Modelling a Murder Case in Probability Theory

René Mellema, r.mellema.1@student.rug.nl
Supervisors: Bart Verheij *

Abstract

Reasoning with evidence is an important task that most humans engage in on a daily basis. It is especially important in criminal trials, because the lives of humans are at stake here. For this kind of reasoning, three approaches of modelling evidential reasoning are used: the argumentative, narrative and probabilistic approaches. A new method for integrating these three approaches was proposed by Verheij (2014). In this paper, the method is tested using a case study of the Anjum murders, a Dutch murder case. It is also compared to a Bayesian approach of Vlek et al. (2014) for evidential reasoning. The new method could model the case comparatively to Vlek et al. (2014).

1 Introduction

Reasoning about the world is an activity that most humans take part in on a daily basis. This is most certainly the case for detectives, lawyers and judges, whose goal it is to draw conclusions about events from evidence in criminal trials. For reasoning about evidence, three different normative frameworks can be used. These are the argumentative, narrative and probabilistic frameworks (Kaptein et al., 2009; Dawid et al., 2011b; Verheij et al., to appear).

The *argumentative* approach is based on arguments that reason from the evidence to the conclusions. For example, a prosecutor can use this view to reason from the evidence, like DNA evidence, to the conclusion that someone committed a crime, building up arguments for and against his case as he goes along. For a long time, this has been done using inference charts, such as those used by Wigmore (1931).

The *narrative*, or story-based, approach is based on scenarios that tell a story about what might have happened. According to Pennington and Hastie (1993), this is also the approach that most jurors use when reasoning about a case. These stories incorporate the evidence into a coherent whole that most people can follow and reason about, instead of the evidence presented before the court. The stories can be built out of smaller stories concerning evidence.

The last approach is the *probabilistic* approach. It is based around probabilities for certain events. This is an approach that is mostly used in forensic science (Aitken, 1995) to describe probabilities and odds. This approach is also used for the construction of Bayesian Networks (Dawid et al., 2011a) for reasoning about evidence.

All of these approaches have their own strengths and weaknesses. For example, the argumentative approach only deals with establishing truth, without gradations between true and not true, which the probabilistic approach has. On the other hand, the probabilistic approach is hard to apply to certain pieces of evidence, like a testimony of a witness, where the narrative approach can be effectively used to capture the global coherence and integrate it with evidence. The narrative approach has the problem that it has the tendency to favour a good story over a true story (Pennington and Hastie, 1993). This is a problem the argumentative approach has solved using argumentation schemes, which can be used to determine if an argument has all the necessary elements.

Because of this, different attempts have been made to create combined methods, utilising parts of the different approaches. Examples of these are Vlek et al. (2014), which combines narrative methods and Bayesian networks, and Bex et al. (2007) which combines the narrative and argumentative

*University of Groningen, Department of Artificial Intelligence

approaches. Verheij (2014) gives a new perspective for integrating all three approaches that is based upon standard probability theory. In the article this perspective is illustrated using an example robbery case, but it is not tested with a real case. In this article, the perspective will be evaluated using a case study of the Anjum case.

Because the method combines the multiple frameworks, it is expected that it has the following properties:

1. The method prevents tunnel vision.

   Tunnel vision is often a problem where there is only focus for one possible explanation. By explicitly allowing for multiple scenarios, it is expected that the method will stop tunnel vision.

2. The method prefers a true story over a good story.

   Sometimes a good, or well constructed, story is being chosen as the truth over a true story (Pennington and Hastie, 1993). Because the method requires the modeller to make all his assumptions explicit, it is expected that the true story will be chosen instead of the best written story.

3. The method is not dependant on subjective probabilities while keeping the gradation of degrees of uncertainty from probability theory.

   A large problem for Bayesian Network approaches is their dependence on unknown probabilities. By using relative probabilities instead of absolute probabilities we hope to circumvent this problem.

4. The method does not allow for stories that disagree with evidence.

   A story that disagrees with evidence can never be true, so it would be preferred if the system would alert us if this is the case.

On these points, the method will be compared against the method described in Vlek et al. (2014), who used the same case to test their method.

Next, the Anjum case will be introduced. Then the methods from Vlek et al. (2014) and Verheij (2014) will be explained. This will be followed by the model of the case study. Then the perspective will be evaluated and compared against Vlek et al. (2014). The article will end with a conclusion.

2 The Anjum case

In this section the murder case will be introduced. This case is the murder of Leo de Jager, which is part of the so-called 'Anjum murders'. Marjan van der E., owner of the boarding house where Leo was killed, was convicted for the murder. Legal scholars from the 'Project Gerede Twijfel' (Project Reasonable Doubt) [1] reinvestigated the case. This project is an innocence project conducted by scholars from the VU Amsterdam and the University of Maastricht.

The primary source for the information for this case study is Vlek et al. (2014), using Crombag and Israëls (2008) as a second source. This will mean that our results are influenced by the ideas presented in these papers. Since our aim is to model a complex case with this new approach and not to look at the case objectively, this should pose no problem.

Since Vlek et al. (2014) only models two of the four scenarios presented by Crombag and Israëls (2008), we also focus on those two scenarios. We will also use the same fictional names for most persons involved as Vlek et al. (2014) and Crombag and Israëls (2008). Only the real name of the prime suspect, Marjan van der E., is used.

2.1 Introduction

On the evening of the 24th of December, 1997, Evert Beekman, together with Jaap Kuilstra, came into the police station to report a murder. Beekman said that he had seen the dead body of Leo de Jager on the property of Marjan van der E.. Beekman also helped the police dig up the remains of the body Herre Sturmans on the same terrain. We will only focus on the first murder.

2.2 The persons involved

Marjan van der E. was a proprietor of a boarding house in Anjum. Leo lived in a small house of Marjan in Moddergat and helped her with chores around the boarding house. Beekman sold timber

[1] `www.projectgeredetwijfel.nl` (in Dutch)

in Anjum and he knew Marjan because she had placed orders with him in the past. Some other persons of interest are Marga Waanders, Eef Tasman and Jaap Kuilstra. Waanders was a friend of Marjan who came over to take care of the boarding house over Christmas and New Years. Tasman used to do some administrative work for Marjan. Finally, Kuilstra was a friend of Beekman whom Beekman told about the murder, after which Kuilstra advised him to go to the police.

Marjan, Beekman and Kuilstra also had a cannabis farm on the attic of Marjan's barn. This farm had been exposed by the police a week prior, and Marjan was a suspect. However, she had claimed to have rented the barn to someone else and had promised that she would show the police the contract of this agreement.

2.3 The order of events

Beekman reported that Marjan had told him that she had killed someone around 7 pm. She returned to the boarding house and Beekman followed her soon after. There he met Waanders and talked to her for a few minutes, while Marjan cleaned the hallway; Leo was said to have puked in there and they were not allowed to see. Beekman later went into the hallway and found Marjan scrubbing the floor. He saw a trail of blood in the hallway which he suspected to be from the back of a head as a body was dragged to the front door.

Marjan took Beekman to the front door, where he saw a dead body laying outside, underneath a tent canvas. He recognised this body as the body of Leo de Jager and noticed that the head had been bashed in at six or seven different places. After that Beekman went home.

Later on, Beekman said that he had helped Marjan wrap the body in a piece of canvas, and that he went back to the boarding house at 2 am to help Marjan drag the body to the front yard. According to Kuilstra, Beekman should not have told this part to the police, so Beekman omitted this part the first time.

First Marjan and Waanders were arrested and interrogated. At first, Marjan seemed too confused to say anything of value. Even later on, she kept claiming that she had not killed or drugged Leo. Waanders, on the other hand, did give information about what happened. Leo was at the boarding house when she arrived in the afternoon, and she last saw him at 6 pm while he was talking to Marjan. Marjan was trying to convince Leo to sleep in one of the rooms of the boarding house.

On later days, Waanders gave even more information. When she saw Leo on the afternoon of the 23th of December, he seemed under the influence of something. She also mentioned she found him in the barn at some point after which she took him back to the house. This agitated Marjan, who gave him a glass of warm water and jenever (Dutch gin), which she called a grog.

When Waanders called Marjan for dinner, she did not show up. When Waanders went into the hallway to see what Marjan was doing, she saw that Marjan was giving Leo another grog. She also claimed to have 'images' of Marjan hitting Leo, but she was not sure whether she really saw this. Waanders did explicitly state that she did not see any blood stains. She said that she did remember Beekman coming over. She also remembered two shadows outside, possibly Marjan and Beekman. Later in the evening, Marjan and Waanders went for a walk with their dogs and they had a drink together.

2.4 The evidence

Statements made by Beekman, Waanders, Marjan and Kuilstra are all used as evidence in this case. Their main points can be found in the description above. When additional information is of interest for the model, it will be introduced as soon as it is relevant. The rest of this section will be key evidence that is not a testimony.

When the police investigated the boarding house on the 25th of December, they found blood stains in multiple places, most of which were in the hallway. They also found a wad with a bloody knot of hair in the trash can in the room of Waanders. Also, two hammers were found in the barn, a large one and a smaller one, with watery bloodstains on them. Both bloodstains had a DNA match with Leo, but the profiles drawn from both were incomplete. There was a probability of less then one in a million that the blood in the hallway came from a random other person than Leo. For the blood on the hammer, this probability was 1 in 100 for the hammer hand and 1 in 1700 for the hammer handle. This latter estimation has been disputed by

another expert.

Three empty strips of Temazepam, a drug used to treat insomnia, were found in the trash in the kitchen, as well as a strip which had ten empty capsules that were cut open. More empty capsules were found in the trash as well as a medicine bottle in the name of Leo de Jager. During an autopsy of Leo's body, high concentration of both Temazepam and alcohol were found in Leo's blood. The Temazepam levels were far more than the advised amount for daily use and even far above the toxicity level. Although Temazepam is not lethal it will cause strange behaviour in the patient. A pathologist concluded that the cause of death was a heavy blow to the head, which caused a fractured skull.

Another part of the evidence concerned a bank fraud in which Marjan was supposedly involved. For this fraud she used Leo's account, which gave her an additional motive. Since Vlek et al. (2014) leaves this out, we will as well, to make it easier to compare the two methods.

2.5 The two scenarios

In Crombag and Israëls (2008), four scenarios are proposed. We will only focus on their second and third scenario. The first of these two scenarios has Marjan as the culprit with Beekman as an accomplice, while the second one has the roles reversed. The view of the judiciary mostly agrees with the first of these two scenarios.

The first scenario matches the order of events above. Marjan drugged Leo, got him to sign the forms and killed him around 7 'o clock. Then she went to get Beekman, who helped her later in the evening to drag the body into the front yard.

In the second scenario, Marjan also drugged Leo to get him to sign the forms, but she did not kill him. In this scenario, Beekman killed Leo when Marjan and Waanders walked their dogs around midnight and then also dragged his body into the front yard of the boarding house to hide it there. He left the body in this visible spot because the women returned after their walk.

The conclusion from Crombag and Israëls is that the murder was committed by Beekman with help from Marjan, so they believe that the second scenario is true. They reach this conclusion because the other theories they proposed, one of which is based on the judiciaries' view, all have more problems that cannot be explained. For example, it is unlikely that Marjan would have killed Leo within meters distance of Waanders who said to have noticed nothing.

3 The methods

For the following sections, an example will be used to explain the two methods in a concrete way. These examples are based on the following example case:

Someone's house was broken into while she was asleep. When she went to check what was going on, she was confronted by a man with a pistol who shot her in her leg before running away.

The police had three possible suspects, who all had their houses searched. In the house of the first and third suspects guns were found. After a ballistic rapport had been made, it was found that the gun of the first suspect matched the bullet found in the leg of the owner of the house. After this the first suspect confessed and told the police the location of the stolen items.

3.1 The method from Vlek et al. (2014)

The method proposed in Vlek et al. (2014) utilises narratives to build Bayesian networks. A primary goal of this method is to capture the global coherence of a narrative into a Bayesian network.

A Bayesian network is a directed acyclic graph that represents the dependencies among variables in a domain (see Russell and Norvig, 2010, chapter 14 for more information). This is normally done by directing nodes in a causal direction. This allows for a representation of a full joint probability distribution but more concisely. Each node is annotated with a conditional probability distribution.

According to Fenton et al. (2013), Bayesian networks are well suited to model legal arguments. This is because they allow us to visualise the causal relationships between between hypotheses and different pieces of evidence, which is especially useful in a complex argument. A Bayesian network is also a powerful tool because of its underlying calculus that can be used to determine the probability of

uncertain events on the basis of evidence (Fenton et al., 2013).

In order to make the construction of these networks for legal reasoning easier, Fenton et al. (2013) proposed a set of often recurring patterns, which they called idioms, to be used in the construction of Bayesian networks. In Vlek et al. (2014), more idioms are proposed that are meant to capture the global coherence of a narrative into a Bayesian network. The following description is adapted from Vlek et al. (2014).

The first of these idioms is the *scenario idiom*. This idiom is used to represent a scenario in the Bayesian network. Its goal is to capture the global coherence of a narrative. This idiom works by creating a scenario node that has a connection to all of the nodes that are part of this scenario. Then the probabilities of the child nodes are set in such a way that if the scenario node is true, all of its elements should also be true.

The second idiom is the *subscenario idiom*. This is used to represent sub elements of the narrative. The idiom is similar to the scenario idiom, but it is always part of an encompassing scenario. The probability table of a subscenario should always be the same as the probability table of the node it replaces.

The third idiom is the *variation idiom*. This is used to model cases where there are different possibilities but it does not matter which of the possibilities is true. For example, we know that Leo was killed, but we do not know the murder weapon. In that case the variation idiom could be used to model multiple murder weapons without influencing the outcome of the model.

The fourth and last idiom is the *merged scenario idiom*. This idiom explains how you can combine multiple scenario idioms into one network. This is done by creating a new node that makes sure that no more and no less than one scenario is true. It also has rules for combining the scenarios if they share one or more nodes.

The design method further consists of a method for unfolding nodes when desired. When a node is unfolded, the old node is replaced by an instance of the subscenario idiom. This subscenario can then contain more information, an instance of the variation idiom or nodes with direct links to evidence.

If we were to apply this methodology to our example, we would have three instances of the scenario idiom, one instance per hypothesis. The instance for the first hypothesis will be explained, the instances for the other hypotheses are similar. The full network can be found in Figure 2. The network was built using the tool GeNie 2.0[2].

Initially, the network would consist of one scenario node, 'Suspect 1 is robber', which would have a connection to all nodes that are part of the scenario. This model can be seen in Figure 1.

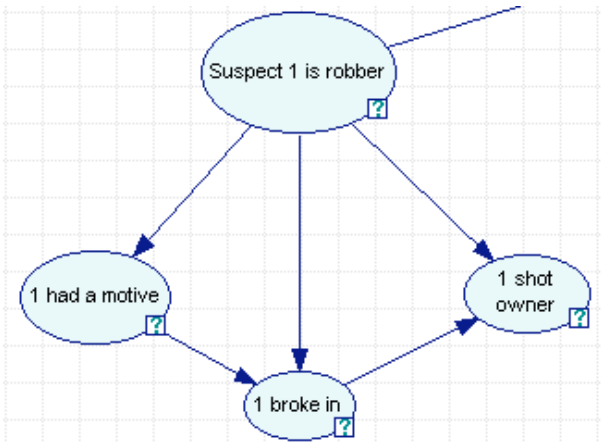

Figure 1: The initial first scenario

Now we can see which nodes need unfolding. We should also unfold the '1 shot owner' node, since we have evidence supporting subparts of this. Therefore we add a node '1 had gun' and '1's gun was used'.

For the other suspects, one goes through a similar process, and then all the scenarios are merged into one. This would be done by adding a constraint node that says that only one scenario can and has to be true, which would have connections from all scenario nodes. All the nodes that occur in multiple scenarios should be part of the network only once. The final network can be found in Figure 2.

To complete the model, we had to estimate probabilities that were not known, like the probability that Suspect 2 broke into the house while he was not the culprit and did not have a motive. Probabilities like these are necessary for the network, but nearly impossible to give a good estimation of.

[2]This tool can be downloaded at `https://dslpitt.org/genie/`

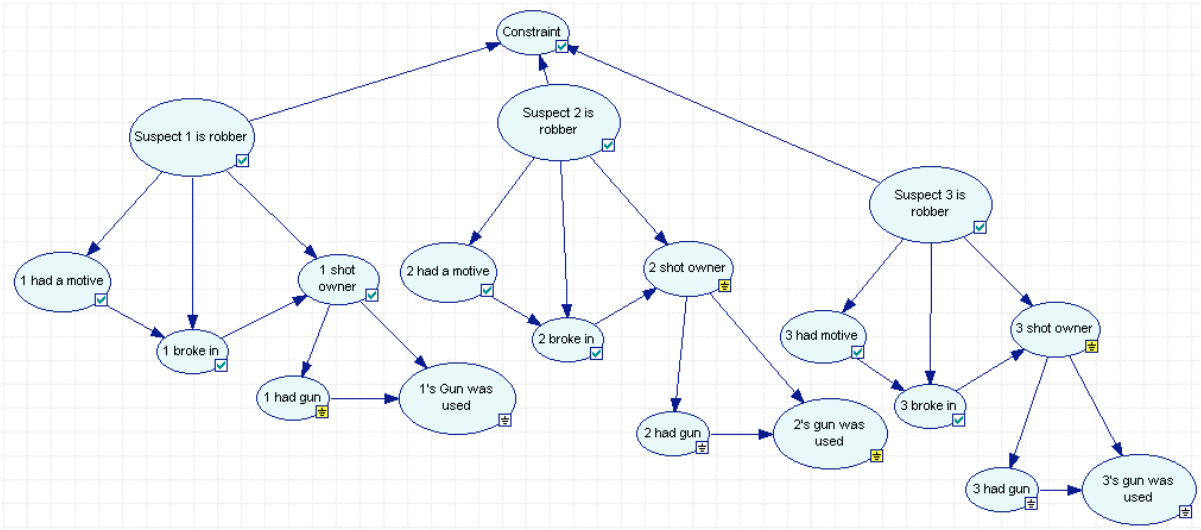

Figure 2: The full example network

3.2 Model of Anjum case

In the network developed to model the Anjum case, we find that the method from Vlek et al. (2014) helps against a good story pushing out a true story and it helps as a heuristic for finding the relevant variables for a Bayesian network. It does however require subjective probabilities, and specifying subjective probabilities leads to the illusion that these numbers are known.

What works especially well for this case is that the evidence can be added piece by piece, so we can see how the probabilities change for this case. We can see that at first Marjan is more likely to have committed the murder, but that at the end the probabilities flip in her favour and that it is more likely that Beekman killed Leo. These probabilities are however subjective. Over all, we find that the method from Vlek et al. (2014) models the Anjum case well.

3.3 The method from Verheij (2014)

Since it only uses classical logic and probability theory, the formal background for the method of Verheij (2014) does not go beyond introductory textbooks on logic and probability theory.

In the method, scenarios are modelled as alternate hypotheses, which are mutually exclusive. The scenarios are supported by arguments, which have strengths. These strengths are measured by numbers that behave like conditional probabilities. The strength for all arguments does not have to be known, since it is impossible to get a reasonable strength for all arguments.

The method uses a logical language L, which has binary connectives $\wedge$ for logical conjunction, and $\vee$ for logical disjunction. It also has the unary connective $\neg$ for logical negation. The set of sentences L is then constructed inductively from the non-empty set of propositional Boolean variables P. A *reason* is a pair of sentences, (φ, ψ), where $\varphi \in L$ and $\psi \in L$. The *strength* of a reason is measured as the conditional probability $\mathrm{P}(\psi|\varphi)$. Here P is a function that obeys the properties of a standard probability function. Therefore, $\mathrm{P}(\psi|\varphi)$ is only defined when $\mathrm{P}(\varphi) > 0$.

When a reason has a strength of 1, it is *conclusive*. When a reason is conclusive, the conclusions ψ are certain given the premises φ. When the strength of a reason is positive, but smaller than 1, the reason is *defeasible*. A reason (φ, ψ) can become weaker or stronger if the reason's premises are extended to $(\varphi \wedge \chi, \psi)$, where $\chi \in L$. If the reason is defeasible, the reason's strength can be weakened to defeat, which means that $\mathrm{P}(\psi|\varphi \wedge \chi) = 0$. Each defeasible reason also has an associated defeasible reason $(\varphi, \neg\psi)$ for the opposite conclusion. The combined strength of these reasons is 1. Therefore, if one is weak, the other is strong.

Once we have a statement of the kind $\mathrm{P}(H_i|E_1 \wedge E_2 \wedge \ldots \wedge E_n) = 1$ and no new relevant evidence can be found it can be said that the case is solved beyond a reasonable doubt.

In the example, we have three different scenarios, H_1, H_2 and H_3, one for each of the suspects. We also have the following list of evidence:

E_1 The fact that there was a burglary

E_2 Gun found at Suspect 1

E_3 Gun found at Suspect 3

E_4 Ballistic rapport

E_5 Suspect 1's testimony

We write the argument that the first suspect is the burglar given the fact that there was a burglary as (E_1, H_1). Its strength is written as $\mathrm{P}(H_1|E_1)$. There is no reason to assume the guilt of one suspect over another, so we say that the relationship between the arguments is undetermined.

We also have the expectation that the burglar has a gun (G). After the house search, two guns were found, at the houses of Suspect 1 (E_2) and Suspect 3 (E_3). This means that the relation between the arguments for these suspects is still undefined. We can however assume that it is less likely that suspect 2 committed the crime, so $\mathrm{P}(H_2 \wedge G|E_1 \wedge E_2 \wedge E_3) < \mathrm{P}(H_i \wedge G|E_1 \wedge E_2 \wedge E_3)$ for $i = 1$ or 3.

If we also add the ballistic rapport (E_4), we now have a reason to assume that suspect 1 did it. We write this as $\mathrm{P}(H_1 \wedge G|E_1 \ldots E_4) > \mathrm{P}(H_i \wedge G|E_1 \ldots E_4)$ for $i = 2$ or 3. If we now add the testimony, we can assume a conclusive argument for H_1. We would write this as $\mathrm{P}(H_1 \wedge G|E_1 \ldots E_5) = 1$. This means we are convinced that suspect 1 is the burglar.

3.3.1 Overview

In the following section φ and ψ can stand for complex expressions including conjunctions and disjunctions.

$\mathrm{P}(\psi\|\varphi)$	The strength of the argument of ψ supported by φ.
$\mathrm{P}(\psi) = 0$	ψ is defeated.
$\mathrm{P}(\psi) = 1$	ψ is conclusive.
$\mathrm{P}(\psi\|\varphi \wedge \chi) > \mathrm{P}(\psi\|\varphi)$	The argument $(\varphi \wedge \chi, \psi)$ is stronger than the argument (φ, ψ).

4 Case Study

Now we will take a look at the case from the perspective of the method from Verheij. First we will define the two scenarios and then we will enumerate the evidence and the expectations.

As before, we will only be looking at the scenario where Marjan van der E. killed Leo on her own and the scenario where Beekman killed Leo with help from Marjan. We will call these scenarios H_1 and H_2 respectively.

A piece of evidence is written as E_i. There is no distinction between a testimony and other types of evidence. A complete table of evidence can be found in Table 1a. The expectations are given a one letter name. A list of all the expectations can be found in Table 1b.

The first thing we assume is that the hypotheses cannot both be true at the same time. We write this as $\mathrm{P}(H_1 \wedge H_2) = 0$. We also assume that both of the hypotheses can be true: $\mathrm{P}(H_1) > 0$ and $\mathrm{P}(H_2) > 0$, meaning that the strength of those arguments is larger than 0. This means that both hypotheses are possible. We also assume that one of these hypotheses is true: $\mathrm{P}(H_1 \vee H_2) = 1$. This is similar to the merged scenario idiom from Vlek et al. (2014), since we are preparing to compare the two scenario's. We do this at the start instead of at the end however, since we will want to compare subparts of the scenarios with each other.

We will start with supporting some of the expectations. We can support the assumption that Leo had a high concentration of alcohol and Temazepam in his blood (A) with the toxicological report (E_{16}) and the empty strips of Temazepam (E_{14}). This makes the argument stronger: $\mathrm{P}(A|E_{16} \wedge E_{14}) > \mathrm{P}(A)$.

We can also use an expectation to support another expectation, by including its premises and

Table 1: A list with all the relevant identifiers

(a) List of Evidence

E_1	Beekman's testimony
E_2	Waanders' testimony
E_3	Marjan's testimony
E_4	Kuilstra's testimony
E_5	Pier's testimony
E_6	Leo's body in the front yard
E_7	The prepared hole in the front yard
E_8	Tracts of dragged feet in gravel
E_9	Bloodstains found in hallway
E_{10}	DNA match with bloodstains and Leo
E_{11}	Bloody hair knot in Waanders' trashcan
E_{12}	Large hammer in barn
E_{13}	DNA match on large hammer
E_{14}	Empty strips of Temazepam
E_{15}	Autopsy report
E_{16}	Toxicological report
E_{17}	Marijuana farm in Marjans barn
E_{18}	Leo's signature and false contracts

(b) List of Expectations

A	Leo had a high concentration of Temazepam and alcohol in Leo's blood
B	The blood trail in the hallway and blood on the steps in front of the house
C	Leo died because of a heavy blow to the head
D	Marjan drugged Leo
G	Leo asked Pier to get him around 20:00 – 20:30
H	Leo was killed with a hammer
M	Leo was a front for the Marijuana farm
N	Marjan and Waanders walked the dogs around midnight
S	Leo was killed with a stone
T	Leo was dead before 18:30
L	Beekman was at the pension around 18:30

conclusion in the premises of the argument we want to support. If we want to support the expectation that Leo was drugged by Marjan (D), we make that the conclusion and add the testimony of Beekman (E_1) and the expectation that Leo was under influence (A) to the premises: $\mathrm{P}(D|A \wedge E_1 \wedge E_{16} \wedge E_{14}) > \mathrm{P}(D|E_1)$. Now we have said that the argument for Marjan drugging Leo gets stronger if we include the expectation that Leo was actually drugged.

This argument is the same as the subscenario called 'M drugged L' in Vlek et al. (2014, see section 4.2.2). We can see that the method from Verheij (2014) leads to a smaller representation than the method from Vlek et al. (2014). It takes up less space on paper and does not require background information like a conditional probability distribution.

The system also allows us to model variations, which we can show with the discussion about the murder time. Beekman tells us that Leo was killed before 18:30 (T) and Pier tells us that Leo was still alive between 20:00 and 20:30 (G). Modelling this discrepancy can be done by stating that T and G cannot both be true: $\mathrm{P}(T \wedge G) = 0$.

With this, we can start building the arguments for H_1 and H_2. The first thing we add is the location of the body of Leo (E_6). This is not distinguishing between the hypotheses, since anybody could have placed that body there. The relation between $\mathrm{P}(H_1|E_6)$ and $\mathrm{P}(H_2|E_6)$ is undefined. This does not change when we add the hole that was prepared in Marjan front yard (E_7), which was arranged for by Marjan, so the relationship between $\mathrm{P}(H_1|E_6 \wedge E_7)$ and $\mathrm{P}(H_2|E_6 \wedge E_7)$ is still undefined. Since Beekman dragged the body together with Marjan in both scenarios, also adding the tracts in the gravel (E_8) does not change this relation.

The same thing goes for E_9 through E_{13}. None of them give a clear case for one side or the other. Instead of focusing on the evidence, we will focus our

attention on the expectations. Among the expectations is the case of the murder weapon. We have the expectation S, which states that the murder weapon is a stone. This expectation is supported by the testimonies of Beekman (E_1) and Kuilstra (E_4), and it is a plausible theory, so $\mathrm{P}(S|E_1 \wedge E_4) > 0$. We also have the expectation that Leo was killed with a hammer (H). We know that one of these is true, but not both at the same time, so we write $\mathrm{P}(S \wedge H) = 0$.

There is evidence for the second expectation, but not for the first one. This evidence is the blood on the hammer that was found (E_{12}), which was Leo's (E_{13}). We therefore can say that $\mathrm{P}(S|E_1 \wedge E_4 \wedge E_{12} \wedge E_{13}) < \mathrm{P}(H|E_1 \wedge E_4 \wedge E_{12} \wedge E_{13})$. We can even be so sure about this that we can say that the argument that Leo was killed with a stone is weakened until defeat, because the blood on the hammer gives us clear evidence that Leo was killed that way.

$$\begin{aligned} \mathrm{P}(S|E_1 \wedge E_4 \wedge E_{12} \wedge E_{13}) &= 0 \\ \mathrm{P}(H|E_1 \wedge E_4 \wedge E_{12} \wedge E_{13}) &= 1 \end{aligned}$$

This is similar to the subscenario 'M killed L' in Vlek et al. (2014, see section 4.2.6), specifically the 'M hit L on the head with a hammer/stone' variation.

Another interesting expectation is the earlier discussed case of the time of the murder. Since this is one of the few points were the scenarios differ, it is important to model this. In support of the expectations that Leo was killed before 18:30 (T) are Beekman's (E_1) and Kuilstra's (E_4) testimonies, which directly conflicts with the expectation that Leo asked Pier to get him around 20:00–20:30 (G), which is directly supported by Pier's testimony (E_5). Pier's testimony however conflicts with telephone records, which say that Pier called the pension around that time, not the other way around. This makes it so that T is more likely than G. $\mathrm{P}(T|E_1 \wedge E_4 \wedge E_5) > \mathrm{P}(G|E_1 \wedge E_4 \wedge E_5)$. Another problem with T however, is that Waanders did not report that she noticed anything that might point in the direction of murder, which she probably would have noticed if Leo was murdered before 18:30, since she was in the house the whole time.

We also suspect that Waanders and Marjan took their dogs for a walk around midnight (N). This would have given a potential murderer the time to kill Leo without anyone noticing. This does mean that Marjan could not have killed Leo, but we will ignore that for now. N is supported by Waanders testimony (E_2), which makes it a stronger argument: $\mathrm{P}(N|E_2) > \mathrm{P}(N)$. If we combine this information with the discussion about the time of the murder in the last paragraph, we can conclude that Leo was still alive at 18:30, since later on the evening would have been a better time to stage the murder, $\mathrm{P}(G \wedge N|E_5 \wedge E_2 \wedge E_1 \wedge E_4) > \mathrm{P}(T|E_5 \wedge E_2 \wedge E_1 \wedge E_4)$. This argument is however not decisive, so we will keep both options open.

We also have the blood trail that was reported by Beekman (B). The police found blood in the hallway of the boarding house (E_9), but this did not correspond with the blood that Beekman reported. Therefore, we can assume that the blood Beekman talked about was non existent. Now we can say that $\mathrm{P}(B|E_1 \wedge E_9) = 0$.

If we now combine all this information, we would get the full argument for the case. We will start with adding the argument for Marjan drugging Leo (D). This argument does not favour one hypothesis over the other, so the relative strengths stay undetermined.

If we add the arguments for the time of the murder (T and G), we have to make both arguments different, since H_1 does not agree with a theory where Leo was killed when Marjan and Waanders were walking their dogs. However, as mentioned before, this theory is preferred over the other theory. Therefore one argument now holds T among the conclusions while the other as G as a conclusion.

If we now add all the information, we can see that B is among the conclusions for H_1. As we also saw, B conflicts with the blood found in the hallway of the pension (E_9). Probability theory tells us that if $\mathrm{P}(\varphi) = 0$ then $\mathrm{P}(\varphi \wedge \psi) = 0$, regardless of the probability of ψ. This also means that the argument for H_1 has a strength of 0, since it contains $(E_1 \wedge E_9, B)$, which had a strength of 0. Assuming that there are no unknowns, we can now conclude that the argument for H_2 is conclusive, since one of the arguments need to be true. The full theory can be found below.

$$
\begin{aligned}
&\mathrm{P}(H_1 \wedge H_2) &&= 0\\
&\mathrm{P}(H_1 \vee H_2) &&= 1\\
&\mathrm{P}(T \wedge G) &&= 0\\
&\mathrm{P}(T \vee G) &&= 1\\
&\mathrm{P}(G \wedge N | E_1 \wedge E_2 \wedge E_4 \wedge E_5) &&> \mathrm{P}(T | E_1 \wedge E_2 \wedge E_4 \wedge E_5)\\
&\mathrm{P}(B | E_1 \wedge E_9) &&= 0
\end{aligned}
$$

$$
\begin{aligned}
E &= E_1 \ldots E_{18}\\
P_1 &= \mathrm{P}(H_1 \wedge T \wedge D \wedge M \wedge A \wedge H \wedge B | E)\\
P_2 &= \mathrm{P}(H_2 \wedge G \wedge D \wedge M \wedge A \wedge H \wedge N | E)\\
P_1 &= 0\\
P_2 &= 1
\end{aligned}
$$

We deem it more probable that Beekman was the murderer than Marjan. This is the same conclusion that was reached in Crombag and Israëls (2008), but different from what the court found. We are so certain about this conclusion that we are willing to accept this as truth and write of the first scenario as fiction.

5 Discussion

The criteria for the analysis were:

- Does it stop tunnel vision?
- Does it allow a good story to push out a true story?
- Is it dependant on subjective probabilities?
- Does it allow for internally inconsistent stories?

The analysis will now be evaluated and compared with the method from Vlek et al. (2014).

5.1 Tunnel vision

Since the method does not require two distinct scenario's, the method does not prevent tunnel vision. The method does however work fine if multiple scenario's are available, so it does not force the modeller into tunnel vision. Because the method forces the modeller to make everything explicit, it might still help.

This is different from the method from Vlek et al. (2014), which requires the modeller to present multiple alternatives. However, as Vlek also notes, both methods require the scenarios to be defined beforehand. More research should go into testing if the methods also work correctly if the scenarios are to be found during the gathering of evidence, as happens during a police investigation.

It is also debatable how well the method could potentially prevent tunnel vision, since tunnel vision is a problem of the modeller, not the model itself. If the modeller suffers from tunnel vision, she will put most of her attention into only one of the scenario's, so forcing multiple scenario's will then lead to one or more scenario's that are not taken seriously. A better way to combat tunnel vision could be achieved if the modeller were to model scenario's that were created by multiple persons with different opinions on the case.

5.2 Good story vs. a true story

Since both Marjan and Beekman deny that they committed the murder, it is not possible to know which story is true. Since Crombag and Israëls (2008) reach the conclusion that Beekman is the murderer with help from Marjan, we will use their analysis as the correct one.

It is likely that a good story pushed out a true story in court, since the two scenario's are both supported by most of the evidence. The only real problem with the first scenario is that there was no blood trail found in the hallway. The rest of the evidence does not sufficiently point in one direction or the other. This can also be seen in the model, where we do not reach a conclusion until we add the inconsistency of the blood trail.

The same conclusion is reached by Vlek et al. (2014), where the bloody trail is also the deciding factor between the two scenario's. It is possible that the judge suffered from tunnel vision because only evidence against Marjan was presented, but nothing against Beekman, because he was no longer a suspect at that point.

5.3 Subjective Probabilities

In the model, exact numerical probabilities were not used, only relative strengths for arguments. So

the method does not rely on numerical probabilities. It does rely on the modellers intuitions about the arguments. This can be solved by techniques for assessing the relative strengths of arguments.

This is different in the method of Vlek, where everything needs to be explicitly quantified. This gives the illusion that some of these probabilities are known, where most of them are not.

A good example of this are the priors, which need to be estimated in a Bayesian Network. In the method of Verheij (2014), these priors are not modelled, since the relations between the scenario's are undefined.

5.4 Stories that disagree with evidence

The method does not allow for stories that disagree with evidence. As soon as one of the arguments, for example on that supports an expectation, has a strength of 0, any argument that contains the same elements is also defeated. This means that less important disagreement could defeat an argument, but it also means that a story that does not seem to have disagreements will be defeated if it has them.

This can be seen at the end of the model, where we have the statement $\mathrm{P}(B|E_1 \wedge E_9) = 0$, after which we were forced to reject the first scenario in favour of the second, since the first scenario contained a disagreement.

The method of Vlek et al. also punishes incongruent stories, but it does not completely write them off. This can be seen in section 4.7 of Vlek et al. (2014), which goes over the conclusions from the network. At first the network prefers the story where Marjan killed Leo, until the lack of bloodtrail is added, when the probabilities flipped, and Beekman became the more likely murderer. The method from Verheij treats the case where Marjan killed Leo as defeated, while the method from Vlek still leaves it as a unlikely possibility. This leads to a more nuanced view of disagreements, but it also means that they may go undetected.

5.5 Overall

In conclusion, the method works to model a complex case. It also works comparable to the method from Vlek et al. (2014), but is less dependent on subjective probabilities and does not allow for internally inconsistent stories at all.

A downside of the method is that it does not show the normal argument structure that argumentative approaches, like Bex et al. (2010) and Timmer et al. (2014) do. This makes the arguments harder to follow and less easy to read. A graphical notation can be used alongside the current notation to alleviate this problem.

6 Conclusion and further research

In this paper a method for modelling criminal cases was evaluated using a complex case. It was also compared with a different approach using Bayesian Networks for the same case. The method focuses on building arguments from narratives and accessing their relative strengths.

The method does not allow a good story to push out a true story and does not depend on subjective probabilities, only on the modellers own intuitions about arguments. It punishes stories that are incongruent with evidence so badly that they are not allowed in the current framework.

The case study showed that a complex case can be accurately modelled in the new methodology, based on an earlier analysis (Vlek et al., 2014; Crombag and Israëls, 2008). It also showed that it can do this comparatively to the Bayesian Network approach of Vlek et al. (2014).

Although it does not have the clear distinction of idioms as Vlek et al. (2014), it can model a lot of the same narrative idioms. Scenarios are modelled as hypotheses, and constraints can be placed on the truth values of these hypotheses, comparing to the scenario and merged scenario idioms from Vlek. Also the variation idiom can be modelled by using constraints, but on expectations instead of hypotheses, as can be seen in the case study.

Still, a lot of work needs to be done. Currently assessing the relative strengths of arguments has to be done by hand. It would be better if this could be done methodologically, so the preferences of the modeller have less of an effect on the outcome. This would help to prevent tunnel vision during the process of modelling a narrative.

Another interesting line of research would be

combining the new method with argumentative frameworks, like the formal method of Bex (2009) or the method from Timmer et al. (2014), who constructs arguments in the ASPIC+ (Modgil and Prakken, 2013) framework from Bayesian Networks.

References

C. G. G. Aitken. *Statistics and the Evaluation of Evidence for Forensic Scientists.* Wiley, Chichester, 1995.

F. Bex. *Evidence for a Good Story.* PhD thesis, Rijksuniversiteit Groningen, 2009.

F. Bex, H. Prakken, and B. Verheij. Formalising argumentative story-based analysis of evidence. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 1–10. Rijksuniversiteit Groningen, Groningen, 2007.

F. Bex, P. J. van Koppen, H. Prakken, and B. Verheij. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, 18(2):123–152, 2010.

H.F.M. Crombag and H. Israëls. *Moord in Anjum: Te veel niet gestelde vragen.* Boom Juridische Uitgevers, Den Haag, 2008.

P. Dawid, D. Schum, and A. Hepler. Inference networks: Bayes and Wigmore. In *Evidence, Inference and enquiry*, chapter 5, pages 119–150. Oxford University Press, Oxford, 2011a.

P. Dawid, W. L. Twining, and M. Vasilaki, editors. *Evidence, Inference and Enquiry.* Proceedings of the British Academy. Oxford University Press, Oxford, 2011b.

N. Fenton, M. Neil, and D. A. Lagnado. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science*, 2013.

H. Kaptein, H. Prakken, and B. Verheij, editors. *Legal Evidence and Proof.* Ashgate Publishing Group, Aldershot, 2009.

S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 2013.

N. Pennington and R. Hastie. Reasoning in explanation-based decision making. *Cognition*, 49:123–163, 1993.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach.* Pearson Education, third edition edition, 2010.

S. T. Timmer, J. J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. *Extracting Legal Arguments from Forensic Bayesian Networks*, volume 271, pages 71–80. Legal Knowledge and Information Systems, 2014.

B. Verheij. To catch a thief with and without numbers: Arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk*, 13(3-4):307–325, 2014.

B. Verheij, F. Bex, S. T. Timmer, C. S. Vlek, J. J. C. Meyer, S. Renooij, and H. Prakken. Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, to appear.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Building Bayesian networks for legal evidence with narratives. *Artificial Intelligence and Law*, 22(1):375–421, 2014.

J. H. Wigmore. *The Principles of Judicial Proof; or, The Process of Proof as Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials.* Boston, Little, Brown, and Company, second edition, 1931.