

Unfolding Crime Scenarios with Variations: A Method for Building a Bayesian Network for Legal Narratives

Charlotte S. VLEK^{a,1}, Henry PRAKKEN^{b,c}, Silja RENOUIJ^b and Bart VERHEIJ^{a,d}

^a*Institute of Artificial Intelligence, University of Groningen*

^b*Department of Information and Computing Sciences, Utrecht University*

^c*Faculty of Law, University of Groningen*

^d*CodeX, Stanford University*

Abstract Legal reasoning can be approached from various perspectives, traditionally argumentation, probability and narrative. The communication between forensic experts and a judge or jury would benefit from an integration of these approaches. In previous papers we worked on the connection between the narrative and the probabilistic approach. We developed techniques for representing crime scenarios in a Bayesian network. But for complex cases, the construction of a Bayesian network structure using these techniques remained a cumbersome task.

In this paper we therefore propose a method called *unfolding a scenario* and a representation for small variations within a scenario. With these tools, a Bayesian network can be built up step by step, gradually adding more details. The method of unfolding a scenario is intended to support the process of building a Bayesian network, additionally resulting in a well-structured graphical structure.

Keywords. Legal reasoning, Bayesian networks, Narrative

1. Introduction

Legal reasoning involves establishing what may have happened around a supposed crime based on the collection of evidential data. Traditionally there are three main approaches to legal reasoning: argumentation, probability and narrative [1]. An integration of these three approaches can enable a better communication between forensic experts (who reason with probabilities) and judges and juries (who reason with arguments and narrative) [2]. In previous papers [3,4] we have worked on the combination of probability and narrative, proposing techniques to enable the representation of a crime scenario in a Bayesian network. In this paper, we continue this research by proposing a method for gradually constructing the Bayesian network model of a scenario.

Judges and juries tend to think in terms of stories [5]. A crime scenario, a story about what happened around a supposed crime, can help to take a holistic point of view: it provides a context for the evidence. Narrative and law have been studied by a number of

¹Corresponding Author: Charlotte S. Vlek, University of Groningen, E-mail: c.s.vlek@rug.nl.

researchers, see e.g. [6,5,7]. Recently, Bex proposed a hybrid theory for reasoning with narrative and argumentation in legal cases [8,9].

Several forensic scientists have proposed Bayesian networks as a general method for reasoning with evidence in legal cases and a number of researchers have worked on Bayesian networks and law [10,11,12,13]. A Bayesian network is a representation of a joint probability distribution: the variables in a domain are represented as nodes in a graph. Arrows between nodes show dependencies between the variables. By observing evidence, the result of this on the probabilities of the hypotheses can be calculated. Despite known difficulties with the application of Bayesian networks, we believe that a Bayesian network can prove to be very useful in comparing various hypotheses. We aim to model multiple hypotheses in one network and compare how well the evidence supports each. In this paper we focus on the graphical structure for such a network, leaving aside the issue of finding the underlying probabilities.

Recently, Fenton, Neil and Lagnado [14] proposed to compile a list of *legal idioms*: substructures that can be reused in various Bayesian networks. In our project,² we intend to develop narrative idioms for modeling crime scenarios in a Bayesian network. In our previous papers [3,4] we developed three narrative idioms, enabling the representation of multiple scenarios in a Bayesian network. In this paper we will develop a fourth narrative idiom, intended to model small variations in a scenario. Furthermore, using all four idioms we propose a technique for constructing a Bayesian network, called *unfolding a scenario*. This is inspired by Hepler, Dawid and Leucari [10], who proposed to use Object Oriented Bayesian networks for a stepwise construction of the graph of a Bayesian network. In this paper, we extend these ideas from a narrative perspective, enabling the construction of a Bayesian network in which gradually more details can be specified.

The contributions of this paper are twofold: (1) we develop an idiom for modeling variations in a crime scenario and (2) we propose a method for gradually constructing the graphical structure of a Bayesian network for a crime scenario. We hope to support the modeler's task, leading to a more structured Bayesian network.

The remainder of this paper is organized as follows: first we discuss some preliminaries in Section 2. Then a brief treatment of the ideas from our previous papers follows in Section 3. Section 4 forms the body of this paper: there the new variation idiom and the concept of unfolding is introduced. In Section 5, the position of this work relative to related research is discussed. The paper concludes with a conclusion.

2. Preliminaries

A Bayesian network consists of a directed acyclic graph and conditional probability tables. It allows for a compact representation of a joint probability distribution using the independence relations that exist between the variables. The nodes in the graph represent variables and when there is no arrow between nodes, the variables are either independent or conditionally independent. A conditional probability table for each node gives for each value of the node the probability that it occurs, conditioned on the values of its direct predecessors.

Figure 1 shows an example of a Bayesian network. There are three variables represented as nodes. Each of the nodes can have a number of values. Throughout this pa-

²project website: www.ai.rug.nl/~verheij/nwofs

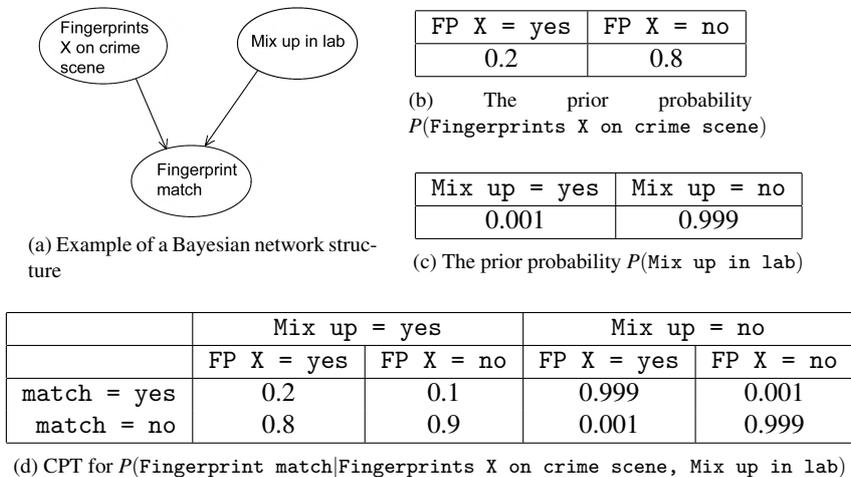


Figure 1. An example of a Bayesian network: graph and probability tables

per the nodes have binary values (yes and no). The tables in Figure 1 are probability tables for each of the nodes. Note that the probabilities for the two top nodes are not conditional: they have no predecessors so tables (b) and (c) contain *prior probabilities*.

Bayesian networks can be used to evaluate how likely a hypothesis is given the evidence. When a value of a node is observed (evidence is found), the node is *instantiated*. Inference in the network leads to updated posterior probabilities given all observed values. From the network, any prior or posterior probability of interest can be calculated. Various tools are available for working with Bayesian networks, for example GeNIe 2.0.³

3. Three narrative idioms

In previous papers [3,4], we developed three narrative idioms for representing crime scenarios in a Bayesian network: the scenario idiom, the subscenario idiom and the merged scenarios idiom. In this section, we review these idioms with examples from the Anjum murder case, in which Marjan was convicted for the murder of Leo de Jager [8,15].

3.1. The scenario idiom

The purpose of the scenario idiom is to model a scenario such that it is captured as a coherent whole. Consider an example from the Anjum case: ‘Marjan hit Leo on the head with a hammer, resulting in Leo’s blood in the hallway and Leo’s death.’ To model this scenario with the scenario idiom, each state or event in the scenario is represented as a node in the structure (see Figure 2a). This amounts to three nodes: ‘M hit L in the head with a hammer’, ‘L’s blood in the hallway’ and ‘L died’. There are connections between these events, such as between Marjan hitting Leo on the head and Leo dying, drawn as arrows in the structure. Validating all (in)dependencies requires careful examination of

³GeNIe 2.0 is available for free on genie.sis.pitt.edu

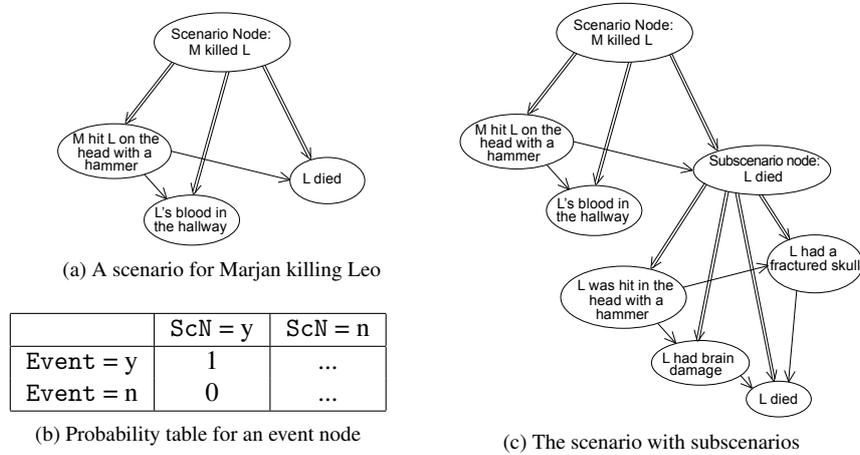


Figure 2. Scenario and subscenario for the Anjum case

the scenario, see [16] for approaches to catching the structure of a Bayesian network. Finally, an extra node called the *scenario node* is included to capture the coherence of the scenario. This scenario node is connected to all state and event nodes. These connections are shown as double arrows in the graph to represent that the underlying probability tables are always of the same form, as discussed below. This is purely meant as a visual aid; technically they are not different from the single arrows. The result is in Figure 2a.

The probabilities in the scenario idiom are determined according to two principles:

- If the scenario node is true, all states and events in the scenario must be true;
- A state or event can occur without the scenario node being true.

These principles secure part of the probability table (see the table labeled 2b in Figure 2) for a state or event node: the probability is 1 for a state or event to be true given that the scenario node is true (top left). On the other hand, the probability that a state or event is true given that the scenario node is false remains to be determined.

3.2. The subscenario idiom

In Section 3.1, the event that Leo died was modeled as one node in the structure. However, Leo's death is in itself a small scenario, which we call a subscenario. This subscenario describes how Leo died as a result of being hit with a hammer: 'Leo was hit in the head with a hammer so Leo had brain damage and a fractured skull. Leo died.' This can be modeled with the subscenario idiom, see Figure 2c. The subscenario idiom is similar to the scenario idiom: analogous to the scenario node there is now a subscenario node with connections to each of the events in the subscenario.

Like the scenario idiom, the subscenario idiom consists of state/event nodes, arrows between these nodes and a node at the top connected to all state and event nodes, which is now called the subscenario node. A subscenario is always part of a scenario as shown in Figure 2c. A scenario can have multiple subscenarios and there can be connections between those scenarios. For a scenario with subscenarios, the probabilities are determined according to similar principles like in Section 3.1.

3.3. *The merged scenarios idiom*

By modeling multiple scenarios in one Bayesian network, the various scenarios can be compared. To this end, the merged scenarios idiom was developed. This idiom combines the structures of separate scenarios into one large structure. The key idea is that in order to compare scenarios, we want at most one of them to be true at a time, that is, only when they are indeed mutually exclusive. For example, in the Anjum case there can be a scenario in which Marjan killed Leo, and one in which Marjan's neighbor Beekman killed Leo. In that case, the merged scenarios idiom puts a constraint on the scenario nodes of the two scenarios, such that they cannot both hold at the same time. See [3] for a more detailed explanation of the constraint node.

In a previous version of the merged scenarios idiom from [3,4], the constraint node was connected to the guilt hypothesis nodes of the scenarios. The purpose of this guilt hypothesis node was to summarize the scenarios and combine scenarios that essentially led to the same conclusion. For example: there might be two different scenarios in the Anjum case in which Marjan killed Leo. The scenario nodes for these two scenarios would be connected to one guilt hypothesis: Marjan killed Leo. However, with the developments in this current paper, the guilt hypothesis has become redundant: with the variation idiom from Section 4.1, the two ways in which Marjan may have killed Leo will now be modeled in one scenario. Therefore, in the current version of the merged scenarios idiom, the constraint is on the scenario nodes.

3.4. *A design method in four steps*

In our previous work, we proposed a design method for constructing the graphical structure of a Bayesian network, consisting of four steps: (1) collect all relevant scenarios, (2) model each scenario with the scenario idiom and subscenario idiom, (3) merge the resulting structures and (4) include the evidence in the network. Step (2) can be very complex for large cases. In Section 4 we propose the concept of unfolding a scenario as a systematic approach to this second step in the design method and in Section 4.3 we present a revised version of the above design method.

4. **Unfolding a scenario with variations**

When constructing a Bayesian network for a case, scenarios can consist of several sub-scenarios and even sub-subscenarios. Consider the Anjum case, where the scenario from Section 3 is only a small part of a larger scenario: Marjan had a motive for killing Leo (a subscenario consisting of several nodes), she drugged him (another subscenario) and then killed him (the scenario from Figure 2c with a subscenario about Leo dying).

In this section we propose the concept of unfolding a scenario for constructing a Bayesian network: a procedure for systematically constructing the graph. The idea is to first formulate a crude version of a scenario and then gradually work out the details by elaborating events as subscenarios. For example, in the Anjum case we could first model the scenario as described above (Marjan had a motive, drugged Leo and killed him) and then further elaborate each of these states and events to more detailed subscenarios.

During this process of unfolding a scenario we may run into small variations within a scenario. For this, we first introduce a fourth narrative idiom: the variation idiom (Section 4.1). Then, in Section 4.2 we treat the concept of unfolding a scenario.

	disjunction node = yes		disjunction node = no	
	var 1 = yes	var 1 = no	var 1 = yes	var 1 = no
var 2 = yes	...	1	0	0
var 2 = no	...	0	1	1

Table 1. Conditional probability table for variation 2. The table for variation 1 is analogous.

4.1. The variation idiom

In the Anjum case, the autopsy indicated that Leo must have been hit on the head with an angular object. The police considered two possibilities: either Leo was hit with a stone, or with a hammer. This variation on the murder weapon does not lead to a very different scenario: in any case Leo died due to a hit on his head by Marjan.

When modeling the scenario about Marjan killing Leo, we would like to include the hammer/stone variation within one scenario. Two reasons for including a variation rather than modeling it as separate scenarios, are that (1) it will result in a simpler graphical structure for the network and (2) it will help the modeler with his task: it is much easier to model a small variation at the moment you encounter it as a modeler, than setting up an entirely new scenario structure for each variation of a scenario.

In order to model variations within the structure of our previously introduced narrative idioms, we introduce the variation idiom, shown in Figure 3a. The variation idiom consists of nodes for each variation (the hammer and the stone) and a node describing the disjunction of variation 1 and variation 2, in this case: ‘Marjan hit Leo on the head with a hammer or a stone’. We model the variations in separate binary nodes rather than in one node with all variations as values to enable variations consisting of more than one node. This way, a variation can later on be unfolded into a more detailed subscenario.

The variation idiom has connections from the disjunction nodes to the two (or possibly more) variations. Furthermore, a connection between the variations is required to be able to model that at least one of them holds when the disjunction node holds: if it was not the hammer, it must be the stone. We can then specify the probability tables (see Table 1) such that the variables behave like a disjunction:

- If the disjunction node is false, neither of the variations can be true. This leads to a probability 0 for $P(\text{var 2} = \text{yes} | \text{disjunction node} = \text{no})$ in the table;
- If the disjunction node is true and one variation is false, the other variation must be true. So $P(\text{var 2} = \text{yes} | \text{disjunction node} = \text{no}, \text{var 1} = \text{no})$ is 1.

From this it follows that the following two properties of a disjunction are satisfied: (1) when at least one variation is true, the entire disjunction is true and (2) when both variations are false, the entire disjunction is false.

The disjunction node in the variation idiom can be used as an event in the scenario idiom. This way, when the scenario node is true, the disjunction (‘Marjan killed Leo with a hammer or a stone’) must be true. Which of the two variations (hammer or stone) is the case, does not matter for the rest of the scenario. This construction leaves the concept of coherence of a scenario intact, including the role of the scenario node representing the scenario as a whole. See Figure 3c for the hammer/stone variation in a scenario structure. Note that each variation is itself a small subscenario.

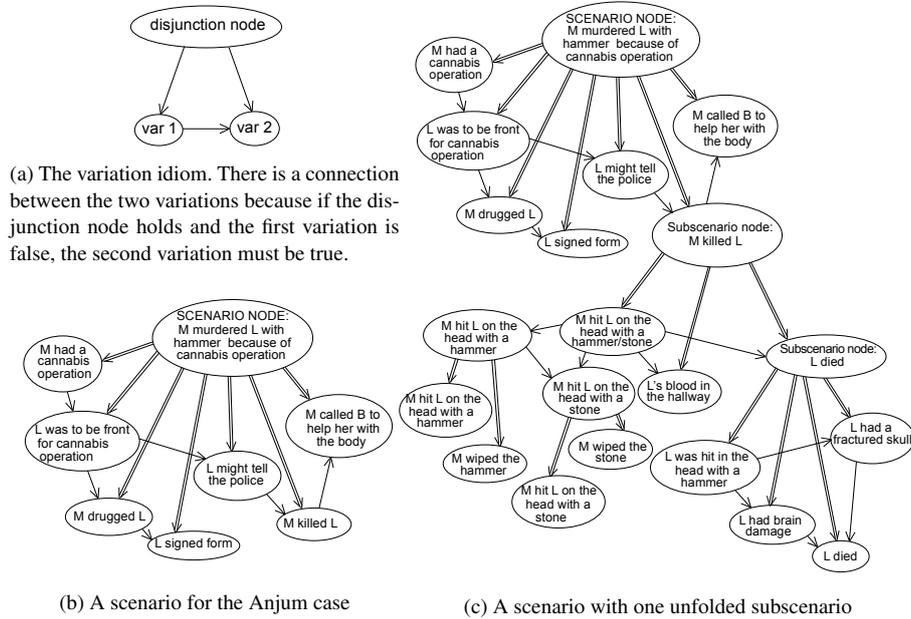


Figure 3. Scenario for the Anjum case with variations and subscenarios

4.2. Unfolding a scenario

The concept of unfolding a scenario relies on the idea that a scenario can have different levels of precision. For many practical purposes, we leave a lot of details out of a scenario: when we tell someone about our restaurant visit, we often do not include much detail about how we paid for the food (asked for the check, took out wallet, paid). In legal cases, often the details are what matters. For example: if some money was stolen while visiting the restaurant, it might be quite interesting to know how we payed: perhaps some money was taken from the wallet at that point. In fact, we might want to specify even further: how did we take the wallet from our pocket: was there something irregular there? However, no legal case requires information about all detailed aspects of the case: it depends on the circumstances which details we are interested in. A similar point was made by Wagenaar, Van Koppen and Crombag in their Anchored Narratives Theory [7]. They argue that many connections in a scenario remain implicit, but can be made more precise by asking for further details. With the technique of unfolding a scenario, we gradually construct the Bayesian network by asking for more details whenever they are relevant.

First a crude scenario is formulated about what may have happened around a crime. For the Anjum case, one scenario is as follows:

Marjan had a cannabis operation. She wanted to use Leo as a front for this cannabis operation. She drugged him because she wanted him to sign a contract. Leo signed the contract. Marjan was worried Leo might tell the police, so Marjan killed Leo with a hammer in the hallway. After this, she went to call Beekman, who helped her to drag the body to the front yard.

This scenario can be modeled with the scenario idiom as shown in Figure 3b.

The next step is to ask which state or event requires elaboration. We already have a subscenario for ‘M killed L’: the scenario from Section 3 including the variation (hammer/stone) will now serve as a subscenario. But why does this event require more details? One reason is that we have no direct evidence for the event ‘M killed L’, but we do have evidence for some of the details: there was a hammer found with blood on it, and a wad with blood and hair on it was found as evidence that a hammer or stone was wiped off. Unfolding this to a subscenario results in a structure as shown in Figure 3c.

For deciding whether a state or event requires unfolding, the following questions should be asked:

1. Is there direct evidence for the state or event? If so, no unfolding is required.
2. Is there relevant evidence for details of a subscenario for this state or event? If so, unfolding is required.
3. Would it be *possible* to find relevant evidence for details in a subscenario for this state or event? If so, unfolding is required.

To summarize, whenever unfolding helps to add more relevant evidence to the structure, it should be done. On the other hand, when there is direct evidence for a state or event, unfolding is not necessary. However, unfolding might still be beneficial: in the example from Figure 3c we chose to unfold the event ‘L died’, even though there was direct evidence for this node (Leo’s dead body). By unfolding it, more evidence could be included: Leo had a fractured skull and suffered brain damage, as the autopsy report supports.

Unfolding a scenario is a top-down process, quite literally. Starting with a high-level description of what happened, a crude scenario is modeled with the scenario idiom. By unfolding the scenario, new substructures are added to the network, always *below* the existing structure, in the sense that arrows point from the existing nodes to the new nodes.

4.3. A design method with unfolding and variations

Based on the design method from Section 3.4 and our unfolding technique from Section 4.2, we propose the following design method for constructing a Bayesian network:

1. Collect: formulate relevant scenarios for the case
2. Unfold: for each scenario, model a crude scenario with the scenario idiom. Then unfold this scenario by repeatedly asking the three questions from Section 4.2. Use the subscenario idiom and the variation idiom to model the unfolding subscenarios. The process of unfolding is finished when the three questions from Section 4.2 indicate that no more relevant evidence can be added to the structure.
3. Merge: use the merged scenarios idiom to merge the scenario structures constructed in the previous step.
4. Include evidence: for each piece of evidence that is available, include a node and construct it to the state or event it supports. Additionally, include nodes for evidential data that is to be expected as an effect of states and events in the structure.

5. Related work

Hepler, Dawid and Leucari [10] proposed to build Bayesian networks for legal cases from modules, employing ideas from Object Oriented Bayesian networks. The authors

emphasize that these modules can be reused for different cases. They identify a number of recurrent structures, such as ‘identification’, ‘contradiction’, ‘conflict’, etcetera. Their stepwise construction of a Bayesian network relies on a similar idea of recurring structures: an accusation of first degree murder should standardly be decomposed into modules ‘a murder occurred’, ‘a felony was committed’ and ‘the accused is the culprit’.

Hepler, Dawid and Leucari’s use of modules is similar to our technique of unfolding a scenario since both techniques aim to build up a Bayesian network from substructures (modules or subscenarios). However, the techniques differ in their perspective on how to build a network using these substructures. Hepler, Dawid and Leucari aim to use standard modules that can be included in the network depending on the type of case. They present no heuristics on which modules to include in the network, but they do give a description of a murder case consisting of the three modules mentioned above.

Van Gosliga and Van de Voorde also worked on a modular construction of Bayesian networks [17]. They proposed a design pattern called Hypothesis Management Frameworks (HMF’s). HMF’s consist of modules with probabilities explicitly included as connecting nodes in the graph. This enables multiple experts to work on the same network simultaneously, including a situation in which one expert draws (part of) the graph, and another specifies the numbers. HMF’s thereby enable frequent changes or extensions of the model, which is especially useful for application domains such as the legal field, where insights might change during the process of investigation. The main contribution of [17] is a new modeling framework that enables a structured building process for constructing a belief network. However, they formulate no procedure for this building process.

In our work, we specify how to gradually build a Bayesian network from a narrative perspective. A crime scenario can provide the required context to distinguish between relevant and less relevant details. Narrative thereby provides the heuristics to decide which modules to unfold. Our narrative perspective enables a view in which a network for a case can be made as detailed as necessary. Directions for a stop condition are given by three questions that help to determine whether further unfolding the structure will lead to more relevant evidence. This is different from [10] and [17], where directions about what modules to include and when the network is detailed enough are lacking. To conclude, our main contribution is (1) a procedure for gradually constructing a Bayesian network based on a narrative perspective and (2) a stop condition: a list of critical questions that can help to determine when a network is detailed enough.

6. Conclusion

In this paper we have extended our previous work on modeling crime scenarios in a Bayesian network with the concept of unfolding a scenario and with the variation idiom. The method of unfolding a scenario is a process to gradually make a scenario more precise, modeling the details that can enhance the network in such a way that more evidence can be included. The concept of unfolding a scenario was inspired by work by Hepler, Dawid and Leucari [10] who proposed to apply Object Oriented Bayesian networks to legal cases. Giving this a narrative twist, we added the idea that refining the network is always possible but not always relevant. We proposed a number of questions that can help decide when a scenario should be further unfolded. We also developed a fourth narrative idiom: the variation idiom. This idiom makes it possible to model a

scenario as a whole, including small variations within the scenario. Both this new idiom and the concept of unfolding a scenario will result in a simpler structure for the resulting Bayesian network.

In the future, the concept of story schemes could help to further develop the technique of unfolding a scenario. A story scheme such as, for example, a scheme for a typical murder case, could then serve as a starting point for unfolding a scenario, after which each element of the scheme will be unfolded to specify what happened around the crime. A potential problem here is that each crime is unique: it remains to be investigated how useful a scheme for a general situation is when dealing with a specific case.

References

- [1] H. Kaptein, H. Prakken, and B. Verheij, eds., *Legal Evidence and Proof: Statistics, Stories, Logic*. Aldershot: Ashgate Publishing Company, 2009.
- [2] B. Verheij, "Integrating argumentation, narrative and probability in legal evidence," in *The Third Workshop on Computational Models of Narrative* (M. Finlayson, ed.), pp. 176–177, 2012.
- [3] C. Vlek, H. Prakken, S. Renooij, and B. Verheij, "Modeling crime scenarios in a Bayesian network," in *Proceedings of the 14th International Conference on Artificial Intelligence and Law*, (New York), pp. 150 – 159, ACM Press, 2013.
- [4] C. Vlek, H. Prakken, S. Renooij, and B. Verheij, "Representing and evaluating legal narratives with sub-scenarios in a Bayesian network," in *2013 Workshop on Computational Models of Narrative* (M. Finlayson, B. Fisseni, B. Löwe, and J. Meister, eds.), (Saarbrücken/Wadern, Germany), pp. 315–332, Schloss Dagstuhl, 2013.
- [5] N. Pennington and R. Hastie, "Explaining the evidence: Tests of the story model for juror decision making," *Journal of Personality and Social Psychology*, vol. 62, no. 2, pp. 189 – 206, 1992.
- [6] W. Bennett and M. Feldman, *Reconstructing Reality in the Courtroom - Justice and Judgment in American Culture*. London: Methuen-Tavistock, 1981.
- [7] W. Wagenaar, P. van Koppen, and H. Crombag, *Anchored Narratives: the Psychology of Criminal Evidence*. Hemel Hempstead: Harvester Wheatsheaf, 1993.
- [8] F. Bex, *Arguments, Stories and Criminal Evidence, a Formal Hybrid Theory*. Dordrecht: Springer, 2011.
- [9] F. Bex, P. van Koppen, H. Prakken, and B. Verheij, "A hybrid formal theory of arguments, stories and criminal evidence," *Artificial Intelligence and Law*, vol. 18, pp. 123–152, 2010.
- [10] A. Hepler, A. Dawid, and V. Leucari, "Object-oriented graphical representations of complex patterns of evidence," *Law, Probability & Risk*, vol. 6, pp. 275–293, 2004.
- [11] J. Keppens, "On extracting arguments from Bayesian network representations of evidential reasoning," in *The 13th International Conference on Artificial Intelligence and Law* (K. Ashley and T. van Engers, eds.), (New York), pp. 141 – 150, ACM, 2011.
- [12] D. Lagnado, N. Fenton, and M. Neil, "Legal idioms: a framework for evidential reasoning," *Argument and Computation*, 2012.
- [13] F. Taroni, C. Aitken, P. Garbolino, and A. Biedermann, *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester: Wiley, 2006.
- [14] N. Fenton, M. Neil, and D. Lagnado, "A general structure for legal arguments using Bayesian networks," *Cognitive Science*, vol. 37, pp. 61 – 102, 2013.
- [15] H. Crombag and H. Israëls, *Moord in Anjum. Te veel niet gestelde vragen*. Den Haag: Boom Juridische uitgevers, 2008.
- [16] F. Jensen and T. Nielsen, *Bayesian Networks and Decision Graphs*. New York: Springer, 2007.
- [17] S. van Gosliga and I. van de Voorde, "Hypothesis management framework: A flexible design pattern for belief networks in decision support systems," in *6th Bayesian Modelling Applications Workshop on UAI 2008, Helsinki, Finland*, July 2008.