour is judged as very unlikely (only 2% of the simulated trials).

To assess the CPT in case $H = h_2$, the investigator also evaluates the journey from the suspect's work to his house. In the 80% of the simulated trials, the time needed would easily allow the suspect to be home by the news beginning. This result is used to build Table 2.

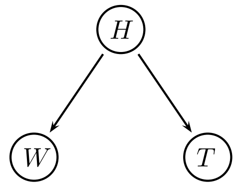

Figure 2: A working hypothesis H on a suspect's guilt, a witness testimony T, and the statements of the suspect's wife, W

	$H = h_1$	$H = h_2$
$W = w_1$	0.98	0.20
$W = w_2$	0.02	0.80

Table 2: The CPT for the suspect's wife testimony W, given the hypothesis H

The Bayes theorem can be directly used to infer about H, but even in this simple case, a specialized BN algorithm can be used to accomplish the calculations more straightforwardly.

The woman provides his husband with an alibi, i.e. $W = w_2$: the new available evidence is $\mathbf{E}_2 = \{T = t_1, W = w_2\}$, so that $\Pr(H = h_1|\mathbf{E}_2) = 0.0698$. Notice that the two testimonies have not the same impact on the hidden hypothesis, yet being on the same level within the network structure. This circumstance depends on the definition of the CPTs. In fact, the probability distribution of $W|H = h_1$ is almost degenerate, and the observation of the value $W = w_2$ almost makes impossible the hypothesis of guilt.

3.3. **Adding a control evidence.** Since the conflict between the two testimonies, the investigator needs to find other variables in order to check on them. To gain more understanding of the testimony T, the investigator seeks for a *control evidence*, i.e. a variable in the set $\mathrm{ch}(T) \cap \mathrm{ch}(H)$, which is probabilistically influenced by both the testimony and the hidden hypothesis.

The investigator notices a surveillance camera set at a cash dispenser just in front of the crime scene, and finds out that the CCTV video is available.

The original BN can be modified accordingly, to take into account this new variable, as in Figure 3, where the variable A is the observation of the ATM surveillance video. Notice that in this case, the presence of the direct link between T and A is such that these two nodes are not independent, even in case H was known.

Let the possible values for A be: a_1, if the suspect appears in the video, although it is not possible to detect any criminal action; and: a_2, if the suspect is not shown in the video.

The investigator assigns the probabilities of Table 3 to these events. In case that the suspect is actually innocent, then the fact that he is shown in the video becomes independent on the witness testimony. The probability of this event is set to a low value, since the ATM is not on the way from the suspect's work to his house.

Suppose that the person under investigation appears in the video. The BN updates the probability of guilt, given the evidence $\mathbf{E}_3 = \{T = t_1, W = w_2, A = a_1\}$

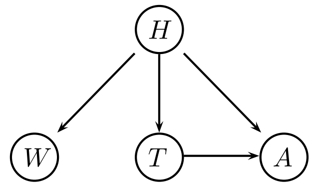

Figure 3: A working hypothesis H on a suspect's guilt, a witness testimony T, the wife statement W, and the video recorded by a CCTV of an ATM nearby the crime scene, A

	$H = h_1$		$H = h_2$	
	$T = t_1$	$T = t_2$	$T = t_1$	$T = t_2$
$A = a_1$	0.9	0.7	0.1	0.1
$A = a_2$	0.1	0.3	0.9	0.9

Table 3: The CPT for the variable A, given the testimony T and the hypothesis of guilt H

gathered by the investigator as $\Pr(H = h_1 \mid \mathbf{E}_3) = 0.4029$. This new evidence increases the posterior probability of guilt, although uncertainty remains on whether the suspect is actually the culprit of the crime. The two testimonies are in conflict, and the control evidence is not enough to explain away this contradiction.

Moreover, the suspect claims to have been framed, and that in fact, yet having cashed some money at that ATM, he is not the culprit of the crime, and the witness declared to recognise him only in order to make him considered guilty. How should the investigator handle this situation?

4. Handling manipulated evidence

4.1. **Modelling external interventions.** The case of external intervention on a variable within a stochastic system is one of the paradigms of causal inference (Holland 1986). Despite many scholars are still working with different approaches, a point of agreement is that causality mechanisms are mimicked by external interventions, which modify the natural behaviour of the stochastic system under study.

Two major contributions to the literature are those of Spirtes et al. (1993) and Pearl (1993), among the first to apply BNs to the study of causality. In order to do so, a new semantic is defined that takes into account the fact that one or more variables are subjected to intervention.

In case of intervention, any direct link between the intervened node and its parents has to be removed. If the link $H \rightarrow T$ is suggestive of a causal mechanism, there is no point in modifying H after that T is set to a given value, since the observation of $T = t$ is not attributable to that causal mechanism, but rather to the intervention.

If we make reference to the forensic case, this feature has relevant practical implications. If evidence is not genuine, removing the link with the hypothesis node avoids inappropriate inference, and the aim of the person who made the intervention is thwarted.

Conversely, the descendants remain dependent on T, either it arose naturally or by intervention. This circumstance has a special relevance when a descendant of the possibly manipulated node is also in the set $\mathrm{de}(H)$, and its origin is not under suspicion (see Figure 4).

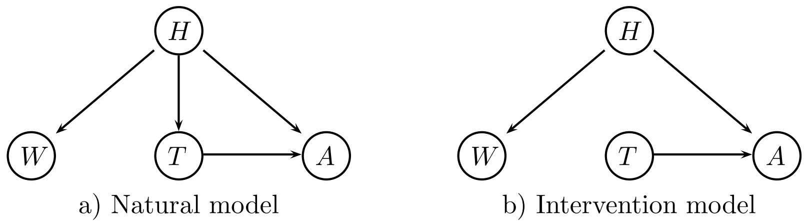

Figure 4: The DAG representation of the external intervention. In case the evidence arose by means of external manipulations, the direct connections between the intervened node T and its parent is removed. The rest of the graph is unchanged

Under the natural model, the observation of $T = t$ modifies the distribution of A both directly and through updating the distribution of the unobservable node H. Therefore, the most likely value of A is the one that is most consistent with the observed value of T, and the distribution of H induced by $T = t$.

However, if T did not arise genuinely, the distribution of A is only modified by T itself, as the distribution of H is not updated, since T and H are marginally independent in the intervention model.

Consequently, using the natural model when the evidence is not genuine assigns a higher probability to values of A that in fact are not as likely to occur. Therefore, the value of A that becomes available after observing T can be in conflict with the previous evidence, suggesting the possibility of manipulation.

For instance, when the genuine model of Figure 4a holds, the event $A = a_2$ becomes unlikely after observing $T = t_1$: $\Pr(A = a_2 \mid T = t_1) = 0.18$. On the contrary, using the intervention model of Figure 4b, the same value becomes much more plausible: $\Pr(A = a_2 \mid\mid T = t_1) = 0.50$. The operator $\mid\mid$ was introduced by Lauritzen (2000), to highlight the fact that the value of T is determined by an external intervention. The probability is calculated conditioning 'by intervention', rather than 'by observation'.

4.2. **Modelling interventions with Augmented DAGs.** Dawid (2002) proposed a unified representation of the two alternative regimes, using a decision theoretic approach based on an Augmented DAG (ADAG). This is a graphical model in which a possibly manipulated variable T is explicitly associated with an external *intervention* variable, F_T, which is used to rule its demeanour. This representation is depicted in Figure 5.

The possible external intervention is modelled as a *decision variable* F_T, represented as a square. Unlike a random node, F_T is not associated with a CPT, as its state is always decided by the experimenter. Therefore, it serves as a switch and it is used to allow the experimenter to activate a given scenario.

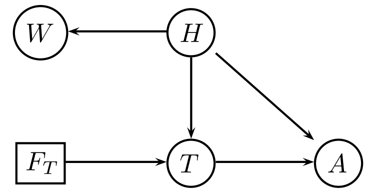

Figure 5: The ADAG representation of the intervention model

The variable F_T takes on the elements of the set $\{\mathcal{T} \cup \emptyset\}$, where $\mathcal{T}$ is the set of values that the possibly intervened node T may assume. When $F_T = \emptyset$, then the intervention is void, and hence T is a random variable governed by its conditional probability distribution $\Pr(T|\text{pa}(T))$.

Conversely, when $F_T = t, t \in \mathcal{T}$, then an intervention occurred. As a result, T becomes a degenerate variable, whence $\Pr(T = t|\text{pa}(T)) = 1$, for every configurations of the variables in $\text{pa}(T)$, so that the parents are not updated by $T = t$.

Back to the simple example of § 3.3, if the observed evidence was genuine, then F_T would be set to $\emptyset$, and the DAG representation implicit in the ADAG of Figure 5 would be the same as that depicted in Figure 4a.

On the contrary, should the investigator believe that the testimony is not genuine, then F_T would be set to the value t_1, and so would T. However, in this case, the knowledge of T should not update the CPT of its parent H. In other words, in case that $F_T \neq \emptyset$, the correspondent DAG is modified as in Figure 4b.

The use of the ADAG translates into a more compact representation of the problem, since both the situations are handled by the intervention node. The CPT of the variable T is then built as in Table 4 and comprises both the natural and the intervention cases.

	$F_T = \emptyset$		$F_T = t_1$		$F_T = t_2$	
	$H = h_1$	$H = h_2$	$H = h_1$	$H = h_2$	$H = h_1$	$H = h_2$
$T = t_1$	0.9	0.3	1	1	0	0
$T = t_2$	0.1	0.7	0	0	1	1

Table 4: The CPT of the possibly manipulated variable T. When the evidence is genuine, the CPT is that specified by the expert; in case of manipulation, the distribution of the variable becomes independent on the other parent, H, and degenerate to the value specified by F_T

Dawid's model has been originally used to deal with standard causal inference problems, where the objective is to estimate the effect of a 'treatment' T over a 'response' A, discarding all the factors, defined as 'potential confounders', which can generate spurious relations between them.

In the situation of Figure 5, a standard causal model would use the observations of the treatment T and of the confounder H to infer the desired causal effect on the unobserved response A. The graph of Figure 5 also encodes the assertions highlighted by Dawid (2002) that allow the identifiability of such causal effect:

$$F_T \perp\!\!\!\perp H, \quad (1)$$

$$A \perp\!\!\!\perp F_T | T, H. \quad (2)$$

As compared to the causal framework, the objective of our analysis is reversed, being to evaluate how the unobservable variable H is modified whether T is genuine or not, after observing the available evidence, including A.

The nature of assumption (1) is not modified, as it makes sense to assume that given that it is known whether the testimony is manipulated or not, no matter what the witness declares, the investigator's uncertainty over the hypothesis of guilt will remain the same.

Assumption (2) simply means that the knowledge of the actual value of H and of the testimony would be sufficient to guarantee that the control evidence A has a clear origin with respect to the testimony, being independent on F_T without the need of any further information.

Obviously, the actual value of H is hidden, and its probability evaluation is the objective of our analysis. However, assuming the validity of assumptions (1) and (2), the investigator takes the responsibility to ensure the absence of other unmeasured factors that can be connected to both A and T, which could confound the inference on H. This feature entitles the investigator to use A in order to check on T.

If the evidence is considered as genuine ($F_T = \emptyset$), the observation of both T and A updates the distribution of H, whereas in the intervention case only the control evidence can modify the distribution of the hypothesis of guilt.

For instance, after the suspect claim of having been framed, we suppose that the origin of the variable T is now unclear to the investigator. The evidence would be $\mathbf{E} = \{F_T = t_1, W = w_2, A = a_1\}$. Using a propagation algorithm on the ADAG of Figure 5, we obtain that $\Pr(H = h_1 \| \mathbf{E}) = 0.1837$, whereas, by definition, using the natural model for which the evidence is $\mathbf{E} = \{F_T = \emptyset, T = t_1, W = w_2, A = a_1\}$ the posterior probability of guilt would be $\Pr(H = h_1 | \mathbf{E}) = 0.4029$, the same inference described in § 3.3.

4.3. **Model assessment: the probabilistic evaluation of the intervention node.** Even more interesting is the possibility of evaluating probabilistically the two competing models:

- m_1: the unclear origin evidence T is in fact genuine;
- m_2: the unclear origin evidence T is manipulated,

conditionally on all the observed variables.

To this aim, it is necessary to define a further specialised version of the ADAG representation, as the one depicted in Figure 6. We term this graph *Model Assessment DAG* (MADAG), and we characterise the model node as a dashed circle. In this case, we define a new random variable M_T, which takes into account the two possibilities described above.

The model node is a root of the graphical representation. This assumption is useful to characterise it as an unobservable conjecture about the data generating process, whose uncertainty is updated by the evidence.

Just like F_T in the ADAG, M_T acts on the possibly intervened node so that the update of its parents is avoided, in case of manipulated evidence (i.e. when model m_2 holds). Yet, since the investigator is in doubt whether the observed evidence is

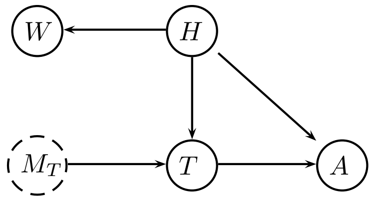

Figure 6: The Model Assessment DAG (MADAG) representation of the problem. The intervention node is now modelled as a random variable, rather than a decision node, in order to take into account the fact that a probabilistic evaluation is needed

genuine or invalidated by some intervention, T must remain a random variable in either case.

The use of the MADAG requires the following operations. First, the investigator must provide a prior CPT for the model node. Again, if the interest is in the evaluation of the weight of evidence with respect to the competing models, it is useful to choose $\Pr(M_T = m_1) = \Pr(M_T = m_2) = 0.5$. Obviously, in case that the investigator has different prior knowledge about the origin of the unclear evidence, it is easy to modify the CPT accordingly.

Second, while the CPTs under the natural model m_1 are provided by the expert, the ones for the intervention model are to be defined so that consistency between the two regimes is maintained. More specifically, the requirement is that, if the origin of a piece of evidence is under suspicion, the investigator cannot update his uncertainty on which is the model that generated the data, by means of the observation of that node only. Consequently, we require that:

$$\frac{\Pr(M_T = m_1 | T = t)}{\Pr(M_T = m_2 | T = t)} = \frac{\Pr(M_T = m_1)}{\Pr(M_T = m_2)},$$

which holds if:

$$\begin{aligned} \Pr(T = t | M_T = m_2) &= \Pr(T = t | M_T = m_1) \\ &= \mathrm{E}_{\mathbf{P}} \left[\Pr(T = t | \mathbf{P} = p, M_T = m_1) \right] \\ &= \sum_{p \in \mathcal{P}} \Pr(T = t | \mathbf{P} = p, M_T = m_1) \Pr(\mathbf{P} = p) \end{aligned} \tag{3}$$

for any $t \in \mathcal{T}$. The set $\mathbf{P} = \{\mathrm{pa}(T) \setminus M_T\}$ includes all the parents of T except the model node, and the average $\mathrm{E}_{\mathbf{P}}$ is taken over all the possible configurations of the variables in $\mathbf{P}$, indicated by the set $\mathcal{P}$.

The model m_2 can be seen as *nested* within m_1 (cfr. O'Hagan 1994): the latter includes the former, and they differ only in the fact that m_2 does not depend on the variables in the set $\mathbf{P}$, whereas m_1 does. By means of the requirement expressed by (3), they are marginally equivalent, marginalisation being over that set. This procedure of setting prior distributions for nested models is consistent with that discussed by Giudici (1996).

The definition of the CPT for the variable T essentially renders the testimony independent on the model node M_T, even if this property cannot be read off by the graph. Moreover, despite condition (2) is assumed to hold, since only T is known and H is unobservable, then $A \not\perp\!\!\!\perp M_T$, which allows to update also the uncertainty over the data generating process when the control evidence is made available.

In the example of § 3.3, $\mathbf{P} = \{H\}$ whence $\mathcal{P} = \{h_1, h_2\}$, and the distribution of T under the natural regime m_1 is that of Table 1b. Therefore, applying (3), the coherent CPT of the variable T is that shown in Table 5.

	$M_T = m_1$		$M_T = m_2$	
	$H = h_1$	$H = h_2$	$H = h_1$	$H = h_2$
$T = t_1$	0.9	0.3	0.6	0.6
$T = t_2$	0.1	0.7	0.4	0.4

Table 5: The CPT of the possibly manipulated variable T. When the evidence is genuine, the CPT is that specified by the expert; in case of manipulation, the distribution of the variable becomes independent on the other parent, H, and the two distributions are marginally equivalent, by definition

Given the observed evidence $\mathbf{E}_3 = \{T = t_1, W = w_2, A = a_1\}$, the probabilities for the unobservable variables are updated so that $\Pr(H = h_1 \,|\, \mathbf{E}_3) = 0.2728$ and $\Pr(M_T = m_1 | \mathbf{E}_3) = 0.4060$.

As compared to the inference obtained using the ADAG, the results derived here are subjected to an additional source of variability, i.e. that related to the model itself. The probability of guilt is a mixture of the natural and of the intervention case, with weights given by the posterior probability of the model node.

In general, it would be appropriate to check on the possibly manipulated node by means of several pieces of evidence. This situation could be easily handled by extending the MADAG of Figure 6 with other nodes whose origin is not under suspicion.

5. More complicated situations

5.1. **More than one manipulated pieces of evidence.** Let us now concentrate on the case where the investigator is not certain about the origin of more than one piece of evidence. Given the increasing complexity of the case and the growing conflict among the evidence, the investigator decides to regard the wife's testimony W as possibly manipulated as well.

In order to assess the latter testimony, the investigator also questions the doorman of the suspect's building to check whether he saw the suspect coming home that night. In Figure 7, this is represented by the variable D, assuming the possible values: d_1, in case that the doorman declares that, unlike his usual habits, the suspect just passed by and ran home; and: d_2, in case that the doorman declares that the suspect stopped by and chatted with him for about 10 minutes, like he usually does.

Since there are two unclear origin variables, there are also two model nodes M_T and M_W, which respectively rule the behaviour of the nodes T and W, just as described in the previous section.

While the CPT for the variable T as a function of its own model node M_T is that of Table 5, the distribution of the variable W, in the natural and in the intervention case, as derived by the application of condition (3) is shown in Table 6. As for the variable D, suppose that the CPT provided by the investigator is that of Table 7.

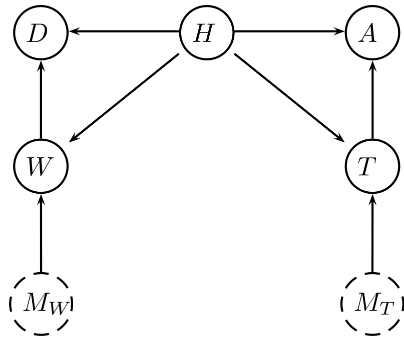

Figure 7: The case goes on: the investigator gathers more pieces of evidence, some of which have uncertain origin. For each possibly manipulated variable, the investigator defines a model node, and seeks for a suitable control evidence

	$M_W = m_1$		$M_W = m_2$	
	$H = h_1$	$H = h_2$	$H = h_1$	$H = h_2$
$W = w_1$	0.98	0.20	0.59	0.59
$W = w_2$	0.02	0.80	0.41	0.41

Table 6: The CPT of the possibly manipulated wife's testimony, under the genuine and the intervention model

	$H = h_1$		$H = h_2$	
	$W = w_1$	$W = w_2$	$W = w_1$	$W = w_2$
$D = d_1$	0.8	0.9	0.2	0.5
$D = d_2$	0.2	0.1	0.8	0.5

Table 7: The CPT for the variable D, representing the doorman's declaration

In case that the suspect is actually guilty, then it is quite unlikely that he had the time to stop by and chat with the doorman. If, in addition, his wife declares that he was home by the beginning of the 6 o'clock news, this event becomes even less likely. Conversely, if the suspect is innocent, depending on his wife statement, there is a higher probability that he actually stops by the doorman.

The structure of Figure 7 encodes the assumption that the observation of only one possibly manipulated evidence is not able to update the prior knowledge on the data generating model.

However, when both the unclear origin nodes are made available, the BN establishes an undirect connection between them, via the nodes H, D and A. Consequently, observing T and W does modify the probabilities of the model nodes, according to how consistent the two pieces of evidence are. Besides, the observation of the control evidence allows to produce a sharper update of the probabilities of the unobservable variables.

Suppose that the investigation leads to the following observed evidence $\mathbf{E}_4 = \{T = t_1, W = w_2, A = a_1, D = d_2\}$, i.e.:

- The witness testifies that he recognises the suspect as the man he saw on the crime scene;
- The suspect wife testifies that by the beginning of the news, she and her husband were watching TV together;
- The suspect is shown on the video recorded at the ATM CCTV, just in front of the crime scene;
- The doorman declares that the suspect stopped by as usual to have a little chat with him.

Using the BN of Figure 7, the investigator obtains that the posterior probability of guilt is $\Pr(H = h_1 \,|\, \mathbf{E}_4) = 0.5160$.

Moreover, it is possible to directly produce a joint evaluation of the unclear origin pieces of evidence, as shown in Table 8.

Data generating process (M)	Origin of T	Origin of W	Posterior probability given evidence $\mathbf{E}_4$
m_1	genuine	genuine	0.121
m_2	genuine	manipulated	0.350
m_3	manipulated	genuine	0.223
m_4	manipulated	manipulated	0.306

Table 8: Probabilistic evaluation of the data generating process, given the available evidence $\mathbf{E}_4$

The most likely model, given $\mathbf{E}_4$ is then $M = m_2$ indicating that the wife testimony is presumably not genuine. Should the investigator not take into account the possibility of observing manipulated pieces of evidence, starting from the same CPTs as for the genuine model, his inference would be that $\Pr(H = h_1 \,|\, \mathbf{E}_4, M = m_1) = 0.1189$. In fact, in this case, the testimony of the wife would be treated as genuine and, together with the doorman's testimony, it would tend to acquit the suspect.

However, the probability of some form of manipulation is around 0.90, as is depicted in Table 8, although the evidence is not sufficient to establish which variable has been intervened on. Therefore, the investigator cannot be satisfied with his comprehension of the case, and looks for other pieces of connected evidence.

5.2. **Evaluating two pieces of evidence by comparison.** We consider finally a very relevant situation that can occur during an investigation: the case where two different pieces of evidence need be compared, before the investigator can assess their actual relevance. This is a typical example involving 'scientific evidence' such as a blood trace left on a crime scene. A sample of the suspect DNA is analysed, but neither the crime sample, nor the suspect sample are relevant *per se*. Usually, if the two samples match, strong evidence against the suspect is found.

Even in this situation, the investigator may be in doubt that one of the two samples has been manipulated, and can easily specialise the BN representation of the investigations, in order to take into account properly of this possibility.

Considering again the previous example, suppose further that a gun is found in the suspect's house. The bullets found on the crime scene are then analysed and compared to a sample of bullets exploded by the suspect's gun, and the two types happen to be compatible.

Since the complexity of the case, the bullets found on the crime scene are supposed as possibly manipulated, to take into account of the possibility that a match between the bullets is not genuine. Moreover, the investigator gets the suspect tested with the paraffin glove method. A BN representation of this problem could be that depicted in Figure 8.

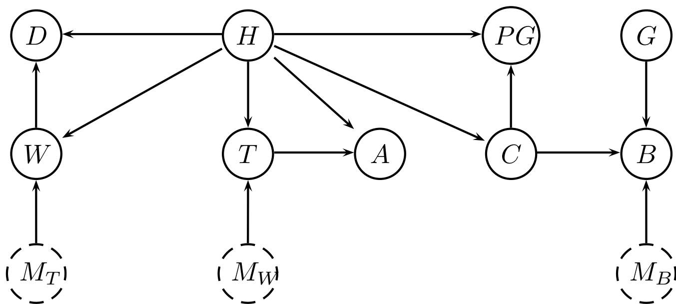

Figure 8: The MADAG for the comparison of two pieces of evidence. The nodes B and G are evaluated jointly, into the node C

The node G indicates the observed characteristics of the bullets exploded by the suspect's gun. The possible values are g_1 and g_2, where the former represents a relatively rare type, whose probability can be estimated by a suitable database as 0.1, and the latter are the remaining types, associated with a 0.9 probability of occurrence.

The bullets found on the crime scene are represented by the node B, whose possible values are b_1 and b_2. If $B = b_1$, then the bullets on the crime scene are compatible with the bullets exploded by the suspect's gun, whereas in case $B = b_2$ they are not.

In other words, the nodes B and G actually measure the same variable (the characteristics of the bullet), only on two different situations, and B is evaluated in comparison to G by means of the 'compatibility' node, C. If $c = c_1$, then there is a high probability of a compatibility between the two types of bullets analysed. Conversely, $C = c_2$ represents a mismatch, i.e. suggests that the two types of bullets do not come from the suspect's gun.

The variable PG indicates the result of the paraffin glove test, taking on the values pg_1, in case the test is positive, i.e. the suspect has shot within the last 5 days, and pg_2 in case the test is negative. For the sake of simplicity, we also assume that the test is very effective.

Suppose for example that the investigator defines the natural distribution of the variable B as in the left half of Table 9.

Under the natural model, if the bullet from the suspect's gun are of the type g_1, in order to claim a compatibility match ($C = c_1$), the bullet found on the crime scene must be of type b_1 with a high probability, say 0.9. Similarly, in order not to have a match ($C = c_2$), there must be a high probability, say 0.999, that the crime

	$M_B = m_1$				$M_B = m_2$			
	$G = g_1$		$G = g_2$		$G = g_1$		$G = g_2$	
	$C = c_1$	$C = c_2$	$C = c_1$	$C = c_2$	$C = c_1$	$C = c_2$	$C = c_1$	$C = c_2$
$B = b_1$	0.900	0.001	0.100	0.999	0.542	0.542	0.542	0.542
$B = b_2$	0.100	0.999	0.900	0.001	0.458	0.458	0.458	0.458

Table 9: The CPT for the variable B, given its parents G and C, for both the natural and intervention models

scene bullets come from a different gun (i.e. $B = b_2$). In case the bullets from the suspect's gun are of the much more common type g_2, a similar reasoning applies.

Consequently, we suppose that the test C is not error-free. However, it can be used reliably: in fact, the probabilities described above, possibly provided by the producer of the measurement devices, imply that its *sensitivity* (the probability that C suggests a match, given that B and G actually match) is around 90%. Similarly, from the values of Table 9 it is possible to calculate back its *specificity* (the probability that C suggests a non match, given that B and G are not compatible) as being around 99%.

Applying condition (3) it is straightforward to derive the intervention distribution, as in the right half of Table 9.

The variable C depends exclusively from the guilt hypothesis. Suppose that the investigator defines the CPT depicted in Table 10.

	$H = h_1$	$H = h_2$
$C = c_1$	0.995	0.001
$C = c_2$	0.005	0.999

Table 10: The compatibility match between the two types of bullets. The value c_1 indicates that the bullets found on the crime scene are compatible with those exploded by the suspect's gun

Even in case $C = c_1$, the investigator cannot be sure that the suspect is guilty, given to the nature of the observed evidence. For this reason, the CPT of C is not degenerate. However, other types of evidence, involving for example DNA measurements in a case of rape, could allow such a result.

Finally, the distribution of the control variable PG is that depicted in Table 11.

	$H = h_1$		$H = h_2$	
	$C = c_1$	$C = c_2$	$C = c_1$	$C = c_2$
$Pg = pg_1$	0.999	0.500	0.100	0.500
$Pg = pg_2$	0.001	0.500	0.900	0.500

Table 11: The CPT for the variable Pg, representing the result of the paraffin glove test

The suspect tests negative to the paraffin glove test, i.e. the evidence is $\mathbf{E}_5 = \{T = t_1, W = w_2, A = a_1, D = d_2, Pg = pg_2, G = g_1, B = b_1\}$. Allowing for the possibility of manipulation, the posterior probability of guilt decreases to just $\Pr(H = h_1 \mid \mathbf{E}_5) = 0.2224$. However, if the investigator considered all the evidence as genuine, the probability of guilt would be over 0.80, given that the bullets types match, which in turn would tend to incriminate the suspect.

As for the data generating process, the posterior results given the evidence $\mathbf{E}_5$ are shown in Table 12.

Data generating process (M)	Origin of T	Origin of W	Origin of B	Posterior probability given evidence $\mathbf{E}_5$
m_1	genuine	genuine	genuine	0.005
m_2	genuine	genuine	manipulated	0.173
m_3	genuine	manipulated	genuine	0.078
m_4	genuine	manipulated	manipulated	0.137
m_5	manipulated	genuine	genuine	0.004
m_6	manipulated	genuine	manipulated	0.343
m_7	manipulated	manipulated	genuine	0.053
m_8	manipulated	manipulated	manipulated	0.207

Table 12: Probabilistic evaluation of the data generating process, given the available evidence $\mathbf{E}_5$

Given the prior probabilities asserted by the investigator, and the pieces of evidence gathered, the most likely models is m_6 (T and B manipulated, and W genuine).

As compared to the previous findings, the introduction of the paraffin glove test and the analysis of the scientific evidence somehow explain away the inconsistencies in the testimonies. The wife's declaration becomes now more coherent, whereas the compatibility of the bullets is essentially due to an intervention.

5.3. **Synthesis of the case investigation.** Table 13 shows an overall summary of the case evaluation, upon varying the different information status. We consider the situation when the investigator is not aware of the possibility that one or more evidence is manipulated, and compare it to the models based on the suitable MADAG representations.

In general, one can appreciate how the inference is changed when the possibility of manipulation is taken into account. In particular, even when only a single piece of evidence is considered (case number 1), the possibility that it has been manipulated is such that the hypothesis of guilt loses strength, since the testimony does not appear to be completely reliable.

Considering the wife's testimony (case number 2), strongly reduces the posterior probability of guilt. This circumstance is certainly dependent on the construction of the CPT for W, and highlights the possible consequences of assigning (almost) degenerate prior probability distributions to some of the evidence.

When the control evidence for the witness testimony is made available, since it is consistent with the observed value of T (case number 3), the posterior probability of guilt increases. However, given the inconsistency between T and W, starting from the prior value of 0.5, the posterior probability that T is genuine is only 0.4061, leading the investigator to question it.

In fact, when we allow for possible manipulation over both T and W (case 4), the BN reacts suggesting that W is actually not genuine, as T and A accord.

Consequently, the posterior probability of guilt is increased up to 0.84 (a huge increase, as compared to the prior value of 0.5).

Moreover, the posterior probability that T is genuine becomes 0.5579, whereas for W it is only 0.1436 (again from the non informative prior value of 0.5). Obviously, since W has no control evidence at the moment, the model associates all the inconsistency with the possibility that it has been manipulated.

When the control evidence D is made available (case number 5), since it agrees with the evidence W, the posterior probability of guilt is again reduced. However, as reported earlier, the total probability that some external intervention has occurred is around 0.90, although it is not clear whether the two questioned variables are genuine or not - the posterior probabilities that T and W are genuine are, respectively, around 0.47 and 0.34.

In case number 6, finally, a scientific evidence is entered and controlled for. Without the possibility of manipulation, since the compatibility test C is positive, the suspect is associated with a probability of guilt of 0.8219. However, since the paraffin glove tends to acquit him, the intervention model responds assigning a posterior probability of guilt of only 0.2224.

Moreover, the external manipulations are identified as the testimony T and the type of bullets B. The posterior probabilities that these two variables are genuine are, respectively, 0.392 and 0.140, whereas for the testimony W it is 0.524, as compared to the common starting point of 0.5.

Table 14 depicts the posterior probability for the hypothesis that each single unclear variable is in fact genuine, given different status of knowledge. As appears clear, the evidence of W is likely to be not reliable, until the scientific evidence B is entered and its origin is questioned.

Evidence gathered	*Posterior probability for $H = h_1$*		
	Manipulable nodes	All genuine	Allow for manipulation
1. t_1	T	0.7500	0.6250
2. t_1, w_2	T	0.0698	0.0400
3. t_1, w_2, a_1	T	0.4029	0.2728
4. t_1, w_2, a_1	T, W	0.4029	0.8420
5. t_1, w_2, a_1, d_2	T, W	0.1189	0.5160
6. $t_1, w_2, a_1, d_2, pg_2, b_1, g_1$	T, W, B	0.8219	0.2224

Table 13: The posterior probability of guilt calculated using the natural models (without the possibility of accounting for not genuine evidence), and the MADAG models, upon varying the manipulable nodes. The prior probability distribution is $\Pr(H = h_1) = \Pr(H = h_2) = 0.5$

6. Discussion

In this paper we showed a methodology to deal with unclear origin variables, within an investigation case. This possibility is ensured by the BN structure that we associated with the problem.

Evidence gathered	Manipulable nodes	*Posterior probability for*		
		$M_T = m_1$	$M_W = m_1$	$M_B = m_1$
1. t_1	T	0.5000	–	–
2. t_1, w_2	T	0.3440	–	–
3. t_1, w_2, a_1	T	0.4061	–	–
4. t_1, w_2, a_1	T, W	0.5286	0.1429	–
5. t_1, w_2, a_1, d_2	T, W	0.4709	0.3440	–
6. $t_1, w_2, a_1, d_2, pg_2, b_1, g_1$	T, W, B	0.3927	0.5244	0.1401

Table 14: The posterior probability for the model nodes, upon varying the informative status. The prior probabilities are set to a non informative distribution: $\Pr(M_T = m_1) = \Pr(M_W = m_1) = \Pr(M_B = m_1) = 0.5$

The very first advantage in using a BN is the fact that the overall judgement on the working hypothesis is articulated into each single relation among the variables.

At first sight, this could be perceived as a drawback, as the investigator may reckon that he is not able to provide a detailed assessment. However, in our opinion, once this task has been performed, the investigator is rewarded, in that this only makes the evaluation of the whole pieces of evidence more straightforward. Besides, it can become clear which variables need be investigated more thoroughly, before a sharpest opinion can be reached.

A second important feature of this method is that it explicitly models the presence of conflicting evidence. Some works in the statistical literature have focused on this matter (Jensen et al. 1991, Jensen 1995), defining a diagnostic statistics, which is able to detect possible conflicts between different pieces of evidence.

The model proposed in this paper provides an alternative way of quantifying inconsistencies in the observed variables, as it directly calculates the posterior distributions for both the working hypotheses and the data generating models.

Third, unlike most standard Bayesian analysis of forensic data, the use of the weight of evidence is not particularly useful in the framework we presented here.

In fact, the most important feature of this measure is its invariance to the choice of the prior distribution for the hypothesis of interest. Since in our case, the working hypothesis is evaluated jointly with the model variable, the weight of evidence changes with the choice of the prior distributions for the model node. For this reason, the evaluation of the posterior distribution of the unobservable variables can be more relevant.

Finally, in our opinion, the framework described in this paper has several possible generalisations to other research areas. The role of evidence becomes absolutely central to our modelling, as the underlying assumptions can be explicitly questioned and validated through both the unclear and the control evidence. This feature could be useful also in areas such as Economics, Physics, Psychology or Medicine.

References

Aitken, C. & Taroni, F. (2004), *Statistics and the valuation of evidence for forensic scientist*, John Wiley and Sons, Chichester, UK.

Cowell, R., Dawid, A., Lauritzen, S. & Spiegelhalter, D. (1999), *Probabilistic Networks and Expert Systems*, Springer-Verlag, New York, NY.

Dawid, A. (1979), 'Conditional independence in statistical theory (with Discussion)', *Journal of the Royal Statistical Society,* B **41**, 1–31.

Dawid, A. (2000), 'Causal Inference without Counterfactuals (with Discussion)', *Journal of the American Statistical Association* **95**, 407–448.

Dawid, A. (2002), 'Influence Diagrams for causal modelling and inference', *Statistical Review* **70**, 161–189.

Dawid, A. & Evett, I. (1997), 'Using a graphical method to assist the evaluation of complicated patterns of evidence', *Journal of Forensic Sciences* **42**, 226–231.

Garbolino, P. & Taroni, F. (2002), 'Evaluation of scientific evidence using Bayesian networks', *Forensic Science International* **125**, 149–155.

Giudici, P. (1996), Learning in Graphical Gaussian Models, *in* J. Bernardo, J. Berger, A. Dawid & A. Smith, eds, 'Bayesian Statistics 5', Oxford University Press, Oxford, UK, pp. 621–628.

Holland, P. (1986), 'Statistics and Causal Inference', *Journal of the American Statistical Association* **81**, 945–968.

Jensen, F. (1995), Cautious Propagation in Bayesian Networks, *in* P. Besnard & S. Hanks, eds, 'Proceeding of the Eleventh Conference on Uncertainty in Artificial Intelligence', Morgan Kaufmann Publishers, San Francisco, CA, pp. 519–528.

Jensen, F. (2001), *Bayesian Networks and Decision Graphs*, Springer, New York, NY.

Jensen, F., Chamberlain, B., Nordahl, T. & Jensen, F. (1991), Analysis in Hugin of data conflict, *in* P. Bonissone, M. Henrion, L. Kanal & J. Lemmer, eds, 'Uncertainty in Artificial Intelligence 6', Elsevier Science Publishers, Amsterdam, Netherlands, pp. 323–329.

Lauritzen, S. (2000), Causal Inference from Graphical Models, *in* O. Barndorff-Nielsen, R. Cox & C. Klüppelberg, eds, 'Complex Stochastic Systems', Chapman and Hall, London, UK.

Lindley, D. (2000), 'The philosophy of Statistics', *The Statistician* **49**, 293–337.

O'Hagan, A. (1994), *Bayesian Inference, volume 2B Kendall's Advanced Theory of Statistics*, Arnold, London, UK.

Pearl, J. (1993), 'Comment: Graphical models, causality, and intervention', *Statistical Science* **8**, 266–269.

Pearl, J. (2000), *Causality*, Cambridge University Press, Cambridge, UK.

Spiegelhalter, D., Dawid, A., Lauritzen, S. & Cowell, R. (1993), 'Bayesian Analysis in Expert Systems', *Statistical Science* **8**, 219–283.

Spirtes, P., Glymour, C. & Scheines, R. (1993), *Causality, Prediction and Search*, Springer-Verlag, New York, NY.