Designing and Understanding Forensic Bayesian Networks using Argumentation

Sjoerd T. Timmer

Printed by: Ipskamp printing
ISBN: 978-90-393-6695-0

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van het Forensic Science programma van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

SIKS Dissertation Series No. 2017-02
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Designing and Understanding Forensic Bayesian Networks using Argumentation

Ontwerp en Uitleg van Forensische Bayesiaanse Netwerken met Argumentatie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 1 februari 2017 des middags te 2.30 uur

door

Sjoerd Themba Timmer

geboren op 7 augustus 1988
te Bulawayo, Zimbabwe

Promotoren:	Prof. dr. J.-J. Ch. Meyer
	Prof. dr. mr. H. Prakken
	Prof. dr. H. B. Verheij
Copromotor:	Dr. S. Renooij

Contents

Chapter 1

Introduction

In recent years, the rise of forensic sciences has posed the legal domain with the challenge of dealing with numerical, probabilistic evidence. The availability of forensic methods, such as DNA analysis, has increased over the last decades. Most of these methods yield results that are accompanied by graded uncertainties. These uncertainties are often expressed numerically by forensic experts, but lawyers, judges and other legal experts have notorious difficulty interpreting such results [Fenton, 2011]. A number of recent legal cases in which forensic evidence was misinterpreted has emphasised the need for a better foundational understanding of the role of probabilities in legal proof [Berger et al., 2011; Sjerps, 2011]. The cases of Lucia de B. in the Netherlands [Derksen and Meijsing, 2009; Meester et al., 2007] and Sally Clark in the UK [Schneps and Colmez, 2013] are well-known examples where probabilistic reasoning has gone wrong with dramatic consequences. The currently ongoing debate about the role of probabilities in legal cases provokes an interest in legal argumentation theory about probabilistic evidence in court. Reasoning under uncertainty is in general a difficult task that easily leads to reasoning errors and miscommunication. The correct application of independence and statistical syllogisms seems to be difficult for those not trained in it.

Legal reasoning with probabilities has gained attention from different scientific fields. Three different approaches [Kaptein et al., 2009] to model and reason with legal evidence have been explored: *probabilistic*, *argumentative*, and *narrative* models of proof. Argumentation has the earliest and strongest connection to legal reasoning [Wigmore, 1913; Anderson and Twining, 1991]. Probabilistic models, and so-called Bayesian networks (BNs) in particular, have recently also gained scientific interest. Narrative models of evidence focus on coherent stories about events and originate from legal psychology [Pennington and Hastie, 1993].

All three approaches mentioned above have their strengths and weaknesses. Argumentation, for instance, is well structured in a way that is easy to follow. However, it has difficulty handling probabilistic (numerical) evidence. Until recently, argumentation models were almost exclusively qualitative. In the past

few years there has been some research on probabilistic argumentation [Pollock, 2001; Hunter and Thimm, 2016]. In probabilistic argumentation numeric degrees of uncertainty can be incorporated in arguments. However, in this thesis we take a different approach by looking at BNs and how these can be explained or constructed using argumentation. BNs, a state of the art probabilistic model, are well understood from a mathematical point of view. How they are to be applied in legal proof is a subject of debate. Especially the strong requirement on the availability of many numerical parameters limits their applicability in real cases. Narrative models, which use stories and scenarios, have a strong emphasis on forming a coherent view, which is very important in legal reasoning, for instance to prevent tunnel vision. However, formalisation of narrative approaches is difficult and is less developed than formalisations of argumentation. The relation between narrative and probability theory is also not well understood.

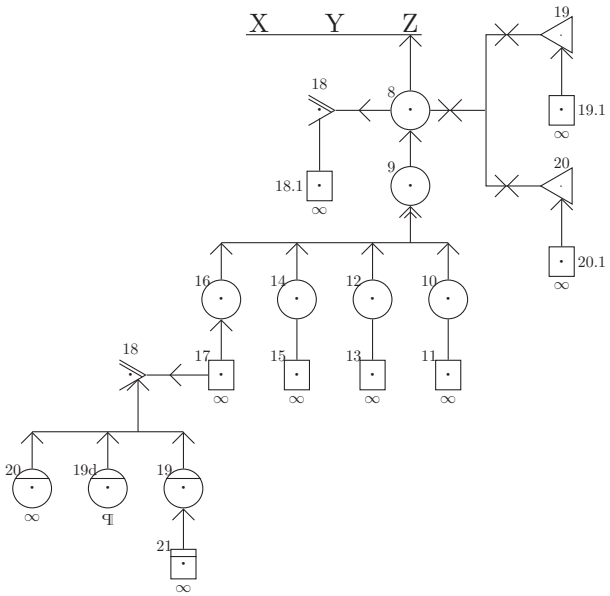
There have been attempts to combine the three approaches. Notable is the work by Bex [2011] who has created argument-narrative hybrid models of evidence. In this thesis we combine argumentation and Bayesian networks. This research is part of a larger research project entitled 'Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios' [Verheij et al., 2016] in which also the combination of narrative approaches with Bayesian networks is studied [Vlek, 2016].

We will now first describe the three approaches in more detail and then proceed with a formulation of research questions that follow from the difficulties that the legal domain is facing with probabilistic evidence. We will take examples from the legal domain on several occasions, but we will argue that the combined models proposed herein are not limited to legal issues. Legal reasoning is a very suitable application domain to test and develop such theories, because in legal reasoning the rules and their exceptions are often more explicitly stated compared to other domains.

1.1 Argumentative models of evidence

Argumentation is a well studied topic in the field of artificial intelligence (AI) [Van Eemeren et al., 2014, chapter 11]. The legal domain has always had a keen interest in the development of argumentation theories because in legal decisions the stakes are often high and the study of correct and reasonable argumentation can expose possible mistakes.

Argumentative approaches to legal reasoning with evidence go back to the work of Wigmore [1913] in the beginning of the 20th century. Wigmore drafted graphical representations of argumentative reasoning, in which propositions can support and attack each other. In these diagrams, statements, intermediate and final conclusions are connected by arrows and other graphical connections. Such a chart can be seen in Figure 1.1. It is usually accompanied by a table of textual



Z	The charge that U killed J.	18(2)	The witness is biased.
8	Revengeful murderous emotion toward J.	19	U and J remaining in daily contact, wound must have rankled.
10	Letter received by priest stating that U already had a family in the old country.	19.1	Witness to daily contact.
11	Anonymous witnesses to 10.	19(2)	The witness is a discharged employee of U.
12	J was author of letter (although it was in a fictitious name).	21	Anonymous witness to 19(2).
13	Anonymous witnesses to 12.	19d	Discharged employees are apt to have an emotion of hostility.
14	Letter communicated by priest to U.	20	Wife remaining there, jealousy between U and J probably continued.
15	Anonymous witnesses to 14.	20.1	Witness to wife remaining.
16	Letter's statements were untrue.	20(2)	The witnesses' strong demeanour of bias while on the stand.
17	Anonymous witnesses to 16.		
18	U's marriage being finally performed, U would not have had a strong feeling of revenge.		

Figure 1.1: Wigmore chart [Wigmore, 1913]. The numbers refer to written descriptions of the statements. Graphical cues can be used to indicate whether the arguments belong to the defence or the prosecution. Different arrow tips indicate support and attack and can even distinguish strong and weak support.

descriptions that tells us what each number refers to. Graphical signals are used to indicate all sorts of information about these statements. Different arrow tips can, for instance, indicate strong or weak support. Horizontal markings in nodes are used to differentiate between prosecution and defence. Further node-shapes can be used to distinguish ancillary from direct evidence. *Ancillary evidence*—in contrast to *direct evidence*—is evidence for a link between other statements. Similar models were later further developed in a movement called *the new evidence scholarschip* [Lempert, 1986; Anderson and Twining, 1991; Schum and Tillers, 1991].

Independently, the field of artificial intelligence (AI) has developed argumentation systems for reasoning with uncertainty. These systems provide models that describe how conclusions can be justified in a way that closely follows the reasoning patterns present in human reasoning. This makes argumentation an intuitive and versatile model for common sense reasoning tasks. The notion of *defeasible* inference [Pollock, 1987] was introduced to deal with uncertain inferences. A defeasible inference is an inference that does not always hold but is merely a general rule with possible exceptions. At first, this was used in combination with informal *argumentation schemes* [Walton et al., 2008]. Argumentation schemes are recurring patterns of reasoning. They can be used to check that all premises of an argument are met and that all possible sources of doubt have been sufficiently investigated. Later, focus was directed towards computational models of argument acceptability, such as the work of Dung [1995] on the acceptability of abstract arguments. This describes how arguments can be evaluated on the basis of an attack relation between them in a way that does not depend on the nature of the argument or the attack. Other formal argumentation approaches, such as DefLog [Verheij, 2003a], Carneades [Gordon and Walton, 2009], and ASPIC+ [Modgil and Prakken, 2014], the latter of which we will further discuss in Chapter 2, are built on these ideas. These techniques find applications in legal reasoning about evidence [Bex et al., 2003; Bex, 2011; Prakken, 2012; Verheij, 2014].

One of the issues with argumentation in AI is that the relation with numerically graded uncertainty is not well understood. Comparing arguments in argumentation is usually not done on the basis of numerical calculation, but, for instance, on the basis of orderings of premises and rules [Modgil and Prakken, 2014]. However, this makes it hard to incorporate probabilistic evidence such as e.g. DNA or fingerprint matching evidence.

The second issue is that causality creates complex interactions if implications are not strict. Two causes of one effect (a sounding alarm for instance) that are a priori probabilistically independent of each other (a burglar and a fire in this same example) become dependent when the shared effect is observed. That is, apprehending the burglar reduces the probability that a fire is the cause of the alarm. Pearl [1988b] identified this issue and proposed a solution. This solution is very much inspired by the way conditional independence is modelled in a BN. These BNs have since gained vast popularity, in part because they also addresses the former issue.

1.2 Probabilistic models of evidence

By using a method that is grounded in probability theory, the above issues with numerical evidence and complex interactions between evidence can be overcome. Probabilistic methods can provide very precise descriptions of uncertainty and (in-)dependencies between stochastic variables. A *stochastic variable* or *random*

variable V can take on exactly one of its mutually exclusive and collectively exhaustive outcome states when observed. Assume for the following examples that the outcomes of variables can be either *true* or *false*. Such variable are referred to as *Boolean* variables. A probability function P() assigns a probability to each outcome of V. For example $\mathrm{P}(\mathtt{V} = \mathtt{true})$ denotes the probability that V turns out to be true when observed. The notation $\mathrm{P}(\mathtt{V} = \mathtt{true} \mid \mathtt{E} = \mathtt{true})$ denotes the probability that V is true 'given' that variable E is known to be true. This conditional probability is mathematically equivalent to $\mathrm{P}(\mathtt{V} = \mathtt{true} \wedge \mathtt{E} = \mathtt{true}) / \mathrm{P}(\mathtt{E} = \mathtt{true})$. Similarly $\mathrm{P}(\mathtt{V} \mid \mathtt{E} = \mathtt{true})$ is the conditional probability distribution of outcomes of V given that $\mathtt{E} = \mathtt{true}$. Evidential strength in a probabilistic setting is often expressed in terms of a *likelihood ratio* (LR). The LR expresses the relation between prior and posterior belief in a hypothesis variable H being true or false upon observing some evidence $\mathtt{E}_1 = \mathtt{true}$:

$$\frac{\mathrm{P}(\mathtt{H} = \mathtt{true} \mid \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false} \mid \mathtt{E}_1 = \mathtt{true})} = \frac{\mathrm{P}(\mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false})} \cdot \frac{\mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{false})}$$

Here, the fraction $\mathrm{P}(\mathtt{H} = \mathtt{true}) / \mathrm{P}(\mathtt{H} = \mathtt{false})$ is usually referred to as the prior odds and $\mathrm{P}(\mathtt{H} = \mathtt{true} \mid \mathtt{E}_1 = \mathtt{true}) / \mathrm{P}(\mathtt{H} = \mathtt{false} \mid \mathtt{E}_1 = \mathtt{true})$ as the posterior odds. The ratio $\mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{true}) / \mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{false})$ is the LR of this evidence. This formula is often viewed as an *update rule* because it can be used to compute posterior probabilities (after observing evidence) from prior probabilities (before observing that evidence). It is noteworthy that *prior* and *posterior* are notions relative to the observation of $\mathtt{E}_1 = \mathtt{true}$. The rule can be applied iteratively on multiple pieces of evidence by using the posterior of the first update as the prior of the second computation. Consider one hypothesis variable H and two observed evidence variables $\mathtt{E}_1$ and $\mathtt{E}_2$. Suppose that after observing $\mathtt{E}_1 = \mathtt{true}$, if it is also observed that $\mathtt{E}_2 = \mathtt{true}$. If we then wish to calculate the posterior odds of H conditioned on $\mathtt{E}_1 = \mathtt{true}$ and $\mathtt{E}_2 = \mathtt{true}$ (denoted as $\mathtt{E}_1 = \mathtt{true} \wedge \mathtt{E}_2 = \mathtt{true}$), we invoke the same update rule again, conditioning all probabilities on the previously processed evidence for $\mathtt{E}_1$:

$$\frac{\mathrm{P}(\mathtt{H} = \mathtt{true} \mid \mathtt{E}_2 = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false} \mid \mathtt{E}_2 = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})} = \frac{\mathrm{P}(\mathtt{H} = \mathtt{true} \mid \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false} \mid \mathtt{E}_1 = \mathtt{true})} \cdot \frac{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{false} \wedge \mathtt{E}_1 = \mathtt{true})}$$

Note how the prior of this update equals the posterior of the first application of the update rule. Also observe that all odds have been conditioned on the existing knowledge $\mathtt{E}_1 = \mathtt{true}$. Not just the posterior and the prior, but also the likelihood ratio must be conditioned on this fact.

Variables $\mathtt{V}_1$ and $\mathtt{V}_2$ are said to be *conditionally independent* given $\mathtt{V}_3$ when observing $\mathtt{V}_1$ does not change the posterior probability distribution of $\mathtt{V}_2$ given $\mathtt{V}_3$. Mathematically this is expressed as $\mathrm{P}(\mathtt{V}_1 \mid \mathtt{V}_3) = \mathrm{P}(\mathtt{V}_1 \mid \mathtt{V}_3 \wedge \mathtt{V}_2)$. When the two pieces of evidence $\mathtt{E}_1$ and $\mathtt{E}_2$ are independent (conditioned on the hypothesis H),

the likelihood ratio can be simplified to:

$$\frac{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{false} \wedge \mathtt{E}_1 = \mathtt{true})} = \frac{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{false})}$$

and the two likelihood ratios of the individual observations are then multiplied, resulting in the following simplified calculation:

$$\begin{aligned}\frac{\mathrm{P}(\mathtt{H} = \mathtt{true} \mid \mathtt{E}_2 = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false} \mid \mathtt{E}_2 = \mathtt{true} \wedge \mathtt{E}_1 = \mathtt{true})} = \frac{\mathrm{P}(\mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{H} = \mathtt{false})} \\ \cdot \frac{\mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{E}_1 = \mathtt{true} \mid \mathtt{H} = \mathtt{false})} \\ \cdot \frac{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{true})}{\mathrm{P}(\mathtt{E}_2 = \mathtt{true} \mid \mathtt{H} = \mathtt{false})}\end{aligned}$$

In practice, it is easy to forget that evidence was already accounted for in the prior and then unwittingly disregard the effect of this evidence on the likelihood ratio. When evidence is accounted for in the prior, the LR must be calculated conditional on this evidence. When this is neglected the evidence is effectively counted twice. This can lead to significant miscalculations.

It has been shown before that Bayesian updating is not always treated with the prudence it needs. A number of statistical fallacies concerning Bayesian updating have been identified [Diaconis and Freedman, 1981; Evett, 1995; Thompson and Schumann, 1987; Kahneman, 2011] throughout the literature. These fallacies, especially the *transposed conditional* fallacy, are commonly made in legal cases [Berger et al., 2011]. This fallacy—closely related to the *base rate neglect* or *prosecutors fallacy*—occurs when the probability of the evidence ($\mathbf{e}$) given a certain scenario ($\mathbf{s}$) about the crime $\mathrm{P}(\mathbf{e} \mid \mathbf{s})$, reported by the forensic expert, is mistaken for the probability of that scenario given the evidence $\mathrm{P}(\mathbf{s} \mid \mathbf{e})$. Consider, for example, a forensic expert comparing a DNA trace to a sample drawn from a particular suspect. If the scientist reports that the DNA matches, then this will often be accompanied by a numerical statement of the likelihood ratio of this match. Suppose that the expert testifies that "the probability of finding this match is a million times more likely if the suspect is the source, than if a random other person is the source". That is,

$$\frac{\mathrm{P}(\text{there is a match} \mid \text{suspect is the source})}{\mathrm{P}(\text{there is a match} \mid \text{suspect is not the source})} = 1000000$$

It is in this case tempting to conclude that the chance that the suspect did not commit the crime equals 1 in a million. However, this last step is not justified. Here we transposed the odds of the evidence given the hypothesis

$$\mathrm{P}(\text{there is a match} \mid \text{suspect is not the source})$$

as testified by the expert, for the probability of the hypothesis given the evidence

$$\mathrm{P}(\text{suspect is not the source} \mid \text{there is a match})$$

Other fallacies include the *defence fallacy*, in which an unrealistically large reference group is taken into account. In the defence fallacy it is (often implicitly) assumed that all members of a large population, for instance all people who live within a certain radius from the crime scene, are equally likely to have committed the crime. In the above example, to derive at the correct posterior probability of the suspect being the source, the reported likelihood ratio needs to be multiplied by the prior probability of the suspect being the source. The defence (hence the name of the fallacy) could argue that any person in some large population is equally likely to be the source of that DNA sample. However, this does not result in a realistic estimation of the prior probability that the specific suspect is the source, because usually not all people that live in the designated area are equally likely to leave DNA at the crime scene.

Furthermore, mistakes concerning dependence and independence between variables are not uncommon. Dependence and independence between variables is a particularly easy source of mistakes. Take, for instance, the case of Sally Clark, where the probability of two children dying of sudden infant death syndrome (SIDS) was estimated by multiplying the probability of one child dying of SIDS by itself. In reality the occurrences of SIDS in two children of the same parents are not independent at all because the death could have been caused by a genetic defect in the first child, which the second child is then likely to have as well.

When dealing with large amounts of evidence where some independence information is available, evidence can be organised in a Bayesian network (BN) [Jensen and Nielsen, 2007]. In a BN, dependencies between variables can be modelled such that the above mistakes are easier to spot and prevent. At the same time, independence, where applicable, is exploited to create a more efficient representation. More technical details of this model will be discussed in Chapter 2. BNs are one of the tools that have been put forward in the probabilistic reasoning domain. BNs have been used as a tool to model forensic evidence, and even complete legal cases. Schum and Kadane were early adopters of BNs to model legal cases [Kadane and Schum, 1996]. They modelled the famous Sacco and Vanzetti case as a Bayesian network containing variables about events as well as intentions and beliefs held by the suspects. More recently Fenton et al. [2013], Lagnado et al. [2012] and Hepler et al. [2007] have modelled (part of) legal cases in BNs. Vlek et al. have investigated how legal BNs can be built [Vlek et al., 2013, 2014] using scenarios.

A disadvantage of numerical models is that the construction of a BN is a difficult task. Especially for large networks, the number of numerical parameters that is required tends to grow fast, even though this is limited compared to the number of parameters in the full joint probability distribution where all variables depend on each other. Not alleviating this issue is the fact that the BN method requires exact, numeric estimates for all parameters for the method to be usable.

Obtaining reliable estimates of such a large number of parameters is in many cases unrealistic. A solution for this is to start with rough estimates for the unknown probabilities and perform a sensitivity analysis [Van der Gaag et al., 1999] to determine the effect of these estimations on the final result.

Secondly, even after successfully constructing a BN, its structure and reasoning results are not always easy to understand and explain to legal experts [Reddy et al., 2014]. Although BNs allow for reasoning about probabilistic evidence, which may come naturally to forensic experts familiar with mathematics and statistics, non-mathematical experts are prone to many reasoning fallacies when it comes to probabilities [Kahneman, 2011]. Moreover, the graphical structure of a BN, though intuitive for those familiar with the formalism, is easily interpreted incorrectly [Dawid, 2010]. Legal experts are not always used to probabilistic reasoning, such as calculating posterior from priors using likelihood ratios or BN evidence propagation. They are usually more familiar with argumentative models of proof. The different backgrounds of legal and forensic experts create a communication gap between them which may hamper the communication and interpretation of forensic evidence in court.

1.3 Narrative models of evidence

A third model of representing and reasoning with uncertainty exists, which uses *stories* or *scenarios* to model the discourse of events. This is often called the *narrative* approach [Wagenaar et al., 1993]. This thesis will focus on combinations of the argumentative and the probabilistic approaches, but to be able to understand the context of this research it is beneficial to briefly review this third approach as well.

A story is often modelled as a sequence of events that somehow cohere. Stories create expectations of what may happen in certain situations. The narrative approach focusses on globally coherent accounts of the events. Competing hypotheses are formulated in an attempt to explain all evidence. The narrative approach emphasises the use of multiple scenarios. In legal reasoning about evidence this can have the advantage that it forces investigators to formulate alternative hypotheses. This can avoid *tunnel vision*, which occurs when the investigators focus too much attention on the most credible scenario and fail to explore alternative scenarios, by trying to prove that scenario rather than searching for evidence that might disprove it.

A drawback of narrative models of evidence is that they are less formally well-developed. Another drawback comes from the fact that people tend to believe a story more easily if it is well-told and nicely fits within their expectations.

Links between argumentation and narrative [Bex, 2011] and narrative and probabilistic models [Vlek, 2016] have been studied in the field of artificial intelligence and law.

1.4 Research questions

The difference between argumentative and probabilistic reasoning is more than a difference in the way evidence is modelled. There is also a difference in the way people with scientific and legal backgrounds reason with this evidence. Forensic experts usually have a scientific background, where probabilistic reasoning is more common. As a result, they suffer less from common probabilistic reasoning fallacies than legal experts, who are more inclined to argumentative reasoning [Evett, 1995; Thompson and Schumann, 1987]. The differences in background between legal and forensic experts create a communication gap, which we believe can be bridged if a better understanding of the relation between the two kinds of reasoning is developed.

In this thesis we address the topic of argumentation for the design and understanding of BNs. We have the following two main research topics:

Topic 1 How can argumentation techniques aid the understanding of Bayesian networks that model legal evidence?

Topic 2 How can argumentation techniques aid the design of Bayesian network models of legal evidence?

Accordingly, we will address both main research topics by investigating translation methods from one formalism to the other. Integration methods that combine arguments with probabilities have also been proposed [Hunter, 2013; Dung and Thang, 2010; Li et al., 2012]. However, in this thesis we focus on translation methods because a translation method can potentially contribute to solving the underlying communication problem between experts from different backgrounds. This results in *explanatory argumentation*, in which arguments are used to explain probabilistic reasoning. Note that 'translation' in this case is not a lossless conversion because, at least from quantitative to qualitative representations, information is lost by definition. To help the understanding of BNs, we propose methods to extract arguments from BNs. To help the design of BNs we propose a method to translate argumentation schemes to BN fragments. In particular we will address the following questions:

Research question 1 Can explanatory arguments be extracted from a given Bayesian network?

Research question 2 Can argumentation schemes for evidential reasoning be modelled in a Bayesian network?

These can each be divided in a number of subquestions. Regarding the extraction of explanatory arguments we discuss the following subquestions:

1.1 Can Bayesian inference be translated to argumentative inference rules?

1.2 Can probabilistic arguments be constructed that respect the dynamic interactions present in Bayesian networks?

1.3 What is the computational complexity of a translation from Bayesian network to arguments via rules?
1.4 Does a translation method impose limiting constraints on the Bayesian networks that are applicable as inputs to such a system?

To construct arguments on the basis of a BN we first identify defeasible inference rules from the network. We will see in Chapter 3 that exhaustive chaining of inference rules yields interesting argumentative results but quickly becomes infeasible when the input network is larger. In Chapter 4 we subsequently study a more integrated method to identify argumentative structures in Bayesian networks. Rules are no longer enumerated and instead the structure of possible arguments is first explored after which the signs and strengths of the required inference steps are determined.

Argument extraction can help us develop a better understanding of the relation between probabilistic and argumentative reasoning by providing a translation from a BN to arguments. In this way it becomes possible to explain which arguments can and which arguments cannot be made on the basis of particular probabilistic information.

To answer the second main question we look at the following detailed subquestions regarding the construction of BNs using argumentation schemes:

2.1. Can argumentation schemes be captured by Bayesian networks?
2.2. How are critical questions to be incorporated in such a BN (fragment)?
2.3. What are the representational complexity implications of the answers to the above?
2.4. What properties does an argumentation scheme need to possess to be incorporated in a Bayesian network?

In Chapter 5 we show how argumentation schemes and critical questions can be modelled in a BN. We illustrate this with an argumentation scheme for testimony evidence. We show two methods to do this and we propose a way to integrate these methods, inheriting advantages from both.

1.5 Outline of this thesis

In the next chapter (Chapter 2) we provide the reader with necessary technical details of both argumentative and probabilistic models used in the remainder of the chapters. In Chapter 3 we present a first attempt at extracting arguments from BNs. In this chapter we present a method to extract defeasible inference rules from BNs to construct arguments by exhaustive enumeration of rule combinations. One of the disadvantages of that method is the exponential explosion of the number of arguments, many of which turn out to be redundant or irrelevant for the case. Nevertheless, a number of key observations about the nature of ar-

guments and probabilities are made and this provides a basis for the argument extraction method that is subsequently introduced in Chapter 4. The ideas described in Chapter 3 are based on earlier publications [Timmer et al., 2013, 2014; Verheij et al., 2016].

Even with a number of proposed optimisations the method in Chapter 3 generates too many arguments to be adequate from both a computational and an explanation point of view and is therefore limited to small example networks. To overcome this limitation, in Chapter 4 we introduce a more efficient method that inherits concepts from the previous chapter but generates arguments in a different way. This introduces a novel approach to the explanation of BNs and we apply this method to our legal setting. We introduce the notion of a *support graph*, which provides insight into the relations modelled in a Bayesian network. We show how these support graphs can be constructed and how they can be used for the argumentative analysis of a case modelled by a Bayesian network. This results in numerically backed arguments based on probabilistic information modelled in a BN. This idea was first introduced in [Timmer et al., 2015b,e] and its formal properties were first shown in [Timmer et al., 2015a]. In [Timmer et al., 2015d] we further developed the argument generation in this method. The work presented in Chapter 4 and published in [Timmer et al., 2016] further formalises this.

Based upon the results from Chapters 3 and 4 we formulate an answer to research question 1.

In Chapter 5 we address research question 2 about designing BNs. We propose a method to translate argumentation schemes to BN fragments. We discuss other work on this in the literature and establish that all existing methods can be classified as either of two main approaches. We highlight disadvantages of choosing one method over the other and propose a method that combines positive features of both, thereby creating a middle road. Chapter 5 extends work published in [Timmer et al., 2015c].

We discuss how the research presented in this thesis relates to other research endeavours in Chapter 6. In particular we discuss how the research relates to other integration approaches that combine argumentation and probabilities. In Chapter 7 we summarise the problem, the approach and the results. Subsequently we answer the research questions and we suggest ways in which the proposed formalisms and methods can be extended in future research.

Chapter 2

Preliminaries

In this chapter, we review technical details of argumentation and Bayesian networks required in the following chapters. We establish background knowledge and notational conventions.

2.1 Argumentation

As discussed, argumentation as a subject of scientific research can be divided into formal and informal approaches. In informal argumentation, argumentation schemes were introduced as a method to structure arguments. For instance, in legal and medical cases, it is important to draw the right conclusions and for that purpose it is important to scrutinise any argument that is involved. Argumentation schemes are particularly useful to identify possible sources of doubt in arguments. An argumentation scheme is a general pattern of argumentation that can be invoked for a particular (legal, medical or other) case. Since these patterns usually have exceptions—circumstances under which they do not apply—they are often accompanied with so-called critical questions that expose how argumentation schemes can be attacked. Argumentation schemes have been identified for various types of (legal) evidence. In Chapter 5 argumentation schemes *from position-to-know* will be discussed, which are of particular interest in the legal domain because they can be used to model testimony evidence, including expert and witness testimonies (after Prakken [2014], adapted from Walton et al. [2008] and inspired by Kadane and Schum [1996]):

W is in the position to know about H
W testifies that H
Therefore, presumably, H
Critical Questions:
1. Veracity: is W sincere?
2. Objectivity: Did W's memory function properly?
3. Observational sensitivity: Did W's senses function properly?

This scheme is abstract in the sense that it can be applied to specific testimonies by substituting W and H for actual witnesses and facts. Variations on this scheme exist for different kinds of testimonies, such as those from experts or eye witnesses. The argumentation scheme from position-to-know is just one of many that have been proposed. For instance, schemes have been constructed for *arguments from analogy*, *arguments from causation* and *arguments from excluded alternatives*, but also for practical arguments such as those *from undesired consequences* or *danger appeal arguments* [Walton et al., 2008]. Walton et al. [2008] list a total of 60 argumentation schemes divided into 4 categories each consisting of 4 or 5 subcategories. Each of these schemes has different critical questions.

Argumentation schemes are not limited to legal reasoning since they can capture any situation in which we wish to abstract from a certain pattern of reasoning. For all these cases appropriate critical questions can be formulated.

Critical questions point towards exceptional circumstances under which the *normal* inference from the evidence to the hypothesis does not apply. This kind of exception is sometimes called an *undercutter* of the argument. Sometimes critical questions also point to ways in which evidence or the hypothesis itself can be refuted ("are there other sources suggesting not-H" would be an example), but these are not relevant for the work presented in Chapter 5. Hence, we do not discuss them further here.

Argumentation schemes and critical questions can guide the construction of arguments and counterarguments because they provide guidelines for common types of evidence.

Formal argumentation is studied in the field of artificial intelligence. In formal argumentation the attack relation between arguments is often abstracted from the contents of the arguments. That is, no assumptions are made regarding the nature of arguments or what the premises, conclusions and inference rules consist of. For this purpose Dung [1995] introduced the notion of an abstract argumentation framework:

Definition 2.1 (Abstract argumentation framework [Dung, 1995] using the terminology of Modgil and Prakken [2014])**.** *An abstract argumentation framework is a pair* $(\mathcal{A}, \mathcal{D})$*, where* $\mathcal{A}$ *is a set of arguments and* $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary relation of defeat.*

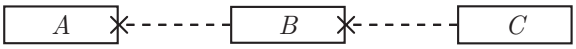
Dung's theory of abstract argumentation can be used to evaluate the acceptability status of such abstract arguments. A number of ways to formalise this has been introduced.

Definition 2.2 (Dung extensions, after Modgil and Prakken [2014])**.** *For any argument A, A is acceptable with respect to some set of arguments S iff any argument B that defeats A is itself defeated by an argument in S. A set of arguments S is conflict free if none of the arguments in S defeat each other. Then a conflict free set of arguments:*

- *S is an admissible extension iff $A \in S$ implies A is acceptable w.r.t. S;*
- *S is a complete extension iff $A \in S$ whenever A is acceptable w.r.t. S;*
- *S is a preferred extension iff it is a set inclusion maximal complete extension;*
- *S is the grounded extension iff it is the set inclusion minimum complete extension;*
- *S is a stable extension iff it is preferred and every argument outside S is defeated by at least one argument that is in S.*

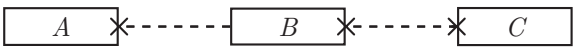
Note that in ASPIC+ the term *defeat* is used for successful *attack* and that the Dung abstract argumentation framework is populated with this defeat relation, but that Dung uses the term *attack* for this purpose. We stick to the ASPIC+ convention for the use of defeat as successful attack.

Example 2.3. *Consider the arguments A, B and C with the following defeat relation:*



So, A is defeated by B, but B is in turn defeated by C. Argument C is said to reinstate A. The set $\{A, C\}$ is conflict free because there is no defeat between A and C. This set is also a complete extension because the only argument (B) that defeats an argument in the set (A) is itself defeated by an argument (C) in the set. There are other admissible extensions such as $\emptyset$ and $\{C\}$, but these extensions are not complete because A is also acceptable with respect to $\{C\}$ and C is acceptable with respect to $\emptyset$. In fact, the set $\{A, C\}$ is the only complete extension, which is therefore also the only preferred and grounded extension. As can be seen from the defeat graph, $\{A, C\}$ defeat B and therefore this extension is also stable.

Now consider the same arguments but with the following defeat relation:



Now, B also defeats C. It can be verified that $\{A, C\}$ is still a preferred and stable extension but now $\{B\}$ is also stable and preferred; i.e., it is an admissible extension because B defends itself against the defeat from C by defeating C itself, and it is a complete extension because neither A nor C can be added to it without

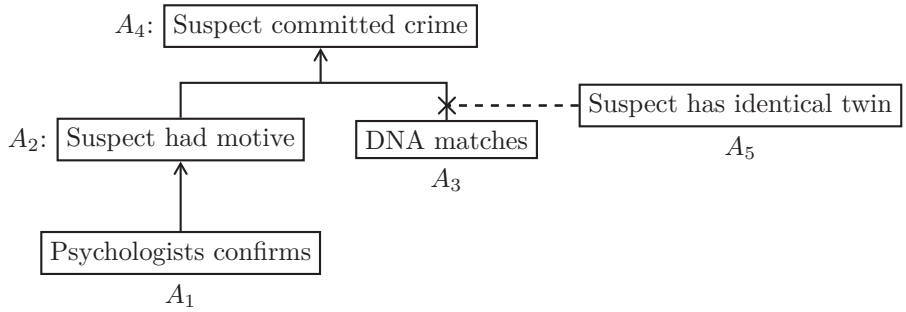


Figure 2.1: An example of arguments shown as a graph. There is one undercutting counterargument (A_5).

introducing a conflict. Finally, $\{B\}$ is a stable extension because it defeats both A and C. The grounded extension in this case is $\emptyset$. The empty set is a complete extension (in contrast to the previous example) because none of the arguments is acceptable with respect to the empty set.

The interpretation of these extensions is that in the first example a rational reasoner should believe A and C but not B. Because nothing defeats C it should be believed. Because C defeats B it follows that B cannot be believed and therefore A must again be believed. In the second example there is no unique solution. A careful reasoning will believe none of the arguments (i.e., the empty set, which is the grounded extension). However the two stable and preferred extensions $\{B\}$ and $\{A, C\}$ pose consistent sets of beliefs that can be held and the arguments in which defend each other against defeat from other arguments.

Formalisms such as ASPIC+ instantiate this framework by describing what the arguments are and how the defeat relation arises. The idea that rules can have exceptions, such as those pointed to by critical questions in the case of argumentation schemes, is also pivotal in many formal argumentation approaches.

Several formal argumentation systems—such as ASPIC+, defeasible logic programming and assumption based argumentation [Hunter, 2014]— work with *inference rules*, sometimes simply called *rules*, where premises justify conclusions. However, in common-sense reasoning (such as in legal cases) inferences are usually not strict. This is partially solved by the introduction of *defeasible* inference rules [Pollock, 1987]. A defeasible inference rule (as opposed to a *strict* inference rule) can have exceptions. One way to look at defeasible inference rules is as formalisations of argumentation schemes. Critical questions to a scheme can then be regarded as pointers to exceptions of the corresponding defeasible inference rule.

An argument applies a rule to a number of premises to derive a conclusion. The premises must satisfy the antecedents of the rule and the conclusion will be the consequent of the rule. This process is often referred to as *inference*. The premises can themselves be argued for or against; this puts arguments in a hierarchical

structure where subarguments are used to derive the required premises to draw the final conclusion.

This idea is shown in Figure 2.1, where a number of argumentative statements are shown in a graph structure. Arrows are used to indicate where inference rules are used to link statements together into arguments. For instance, in the left branch an argument (A_1) with three nested subarguments (A_1, A_2 and A_3) connected by two rules is shown. In this argument, from a psychological report it is derived that the suspect had a motive and together with a DNA match this is reason to believe that the suspect committed the alleged crime.

Undercutting and rebutting attacks between arguments with defeasible rules have been distinguished [Pollock, 1994, 1987]. A rebuttal attacks the conclusion of an argument, whereas an undercutter directly attacks the inference. An undercutter exploits the fact that a rule is not strict by posing one of the exceptional circumstances under which it does not apply.

Different formalisations of arguments and the ways in which they attack each other exist, see for example [Modgil and Prakken, 2014; Simari and Loui, 1992; Vreeswijk, 1997; Verheij, 2003a; Besnard and Hunter, 2009; Dung et al., 2009]. Where applicable, we will stay as close as possible to the ASPIC+ framework [Modgil and Prakken, 2014] for structured argumentation because it is a very flexible framework that models both the internal structure of arguments and the inter-argument attack and defeat relations. ASPIC+ features defeasible inference rules and an abstract notion of argument ordering which we use to express probabilistic inference in both Chapters 3 and 4.

In ASPIC+ a logical language ($\mathcal{L}$) describes the basic elements that can be argued about. Over this language a generalised notion of negation is defined which is called *contrariness*. This generalises negation in the sense that it does not require symmetry. As described by Pollock [1987], defeasible rules ($\mathcal{R}_d$ in ASPIC+) are differentiated from strict rules ($\mathcal{R}_s$ in ASPIC+) because defeasible rules allow for the existence of exceptions. A naming function $n(r)$ maps defeasible rules to well formed formulas of the language. This means that defeasible rules themselves can be argued about to determine the exceptions to those rules.

The logical language, the rules, and the function n together form a so-called *argumentation system*:

Definition 2.4 (Argumentation system, after Modgil and Prakken [2014])**.** *An argumentation system is a tuple* $AS{=}(\mathcal{L}, ^-, \mathcal{R}, n)$ *consisting of*

- *A logical language* $\mathcal{L}$*;*
- *A contrariness function* $^- : \mathcal{L} \mapsto 2^{\mathcal{L}}$ *over this language that assigns incompatible elements, which is a generalised form of negation;*
- *Rules* $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *such that* $\mathcal{R}_d$ *are defeasible rules of the form* $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ *and* $\mathcal{R}_s$ *are strict rules of the form* $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ *(where* φ, φ_i *are meta-variables ranging over wff in* $\mathcal{L}$*) and* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$*;*
- *A naming function* $n : \mathcal{R}_d \mapsto \mathcal{L}$*.*

To reason with the language and the rules, a knowledge base is required. Similar to how defeasible rules and strict rules are separated, presumed knowledge ($\mathcal{K}_p$) is kept separate from necessary knowledge ($\mathcal{K}_n$).

Definition 2.5 (Knowledge base, after Modgil and Prakken [2014])**.** *In an argumentation system $AS = (\mathcal{L}, ^-, \mathcal{R}, n)$, a knowledge base is a set $\mathcal{K} \subseteq \mathcal{L}$. This set is divided in necessary, or axiomatic knowledge $\mathcal{K}_n$, and presumed knowledge $\mathcal{K}_p$ for which $\mathcal{K}_n \cap \mathcal{K}_p = \emptyset$ and $\mathcal{K}_n \cup \mathcal{K}_p = \mathcal{K}$.*

The idea is that $\mathcal{K}_n$ cannot be disputed, whereas $\mathcal{K}_p$ can. The combination of an argumentation system and a knowledge base forms an argumentation theory.

Definition 2.6 (Argumentation theory, after Modgil and Prakken [2014])**.** *An argumentation theory AT is a tuple $(AS, \mathcal{K})$ where AS is an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ $\mathcal{K}$ is a knowledge base.*

An argument is then defined with respect to such an argumentation theory.

Definition 2.7 (Argument, after Modgil and Prakken [2014])**.** *Given an argumentation system AS and a knowledge base $\mathcal{K}$, an argument A is one of the following:*

- *ψ if $\psi \in \mathcal{K}_n$, and we define*
  $\text{Prem}(A) = \{\psi\}$
  $\text{Conc}(A) = \psi$
  $\text{Sub}(A) = \{\psi\}$
  $\text{TopRule}(A) =$ *undefined*
  $\text{ImmSub}(A) = \emptyset$
  $\text{DefRules}(A) = \emptyset$
- *$A_1, \ldots, A_n \to \psi$ if $A_1, \ldots, A_n$ are arguments such that there is a strict rule $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \to \psi$ in $\mathcal{R}_s$, and we define*
  $\text{Prem}(A) = \text{Prem}(A_1) \cup \ldots \cup \text{Prem}(A_n)$
  $\text{Conc}(A) = \psi$
  $\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\}$
  $\text{TopRule}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \to \psi$
  $\text{ImmSub} = \{A_1, \ldots, A_n\}$
  $\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n)$
- *$A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there is a defeasible rule $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$, and we define*
  $\text{Prem}(A) = \text{Prem}(A_1) \cup \ldots \cup \text{Prem}(A_n)$
  $\text{Conc}(A) = \psi$
  $\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\}$
  $\text{TopRule}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi$
  $\text{ImmSub} = \{A_1, \ldots, A_n\}$
  $\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n) \cup \{\text{TopRule}(A)\}$

In an argument A, ψ is referred to as the conclusion of A and sometimes written as $\text{Conc}(A)$. In the example in Figure 2.1 (and also in other figures throughout this thesis) the conclusions of the arguments A_1 through A_5 are written inside the nodes of the graph. The last-applied rule is referred to as the *top-rule* and written as $\text{TopRule}(A)$. In Figure 2.1, for instance, the top-rule of A_4 is *Suspect had motive, DNA matches* $\rightarrow$ *Suspect committed crime.* The notation $\text{ImmSub}(A)$ is used for the set of immediate subarguments of A. In Figure 2.1 A_2 and A_3 are the immediate subarguments of A_4. By subarguments we refer to the argument itself, its immediate subarguments, the immediate subarguments of its immediate subarguments, and so forth. For example, the set of subarguments of A_4 in Figure 2.1 consists of A_1, A_2, A_3 and A_4. By *premises* ($\text{Prem}(A)$) of an argument, we mean all subarguments that are an item from the knowledge base. Again looking at A_4 from Figure 2.1, $\text{Prem}(A_4) = \{A_1, A_3\}$.

Arguments can be attacked in three different ways. Note that attack does not necessarily result in defeat in the abstract argumentation framework. We will later discuss when attack results in defeat. An argument can be *undermined* by attacking it on an ordinary premise. An argument can be *rebutted* by conflicting conclusions when its conclusion is inferred by a defeasible rule, and it can be *undercut* by attacking it on the application of a defeasible rule.

Definition 2.8 (Argument attack, after Modgil and Prakken [2014])**.** *Argument A can attack another argument B on $B' \in \text{Sub}(B)$ in one of three ways:*

- *Argument A undermines argument B on B' iff $\text{Conc}(A) \in \overline{\varphi}$ for some $\varphi = B' \in \text{Sub}(B)$ that is a premise from $\mathcal{K}_p$.*
- *Argument A rebuts argument B on B' iff $\text{Conc}(A) \in \overline{\varphi}$ for some $\varphi = \text{Conc}(B')$ that is the consequent of the defeasible top-rule used in subargument $B' \in \text{Sub}(B)$.*
- *Argument A undercuts argument B on B' iff $\text{Conc}(A) \in \overline{n(r)}$ for the defeasible top-rule r that is applied in some subargument $B' \in \text{Sub}(B)$.*

In ASPIC+ attack does not imply defeat as in Definition 2.1. To determine which arguments defeat each other a preference ordering ($\preccurlyeq$) on arguments is required, for which we denote the strict version of the ordering as $A \prec B$ when both $A \preccurlyeq B$ and $A \not\succcurlyeq B$. ASPIC+ assumes an abstract version of this argument ordering but does not prescribe what ordering should be used. Such an ordering can, for instance, be defined on the basis of an ordering of the defeasible rules. The weakest-link principle [Modgil and Prakken, 2013], for instance, poses one suitable option to order arguments. This orders arguments by the strength of the defeasible rules.

Definition 2.9 (Argument defeat, after Modgil and Prakken [2014])**.** *Argument A defeats argument B iff*

- *A undercuts B on some subargument B' of B, or*
- *A undermines B on some subargument B' of B and $A \not\prec B$, or*
- *A rebuts B on some subargument B' of B and $A \not\prec B$.*

An undercutter always results in defeat because it is in a sense an asymmetric attack, since there is no negation between statements but rather between a statement and a rule. For other attacks the preference ordering determines which arguments result in defeat.

Definition 2.10 (Argumentation framework)**.** *An abstract argumentation framework corresponding to an argumentation theory $(AS, \mathcal{K})$ is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{A}$ is the smallest set of all finite arguments constructed from $cl\mathcal{K}$ in AS satisfying Definition 2.7 and $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by Definition 2.9.*

In this way, ASPIC+ generates an abstract argumentation framework in which Dung extensions can be calculated.

2.2 Bayesian networks

Probabilistic models, such as Bayesian networks (BNs), are based on the notion of random variables. A random variable $\mathtt{V}$ has a number of mutually exclusive and collectively exhaustive outcome states, denoted by vals($\mathtt{V}$). In general, the number of outcomes per variable is not limited, but Bayesian networks (BNs) assume that variables have a finite set of outcomes and, for simplicity, we will further assume throughout the rest of this thesis that variables are boolean valued if not explicitly stated otherwise.

As a notational convention we will use capital letters for variables and lower case for outcomes. When possible we will stick to lower $\mathtt{v}$ and $\neg\mathtt{v}$ as the two complementary outcomes of a variable $\mathtt{V}$. When confusion is not possible we may use $\mathtt{v}$ as a notational shortcut for the assignment $\mathtt{V} = \mathtt{v}$. That is, we write P($\mathtt{v}$) for the probability that variable $\mathtt{V}$ takes on the value $\mathtt{v}$. Note that this is different from P($\mathtt{V}$), which is not a probability but a probability distribution, i.e. a function that assigns a probability to every possible outcome of $\mathtt{V}$. In many cases we will also use short phrases for variable names such as `Suspect_guilty`. Again, the first capital letter indicates that this is the name of a variable. For such variables we will often use the states *true* and *false* as the two possible outcomes.

The notation P($\mathtt{v} \mid \mathbf{e}$) denotes the probability of $\mathtt{v}$ 'given' $\mathbf{e}$, i.e., P($\mathtt{v} \wedge \mathbf{e}$)/ P($\mathbf{e}$). As we have already noted, P($\mathtt{V} \mid \mathbf{e}$) denotes the conditional probability distribution of outcomes of $\mathtt{V}$ given the evidence $\mathbf{e}$. Similarly, P($\mathtt{V} \mid \mathbf{E}$) is used to denote a set of conditional probability distributions of $\mathtt{V}$ for all outcomes of $\mathbf{E}$.

A Bayesian network (BN) [Pearl, 1988a; Jensen and Nielsen, 2007] contains a directed acyclic graph (DAG) in which nodes correspond to random variables. In this graph Cld($\mathtt{V_i}$) and Par($\mathtt{V_i}$) will denote the children and parents respectively of a variable $\mathtt{V_i}$ in the graph. Cld($\mathbf{V'}$) (or Par($\mathbf{V'}$)) will likewise denote the union of the children (or parents) of variables in a set $\mathbf{V'} \subseteq \mathbf{V}$.

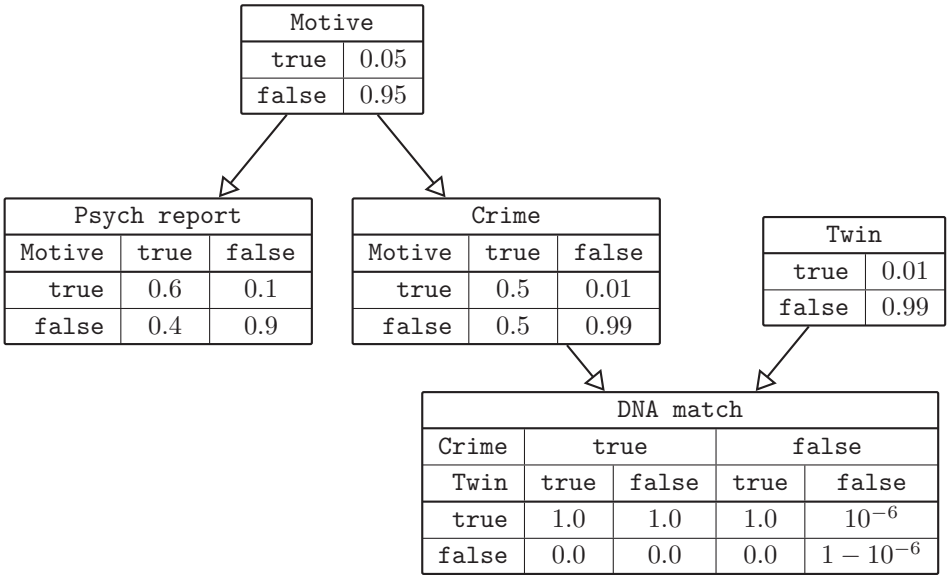


Figure 2.2: A small BN concerning a criminal case. The conditional probability distributions are shown as tables inside the nodes of the graph.

Following common graph terminology, we use the terms *ancestor* for the node itself, the parents, the parents of parents and so forth. Similarly, *descendants* of a node include the node itself, children, children of children etc. Co-parents of a node are those variables that share a common child. That is, co-parents of $\mathtt{V_i}$ include any node $\mathtt{V_j}$ for which the intersection of $\mathrm{Cld}(\mathtt{V_i})$ and $\mathrm{Cld}(\mathtt{V_j})$ is non-empty.

Variables have a number of mutually exclusive and collectively exhaustive outcomes: upon observing the variable, exactly one of the outcomes will become true. The probabilities of these outcomes are specified in conditional probability tables (CPTs) that are associated to each variable. These tables specify the conditional probabilities of the outcomes of the associated variable conditioned on the outcome configuration of all the parents in the graph.

Definition 2.11 (Bayesian network)**.** *A Bayesian network is a pair* (G, P)*. G is a directed acyclic graph* $(\mathbf{V}, \mathbf{E})$*, in which* $\mathbf{V}$ *is the set of nodes and* $\mathbf{E}$ *is a set of ordered pairs of variables* $(\mathtt{V_i}, \mathtt{V_j})$ *such that* $\mathtt{V_i} \in \mathrm{Par}(\mathtt{V_j})$*.* P *is a probability function which specifies for every variable* $\mathtt{V_i}$ *the probabilities of its outcomes conditioned on its parents* $\mathrm{Par}(\mathtt{V_i})$ *in the graph.*

Example 2.12. *An example of a BN is shown in Figure 2.2. This example concerns a criminal case with five variables describing how the occurrence of some crime correlates with a psychological report and a DNA matching report. The variables* `Motive` *and* `Twin` *model the presence of a criminal motive and the existence of an identical twin. The latter can result in a false positive in a DNA matching*

test.

A BN is a compact representation of a joint probability distribution. By using the DAG to code the independences among the variables, the joint distribution factorises into the local conditional distributions present in the node tables:

$$\mathrm{P}(\mathbf{V}) = \sum_{\mathtt{V_i} \in \mathbf{V}} \mathrm{P}(\mathtt{V_i} \mid \mathrm{Par}(\mathtt{V_i}))$$

Just as in the case with simple likelihood ratios, inference algorithms can be used to calculate any prior or posterior probability distribution over the variables in the network. This is called *evidence propagation*. Given a BN, observations can be entered by *instantiating* variables; this update is then propagated through the network, which yields a posterior probability distribution on all other variables, conditioned on those observations. If we are interested in a variable $\mathtt{V}^\star$, for instance, we can calculate the probability $\mathrm{P}(\mathtt{V}^\star = \mathtt{true} \mid \mathbf{e})$ for this variable conditioned on some evidence $\mathbf{e}$. The term *probabilistic inference* is also used for this process of calculating probabilities in a BN, but we stick to *propagation* to prevent confusion with argumentative inference. In Chapter 5 we further discuss similarities between these two kinds of inference. Note that in Pearl's original evidence propagation algorithm [Pearl, 1988a] observations were quite literally propagated along the edges of the graph because updates were processed by passing messages between parents and children in the graph. Although more efficient propagation algorithms [Lauritzen and Spiegelhalter, 1988] do not use this, the message passing algorithm provides an intuitive interpretation of the process of evidence propagation in BNs.

A number of software tools are available for modelling BNs, propagating evidence, and learning BNs from data. Although the task of evidence propagation is in general NP-hard [Cooper, 1990], most existing algorithms exploit sparseness of the graph to achieve fast computations for practical applications. In particular, such algorithms become polynomial for singly connected graphs [Pearl, 1988a].

Definition 2.13 (Singly connected)**.** *In a directed acyclic graph, a loop is a cycle in the undirected version of that graph. A graph* $(\mathbf{V}, \mathbf{E})$ *is singly connected iff it contains no loops. Otherwise, it is said to be multiply connected.*

In a BN, the directions of the arrows have no distinct meaning on their own, but collectively they constrain the conditional independences between variables. This notion of conditional independence is a dynamic concept that changes with the set of variables that is *instantiated*. A variable is said to be instantiated when one of its outcomes has been observed. In the example, observing the outcome of the `Motive` variable will make the psychological report independent of the `Crime` variable. On the other hand, observing `DNA match` will make the `Crime` and `Twin` variables dependent, which they were not before. The concept of d-separation is used to capture this dynamics and is itself defined in terms of *blocking* variables and *chains* that can be *active* or *inactive* depending on the set of instantiated

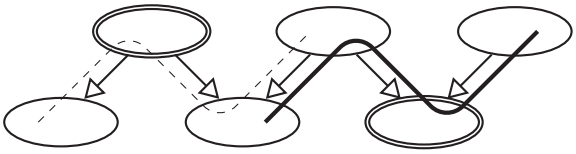


Figure 2.3: An example of an active (bold) and an inactive (dashed) chain in a BN with respect to two observations (visible by double lined nodes).

variables.

Definition 2.14 (Chain)**.** *A path in a graph is simple iff it contains no node more than once. A chain in a DAG is a simple path in the underlying undirected graph.*

A variable is a *head-to-head* node with respect to a particular chain of variables iff it has two incoming edges on that chain.

Definition 2.15 (Head-to-head node)**.** *A variable* $\mathtt{V_i}$ *is a head-to-head node with respect to a particular chain* $[\ldots, \mathtt{V_{i-1}}, \mathtt{V_i}, \mathtt{V_{i+1}}, \ldots]$ *in a DAG* $G = (\mathbf{V}, \mathbf{E})$ *iff both* $(\mathtt{V_{i-1}}, \mathtt{V_i}) \in \mathbf{E}$ *and* $(\mathtt{V_{i+1}}, \mathtt{V_i}) \in \mathbf{E}$*. That is, it has two incoming edges on that chain.*

Blocking is defined in terms of these head-to-head nodes.

Definition 2.16 (Blocking chain)**.** *A variable* $\mathtt{V}$ *on a chain* c *blocks* c *iff either*

- *it is an uninstantiated head-to-head node without instantiated descendants, or*
- *it is not a head-to-head node with respect to* c *and it is instantiated.*

A chain is active iff none of its variables is blocking it. Otherwise it is said to be inactive.

Chains (and the existence of blocked chains in particular) describe the independence relation that is represented by the DAG of the BN.

Definition 2.17 (D-separation)**.** *Sets of variables* $\mathbf{V}_A \subseteq \mathbf{V}$ *and* $\mathbf{V}_B \subseteq \mathbf{V}$ *are d-separated by a set of variables* $\mathbf{V}_C \subseteq \mathbf{V}$ *iff there are no active chains between any variable in* $\mathbf{V}_A$ *and any variable in* $\mathbf{V}_B$ *given instantiations for variables* $\mathbf{V}_C$*.*

Precisely this graphical representation of independence is what makes the BN model very powerful but also sometimes hard to interpret. Two variables in the graph are *independent* when there is no *active* chain (undirected path) between them. A chain becomes inactive when a node on the chain blocks it. A head-to-head node (such as the bottom rightmost node in Figure 2.3) blocks the chain when none of its descendants (including the node itself) are observed. Other nodes

do not block the chain unless they are themselves observed.

To complicate things further, the arrows in the BN have no significance beyond their use to describe independence as just explained. A clear intuition of what an arrow represents—and specifically what the direction of an arrow signifies in the context of the application—is missing. A BN can be used to model causal relations, but it is not necessary to model the direction of the causality by the direction of the edges. Often different graphical models can represent exactly the same independence relation. This serves to illustrate that the directions of the edges in the BN model do not have a clear intuition [Dawid, 2010].

The notions of a Markov blanket (MB) and Markov equivalence [Verma and Pearl, 1991] are useful concepts with regards to independence properties.

Definition 2.18 (Markov blanket)**.** *Given a BN graph, the Markov blanket* $\mathrm{MB}(\mathtt{V_i})$ *of a variable* $\mathtt{V_i}$ *is the set*

$$\mathrm{Cld}(\mathtt{V_i}) \cup \mathrm{Par}(\mathtt{V_i}) \cup \mathrm{Par}(\mathrm{Cld}(\mathtt{V_i})) \setminus \{\mathtt{V_i}\}$$

That is, the parents, children and parents of children of $\mathtt{V_i}$ *(but excluding* $\mathtt{V_i}$ *itself).*

Given the Markov blanket (MB) of a node, this node is independent of the rest of the network. This is an interesting property because it separates the variables that are directly related to the variable in question from the variables that are only indirectly influencing it. Markov equivalence can be defined by means of the following characterising theorem from Verma and Pearl [1991]:

Definition 2.19 (Immorality, from Andersson et al. [1997])**.** *Given a BN graph, an immorality is a tuple* $(\mathtt{V_a}, \mathtt{V_c}, \mathtt{V_b})$ *of variables such that there are directed edges* $(\mathtt{V_a}, \mathtt{V_c})$ *and* $(\mathtt{V_b}, \mathtt{V_c})$ *in the BN graph but no edges* $(\mathtt{V_a}, \mathtt{V_b})$ *or* $(\mathtt{V_b}, \mathtt{V_a})$*.*

Definition 2.20 (Graph skeleton)**.** *The skeleton of a directed graph is the underlying undirected graph.*

Theorem 2.21 (Markov equivalence, from Jensen and Nielsen [2007])**.** *Two BN graphs are said to be Markov equivalent if and only if they have the same skeleton, and the same set of immoralities.*

Markov equivalence is the concept that captures when the directions of edges do and when they do not make fundamental differences in the model. Sometimes, reversing the direction of one or more edges in the BN can be done without fundamentally changing the model. Two Markov equivalent graphs capture exactly the same independence relation among their variables.

The fact that immoralities play an important role in the determination of Markov equivalence corresponds to the fact that directions of edges (and head-to-head connections in particular) play an important role in intercausal interactions.

Head-to-head nodes can capture intercausal interactions, which are important to take into account in reasoning under uncertainty [Pearl, 1988b]. These interactions occur when two variables can cause the same reaction. In our example `Crime` and `Twin` can both cause the DNA to match. The two causes are independent except for the fact that they share a common effect. When the DNA match is not observed they are independent of each other, but once the match is observed they become dependent. However, the dependency constitutes a negative correlation, even though the `DNA match` variable features positive correlations with both parents. This is the case because observing the existence of a twin is sufficient to explain the observation and therefore *explains away* the alternative cause. In a sense, no further explanation for the DNA match is expected. When the intercausal interaction creates a stronger correlation this is called *explaining in*.

To illustrate this, suppose we consider the suspect and his twin as two causes of a profile match. This corresponds to the rightmost three nodes in Figure 2.2. Assume for now that this shared effect has been observed. Either of the two causes (the suspect committing the crime and the existence of a (guilty) identical twin) could explain the presence of the effect (the observed match). Therefore, the observation of the effect increases our belief in either of the two causes. If a twin is then discovered to exist, the profile match can then be explained by the twin committing the crime and the belief in the suspect's guilt drops. What happens is that one cause explains away the other. Although the nature of the interaction need not be causal when two edges converge in one node, we refer to all these interactions as *intercausal* because this is a mechanism that happens often when causes interact. To be able to represent such interactions is a strong requirement to adequately model causality.

2.3 Other notational and graphical conventions

In this thesis several distinct types of graphs are treated. Primarily, there is the graph of the BN. In Chapter 4 we will introduce the notion of a *support graph*. The collection of arguments connected by rules can be seen as a multi-graph (a graph in which edges are directed from a set of nodes to another set of nodes rather than from a single node to another single node). Common terminology for arguments dictates that ancestors are referred to as *subarguments* and parents as *immediate subarguments*. For the children and descendants in a graph, however, there is no distinct argumentative term. Hence, we will use standard graph terminology when required.

To notationally distinguish between nodes in different types of graphs we will write nodes $\mathtt{V_i}$ for nodes in the BN graph, N_i for nodes in the support graph and A_i for arguments. For both BN and support graph nodes we will use the notions of parents ($\text{Par}(X)$), children ($\text{Cld}(X)$), ancestors ($\text{Ancestors}(X)$) and

descendants (Descendants(X)). We also follow the convention to denote nodes by capital letters and assignments by lower case letters. We distinguish solitary nodes and assignments from sets of nodes and assignments by using a boldface font for sets.

In figures we will use oval nodes for BNs, rounded rectangles for nodes in the support graph and rectangles for arguments. Furthermore, we will indicate which variables are observed by presenting them with double outlines in figures. We similarly highlight support graph nodes for observed variables and arguments in the knowledge base.

Chapter 3

Extracting arguments from Bayesian networks: a first attempt

In this chapter we investigate how arguments can be extracted from Bayesian networks (BNs), by identifying inference rules based on the structure and probabilities in the BN. This results in a method to build arguments from BNs.

The extracted arguments reflect the knowledge that is represented by the BN. Such a method can facilitate the explanation of forensic BNs, which is one of the main goals of this thesis. To extract arguments from BNs, several steps are needed. The first step is to identify variable assignments as the basic elements that the argumentation can use. When the language of the argumentation has been established, inference rules can be extracted from the BN. To do so, a measure of strength is required such that a list of rules can be identified that provide sufficient support to reason from observed variable assignments to assignments of other variables and to rank such rules. Those rules can be combined into arguments using an instantiation of the ASPIC+ framework for structured argumentation. We illustrated the effects of different choices of strength measure with examples. How these arguments are useful in a legal setting is also discussed.

The method in this chapter is a precursor to the method in the next chapter in that the same intuitions about rules and inferential strength will be built upon in that chapter. The focus of the current chapter is to establish a formal connection between BNs and argumentation, whereas the next chapter will focus on a more efficient way to produce arguments from a BN.

3.1 Introduction to argument extraction

To automatically extract arguments from BNs it is necessary to be able to identify *inference rules*. It is noteworthy that the term *inference* is used in both argumentation and BNs, although it is used slightly differently. In BNs inference is the process of computing probabilities from the network. In argumentation inference typically is the process of drawing conclusions from premises. The possibility to make such an inference can be represented by an *inference rule* (sometimes simply referred to as a *rule*). An inference rule has a number of premises and a conclusion. In our case these will all be assignments to variables. At the core of our argument extraction method lies the idea that inference rules correspond to probabilistic correlations. When two variables share a strong probabilistic correlation, then knowledge of the outcome of one variable yields strong support for one of the outcomes of the other variable.

It is in the very nature of a joint probability distribution that the strength and sign of a correlation between variables can heavily depend on surrounding observations. That is, sometimes additional information can reverse or weaken the influence between variables. We note that this resembles undercutting, and we therefore define an undercutter as a variable assignment that reduces the strength of a correlation. A rule is then undercut by any further observation that weakens the influence between the premises and the conclusion of the rule.

3.2 A running example

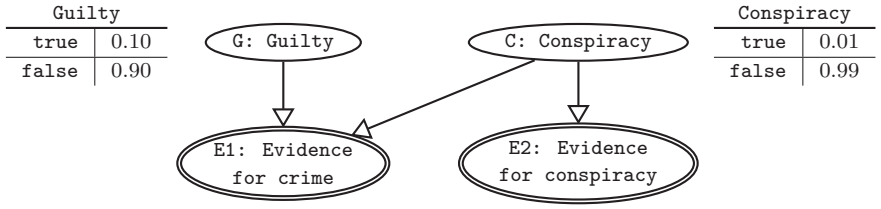
In order to illustrate our approach we use a single running example in this chapter. As can be seen below, we choose an example that satisfies the following requirements:

- The network is relatively small for illustration purposes;
- It has at least one head-to-head connection so that we can demonstrate how intercausal reasoning is captured;
- It does not need to be very realistic and can represent a legal case at a fairly abstract level.

To be able to demonstrate all aspects of BN reasoning we should have a few variables, connected by edges, and at least one head-to-head connection. For this purpose we use a fictional legal case.

3.2.1 The case

For this example we are not really bothered with the details of the crime and the precise evidence, but just presume that some evidence was found and that it correlates to a crime as described. Suppose that, besides the crime, there is another explanation for this evidence, namely that there is a conspiracy against the suspect



Evidence of crime

Guilty	true		false	
Conspiracy	true	false	true	false
true	0.90	0.80	0.90	0.001
false	0.10	0.20	0.10	0.999

Evidence of conspiracy

Conspiracy	true	false
true	0.90	0.10
false	0.10	0.90

Figure 3.1: A legal example of a BN in which a crime with a possible conspiracy has been modelled. Double outlines indicate instantiated variables. Both of these evidence variables are observed to be true in this example.

and that he is being framed. Evidence for such a conspiracy is modelled too.

A network that can represent this case is displayed in Figure 3.1. It contains four variables describing the following events:

G/Guilty: the crime was committed by the suspect as described by the prosecution.

C/Conspiracy: someone tampered with the evidence after it was recovered from the crime scene.

E1/Evidence of crime: incriminating evidence is found during the investigation.

E2/Evidence of conspiracy: further evidence is found that suggests that the evidence E1 was tampered with.

Note that we use the long names (Guilty, Evidence of crime, …) and abbreviations (G, E1, …respectively) interchangeably to denote the same variable. This is done to eliminate clutter in figures with many variables but provide more clarity where space allows it.

In this example the occurrence of the crime is modelled by the variable Guilty. This is indeed directly connected to Evidence of crime, which represents the evidence that the crime occurred as described by the Guilty node. Similarly an arrow is drawn in the BN from Conspiracy to Evidence of conspiracy. Evidence for something is clearly correlated to the event for which it is evidence. The arrows from the Guilty and Conspiracy nodes are drawn towards the evidence nodes, creating a head-to-head connection. This is done to make Guilty and Conspiracy independent causes of E1. Independent causes, as the name suggests, are probabilistically independent but they become dependent when the common effect has been observed. This is exactly what happens with a head-to-head connection. The edge from Conspiracy to Evidence of Conspiracy (E2) is drawn in this direction because a conspiracy causes E2 to be true. There is no technical reason why this

arrow could not have been drawn the other way around, because reversal of this arrow does not create or remove immoralities (as described in Chapter 2). We adhere to the tradition of following the causal or temporal direction of events when designing a BN by hand [Jensen and Nielsen, 2007]. Another reason to do so is that it is often more intuitive to estimate probabilities for effects conditioned on causes (such as P(E2 | C)) than to estimate probabilities of causes given effects (P(C | E2)).

For this example the conditional probability tables (CPTs) have been filled with fictional probabilities. We have assumed that the probability of guilt is 0.1 without further evidence. The prior probability of a conspiracy is ten times lower at 0.01. The conditional probabilities are chosen such that the variables correlate as one would expect from the description of the example. For E2 this means a probability of 0.9 if there is a conspiracy and 0.1 otherwise. The evidence for a crime (E1) has two parents and is therefore conditioned on both Guilty and Conspiracy. In the case that the suspect is indeed guilty, the probability of finding this evidence is estimated at 0.9 or 0.8 depending on whether there was also a conspiracy. If there is a conspiracy, then the probability of finding this evidence is set to 0.9 regardless of whether the suspect actually committed the crime. If the suspect is not guilty and there is no conspiracy, then the probability of finding this evidence is very low (0.001). This means that the evidence is incriminating to some extent.

3.2.2 An argument model of the same scenario

Parallel to the two notions of inference there are two other notions in probabilistic and argumentative reasoning that share some similarities. These are *explaining away* and *undercutting*. In the probabilistic model, besides the crime, there is another explanation for the evidence, namely that there is a conspiracy against the suspect and that he is being framed. Either explanation could bring about the evidence. This is modelled in the BN by a head-to-head connection. When the common descendent (E1 in our running example) is first observed, the posterior probability of both G and C increases. If we then enter evidence for a conspiracy in the BN this explains the evidence, and the belief in the guilt of the suspect drops, i.e. the posterior probability of G becomes a lot lower. This phenomenon is called *explaining away*. The interesting part of such a head-to-head connection is that it can model an induced dependency between independent causes of a common effect. Such a dependency arises only after evidence has been entered. In a Bayesian network such a relation can be modelled by a head-to-head connection. The strength of the intercausal relation is determined by the probabilities in for the head-to-head node.

In argumentation, *undercutting* models a related construction. An undercutter is a circumstance under which an inference changes in strength. This is remarkably similar to the role of explaining away. That is, observation of some evidence leads us to believe some conclusion, but another explanation of that evidence may undercut this inference. We can show this in the example. There are two pieces

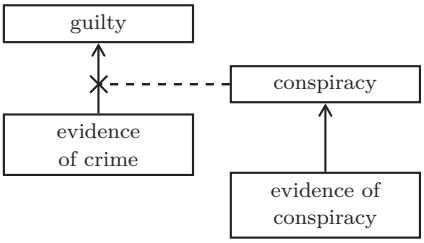


Figure 3.2: Inferences in argument graphs are displayed as normal arrows and the undercutting attack is visualised by a cross-headed arrow.

of evidence that can be represented by information in the knowledge base of the argumentative approach. Both of the corresponding evidence variables (`E1` and `E2`) are probabilistically correlated to the variable for which they are evidence (`G` and `C` respectively) so we expect to find some rules that can be used to derive from that knowledge base an argument for the guilt of the suspect and an argument for a conspiracy. In the BN there is the typical explaining away effect from the conspiracy to the guilt of the suspect. In an argumentation system this can be modelled by undercutting the argument for guilt. All of this is shown in the form of an argument graph in Figure 3.2. It shows two straightforward inferences (from evidence to a corresponding conclusion) where the second conclusion undercuts the first by giving an alternative explanation for the evidence.

3.3 Towards argument extraction

The similarities between argumentative and probabilistic inference and between probabilistic explaining away and argumentative undercutting are the basis for our argument extraction method. Recall that an ASPIC+ argumentation system must have the following elements:

- A logical language $\mathcal{L}$ with:
  - A contrariness function $\bar{\cdot} : \mathcal{L} \mapsto 2^{\mathcal{L}}$;
  - A naming function $n(\cdot) : \mathcal{R}_d \mapsto \mathcal{L}$;
- A set of strict rules $\mathcal{R}_s$;
- A set of defeasible rules $\mathcal{R}_d$.
- Sets of presumed and necessary knowledge $\mathcal{K}_p$ and $\mathcal{K}_n$ for which $\mathcal{K}_p \cap \mathcal{K}_n = \emptyset$
- A preference ordering $\preccurlyeq$ on arguments

The first thing that is required to construct any sensible argument is the specification of a logical language of elements that can be argued about. A natural choice for this logical language is to take assignments to variables as the atomic elements of our argumentation, because we wish to translate BNs to arguments. To build a formal argumentation system we go through a number of steps. First, from the BN rules are extracted. With those rules arguments can be built. From the same BN,

undercutters are extracted that can be used to specify how arguments attack each other. All of these together lead to an ASPIC+ formalisation of argumentation that has at least the following elements:

Definition 3.1 (BN Argumentation System (Preliminary))**.** *Suppose a BN is given with nodes* $\boldsymbol{V}$*. Let* $vals(\mathtt{V_i})$ *denote the values that variable* $\mathtt{V_i} \in \boldsymbol{V}$ *can take on and let* $\boldsymbol{e}$ *denote the evidence. We instantiate the ASPIC+ framework as follows:*

$$\mathcal{L} = \left(\bigcup_{\mathtt{V_i} \in \boldsymbol{V}} \bigcup_{\mathtt{v_{ij}} \in vals(\mathtt{V_i})} \{\mathtt{V_i} = \mathtt{v_{ij}}\} \right) \cup \{n(r) | r \in \mathcal{R}_d\}$$
$$\overline{\mathtt{V_i} = \mathtt{v_{ij}}} = \{\mathtt{V_i} = \mathtt{v_{ik}} \,|\, j \neq k\}$$
$$\mathcal{R}_s = \emptyset$$
$$\mathcal{K}_p = \emptyset$$
$$\mathcal{K}_n = \{\mathtt{V_i} = \mathtt{v_{ij}} | (\mathtt{V} = \mathtt{v_{ij}}) \textit{ occurs in } \boldsymbol{e}\}$$

Note that this does not yet specify the set of defeasible rules $\mathcal{R}_d$ and the contrariness of rules $\overline{n(r)}$. These are to be extracted from the BN.

As elements (antecedents and consequent) of our rules we take propositions that are assignments $\mathtt{V} = \mathtt{v_i}$ of values to variables from the BN. Furthermore, the language contains names for rules ($n(r)$) because ASPIC+ requires that rules map to elements of the language. Since these rules are clearly different from variable assignments we add them to the language separately. Contrariness for variable assignments is defined as other assignments to the same variable. Rules and their contrariness will be defined later. Instantiated variables in the BN are put in $\mathcal{K}_n$ because that guarantees that no defeasible argument for the other outcome can overrule something that has been observed.

The aim of the current chapter is to provide suitable definitions of defeasible rules $\mathcal{R}_d$, of contrariness for rules $\overline{n(r)}$ and rule strength such that argument strength can be determined. A number of iterative improvements will be introduced which lead to a final definition of a BN argumentation system in Definition 3.22. At this point we have already determined the language and the associated contrariness relation for variable assignments. For rule names $n(r)$ the contrariness relation will be defined later. We opt to put the probabilistic rules in $\mathcal{R}_d$ because they can be overturned by further observations.

In the following sections we define how both rules and exceptions to rules are extracted from the BN, we introduce a number of iterative refinements of that method and we discuss the effects of the choices that we make on the way and we illustrate the results on the given example case.

3.3.1 Rules and rule strengths

The first step in constructing arguments is to define which inference rules apply and how they are ordered. We extract inference rules from a BN by looking at a measure of rule strength that is based on conditional probabilities. Evidence

entered in the BN is maintained and used during the extraction of rules, which guarantees that the strengths of the rules are calculated in the context of the available evidence. Due to the dynamic nature of BNs, dependencies and independences (and therefore also any measure of inferential strength that is based on probabilities) may change with every new observation. Therefore it is important to assess the strength of rules in the context of the available evidence.

Candidate rules for every consequent $\mathtt{V} = \mathtt{v_i}$ are identified by enumerating value assignments to all non-empty subsets of parents, children and parents of children of $\mathtt{V_i}$ in the BN graph. Parents and children can have a direct effect on the consequent node because there is a statistical correlation between those nodes. Parents of children can have an intercausal relation with the consequent. This often takes the form of explaining away, which means that they have a negative effect on that node. Recall from Chapter 2 that these nodes together (parents, children and parents of children) are the Markov blanket (MB) of a node and form the minimal set that suffices to explain statements about that node. This is the case because given its MB a node is conditionally independent of any other node and therefore cannot be influenced by or correlated to those nodes.

For each candidate rule we compute its strength using an incremental measure of strength. For this, an inferential strength measure can be used [Crupi et al., 2007]. In this case we take inspiration from Keynes' measure of confirmation [Keynes, 1921]. Other probabilistic measures to evaluate rule strength exist. In Section 3.6 we show how these measures compare. For now we use a variation of the following confirmation measure to determine the strengths of rules. We first introduce Keynes' measure of confirmation. Such a measure cannot directly be used as a measure of strength, which we show by means of an example, after which we introduce our final measure of strength that is inspired by Keynes' measure of confirmation.

Definition 3.2 (Keynes' confirmation measure, adapted from Crupi et al. [2007])**.** *Given a defeasible rule $r : p_1, \ldots, p_n \Rightarrow c$ with premises p_1 through p_n and conclusion c, the confirmation measure of the rule r is defined to be:*

$$\text{confirmation}\,(r) = \frac{\mathrm{P}\,(\,c \mid p_1 \wedge \ldots \wedge p_n\,)}{\mathrm{P}\,(\,c\,)}$$

Note that if we use probabilistic confirmation as rule strength, this definition allows rules with a strength that is less than one, in which case the premises $p_1, \ldots, p_n$ actually make the conclusion c less likely. For binary valued variables, which we assume throughout this thesis, it is in such cases always the case that the rule with the negation of the conclusion $(p_1, \ldots, p_n \Rightarrow \neg c)$ has strength greater than one in such cases. For this reason rules with a strength less than one will be filtered.

Example 3.3. *Let us look at the example BN from Figure 3.1. Take, for instance, the variables* `E1` *(*`Evidence of crime`*) and* `G` *(*`Guilty`*) and the rule*

$r_1 : \texttt{E1} = \texttt{true} \Rightarrow \texttt{G} = \texttt{true}$. *The strength of this rule is:*

$$\text{confirmation}(r_1) = \frac{\text{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true})}{\text{P}(\texttt{G} = \texttt{true})} \simeq \frac{0.899}{0.100} = 8.99$$

The rule with the opposite conclusion ($r_2 : \texttt{E1} = \texttt{true} \Rightarrow \texttt{G} = \texttt{false}$) indeed has a confirmation less than one:

$$\text{confirmation}(r_2) = \frac{\text{P}(\texttt{G} = \texttt{false} \mid \texttt{E1} = \texttt{true})}{\text{P}(\texttt{G} = \texttt{false})} \simeq \frac{0.101}{0.900} = 0.112 < 1$$

Recall that the correlation between variables may depend on observations of other variables. The above definition of confirmation, however, does not take this into account and hence using it directly as the measure of strength is undesirable. In the probabilistic setting some variables may have been entered as observed evidence, and this collection of observations typically affects the ways in which other variables (including those that occur in the rule) influence each other. Actually, the confirmation measure, as well as the strength measure of a rule, can depend heavily on the other observations in the BN. For this reason it is necessary that observations in the BN are taken into account when calculating the strength of a probabilistic rule. A naive way to do this would be to condition the entire confirmation measure expression on the available evidence. Given evidence ($\boldsymbol{e}$), and again a defeasible rule $r : p_1, \ldots, p_n \Rightarrow c$ with premises p_1 through p_n and conclusion c, the strength of the rule r could be defined as:

$$\text{naive\_strength}\,(r) = \frac{\text{P}\,(\,c \mid p_1 \wedge \ldots \wedge p_n \wedge \boldsymbol{e}\,)}{\text{P}\,(\,c \mid \qquad\qquad\quad \boldsymbol{e}\,)}$$

However, conditioning all probabilities in the definition of the strength on the full collection of evidence leads to counterintuitive results, as can be demonstrated by the following example.

Example 3.4. *Consider again the BN from Figure 3.1. Now, let us assume that the evidence consists of the single observation that* $\texttt{Evidenceofcrime} = \texttt{true}$. *Using the above naive strength measure would result in the following strength for the rule* $r_1 : (\texttt{E1} = \texttt{true}) \Rightarrow (\texttt{G} = \texttt{true})$*:*

$$\frac{\text{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true} \wedge \texttt{E1} = \texttt{true})}{\text{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true})} = \frac{\text{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true})}{\text{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true})} = 1$$

That is, we see that the effect of the premise ($\texttt{E1} = \texttt{true}$*) is obscured by the fact that the premise is already in the evidence: once the effect of evidence for* $\texttt{E1}$ *is processed, it no longer affects* $\texttt{G}$*. Instead, we are more interested in how the probability of* $\texttt{G}$ *will change as a result of obtaining evidence for* $\texttt{E}$*.*

More generally speaking, the issue of evidence obscuring the effect of premises on conclusions will always arise when either a premise or the conclusion concerns

variables for which observations have been entered. The strength of a rule should reflect the effect that the premises can have on the conclusion. However, this effect cannot be measured in a context in which these premises are already present. Therefore, the strength of a rule should be evaluated in the context of evidence excluding observations of variables that occur in that same rule. This measures the counterfactual change in the probability of the conclusion, if the premises were to be observed, given that all other evidence is unaltered. For this purpose we define a notion of *context* for rule strength evaluation.

Definition 3.5 (Context of a rule)**.** *Let* $\mathbf{V}_e \subseteq \mathbf{V}$ *be the set of instantiated variables, and let* $\boldsymbol{e}$ *be the conjunction of assignments to the variables in* $\mathbf{V}_e$*. Consider the candidate rule* $r : (\mathtt{V_1} = \mathtt{v_1}), \ldots, (\mathtt{V_n} = \mathtt{v_n}) \Rightarrow (\mathtt{V_c} = \mathtt{v_c})$*. The context* $\boldsymbol{e}\backslash r$ *of rule* r *is an assignment to the variables* $\mathbf{V}_e \setminus \{\mathtt{V_1}, \ldots, \mathtt{V_n}, \mathtt{V_c}\}$ *such that* $\boldsymbol{e}\backslash r$ *is logically consistent with* $\boldsymbol{e}$*.*

That is, the context equals the evidence except those assignments that assign values to variables that occur in the rule itself.

Example 3.6. *To evaluate the strength of the rule* $r : (\mathtt{E1} = \mathtt{true}) \Rightarrow (\mathtt{G} = \mathtt{true})$ *with the evidence* $\boldsymbol{e} = (\mathtt{E1} = \mathtt{true} \wedge \mathtt{E2} = \mathtt{false})$ *the context* $\boldsymbol{e}\backslash r$ *is* $(\mathtt{E2} = \mathtt{false})$*.*

The notion of context leads to the final definition of rule strength. The strength of a candidate rule $p_1, \ldots, p_n \Rightarrow c$ is evaluated in a particular context $\boldsymbol{e}\backslash r$.

Definition 3.7 (Contextual strength measure)**.** *Given evidence* $\boldsymbol{e}$ *for* $\mathbf{V}_e$ *and a defeasible rule* $r : p_1, \ldots, p_n \Rightarrow c$ *with premises* p_1 *through* p_n *and conclusion* c*, the strength of the rule* r *is defined with respect to the evidential context* $\boldsymbol{e}\backslash r$*:*

$$\text{strength}\,(r, \boldsymbol{e}) = \frac{\mathrm{P}\,(\,c \mid p_1 \wedge \ldots \wedge p_n \wedge \boldsymbol{e}\backslash r\,)}{\mathrm{P}\,(\,c \mid \qquad\qquad\quad \boldsymbol{e}\backslash r\,)}$$

All actual evidence in the BN is thus considered during the calculation of the strengths, excluding assignments to the variables that occur in the rule itself.

Example 3.8. *Consider again the BN from Figure 3.1. Let us now assume that the evidence consists of the following two observations*

$$\begin{aligned} &\mathtt{E1} = \mathtt{true} \\ &\mathtt{E2} = \mathtt{true} \end{aligned}$$

Using the strength measure as defined above results in the following strength for the rule $r_1 : (\mathtt{E1} = \mathtt{true}) \Rightarrow (\mathtt{G} = \mathtt{true})$*:*

$$\frac{\mathrm{P}(\mathtt{G} = \mathtt{true} \mid \mathtt{E1} = \mathtt{true} \wedge \mathtt{E2} = \mathtt{true})}{\mathrm{P}(\mathtt{G} = \mathtt{true} \mid \mathtt{E2} = \mathtt{true})} \approx 5.41$$

With this measure of rule strength we can define the set of rules that are to be used in the argumentation system. We choose to allow only the rules with a strength greater than one. In that case the probability of the consequent increases when the antecedents are observed and the rule must be considered to have some inferential power. We choose the threshold of one because rules with strength one actually describe independence between the premises and the conclusion. Rules with a strength below one describe a negative effect of the evidence, in which case there is another rule which has a greater than one strength, for the opposite outcome of the conclusion.

We now address the question which sets of variable assignments to take for the premises and the conclusions of rules. As we have said before, rules are enumerated that have a single variable assignment as their conclusion and for which premises assign outcomes to subsets of the MB of the node associated with the consequent. The reason to do this is to limit rules to premises that are directly relevant. A first version of a method would be to take every possible subset of variable assignments as the premises of a rule and match those with every possible conclusion assignment. However, the number of rules that are created in this way grows extremely fast with larger networks and many of these rules are nonsensical.

Example 3.9. *When constructing a set of premises for a defeasible rule, taking subsets of all possible variable assignments to* $\mathtt{V}_1, \mathtt{V}_2, \ldots$*, can result in the following undesirable situations:*
rules with conflicting premises *such as* $\mathtt{v}_3, \neg\mathtt{v}_1, \neg\mathtt{v}_3 \Rightarrow \mathtt{v}_4$
rules without premises *such as* $\Rightarrow \mathtt{v}_4$
rules with a conflicting conclusion *such as* $\mathtt{v}_2, \mathtt{v}_3 \Rightarrow \neg\mathtt{v}_2$
rules with a redundant conclusion *such as* $\mathtt{v}_2, \mathtt{v}_3 \Rightarrow \mathtt{v}_2$

Note that these examples are already limited to rules with conclusions with a single variable assignment. This is because a rule with a conjunction as its conclusion does not add any value, in terms of derivable arguments, on top of the individual rules for the same outcomes. This will become clear when we demonstrate how arguments can be built using the extracted rules. For premises, it should be clear that limiting rules to have exactly one premise reduces the expressiveness of the model significantly. In probabilistic reasoning it is often the case that a particular combination of variable assignments achieves a different effect than what is to be expected on the basis of these individual variables' assignments.

Besides the counterintuitive rules in the example above, there are more rules that can be eliminated to reduce the number of possible rules without limiting what is derivable by arguments built from those rules.

Example 3.10. *Consider variable* `G` *from Figure 3.1 as a possible conclusion. Rules with premises that assign values to* `E1` *naturally capture the possible inference from evidence to hypothesis. Rules with premises that assign values to* `C` *capture*

the possibly weakening influence of a conspiracy on this influence (i.e., explaining away). Evidence for E2*, naturally, has a similar weakening effect on* G*. However, this effect is the composition of the direct effect of* E2 *on* C *and the aforementioned explaining-away effect of* C *on* G*. A rule from* E2 *directly to* G *is therefore redundant.*

In probabilistic reasoning, observations of E2 *are also propagated from* E2*, via* C *and* E1 *to* G*. Therefore, an argument from* E2 *to* G *via* C *and explaining away the effect* E1 *is actually a more accurate representation of the probabilistic reasoning than an argument from* E2 *directly to* G*.*

To generalise this, we take advantage of the notion of a Markov blanket. Recall that the MB of a node consists of parents, children and parents of children of that node and that given its MB this node is probabilistically independent of all other nodes in the graph. This means that we can limit the search for appropriate premises for a conclusion about node $\mathtt{V_c}$ to its MB. As an effect, in the argumentation, arguments with conclusions about $\mathtt{V_c}$ have immediate subarguments for conclusions about the MB of $\mathtt{V_c}$. This intuitively corresponds to what is modelled in the BN since in a probabilistic setting influence is also propagated along the edges of the BN.

Definition 3.11 (Rules)**.** *Given a BN with variables* $\boldsymbol{V}$ *and evidence* $\boldsymbol{e}$ *for variables* $\mathbf{V}_e \subseteq \mathbf{V}$*, let* c *be an assignment to some variable* $\mathtt{C} \in \boldsymbol{V}$ *in the BN and* $\mathbf{p}$ *be a conjunction of assignments to a non-empty subset of nodes in the MB of* C*.* $\mathcal{R}_d$ *consists of all rules of the following form:*

$$r_i : (\boldsymbol{p} \Rightarrow \mathtt{c}) \quad \textit{such that} \quad \textit{strength}(r_i, \boldsymbol{e}) > T$$

where T *is a threshold.*

Appropriate values for the threshold value T depend on the measure of strength that is used. For the measure that we use in all examples a value of 1.0 intuitively provides a boundary between a (possibly tiny) positive and a negative effect. This is because the measure of strength expresses the change in probability as a fraction. When this fraction is greater than one, there is a positive effect of the premises on the conclusion. A threshold of one, in fact, distinguishes increasing from decreasing posterior probabilities. Note that at this point we allow rules to draw conclusions for all nodes, including the nodes in $\mathbf{V}_e$. This will be addressed in Section 3.4.

Summarising, we must meet the following criteria when constructing rules:

- conclusions assign an outcome to a single variable;
- antecedents are conjunctions of assignments to the MB of the variable assigned to in the conclusions;
- antecedents are logically consistent with each other;
- as a whole, the rule has sufficient strength.

The above considerations are put together in Algorithm 3.1, which constructs a set of what we call *probabilistic inference rules* to be used as defeasible inference rules

in the arguments. This algorithm works by exhaustively checking all possibilities.

Note that this is computationally not the most efficient way to iterate over all possible rules but Algorithm 3.1 is written for clarity and intended to illustrate the process. In terms of complexity classes, no substantial improvement is possible since both enumerating subsets (of the Markov blanket) and the propagation of probabilities are problems that are known to grow exponentially with the size of the networks [Cooper, 1990].

Example 3.12. *Given our running example, again presume that* `Evidence of crime` *and* `Evidence of conspiracy` *have both been observed. Even though the example is considerably small and we have limited the set of premises of rules to subsets of the MB, it turns out that 62 rules can be extracted. Using the following examples we highlight a number of observations about these rules.*

r_{120} : (`EvidenceOfCrime` = `true`) ⇒ (`Guilty` = `true`)

r_{118} : (`EvidenceOfCrime` = `false`) ⇒ (`Guilty` = `false`)

r_{128} : (`Conspiracy` = `true`) ⇒ (`Guilty` = `false`)

r_{61} : (`EvidenceofConspiracy` = `true`) ⇒ (`Conspiracy` = `true`)

r_{108} : (`EvidenceOfCrime` = `true`), (`Conspiracy` = `false`) ⇒ (`Guilty` = `true`)

r_{67} : (`EvidenceofConspiracy` = `true`), (`Guilty` = `false`) ⇒ (`Conspiracy` = `true`)

r_{142} : (`Conspiracy` = `true`) ⇒ (`EvidenceOfCrime` = `true`)

Note that rule numbers exceed the number of 62 possible rules because our implementation discards candidate rules with strengths below the threshold after they

function `RuleExtraction(`G`)`:
- **Input:** A BN with the graph $G = (\mathbf{V}, \mathbf{E})$ for variables $\mathbf{V}$ and edges $\mathbf{E}$
- **Input:** Evidence $\boldsymbol{e}$
- **Output:** a list of rules satisfying Definition 3.11
- ruleList = $\emptyset$
- **foreach** $\mathtt{V_c} \in \mathbf{V}$ **do**
  - **foreach** *outcome* `c` *of* $\mathtt{V_c}$ **do**
    - **foreach** *non-empty* $\mathbf{P} \subseteq$ `MarkovBlanket(`$\mathtt{V_c}$`)` **do**
      - **foreach** *assignment* $\mathbf{p}$ *to* $\mathbf{P}$ **do**
        - **if** strength$((\mathbf{p} \Rightarrow \mathtt{c}),\ \boldsymbol{e}) > 1$ **then**
          - add the rule $\mathbf{p} \Rightarrow \mathtt{c}$ to ruleList
        - **end**
      - **end**
    - **end**
  - **end**
- **end**
- **return** *ruleList*

Algorithm 3.1: Rule extraction algorithm.

have been numbered. Also observe that rules are being constructed for antecedents that will never be satisfied. In this case, rule r_{118}, for instance, presumes that no evidence for a crime was found. Even though this rule is probabilistically correct, it can never be applied in an argument and will therefore remain unused in later stages of argument construction. This is something that will be addressed in Chapter 4. The listed rules have the following strengths:

$$\begin{aligned}\text{strength}(r_{120}) &= 5.419\\ \text{strength}(r_{118}) &= 1.086\\ \text{strength}(r_{128}) &= 1.964\\ \text{strength}(r_{61}) &= 4.977\\ \text{strength}(r_{108}) &= 9.888\\ \text{strength}(r_{67}) &= 9.779\\ \text{strength}(r_{142}) &= 6.033\end{aligned}$$

This corresponds to common intuitions about the evidential force of the different pieces of evidence. For instance rule r_{67} is considerably stronger than r_{61}.

3.3.2 Exceptions to rules

Since we extract defeasible rules, we wish to know under which circumstances there are exceptions to those rules. As already mentioned, undercutting in the argumentation system can be based on explaining away in the BN. In a BN setting, inferences can often be contradicted or weakened by observing further evidence, particularly in the case of explaining away. Therefore, we identify undercutting variable assignments by checking if the measure of strength drops below a given threshold when conditioning on a potentially undercutting variable assignment.

Definition 3.13 (Exceptions to rules)**.** *Given a rule $r : p_1, \ldots, p_n \Rightarrow c$, evidence $\boldsymbol{e}$, and a threshold T. Let p_1 be the assignment $\mathtt{V_1} = \mathtt{v_1}$, and p_2 the assignment $\mathtt{V_2} = \mathtt{v_2}$ and so forth. Let c be the assignment $\mathtt{V_c} = \mathtt{v_c}$. Exceptions to this rule are defined as:*

$$\overline{n(r)} = \left\{ (\mathtt{V_u} = \mathtt{v_u}) \in \mathcal{L} \;\middle|\; \begin{array}{l} \mathtt{V_u} \notin \{\mathtt{V_1}, \ldots, \mathtt{V_n}, \mathtt{V_c}\}, \textit{ and} \\ \text{strength}(r, \boldsymbol{e} \wedge (\mathtt{V_u} = \mathtt{v_u})) \leq T \end{array} \right\}$$

A straightforward choice for the threshold is to use the same value as for the selection of rules. When strength equals 1.0, it actually encodes independence between premises and the conclusion, which should be considered a successful undercut. This resembles, to some extent, the concept of *subproperty defeat*, which was introduced by Pollock [1995].

An algorithm similar to the one above can easily be constructed to exhaustively enumerate all undercutters for all rules. Note that we use the term *undercutter* for a variable assignment that invalidates a rule as well as an argument that uses such an assignment to attack another argument. From the context of the statement it will always be clear whether this refers to an element of the language or to an argument.

Example 3.14. *For the rules in Example 3.12 the following undercutters can be found:*

$$\overline{n(r_{120})} = \{\texttt{Conspiracy} = \texttt{true}\}$$
$$\overline{n(r_{118})} = \{\texttt{Conspiracy} = \texttt{true}\}$$
$$\overline{n(r_{128})} = \{\texttt{EvidenceOfCrime} = \texttt{false}\}$$
$$\overline{n(r_{61})} = \emptyset$$
$$\overline{n(r_{108})} = \emptyset$$
$$\overline{n(r_{67})} = \emptyset$$
$$\overline{n(r_{142})} = \emptyset$$

Not surprisingly, rule 120 is undercut if there turns out to be a conspiracy. The fact that half of these rules do not have undercutters is mainly due to the small size of the example network. Undercutters are taken from variables that do not occur in the rule already, meaning that rules with multiple premises quickly exhaust the possible set of undercutters.

3.3.3 Building arguments using ASPIC+

We now formalise the arguments that can be built using the definitions of rules and undercutters by instantiating ASPIC+. More specifically, we instantiate ASPIC+ in the following way to obtain an argumentation system for probabilistic reasoning in a given BN. All of the ingredients of an ASPIC+ argumentation theory have now been introduced, and with the ASPIC+ formalisation of arguments we can show the arguments that can be built from the rules extracted from a BN.

Definition 3.15 (BN Argumentation System, instantiating Definition 3.1 (tentative))**.** *Suppose a BN is given with nodes $\boldsymbol{V}$. Let $\boldsymbol{e}$ denote the evidence. We define the following instantiation of ASPIC+:*

$$\mathcal{L} = \left(\bigcup_{\mathtt{V_i} \in \boldsymbol{V}} \bigcup_{\mathtt{v_{ij}} \in vals(\mathtt{V_i})} \{\mathtt{V_i} = \mathtt{v_{ij}}\} \right) \cup \{n(r) | r \in \mathcal{R}_d\}$$
$$\overline{\mathtt{V_i} = \mathtt{v_{ij}}} = \{\mathtt{V_i} = \mathtt{v_{ik}} \,|\, j \neq k\}$$
$$\overline{n(r)} = \{\textit{as in Definition 3.13, for any } r \textit{ in } \mathcal{R}_d\}$$
$$\mathcal{R}_s = \emptyset$$
$$\mathcal{R}_d \textit{ as in Definition 3.11}$$
$$\mathcal{K}_p = \emptyset$$
$$\mathcal{K}_n = \{\mathtt{V_i} = \mathtt{v_{ij}} | (\mathtt{V_i} = \mathtt{v_{ij}}) \textit{ occurs in } \boldsymbol{e}\}$$

This extends Definition 3.1 by implementing the defeasible rules, and appropriate contrariness for these rules. The definition is still tentative because later we identify a number of issues with this implementation that we address by some final modifications. Besides variable assignments the language must contain a name for

each rule. This is because the function $n()$ maps rules to names in the language. Since we do not wish to use variable assignments as names for rules, we add a name for every rule to the language. The contraries of a variable assignment are other assignments to the same variable and the contraries to rule names are the undercutters as described above. Furthermore we have modelled observed variable instantiations as necessary knowledge. This is because we do not wish arguments to be attacked on premises. Something that is believed because it was observed should never be defeated by anything that was derived using statistical inference rules. From the strengths of rules, an ordering on rules follows naturally:

Definition 3.16 (Rule ordering)**.** *Given a BN, and evidence* $\boldsymbol{e}$*. Let* r_1 *and* r_2 *be rules. These rules are ordered as follows:*

$$r_1 \leq r_2 \text{ iff } \text{strength}(r_1, \boldsymbol{e}) \leq \text{strength}(r_2, \boldsymbol{e})$$

A preference ordering on arguments follows straightforwardly from the preordering $\leq$ on rules $\mathcal{R}_d$ using the weakest-link principle of Prakken [2010].

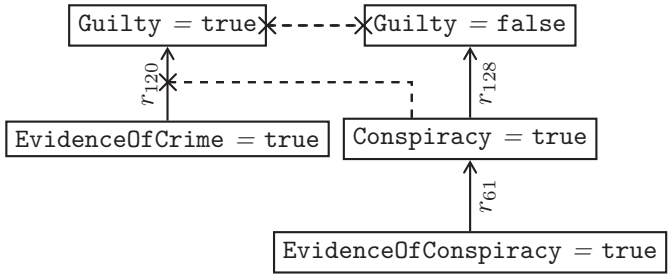
Definition 3.17 (Weakest link ordering, adaptation from Prakken [2010])**.** *Let* A *and* B *be two arguments, where* DefRules(A) *and* DefRules(B) *denote the sets of defeasible rules used in those arguments. Let* $<_s$ *be a partial ordering on sets of rules such that* $S_1 <_s S_2$ *iff there exists an* $r_1 \in S_1$ *for which* $r_1 < r_2$ *for all* $r_2 \in S_2$*. Then* $A \prec B$ *iff*

- B *is strict and* A *is defeasible, or*
- B *is defeasible and* DefRules(A) $<_s$ DefRules(B)

This ordering prefers argument A over B iff at least one rule in B is weaker than all rules in A. That is, competition between arguments is decided by the weakest rule applied in the arguments. The argument with the weakest rule is defeated by the other. We choose to use this ordering because this kind of comparison on the weakest inferential step is also characteristic for probabilistic reasoning. Intuitively, when making a number of inferential steps the argument should be evaluated by the strength of the weakest of those steps. In the argumentation literature this ordering was shown to be a *reasonable* ordering [Prakken, 2010], which is one of the sufficient ingredients to make the argumentation system adhere to the *rationality postulates* introduced by Caminada and Amgoud [2007], which guarantee consistency and strict closure of the conclusion sets of extensions. Together with the observations that our argumentation system has no strict rules the fact that our argument ordering is reasonable implies that it satisfies these rationality constraints.

Invoking the ASPIC+ mechanism to construct arguments, we can already show how the rules that have been found can be combined.

Example 3.18. *Consider again the running example. Using the rules presented in Example 3.12 the following arguments can be built (among others):*



Many other arguments are possible from the complete set of rules, but the presented ones are interesting because they argue for the `Guilty` *node. We can also see that the argument* `Conspiracy = true` *undercuts the argument for the conclusion* `Guilty = true` *because a conspiracy undercuts the rule that is used.*

Looking back at Example 3.12 we can easily see that the rules in this case are ordered as follows:

$$r_{120} > r_{61} > r_{128}$$

As we have noted before, the collection of extracted rules gives rise to many more arguments. These are just a few of the arguments that are potentially interesting from a legal perspective.

Using the ordering on rules, and the resulting ordering on arguments, a defeat relation among these arguments can be determined and ultimately the Dung style argument extensions can be calculated. We will do this for a number of scenarios in Section 3.5.

3.4 Guarding d-separation in argument extraction

The argumentation system described in the previous section extracts rules from which arguments can be built. However, not all combinations of probabilistically extracted rules make sense. One of the reasons that the above system generates incorrect arguments is that it combines rules in every possible way. We know that there are situations in which this is not permissible from a probabilistic standpoint [Pearl, 1988b]. This has to be addressed in order to guarantee correct argumentative results. For this purpose we introduce some final modifications of the above argumentation system.

3.4.1 The problem

Let us illustrate the caveats in the above argumentation system with another example.

Example 3.19. *Given the rules in Example 3.12 and our current ASPIC+ formalisation with evidence for a conspiracy and evidence for a crime, it is possible to construct the argument in Figure 3.3. This example demonstrates, in fact, mul-*

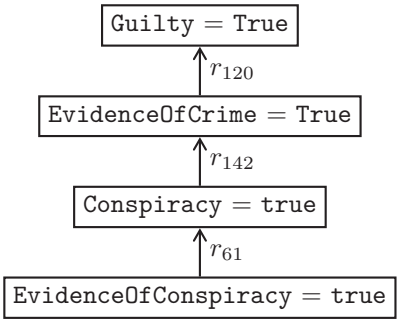


Figure 3.3: Example of a counter-intuitive argument that follows a head-to-head connection in the BN graph.

tiple problems. The first is that the non-premise subargument `EvidenceOfCrime = true` *concerns an instantiated variable in the associated BN. It seems unnatural that the application of rules can result in a conclusion that is already known. Moreover, if* `EvidenceOfCrime`*, would not have been a head-to-head node on the chain between* `Guilty` *and* `Conspiracy`*, then its instantiation would have blocked the chain; the above argument could then still be constructed, but it would connect two probabilistically independent statements.*

The second problem concerns the subsequent application of two rules that are associated with a head-to-head structure in the BN, such as r_{142} *and* r_{61} *in the above example. If we had no evidence for a crime, then the above argument graph still stands (*`EvidenceOfCrime = true` *is not used as a premise) but again connects two statements that are probabilistically independent since the chain between* `Guilty` *and* `Conspiracy` *is blocked due to the uninstantiated head-to-head node* `EvidenceOfCrime`*. If we do have an instantiation for* `EvidenceOfCrime` *then the argument graph reflects an active chain: first, rule* r_{142} *is used to derive from a conspiracy that there is probably evidence of a crime (by itself a valid predictive argument because conspiracies tend to result in planted evidence). Then, from this evidence it is concluded that the suspect is guilty (by itself also a valid argument since evidence for a crime normally suggests that the suspect is guilty). However, the combination of the two rules is fallacious because the evidence of a conspiracy should in fact diminish our belief in the guilt hypothesis. By chaining the rules associated with the edges in a head-to-head connection, we apparantly do not capture the intercausal explaining away effect. This intercausal reasoning step is captured by rule* r_{128} *instead.*

Note that the algorithms for computing probabilities in BNs correctly deal with the above problems by respecting the d-separation properties coded in the graph; posterior probabilities for `Guilty` *indeed reflect an explaining away interaction.*

The above example illustrates that our current argumentation system is incapable of correctly capturing the (in)dependencies and their dynamics (they can change

when evidence changes) coded in the BN graph. It would be desirable if premises (those subarguments directly derived from observations and not from rules) and conclusions of arguments were conditionally dependent given the available evidence. However, our arguments are structured according to the graphical structure of the BN and conditional dependence is not something that can be identified purely on the basis of that graphical structure. Variables that are d-connected in the graph can still be independent due to the specific conditional probability tables for the variables. What can and should be guaranteed is that whenever independence among variables follows from the BN graph, there will be no arguments connecting those variables. This can be done by guaranteeing that only active chains are used in arguments. Now our example serves to identify the following three problems:

- arguments can be constructed that are associated with chains that are inactive due to observed non-head-to-head nodes;
- if an active chain in the BN includes a head-to-head node, then consecutive application of rules associated with the two converging edges for this node does not correctly capture the induced intercausal interaction;
- arguments can be constructed that are associated with chains that are inactive due to head-to-head nodes without observations for at least one descendant.

To solve these problems we will introduce a number of modifications of the above argumentation system that prohibit these situations. First, we simply filter rules that have an instantiated conclusion variable. In that way a chain of arguments can never lead to conclusions that correspond to instantiated variables. As such, we prevent arguments corresponding to chains with instantiated non-head-to-head nodes. Although this also prevents arguments corresponding to chains with head-to-head nodes, which in fact become active when the head-to-head node has been observed, we argued that the intercausal interaction should be captured differently, by applying a rule that directly connects the two parents of the head-to-head node. However, this *intercausal rule* should only be applied if the head-to-head connection is active, so another adaptation is required to prohibit its application otherwise.

Filtering rules is straightforward. The other two issues are harder to address because they concern more global properties of the graph.

3.4.2 Pearl's C-E system

To solve the above problems we take inspiration from Pearl's C-E system, which tackles causality issues in non-monotonic logic using C- and E- labels [Pearl, 1988b]. In the next section we will implement a similar system that is both more general and fully formalised.

In Pearl's C-E system, causal and evidential rules are distinguished using explicit labels. To stick to our example we could distinguish:

$$(\texttt{Guilty} = \texttt{true}) \Rightarrow_c (\texttt{Evidence} = \texttt{true})$$
$$(\texttt{Evidence} = \texttt{true}) \Rightarrow_e (\texttt{Guilty} = \texttt{true})$$

The first should be read as guilt "causes" evidence while the latter tells us that the evidence "is evidence for" guilt. Using these inference rules, conclusions can be derived. Now, a distinction is made between E-believed and C-believed statements. These belief-status labels are written as

$$C(\texttt{Guilty} = \texttt{true})$$
$$E(\texttt{Guilty} = \texttt{true})$$

Using these labels, the applicability of rules is limited. In particular, the following three derivation schemes are allowed for any propositions P and Q:

$$\frac{P \Rightarrow_c Q \quad C(P)}{C(Q)} \qquad \frac{P \Rightarrow_c Q \quad E(P)}{C(Q)} \qquad \frac{P \Rightarrow_e Q \quad E(P)}{E(Q)}$$

Whereas the following derivation is explicitly forbidden:

$$\frac{P \Rightarrow_e Q \quad C(P)}{E(Q)}$$

In other words, causal rules can only derive C-believed conclusions and evidential rules can only be applied to E-believed statements. This means that after an inference in the causal direction, no inference in the evidential direction can be made any more. Even though BNs are not necessarily causally directed, they do have a similar constraint, i.e., after reasoning along the direction of an edge, no inference against the direction of an edge should be allowed unless the chain is explicitly activated by an observation of a descendant of the head-to-head node. In the following section we implement the fact that the fourth derivation is forbidden by filtering labelled rules from the argumentation system accordingly.

3.4.3 Derivation labels for our argumentation system

To prevent reasoning errors such as in Example 3.19, we have to exclude chains of argumentative inference that

- pass through observed variables in the associated BN;
- pass through head-to-head connections in the associated BN;
- use intercausal reasoning steps when no unblocking evidence is present for the variable in the associated BN.

Even if a head-to-head node or a descendant of it is observed (allowing intercausal interactions), we can ignore the active chain through the head-to-head connection itself: we will find the correct influence between the co-parents because there is

a rule directly between them that captures the intercausal reasoning. Excluding the arguments that use any of the above constructions could be done in a post-processing step. However, we choose to apply an on-the-fly approach that is similar to Pearl's C-E system [Pearl, 1988b]. We do this by labelling all intermediate conclusions as either E or C. Statements are C when an inference along the direction of an edge has been made in their derivation. In that case we know that if we never apply a rule against the direction of an edge, we exclude exactly these fallacious reasoning steps.

Taking inspiration from Pearl's C-E system, we now show how our previous ASPIC+ instantiation can be adapted to prevent exactly the fallacious rule applications that we have discussed. The following adjustment to our argumentation system deviates from Pearl's system in two ways:

- We allow inference rules with multiple premises. In particular a rule is considered causal if at least one of its premises assigns an outcome to a parent of the variable from the conclusion and evidential otherwise. In this way, reasoning to a parent (evidential reasoning) is only allowed if no other parents of the antecedent are used to derive that antecedents, which could block the inference in the BN;
- We do not label the rules but the language elements. This means that we have to create multiple copies of the same rule with differently labeled statements. The advantage is that we can suffice with a single set of rules.

We apply C and E labels to guarantee that the resulting arguments correctly follow d-separation. The C labels indicate that the statement was derived by an inference along the directions of an edge. If more than one antecedent was used at least one corresponds to a parent of the consequent node in the BN. Similarly, an E label indicates that the statement was derived without the use of inferences along the direction of an edge. The notations $C(\mathtt{V_i} = \mathtt{v_i})$ (and $E(\mathtt{V_i} = \mathtt{v_i})$) represent that the assignment is C-derived (E-derived). The letters C and E stand for *causal* and *evidential*. Note that this does not imply the presence of a causal relation. When the relations between variables are not causal, other inter-parent interactions can be modelled by the BN which require similar care. Therefore we use the terms causal and evidential reasoning for reasoning along and against the directions of edges in the BN, regardless of whether the BN actually models causal relations or not. This leads to a new definition of the logical language:

Definition 3.20 (Language for C-E argumentation system)**.** *Let* $\mathbf{V}$ *again be the set of variables from a BN and let* $\text{vals}(\mathtt{V_i})$ *denote the possible values that variable* $\mathtt{V_i}$ *can take on. Then,*

$$\mathcal{L} = \left(\bigcup_{\mathtt{V_i} \in \boldsymbol{V}} \bigcup_{\mathtt{v_{ij}} \in vals(\mathtt{V_i})} \{C(\mathtt{V_i} = \mathtt{v_{ij}}), E(\mathtt{V_i} = \mathtt{v_{ij}})\} \right) \cup \{n(r) | r \in \mathcal{R}_d\},$$

for which,

$$\overline{C(\mathtt{V_i} = \mathtt{v_{ij}})} = \overline{E(\mathtt{V_i} = \mathtt{v_{ij}})} = \{C(\mathtt{V_i} = \mathtt{v_{ik}}) \,|\, j \neq k\} \cup \{E(\mathtt{V_i} = \mathtt{v_{ik}}) \,|\, j \neq k\}$$

That is, for every variable assignment there is now a C-labelled and an E-labelled version in the language. For contrariness there is no distinction between the two.

Let us now revise the creation of rules. We do this by first extracting the rules as in Definition 3.15 and then creating multiple versions of each rule such that premises are labelled in all possible correct ways. Note that from a computational point of view this is inefficient, but it allows us to produce the desired formal result. A practical implementation of such a system can ignore this step and simply check the applicability of rules during the generation of arguments at run time, adding no structural computational complexity to the algorithm.

The duplication and filtering of rules is as follows. We introduce for every sufficiently strong probabilistic rule $p_1, \ldots, p_n \Rightarrow v$, a number of labelled versions of that rule. We distinguish antecedents that assign a value to a parent of the variable $\mathtt{V_c}$ corresponding to the consequent, from other antecedents.

Definition 3.21 (Labelled rule set)**.** *Given a BN with graph $G = (\mathbf{V}, \mathbf{E})$. Let r be a rule $p_1, \ldots, p_n \Rightarrow c$ from the original $\mathcal{R}_d$ and let its conclusion c assign an outcome to variable $\mathtt{V_c}$ and its premises $p_1, \ldots, p_n$ assign outcomes to variables $\mathtt{V_{p,1}}, \ldots, \mathtt{V_{p,n}}$. The updated $\mathcal{R}'_d$ contains labelled versions of that rule if and only if:*

Rule constraint 1 *$\mathtt{V_c}$ is not instantiated, and*

Rule constraint 2 *if there exists an antecedent p_i of r for which variable $\mathtt{V_{p,i}}$ assigns an outcome to a co-parent of $\mathtt{V_c}$ with a common child $\mathtt{V_{p,j}}$ (i.e., $(\mathtt{V_c}, \mathtt{V_{p,j}}) \in \mathbf{E}$ and $(\mathtt{V_{p,i}}, \mathtt{V_{p,j}}) \in \mathbf{E}$), then there exists another antecedent p_j that assigns an outcome to $\mathtt{V_{p,j}}$.*

Iff both conditions are satisfied, then multiple fully labelled versions of the rule are generated. A rule exists in $\mathcal{R}'_d$ for every possible permutation of E and C labels that satisfies the following constraints:

Label constraint 1 *if antecedent p_i assigns a value to a variable $\mathtt{V_{p,i}}$ that is a child of $\mathtt{V_c}$, then p_i must be E-derived, and*

Label constraint 2 *the consequent c is C-derived iff there exists an antecedent p_i for which $\mathtt{V_{p,i}}$ is a parent of $\mathtt{V_c}$.*

For all other antecedents (parents and co-parents) the label is free and there exist rules with any combination of E and C labels for those assignments.

In this update, three things happen simultaneously. First, some rules are filtered out (rule constraint 1). This step discards rules that conclude something about a variable that is instantiated. Rule constraint 2 ensures that intercausal reasoning can only occur when there is evidence for a common descendant. This prevents non-justified intercausal reasoning (making the intercausal step without including descendants).

Then multiple labelled versions of the other rules are created. This ensures that reasoning between co-parents happens via the intercausal interaction rather than the combination of a causal with an evidential step. The intuition behind this labelling scheme is that derivations that use information from parents in the BN are labelled C (according to Label constraint 2) and that reasoning from a node upwards to a parent requires an E status (according to Label constraint 1). This is schematically depicted in Figure 3.4 together with a number of example labellings, including an example of a labelling that is not allowed.

The final ASPIC+ instantiation becomes:

Definition 3.22 (BN Argumentation System (final))**.** *Suppose a BN is given with nodes $\boldsymbol{V}$. Let $\boldsymbol{e}$ denote the evidence. We define the following instantiation of ASPIC+:*

$$\mathcal{L} = \left(\bigcup_{\mathtt{V_i} \in \boldsymbol{V}} \bigcup_{\mathtt{v_{ij}} \in vals(\mathtt{V_i})} \{C(\mathtt{V_i} = \mathtt{v_{ij}}), E(\mathtt{V_i} = \mathtt{v_{ij}})\} \right) \cup \{n(r) | r \in \mathcal{R}_d\}$$

$$\overline{E(\mathtt{V_i} = \mathtt{v_{ij}})} = \overline{C(\mathtt{V_i} = \mathtt{v_{ij}})} = \{C(\mathtt{V_i} = \mathtt{v_{ik}}), E(\mathtt{V_i} = \mathtt{v_{ik}}) \mid j \neq k\}$$

$\overline{n(r)} = \{$*as in Definition 3.13, for any r in $\mathcal{R}_d$*$\}$

$\mathcal{R}_s = \emptyset$

$\mathcal{R}_d$ *as in Definition 3.21*

$\mathcal{K}_p = \emptyset$

$\mathcal{K}_n = \{E(\mathtt{V_i} = \mathtt{v_{ij}}) \mid (\mathtt{V} = \mathtt{v_{ij}})$ *occurs in* $\boldsymbol{e}\}$

$\leq$ *s.t.* $r_1 \leq r_2$ *iff* $\text{strength}(r_1, \boldsymbol{e}) \leq \text{strength}(r_2, \boldsymbol{e})$

This definition reflects the addition of labels as just described. Note that observations are labelled E such that initially evidential reasoning is allowed.

As discussed in Section 3.4.1 chains of inference in a BN can be blocked by observations of non-head-to-head nodes. It is now guaranteed that this kind of blocked chain is never used in the arguments:

Theorem 3.23. *Consider the argumentation system and knowledge base from Definition 3.22. Let* $\text{Conc}(A)$ *assign an outcome to variable* $\mathtt{V_c}$*. Variable* $\mathtt{V_c}$ *is instantiated iff* $\text{ImmSub}(A) = \emptyset$*, i.e., A is a premise argument.*

Proof. If A assigns an outcome to an observed variable, then it must be a premise argument, for no rule has a conclusion that assigns a value to observed variables (by rule constraint 1 in Definition 3.21). For the other direction, the knowledge base $\mathcal{K}_n$ consists solely of assignments to all observed variables. Therefore, if A is a premise argument, it must assign an outcome to an observed variable. □

The second problem that we discussed is that without precautions the inference

rules can be combined to follow head-to-head connections in the BN graph, which can result in undesirable situation of including arguments for conclusions that are probabilistically independent of their premises. The precluded situation is shown in Figure 3.5. The C-E labelling scheme solves this issue:

Theorem 3.24. *Consider the argumentation system and knowledge base from Definition 3.22. No argument A for variable* V *has an immediate subargument A′*

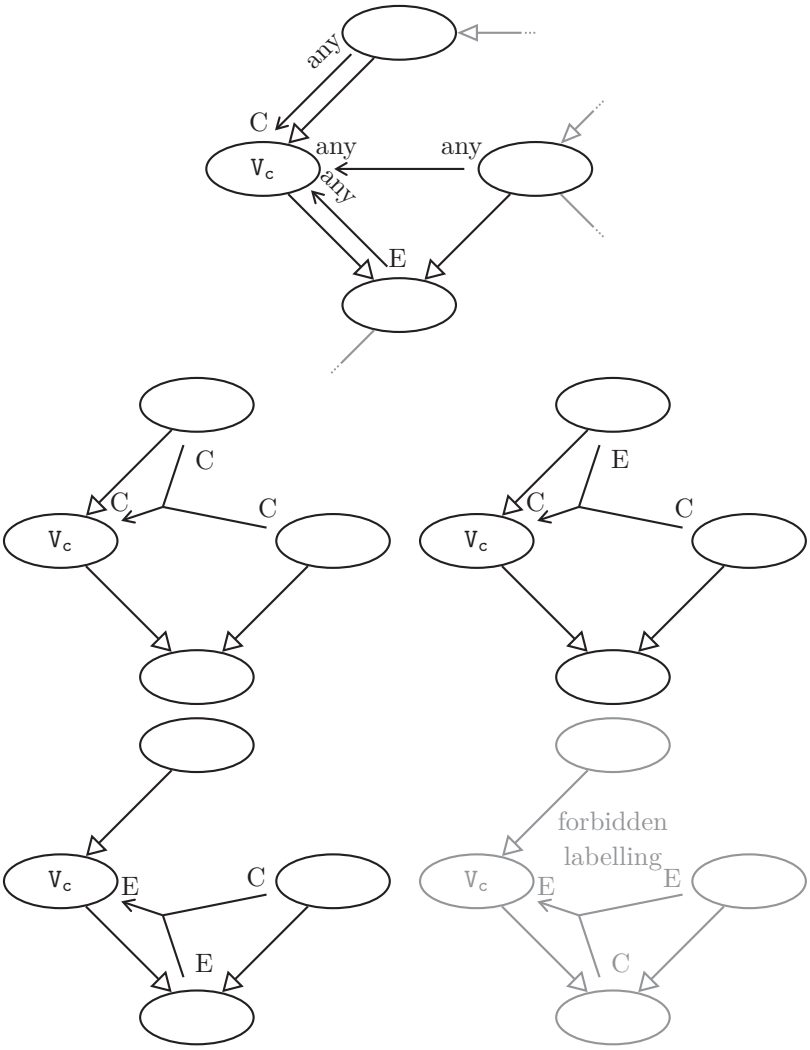


Figure 3.4: Schematic representation of the allowed labels. Triangle arrow heads signify the edges from the BN, and other arrows depict inference rules. The top image summarises the allowed labels for all rules with one premise. Every rule that satisfies this pattern is generated. Below are a number of rules that are allowed and one example of a rule that is specifically not allowed.

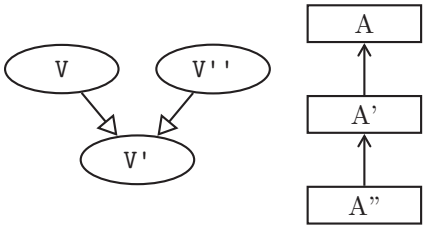


Figure 3.5: The structure of the BN and the arguments as described by Theorem 3.24.

for variable $\mathtt{V'}$ *which in turn has an immediate subargument* A'' *for variable* $\mathtt{V''}$ *when both* $(\mathtt{V}, \mathtt{V}') \in \mathbf{E}$ *and* $(\mathtt{V}'', \mathtt{V}') \in \mathbf{E}$*, forming a head-to-head connection in the BN graph.*

Proof. We prove this by contradiction. Suppose that such an argument exists. Consider the TopRule of argument A, which is of the form $\text{Conc}(A'), \ldots \Rightarrow \text{Conc}(A)$. From label constraint 1 in Definition 3.21 we learn that the conclusion of A' must be E-derived. Then consider the TopRule of A', which is of the form $\text{Conc}(A''), \ldots \Rightarrow \text{Conc}(A')$. From label constraint 2 in that same definition (now applied to the TopRule of A') we know that A' must have a C-derived conclusion. Therefore the described configuration is impossible. □

This theorem states that the argumentation does not follow head-to-head connections in the BN. This proves that arguments do not follow head-to-head connections by traversing the edges along and against the directions of the edges in the BN graph. The intercausal reasoning step that is appropriate in that situation is, of course, enabled by an inference rule directly from $\mathtt{V}''$ to $\mathtt{V}$. This inference is always identified because we included parents of children as possible sources for premises. Such an intercausal reasoning step is only used when at least one common child or descendant of a common child is instantiated. The described configuration of variables and arguments in the following theorem and proof is depicted in Figure 3.6.

Theorem 3.25. *Consider the argumentation system and knowledge base from Definition 3.22. Let* A *be an argument for variable* $\mathtt{V}$ *that has an immediate subargument* A'' *for variable* $\mathtt{V''}$ *such that* $(\mathtt{V}, \mathtt{V}', \mathtt{V}'')$ *forms an immorality in the BN graph. Then, a descendant (including* $\mathtt{V}'$ *itself) of* $\mathtt{V'}$ *is instantiated.*

Proof. Consider rule constraint 2 from Definition 3.21. It tells us that the TopRule of A must have an antecedent for variable $\mathtt{V}'$ and therefore A must have an immediate subargument (call it A') with that conclusion. From label constraint 1 in that same definition we know that the assignment of $\mathtt{V'}$ in argument A' must be E-derived. What remains to be proven is that if argument A' for variable $\mathtt{V}'$

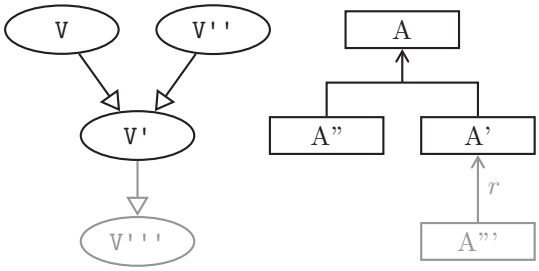


Figure 3.6: The described configuration of variables and arguments in Theorem 3.25.

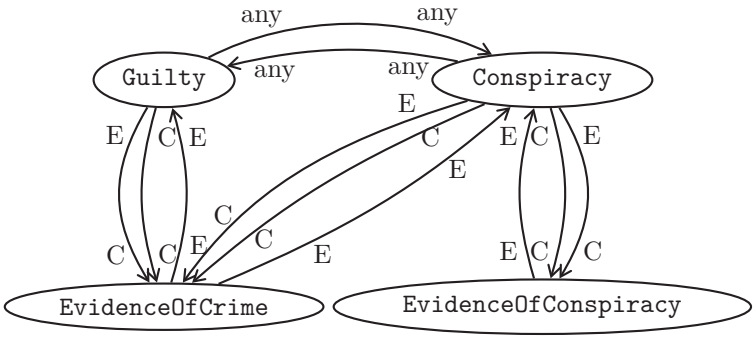


Figure 3.7: Possible labelling of rules. For readability only rules with one premise are shown but the labelling scheme also defines how rules with multiple premises are labelled.

has an E-derived status, then variable $\texttt{V}'$ is instantiated or has an instantiated descendant. This can be proven by induction. The base case occurs when argument A' has no TopRule. In that case the claim is true by definition (3.22). When argument A' applies a rule r, the negation of label constraint 2 dictates that no antecedents of r assign an outcome to a parent of $\texttt{V}'$. Since by Definition 3.11 rules have at least one antecedent we know that argument A' has an antecedent that assigns a value to a child or a co-parent of $\texttt{V}'$. By rule constraint 2 we know that in the latter case it has an antecedent that assigns a value to a child of $\texttt{V}'$ as well. So in either case rule r has an antecedent $\text{Conc}(A''')$ for an immediate subargument A''' that assigns a value to a child $\texttt{V}'''$ of $\texttt{V}'$. By label constraint 1 we know that then $\text{Conc}(A''')$ must have an E-derived status. By invocation of the induction hypothesis the variable $\texttt{V}'''$ must have an instantiated descendant and therefore also variable $\texttt{V}'$ has an instantiated descendant. □

Note that this theorem assumes that V, V' and V'' form an immorality and not simply a head-to-head connection. When there is a head-to-head connection but also a direct edge (so no immorality) then there is no intercausal reasoning step and consequently no limitation on how premises should have been derived.

Example 3.26. *Consider again the example network from Figure 3.1. Figure 3.7 shows a diagram showing all single-premise rules. Although the theorems above hold for all rules, it is easiest to demonstrate on rules that have only one premise, since diagrams like these easily tend to clutter. In this graph the variables are shown together with the ways in which rules from one variable to the other can be labelled. As can be seen, reasoning from* `EvidenceOfConspiracy`*, up to a* `Conspiracy` *is allowed and the resulting assignment will be labelled E. Reasoning from an E-derived statement about* `Conspiracy` *to* `EvidenceOfCrime` *is allowed, but the result is labelled C. This is in itself a fine argument. That is, evidence for a conspiracy predicts that there will also be evidence for a crime. However, reasoning from a C-derived statement about this* `EvidenceOfCrime` *to a statement about the* `Guilty` *variable is no longer allowed since that rule requires an E-derived antecedent.*

Together, the above three theorems show that argumentative chains follow active chains in the BN graph only. First, arguments cannot follow chains blocked by observations. Second, arguments do not traverse head-to-head connections. If a head-to-head connections represents an active chain, arguments must use the—for that purpose explicitly included—intercausal inference step. And, third, such an intercausal reasoning step is only possible if a common descendant is instantiated. This addresses exactly the three problems identified in Section 3.4.1. Together they imply that only active chains can be captured by arguments. Consequently, the variable associated to the premises of any argument are d-connected to the conclusion variable of that argument:

Corollary 3.27. *From Theorems 3.23, 3.24, and 3.25 it follows that all variables associated to premises of an argument are d-connected to the conclusion variable given the observed evidence.*

3.5 Analysis of the running example

To further illustrate the method we consider three different scenarios and show for each of these what the outcomes of the method are and how this compares to the outcome of a normal BN analysis. In the first scenario the investigators have only observed the evidence for the crime. Then we add the observation that no evidence for a conspiracy was found, to show how this changes the argumentation framework. Note that there is a fundamental difference between not observing evidence and observing that the evidence is not there. The former only means that we have not looked for any evidence, whereas the latter means that we have looked and found none. Lastly, we show what happens if the search for a conspiracy turns out positive. The graphical structure of the BN is shown in Figure 3.8. For the conditional probabilities we refer the reader to Figure 3.1 on page 29.

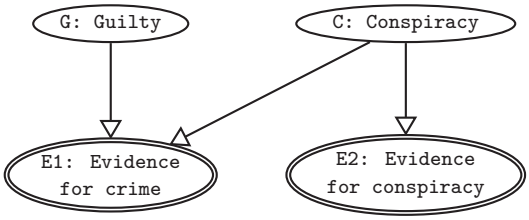


Figure 3.8: The graph from our running example about a crime with a possible conspiracy.

3.5.1 Scenario 1: evidence for a crime

Suppose that the investigators do not yet know what the outcome of the evidence for a conspiracy (`E2`) is and we only have the assignment $\texttt{EvidenceOfCrime} = \texttt{true}$, or $\texttt{E1} = \texttt{true}$ for short. From the network we can extract, among others, the following relevant rules:

$$E(\texttt{EvidenceOfCrime} = \texttt{true}) \Rightarrow E(\texttt{Guilty} = \texttt{true})$$
$$E(\texttt{EvidenceOfCrime} = \texttt{true}) \Rightarrow E(\texttt{Conspiracy} = \texttt{true})$$
$$E(\texttt{Conspiracy} = \texttt{true}) \Rightarrow C(\texttt{EvidenceOfConspiracy} = \texttt{true})$$
$$E(\texttt{EvidenceOfCrime} = \texttt{true}), E(\texttt{Guilty} = \texttt{true}) \Rightarrow E(\texttt{Conspiracy} = \texttt{true})$$

See Appendix A for the full list of extracted and labelled rules. Building arguments from these rules is straightforward. In fact, we can build six arguments given the observation of $E(\texttt{EvidenceOfCrime} = \texttt{true})$; see Figure 3.9. The first two rules can simply be applied to obtain arguments for $E(\texttt{Guilty} = \texttt{true})$ (argument 166) and $E(\texttt{Conspiracy} = \texttt{true})$ (argument 163) respectively.

The latter argument can be used as a subargument to construct the composite argument *"*$E(\texttt{EvidenceOfCrime} = \texttt{true})$ *therefore* $E(\texttt{Conspiracy} = \texttt{true})$ *therefore* $C(\texttt{EvidenceOfConspiracy} = \texttt{true})$*"* (165) using the third rule. Such an argument is usually not considered in a legal setting because it predicts the evidence for a conspiracy. It is, probabilistically speaking, however a solid argument that can be made. For example, we expect to find evidence for a conspiracy because we believe a conspiracy may have happened. The C label also confirms the predictive status of this argument.

Note that formally $E(\texttt{EvidenceOfCrime} = \texttt{true})$ itself is also an argument and that this is a subargument of all arguments that we just mentioned. The other rules can be used to build two further arguments. First, evidence for a crime turns out to support a conspiracy, even if the suspect is guilty. This is because the probability of a conspiracy against a guilty suspect is still higher than the prior probability of a conspiracy. Indeed, a guilty suspect does not exclude the possibility of a conspiracy. Therefore the evidence still has a small effect on the

hypothesis that there was a conspiracy. Again, the predictive rule for evidence for this conspiracy can be applied, resulting in a new argument (172). Note that arguments 172 and 165 have the same top-rule and the same conclusion, but are still distinguished because they use different subarguments.

Figure 3.10 shows the defeat relation between those arguments. Since there is no contradictory information to deal with, no arguments undermine or rebut

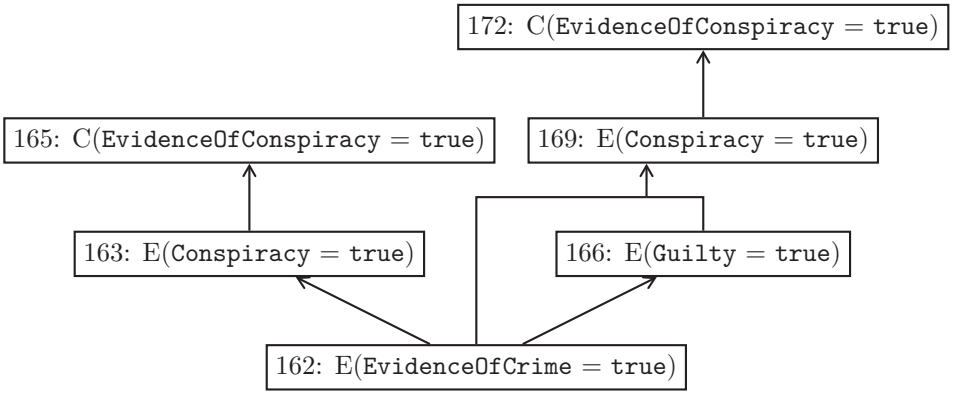


Figure 3.9: Arguments built from the extracted rule base, with observations for `EvidenceofCrime = true` only. Each box represents one argument and the arrows show the rules by which they are connected. Arguments are given a unique number by our implementation that we maintained in these figures for each reference from the main text.

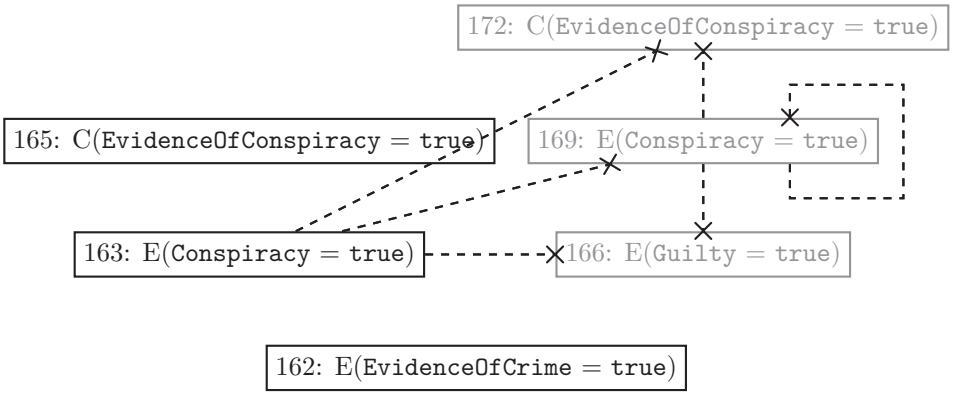


Figure 3.10: The abstract argumentation framework corresponding to Figure 3.9. Both attacks are undercutters that automatically result in defeat. Note that argument 163 and 169 undercut argument 166 on its TopRule. This is direct attack. Any argument that builds on argument 166 is indirectly attacked and this includes argument 169 itself. The unique grounded, stable and preferred extension contains the arguments 162, 163 and 165.

each other. There are however two undercutting arguments. Remember that undercutting corresponds to explaining away in the BN. As we have just discussed, a conspiracy explains the evidence and therefore explains away the guilt hypothesis. However, as the guilty hypothesis does not explain away the conspiracy, the attack does not go the other way around. Observe that both arguments for a conspiracy attack argument 166 for the guilt hypothesis in this way. Also note that Figure 3.10 shows direct attack as well as indirect attack. In ASPIC+ any argument that attacks another argument automatically attacks everything that is derived from the attacked argument as well. This means that, for instance, argument 169 attacks itself by undercutting its own subargument.

In argumentation it is customary to draw conclusions based on these extensions. Grounded, stable and preferred semantics are the most common ways to asses what should be accepted and which hypotheses should be refuted. Recall the definition of argument acceptability: for any argument A, A is *acceptable* with respect to some set of arguments S iff any argument B that defeats A is itself defeated by an argument in S. A complete extension is a set S of arguments such that $A \in S$ whenever A is acceptable w.r.t. S. The arguments 162, 163 and 165 form such a complete extension. Any further addition to the set would result in an argument that is not acceptable with respect to the set.

Recall that the grounded extension consists of a set inclusion minimum complete extension. It expresses what should minimally be accepted. Preferred extensions consist of a set inclusion maximal complete extension. A preferred extension captures a conflict-free point of view that can be defended against attack by other argument. A stable extension is a preferred extension that also defeats all arguments outside this extension. It captures a consistent point of view that is not only defensible but also actively attacks any other argument.

In the case of the current example the complete extension consisting of arguments 162, 163 and 165 is also the unique stable, preferred and grounded extension, which is depicted in Figure 3.10. It is the only extension that is set inclusion minimal and maximal. I.e., no superset or subset is also a complete extension. Proper subsets are not complete because there are arguments outside such extensions that could be added without introducing conflicts. Supersets are not complete because they have defeaters of their own arguments that are themselves not defeated.

What we can conclude is that the reported evidence can provide reasons to believe both the guilt of the suspect and a conspiracy individually. The conspiracy however explains away the guilt hypothesis, whereas the guilt hypothesis does not explain away the conspiracy. This is confirmed by the posterior probabilities from the BN:

$$\begin{aligned} \mathrm{P}(\texttt{G} = \texttt{true}) &= 0.100 & \mathrm{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true}) &= 0.899 \\ \mathrm{P}(\texttt{C} = \texttt{true}) &= 0.010 & \mathrm{P}(\texttt{C} = \texttt{true} \mid \texttt{E1} = \texttt{true}) &= 0.101 \\ \mathrm{P}(\texttt{E2} = \texttt{true}) &= 0.108 & \mathrm{P}(\texttt{E2} = \texttt{true} \mid \texttt{E1} = \texttt{true}) &= 0.180 \end{aligned}$$

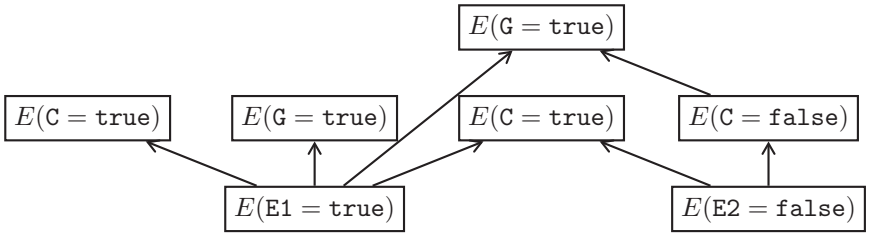


Figure 3.11: Some of the arguments that can be built with observations for EvidenceofCrime = true and EvidenceofConspiracy = false. See Figure A.1 in Appendix A for the full argument graph.

$$\mathrm{P}(\mathtt{G} = \mathtt{true}) = 0.100 \qquad \mathrm{P}(\mathtt{G} = \mathtt{true} \mid \mathtt{E1} = \mathtt{true} \wedge \mathtt{C} = \mathtt{true}) = 0.100$$
$$\mathrm{P}(\mathtt{C} = \mathtt{true}) = 0.010 \qquad \mathrm{P}(\mathtt{C} = \mathtt{true} \mid \mathtt{E1} = \mathtt{true} \wedge \mathtt{G} = \mathtt{true}) = 0.011$$

We observe that the probability of the guilt hypothesis is higher than that of the conspiracy hypothesis. Nevertheless, the argument for the guilt hypothesis is undercut by the conspiracy theory. This is the effect of having an incremental measure of strength. As can be seen from the last two rows of probabilities above, a conspiracy completely nullifies the effect of the evidence on the guilt hypothesis and not vice versa. We must conclude that the relative change is higher. If we are interested in the arguments with the highest posterior probability we could turn to absolute measures of strength, as we will discuss in Section 3.6.

3.5.2 Scenario 2: additional evidence

Suppose that we now add the observation EvidenceOfConspiracy = false (E2 = false) stating that we do not find any evidence that points towards a conspiracy. The original observation of the evidence is still in place and we find similar rules to the ones illustrated above, which is not completely unexpected. In addition we find rules and arguments that build on the newly added observation. For example the following new rules are identified (variable names again shortened as before):

$$\begin{aligned} E(\mathtt{E2} = \mathtt{false}) &\Rightarrow E(\mathtt{C} = \mathtt{false}) \\ E(\mathtt{C} = \mathtt{false}) &\Rightarrow C(\mathtt{G} = \mathtt{true}) \\ E(\mathtt{E2} = \mathtt{false}), E(\mathtt{E1} = \mathtt{true}) &\Rightarrow E(\mathtt{G} = \mathtt{true}) \end{aligned}$$

New arguments can now be built using these rules. Most notably, when combining the two pieces of evidence (using the last rule) the guilt hypothesis is supported stronger than before. See Figure 3.11 for a subset of all possible arguments. For readability we have omitted some predictive arguments that are irrelevant for the case. We observe that $E(\mathtt{Guilty} = \mathtt{true})$ can now be concluded on the basis of E1 = true together with Conspiracy = false.

An analysis of the grounded, stable and preferred extensions can again be done.

Figure 3.12: The selected arguments from Figure 3.11 highlighted according to the two extensions. The grey arguments are outside these extensions.

We do not show the full defeat graph here because the number of defeats results in a cluttered graph that adds no intelligible information. Since the aforementioned effect of a conspiracy explaining guilt away but not vice versa is still in effect, we find two stable and preferred extensions, corresponding with two possible points of view. This also explains the clutter in the defeat graph. Most of the arguments in one of these extensions defeat all arguments in the other (at least those that it does not share). These two extensions correspond to two possible scenarios in the example. Either the suspect is guilty and there is no conspiracy, or there is a conspiracy and no explanation for the absence of evidence for a conspiracy. The fact that there is no explanation for this evidence does not hinder the conclusion that there still may be a conspiracy since we have encoded only a statistical correlation between the conspiracy and the corresponding evidence, not a strict implication. The premise arguments are part of all grounded, preferred and stable extensions. The first stable/preferred extension includes all arguments for $\texttt{Conspiracy} = \texttt{true}$ and the second stable/preferred extension includes all arguments for $\texttt{Conspiracy} = \texttt{false}$ and $\texttt{Guilty} = \texttt{true}$. This is depicted in Figure 3.12. This corresponds to the following conditional probabilities from the BN, which are both higher than the priors for G and C, but that do exclude each other to some extent:

$$\mathrm{P}(\texttt{G} = \texttt{true} \mid \texttt{E1} = \texttt{true} \wedge \texttt{E2} = \texttt{false}) = 0.978$$
$$\mathrm{P}(\texttt{C} = \texttt{true} \mid \texttt{E1} = \texttt{true} \wedge \texttt{E2} = \texttt{false}) = 0.012$$

It may be surprising that while one of these conclusions is highly likely (0.978) and the other is highly unlikely (0.012) they result in competing explanations in the argumentation framework. However, they are both identified as good explanatory

arguments since we used an incremental measure of strength. This means that, although the posterior probability of a conspiracy may be low (0.012), it is still higher than the prior (0.01).

The fact that the two scenarios are identified is the result of the fact that the BN models a logical contradiction. The two observations are, to some extent, incompatible because evidence for a crime suggests that there was a conspiracy while the evidence for a conspiracy turned out to the contrary.

3.5.3 Scenario 3: what if there was a conspiracy

As a last example, we analyse what happens if some evidence suggests a crime but other evidence points towards the presence of a conspiracy. We enter the observation `EvidenceOfConspiracy` = `true` instead of `false`. Using an updated set of rules we find the arguments shown in Figure 3.13.

The new argument graph has many features similar to the ones found in scenario 2 that we do not discuss again. The main difference lies in the assigned outcomes and the strengths of the rules. In fact, since the conspiracy is now confirmed by evidence, the aforementioned conflicting information in the evidence is now resolved and all evidence corroborates towards the same conclusion that the suspect was not guilty. This can be seen from the defeat graph as well which is shown in Figure 3.14. The result may seem rather chaotic but it can be efficiently summarised as: Arguments 165, 171, 172 and 174 defeat argument 164 and therefore also any argument that builds on argument 164, which includes 172, 174 and 177.

The fact that the contradiction in the evidence has disappeared is also reflected in the argument extensions, which can be computed. There is no grounded, stable or preferred argument for the guilt hypothesis any more. There is just one extension, which is both grounded, stable and preferred. This is shown in Figure 3.15.

This extension includes the observed evidence as well as the arguments that a conspiracy has occurred. The fact that only one extension remains corresponds with the fact that the two observed instantiations form a consistent hypothesis about the world. That is, a conspiracy has occurred and therefore evidence that suggests that the suspect is guilty was found, but there is no reason to believe or disbelieve the suspect's guilt. This can also be seen from the posterior probabilities that result from evidence propagation in the BN:

$$\mathrm{P}(\mathtt{G} = \mathtt{true} \mid \mathtt{E1} = \mathtt{true} \wedge \mathtt{E2} = \mathtt{true}) = 0.542$$
$$\mathrm{P}(\mathtt{C} = \mathtt{true} \mid \mathtt{E1} = \mathtt{true} \wedge \mathtt{E2} = \mathtt{true}) = 0.503$$

Again we should emphasise that the argumentation is based on an incremental measure of strength which explains why the conspiracy theory is included in the extension even though its posterior probability is still slightly lower than the guilt hypothesis.

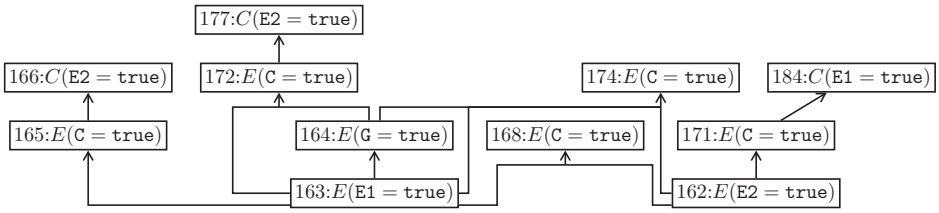


Figure 3.13: Full argument graph for evidence `EvidenceOfCrime` = `true` (`E1`) and `EvidenceOfConspiracy` = `true` (`E2`).

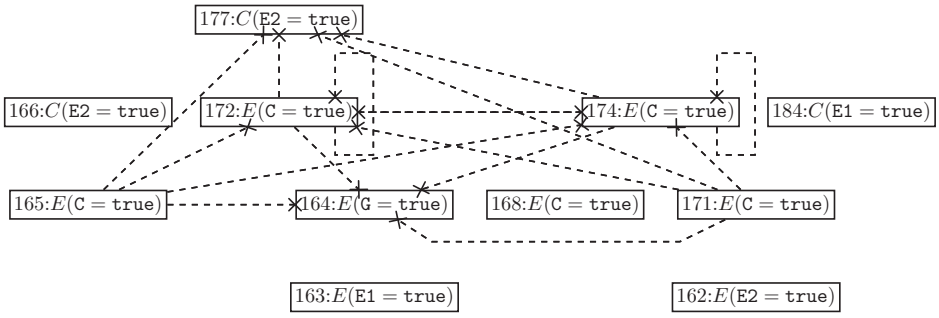


Figure 3.14: The same arguments as those presented in Figure 3.13 now with the defeat relations between them.

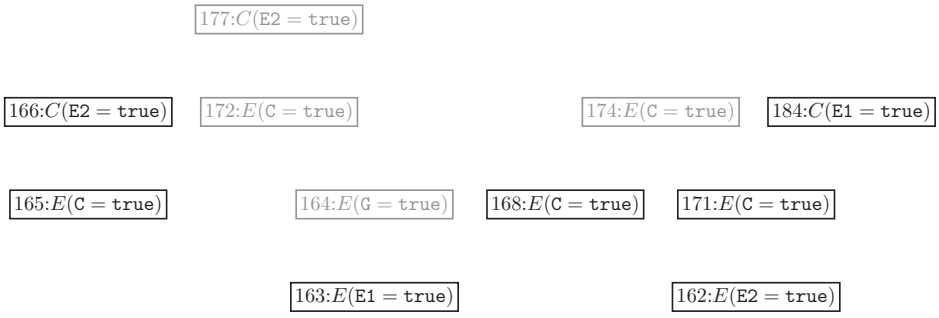


Figure 3.15: The same arguments as those presented in Figure 3.13 now with the unique grounded, stable and preferred extension shown.

3.6 Comparison of strength measures

As we have briefly noted before, various definitions of rule strength can be used. Two classes of heuristics can be distinguished: *incremental* and *absolute* measures of strength. Keynes' measure of confirmation, which we have used so far in all examples, is a good example of an incremental measure of strength since it measures the change in probability of the conclusion that would occur after observing the premises with respect to the prior probability of that conclusion.

Alternatively, we could simply look at the posterior probability of the conclusion given the premises and use that as a measure of rule strength. Such a measure is an absolute measure because the strength of the rule expresses how strongly the conclusion should be believed if the antecedents are observed, but it does not tell whether this belief is in fact the result of those antecedents. Some well-known measures of strength are [Crupi et al., 2007]:

$$\text{Keynes:} \quad \frac{\mathrm{P}(conclusion \mid premise)}{\mathrm{P}(conclusion)}$$

$$\text{Likelihood ratio:} \quad \frac{\mathrm{P}(premise \mid conclusion)}{\mathrm{P}(premise \mid \neg conclusion)}$$

$$\text{Posterior odds:} \quad \frac{\mathrm{P}(conclusion \mid premise)}{\mathrm{P}(\neg conclusion \mid premise)}$$

$$\text{Posterior probability:} \quad \frac{\mathrm{P}(conclusion \mid premise)}{threshold}$$

Using a different incremental measure usually results in the same or at least similar arguments. Although it is not guaranteed that all measures result in exactly the same rule strengths [Crupi et al., 2007], it is not uncommon to have the same arguments and even the same attack relations. For instance, if we repeat the argument generation as described in the previous sections, but replace the Keynes' measure of strength with the likelihood ratio measure, we obtain exactly the same arguments and the same attack relation (and therefore also the same grounded, preferred and stable extensions) in all three scenarios.

If, however, we adopt an absolute measure of strength, such as the posterior probability measure, we obtain entirely different arguments. The major advantage of absolute measures (posterior odds and posterior probability in the above examples), compared to relative measures, is that the application of a strong rule guarantees that the resulting argument corresponds to outcomes that have a high probability in the BN. This is not necessarily the case with incremental measures. Consider, for instance, the likelihood ratio as a measure of strength. It is well known that a high likelihood ratio should not be confused with a high posterior probability. Combined with a sufficiently low prior probability, a high likelihood ratio may result in still low probabilities. Conversely, a high posterior may mis-

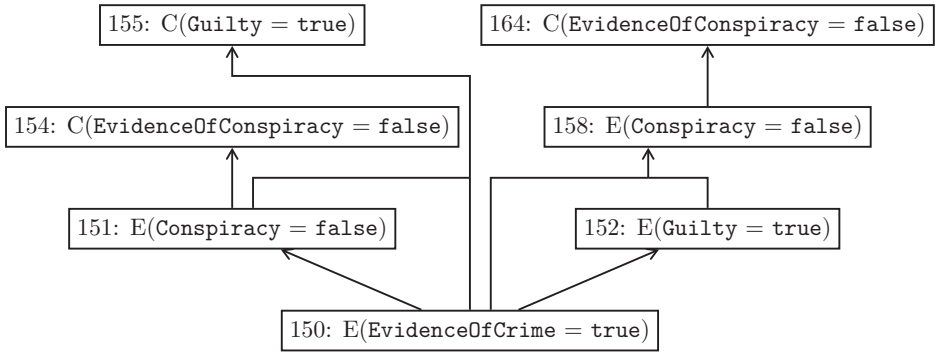


Figure 3.16: Arguments built from the extracted rule base, using posterior probability as the strength measure with threshold of 0.8. Evidence in the BN was entered for `EvidenceOfCrime = true`.

takenly be interpreted as an effect of a strong rule, while the probability of the conclusion is also high without the premises of the rule. That is, even independent premises could seemingly contribute to strong rules. Figures 3.16, 3.17 and 3.18 show the argument graphs resulting from the three scenarios above evaluated with the posterior probability measure.

The most noteworthy feature of these graphs is that the interpretation of what an argument represents has changed from

> *"the subarguments positively affect the probability of the conclusion of an argument"*

to

> *"given the outcomes of the subargument, the probability of the conclusion is high."*

With respect to the aim of this method, which is to explain why certain variables have a high posterior in the BN, it would seem more useful to explain the posterior probabilities in terms of how observations change that probability, rather than restating which variables have high posteriors. However, absolute measures also have a disadvantage, which is that an appropriate threshold may be hard to define. There is no natural threshold for incremental measures—such as 1.0 for the incremental measure above, which captures the boundary between positive and negative influence.

3.7 Discussion

We have formally defined how to extract ASPIC+ arguments from BNs. By doing so we have established a formal connection between concepts from Bayesian reasoning and argumentative reasoning. In particular we have shown the connection

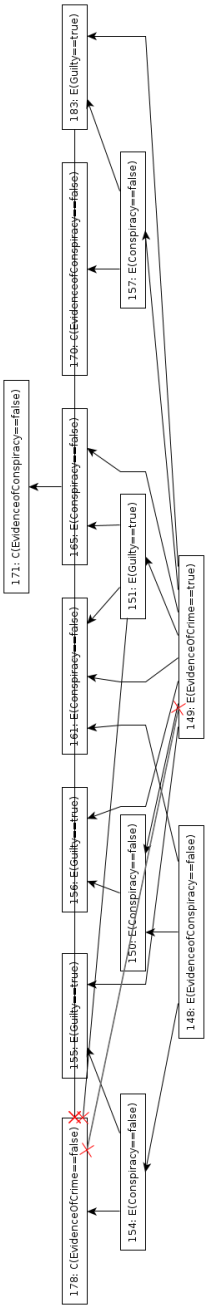


Figure 3.17: Arguments built from the extracted rule base, using posterior probability as the strength measure with threshold of 0.8. Evidence in the BN was entered for `EvidenceOfCrime` = `true` and `EvidenceOfConspiracy` = `false`.

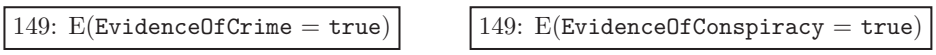


Figure 3.18: Arguments built from the extracted rule base, using posterior probability as the strength measure with threshold of 0.8. Evidence in the BN was entered for `EvidenceOfCrime` = `true` and `EvidenceOfConspiracy` = `true`.

between explaining away and argument attack. By extracting rules and undercutters from BNs we were able to construct arguments that support or attack each other. With an example we showed that the characteristic features of probabilistic inference and explaining away are captured by argumentative inference and undercutters. Using this insight, we proposed an algorithm that extracts inference rules from a given input BN. These rules can then be chained into arguments using ASPIC+. Taking inspiration from Pearl's CE-system, we adjusted such an argumentation system such that it correctly represents d-connected paths in the BN as arguments.

Throughout this method, a measure of inferential strength must still be chosen. We have shown that different options for this measure exist and we have argued that there is a natural distinction between absolute and incremental measures. Given the explanatory purpose of our method, an incremental measure has some natural advantages of absolute measures.

We have demonstrated the proposed method with an example from the legal domain. However, the method is not limited to legal or forensic applications. We have nowhere in the design made any assumptions about the nature of the BN or the variables in it.

A problem that arises with any automated approach is the computational complexity of combining inference rules. There are numerous ways in which rules can be combined. Doing probabilistic inference in BNs is already computationally hard. However, we find that combining rules in all possible ways is in practice far more time consuming than calculating the probabilities required to identify the rules. We have also seen that, even for an input network of four variables, the size of the argument graphs is significantly larger than the size of the input network. For larger networks the size of the argument graph quickly becomes unmanageable. To make the process of finding good arguments more feasible on large scale networks, is the subject of the next chapter.

Chapter 4

Structure guided argument construction using support graphs

In the previous chapter we introduced the notions of probabilistic rules and arguments and a simple algorithm to extract those from a BN. The resulting arguments are intended to explain the inner workings of a Bayesian network (BN) in an argumentation-based way. To this end we created for each variable in the BN potential rules using an assignment to that variable as the consequents and all possible combinations of value assignments to the variables in its Markov blanket as antecedents. For larger networks, however, this algorithm, which exhaustively enumerates every possible probabilistic rule and argument, is computationally infeasible because it examines inferences between all combinations of variable assignments. The number of rules that are identified and subjected to a strength evaluation is (worst case) exponential in the number of BN variables. This is undesirable, both from a computational point of view and from an explanation point of view, because many of the possible inferences are superfluous. Moreover, the algorithm from Chapter 3 does unnecessary work because many of the enumerated antecedents will never be met, resulting in irrelevant rules. Similarly, many arguments constructed in this way are superfluous because they argue for conclusions that are irrelevant to the evidential problem at hand.

In this chapter we limit the number of created rules by first extracting the structure of the desirable arguments. We formalise a two-phase method for extracting probabilistically supported arguments from a Bayesian network. First, from a Bayesian network we construct a *support graph* and, second, given a set of observations we build arguments from that support graph. This eliminates the aforementioned problem of unnecessarily enumerating irrelevant rules and arguments. The extracted arguments can facilitate the correct interpretation and

explanation of the relation between hypotheses and evidence that is modelled in the Bayesian network and more closely follow the reasoning chains in the BN than the arguments extracted by the method proposed in Chapter 3. This new method also provides a new conceptual view on (probabilistic) argumentation in which arguments pro and con a conclusion are always considered collectively.

In Section 4.1 we present a simplified version of the ASPIC+ framework because in this chapter we do not need undercutters for defeasible rules. This is because we extract arguments in which all pro and con reasons are summarised and rules in these arguments have no further exceptions. In Section 4.2 we formally define and discuss support graphs. Using the notion of a support graph we introduce a formalisation of argument construction in Section 4.3. We apply this method in a case study in Section 4.4.

4.1 Introducing a special case of ASPIC+

In the argumentation system that we will introduce there is only one form of attack possible, which is to *rebut* an argument on a conflicting conclusion. Since all premises are certain, arguments cannot be *undermined* and, as discussed, there are also no undercutting arguments possible. This results in simplified definitions of attack and defeat:

Definition 4.1 (Simplified attack (Definition 2.8))**.** *Argument A attacks another argument B on $B' \in \mathrm{Sub}(B)$ iff A rebuts B on B'. Argument A rebuts argument B on B' iff* $\mathrm{Conc}(A) = \overline{\mathrm{Conc}(B')}$.

To determine which arguments defeat each other, a preference ordering $\preccurlyeq$ is required. We denote the strict version of the ordering as $A \prec B$ when both $A \preccurlyeq B$ and $A \not\succcurlyeq B$. Such an ordering is usually defined on the basis of an ordering of the defeasible rules, but in our case it will be based on a notion of strength that is derived from the probabilities in the BN.

Using an argument ordering, some of the attacks result in defeat of the attacked argument.

Definition 4.2 (Simplified argument defeat (Definition 2.9))**.** *Given a collection of arguments $\mathcal{A}$ ordered by an ordering $\preccurlyeq$, a defeat relation $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ among arguments is defined such that: argument A defeats argument B iff A rebuts B on $B' \in \mathrm{Sub}(B)$ and $A \not\prec B'$.*

As before, the set of arguments $\mathcal{A}$ and the defeat relation $\mathcal{D}$ can be used as input to Dung's theory of abstract argumentation [Dung, 1995] as described in Chapter 2 and also done in Chapter 3.

4.2 Support graphs

The process of argument generation can be split in two phases. We first construct a support graph from a BN and subsequently establish arguments from the support graph. In this section we define the support graph and its construction and give an illustration of the construction of a support graph in a small example BN. Moreover, we identify some useful properties of support graphs. The motivation for this graphical transformation from the BN to a support graph is that it abstracts away from the Bayesian network in a way that retains the reasoning chains from the BN. As we will see later, these chains form the skeleton of the arguments, without dealing with evidence yet.

In the previous chapter we developed a method to identify arguments in a BN setting based on exhaustive enumeration of probabilistic rules and rule combinations. A disadvantage of the exhaustive enumeration is the combinatorial explosion of possibilities, even for small models. Using a support graph, we will be able to reduce the number of arguments that needs to be enumerated because only rules relevant to the conclusion of the argument will be considered. We make this more precise in the next section.

4.2.1 Definition

Given a BN and a variable of interest $\mathtt{V}^\star$, the support graph is a template for generating explanatory arguments. It captures the chains in the BN that end with the variable of interest. As such, it does not depend on observations of variables but rather represents the possible structures in arguments for a particular variable of interest in a particular BN. This means that it can be used to construct an argument based on any set of observations, as we will show in the next section. When new evidence becomes available the support graph can be reused (presuming that the variable of interest does not change). This means that the support graph should be able to capture the dynamics in d-separation caused by different observations. Since d-separation is defined on chains we first introduce the notion of a *support chain*.

Definition 4.3 (Support chain)**.** *Given a BN* $((\mathbf{V}, \mathbf{E}), \mathrm{P})$*, a support chain for a variable of interest* $\mathtt{V}^\star \in \mathbf{V}$ *is a sequence of variables that:*

- *follows a simple chain in the BN graph, except that for every immorality* $(\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k})$ *for which* $\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k}$ *is on that chain in the BN graph,* $\mathtt{V_j}$ *is skipped in the support chain;*
- *ends in* $\mathtt{V}^\star$*.*

The intuition behind a support chain is that the effects of observations of a variable in the BN will propagate through the graph and have some influence on $\mathtt{V}^\star$ through the other variables along these chains. From Pearl [1988b] we know that immoralities can create possible intercausal interactions that deserve special at-

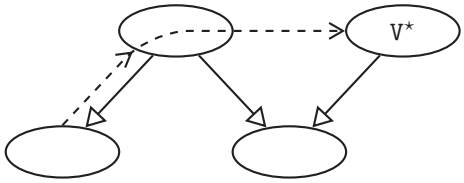


Figure 4.1: Illustration of a support chain for variable of interest $\mathtt{V}^\star$. The BN edges are solid and have open triangle tips. A possible support chain is shown in dashes and with pointy arrow tips.

tention: variables that are only connected through a head-to-head connection are a-priori independent. Information should therefore not be propagated through an inactive head-to-head connection. Additional information can, however, create an intercausal dependency. To explicitly capture the possibility of such an induced intercausal relation, we bypass the immoralities in the support chains and create direct links between parents of a common child. In this way every support chain represents a possibly active chain for some set of observations and any chain that is active for some set of observations is represented by a support chain. An example is shown in Figure 4.1.

To capture all possible ways in which a variable $\mathtt{V}^\star$ can be supported we define the notion of a *support graph*, which can combine multiple support chains. These support chains can be combined in many ways. The definition below defines a family of support graphs that are all valid in the sense that every possible support graph is represented. When used to construct arguments, however, we will see that one specific support graph is exceptionally useful and we provide an algorithm that constructs this (in a sense minimal) support graph.

Definition 4.4 (Support graph)**.** *Given a BN* $((\mathbf{V}, \mathbf{E}), P)$ *and a variable of interest* $\mathtt{V}^\star \in \mathbf{V}$*, a support graph is a pair* $(\mathcal{G}, \mathcal{V})$ *where* $\mathcal{G}$ *is an acyclic directed graph* $(\mathbf{N}, \mathbf{L})$ *with nodes* $\mathbf{N}$ *and edges* $\mathbf{L}$*, and* $\mathcal{V} : \mathbf{N} \mapsto \mathbf{V}$ *associates a variable with every node, such that:*

> $\mathcal{V}(N_1), \mathcal{V}(N_2), \ldots, \mathcal{V}(N_n)$ *is a support chain if and only if* $N_1, N_2, \ldots, N_n$ *is a simple, directed path in* $\mathcal{G}$ *with* $\mathcal{V}(N_n) = \mathtt{V}^\star$*.*

We will call N_i a *supporter* of N_j if N_i is a parent of N_j in the support graph, i.e. there is an edge from N_i to N_j.

If the BN graph is multiply connected, a variable may be reachable in more than one way. In that case, it can be associated with more than one of the nodes in the support graph. To distinguish between nodes in the support graph for the same variable, a mapping $\mathcal{V} : \mathbf{N} \mapsto \mathbf{V}$ is introduced that maps support graph nodes to the corresponding variables. When confusion is not possible we will abuse terminology and call $\mathtt{V_i}$ a supporter of $\mathtt{V_j}$ when we intend to say that N_i is a supporter of N_j for which $\mathcal{V}(N_i) = \mathtt{V_i}$ and $\mathcal{V}(N_j) = \mathtt{V_j}$.

We will later show how active and inactive chains are treated when we use the support graph to construct arguments about the case. Without knowing which variables are instantiated, the paths in the support graph represent all possibly active chains in the BN.

One of the often misleading aspects of BNs is that directions of individual arrows have no inherent meaning. Sometimes an arrow can be reversed without consequences for the implied independence relation. This is captured by the Markov equivalence property that we mentioned before. One of the advantages of support graphs is that they take away this confusing aspect. Indeed, we can prove that Markov equivalent BNs generate the same support graphs.

Proposition 4.5. *Given two Markov equivalent BN graphs G and G', and a variable of interest $\mathtt{V}^\star$, the sets of support graphs are identical for both BNs.*

Proof. Markov equivalent BN graphs have the same skeleton and the same immoralities. Therefore, they must have the same support chains (which follow the skeleton but bypass immoralities). The set of support chains uniquely defines the possible support graphs, which must therefore be equal. □

A trivial support graph can be constructed by simply enumerating simple chains in the BN and creating a path for every such chain in the support graph, which results in a forest with as many components as there are simple chains in the BN and every such component is a linear path. Since the number of simple chains in a BN is of the order $\mathcal{O}(|\mathbf{V}|!)$ this is not feasible, nor desirable, even for small BNs. Instead, we introduce an algorithm that constructs a more concise support graph in which paths with common prefixes are merged. This algorithm is shown in Algorithm 4.1 and illustrated in Figure 4.2.

The support graph construction algorithm, given in Algorithm 4.1, uses the notion of a *forbidden set* of variables to maintain a list of variables that should not be used in further support in that branch. This set is used to prohibit the use of, for instance, cyclical reasoning, or reasoning along a head-to-head connection. Figure 4.2 shows the forbidden sets in the three cases of the algorithm. The forbidden set of a new supporter N_i for variable $\mathtt{V_i}$ always includes the variable $\mathtt{V_i}$ itself, which prevents cyclic traversal of the BN graph and corresponds to the fact that the support graph represents simple chains only.

As we have discussed, in a BN, parents of a common child often exhibit intercausal interactions (such as explaining away, which occurs when a child has positive correlations with both parents but the parents are negatively correlated with each other). More generally, the influence between parents may be weaker or stronger, and, in an extreme case, even have the opposite sign from what we may expect based on the individual influences between the common child and the two parents. Supporting a variable $\mathtt{V_i}$ with one of its children $\mathtt{V_j}$ and then supporting this child $\mathtt{V_j}$ by a parent $\mathtt{V_k}$ would incorrectly chain the inferences through a

function `SupportGraphConstruction(`G, $\mathtt{V}^\star$`):`
  Input: $G = (\mathbf{V}, \mathbf{E})$ is the BN graph with variables $\mathbf{V}$ and edges $\mathbf{E}$
  Input: $\mathtt{V}^\star$ is the variable of interest
  Output: a support graph $\mathcal{G} = (\mathbf{N}, \mathbf{L})$
  $\mathbf{N} := \{N^\star\}$ $\mathbf{L} := \emptyset$
  $\mathcal{V}(N^\star) := \mathtt{V}^\star$
  $\mathcal{F}(N^\star) := \{\mathtt{V}^\star\}$
  `expand(`$\mathcal{G}$,$N^\star$`)`
function `expand(`$\mathcal{G}$,N_i`):`
  Input: $\mathcal{G} := (\mathbf{N}, \mathbf{L})$ is the support graph under construction
  Input: N_i is the support graph node to expand with $\mathtt{V_i} = \mathcal{V}(N_i)$
  foreach $\mathtt{V_j} \in$ `MarkovBlanket(`$\mathcal{V}(N_i)$`)` **do**
    if $\mathtt{V_j} \in \mathtt{Par}(\mathtt{V_i}) \setminus \mathcal{F}(N_i)$ **then** `// case I`
      $\mathcal{F}_{\text{new}} := \mathcal{F}(N_i) \cup \{\mathtt{V_j}\}$
      `AddSupport(`$\mathcal{G}$,N_i,$\mathtt{V_j}$,$\mathcal{F}_{new}$`)`
    end
    else if $\mathtt{V_j} \in \mathtt{Cld}(\mathtt{V_i}) \setminus \mathcal{F}(N_i)$ **then** `// case II`
      $\mathcal{F}_{\text{new}} := \mathcal{F}(N_i) \cup \{\mathtt{V_j}\} \cup \{\mathtt{V_k} | (\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k})$ is an immorality$\}$
      `AddSupport(`$\mathcal{G}$,N_i,$\mathtt{V_j}$,$\mathcal{F}_{new}$`)`
    end
    else if $\mathtt{V_j} \in \mathtt{Par}(\mathtt{V_k}) \setminus \mathcal{F}(N_i)$ *s.t.* $\mathtt{V_k} \in \mathtt{Cld}(\mathtt{V_i})$ **then** `// case III`
      $\mathcal{F}_{\text{new}} := \mathcal{F}(N_i) \cup \{\mathtt{V_j}, \mathtt{V_k}\}$
      `AddSupport(`$\mathcal{G}$,N_i,$\mathtt{V_j}$,$\mathcal{F}_{new}$`)`
    end
  end
function `AddSupport(`$\mathcal{G}$,N_i,$\mathtt{V_j}$,$\mathcal{F}_{new}$`):`
  Get from $\mathcal{G}$ a node N_j with:
    $\mathcal{V}(N_j) = \mathtt{V_j}$ and
    $\mathcal{F}(N_j) = \mathcal{F}_{\text{new}}$
    or create it if it does not exist in $\mathcal{G}$
  Add (N_j,N_i) to $\mathbf{L}$ in $\mathcal{G}$
  `expand(`$\mathcal{G}$, N_j`)`

Algorithm 4.1: Recursive algorithm to construct a support graph while building forbidden sets $\mathcal{F}$. Note that, although the order in which the support graph is constructed is not deterministic, the output is not dependent on the order in which nodes are added to the graph because new nodes do not depend on other branches of the already constructed graph.

head-to-head node in cases where an intercausal interaction is possible. Therefore we ensure that this last step cannot be made, by including any other parents that constitute immoralities with a shared child in the second case in the algorithm. A reasoning step that uses the inference according to the intercausal interaction

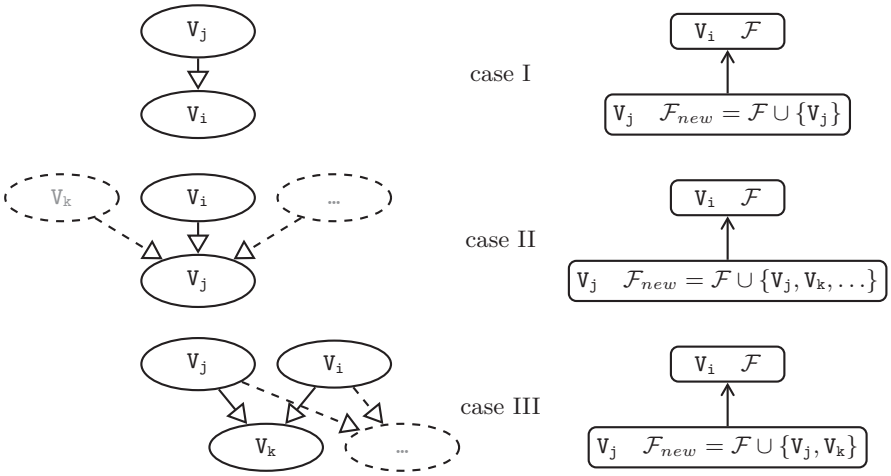


Figure 4.2: Visual representation of the three cases in Algorithm 4.1. A support node for variable $\mathtt{V_i}$ can obtain support in three different ways from a variable $\mathtt{V_j}$, depending on its graphical relation to $\mathtt{V_i}$. Note that every support node N_i is labelled with $\mathcal{V}(N_i) = \mathtt{V_i}$ and $\mathcal{F}(N_i)$.

is allowed by the third case of the algorithm. In terms of the support chains this disallows the traversal of a head-to-head connection that is involved in an immorality and it creates the shortcut between the parents of a common child. Note that the use of the intercausal reasoning step requires evidence to be present for the common child (or descendant). Since the support graph is abstracted from the collection of evidence we allow the step in the support graph, and ensure that the subsequent argument construction verifies that premises and conclusions taken from the support graph are indeed probabilistically dependent.

Theorem 4.6 (Correctness)**.** *Algorithm 4.1 creates a support graph* $\mathcal{G} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$ *for a variable of interest* $\mathtt{V}^\star \in \mathbf{V}$ *of a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *and probability function* P*.*

Proof. We prove this in two parts:

1. $\mathcal{V}$ maps every simple directed path in $\mathcal{G}$ ending with the root to a support chain in G, and
2. for every support chain in G, there is a simple path to the root in $\mathcal{G}$ that is mapped to this support chain by $\mathcal{V}$.

Part 1. Any simple directed path in the support graph is constructed from the steps in Algorithm 4.1 and therefore represents a sequence of nodes in the BN. We need to prove that the sequence of mapped variables is a simple chain in the BN graph where immoralities have been bypassed. By putting the visited variables in the forbidden set $\mathcal{F}$ it is ensured that this sequence is simple. What remains to be shown is that every consecutive pair of support nodes in $\mathcal{G}$ maps to a parent-child

pair or a bypass of an immorality, and that no sequence of support nodes in $\mathcal{G}$ maps to an immorality in G. The former is ensured by the three cases in the algorithm. Every step either goes to a parent or child, creating a parent-child pair in the chain, or to a parent of a child that together form an immorality. The latter (no immoralities remain) follows from the addition of $\mathtt{V_k}$ to $\mathcal{F}$ in the second case of the algorithm that makes it impossible to move to a parent after you move to a child in this sequence.

Part 2. A support chain in G is a simple chain in which immoralities have been bypassed. We need to prove that all such chains have a corresponding directed path in the graph found by the algorithm. We prove this by induction. Suppose that, at some point during the construction, the last part of a support chain starting at $\mathtt{V_i}$ and ending in $\mathtt{V}^\star$ is already represented by a path in the constructed support graph. Then, there is a root N_i in the support graph under construction with $\mathcal{V}(N_i) = \mathtt{V_i}$. The previous variable on the support chain, $\mathtt{V_j}$, is not in $\mathcal{F}(N_i)$ because in that case the support chain would either not be simple or contain an immorality. Therefore $\mathtt{V_j}$ is added in one of the three cases of the algorithm. Given that the end of every support chain $\mathtt{V}^\star$ is added in the first step of the algorithm, this inductively proves that all support chains are found. □

The specific support graph constructed by Algorithm 4.1 has a number of interesting properties that we will discuss later, but we first present a small step-by-step example of this algorithm to familiarise the reader with the method.

4.2.2 Example of construction

Let us now consider the example BN from Figure 2.2 and take `Crime` as the variable of interest $\mathtt{V}^\star$ since, ultimately, that is the variable under legal debate, which models whether or not the crime was committed by the suspect. The construction steps are shown in Figure 4.3. We initiate support graph construction by creating one solitary node $N^\star$ with this variable as its root, i.e. $\mathcal{V}(N^\star) = \mathtt{Crime}$. The forbidden set for this node is simply $\{\mathtt{Crime}\}$ (step 1 in the figure). We then add nodes to the support graph by trying all three extension steps as described above. The `Crime` node has a parent and a child which has another parent, so all three cases apply (exactly once) and we create three supporters in the support graph. First, the `Crime` node can be supported by its parent (`Motive`). This confirms our intuition that the existence of a motive for the suspect affects our belief in the suspect having committed the crime. Secondly, the `Crime` node can also be supported by its child (`DNA_match`) because a match is strong evidence for the suspect's guilt. And thirdly, outcomes of the `Crime` variable may be supported by outcomes of the *parent of a child* node `Twin`. This corresponds to the fact that finding that the suspect has an identical twin explains away the evidence of the DNA match. These three have been added in step 2 of Figure 4.3.

Let us now consider the forbidden sets starting with the last supporter (`Twin`). When using a head-to-head connection in the BN to find support, the common

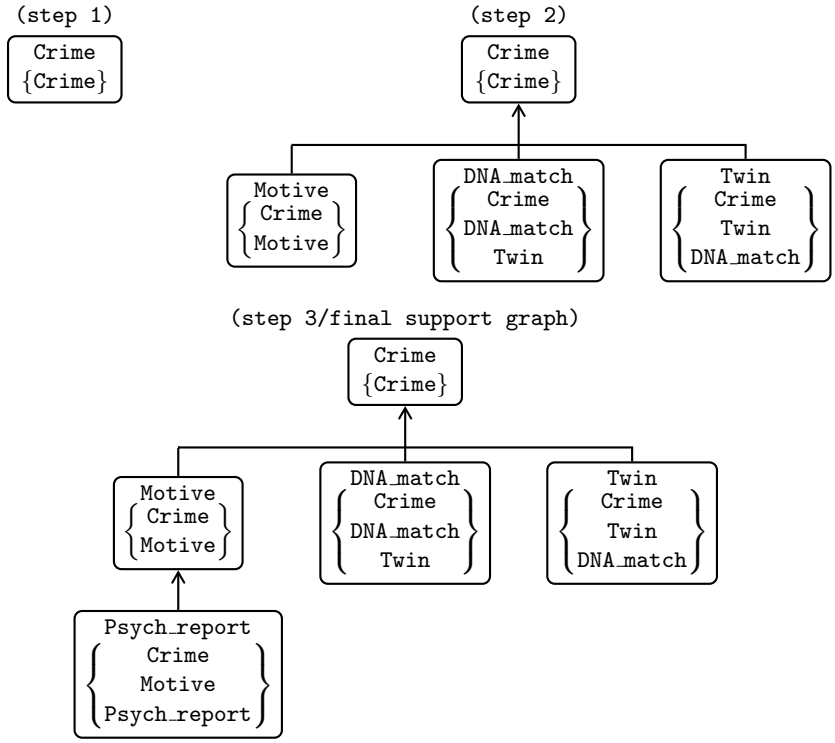


Figure 4.3: The steps in the construction of the support graph corresponding to the example in Figure 2.2 with $\mathtt{V}^\star = \mathtt{Crime}$. For every node N_i we have shown the variable name $\mathcal{V}(N_i)$ together with the forbidden set $\mathcal{F}(N_i)$. Multiple edges $(\mathtt{V_i}, \mathtt{V}), \ldots, (\mathtt{V_k}, \mathtt{V})$ into the same node are represented by a hyperedge.

child is added to the forbidden set, which then becomes {`Crime`, `DNA_match`, `Twin`}. This eliminates any further support because it covers the entire Markov blanket of the `Twin` node. For the second supporter (`DNA_match`), the forbidden set is exactly the same because now the child is the supporter itself (and is added to $\mathcal{F}$ for that reason) and any other parents (`Twin` in this case) are added to the forbidden set as prescribed by the algorithm. Again, the entire Markov blanket of the `DNA_match` variable is covered by the forbidden set and no further support is possible. For the first supporter that we mentioned (`Motive`), however, one additional supporter can be added. The forbidden set of the support graph node for `Motive` that we created will be {`Crime`, `Motive`}. This means that the child `Psych_report` can be used to support outcomes of the `Motive` variable (step 3). This is the result of the fact that the Bayesian network captures the correlation between having a motive and a psychological report on finding this motive. No further support can be added for the `Psych_report` variable and the support graph construction is finished.

4.2.3 Properties of the support graph algorithm

We now describe some properties of our algorithm to construct support graphs that serve to illustrate the way in which support graphs capture an efficient argumentative representation of what is modelled in a BN.

Property 4.7. *Given a BN with $G = (\mathbf{V}, \mathbf{E})$, Algorithm 4.1 constructs a support graph containing at most $|\mathbf{V}| \cdot 2^{|\mathbf{V}|}$ nodes, regardless of the variable of interest.*

Proof. Variables can occur multiple times in the support graph but never with the same $\mathcal{F}$ sets. This set contains subsets of other variables and therefore $2^{|\mathbf{V}|}$ is a strict upper bound on the number of times any variable can occur in the support graph. The total number of support nodes is therefore limited by the expression $|\mathbf{V}| \cdot 2^{|\mathbf{V}|}$. □

The bound given by the above property is a theoretical upper bound. In practice the number of support nodes will often be significantly smaller when the BN graph is not densely connected. In the special case where the BN is singly connected (definition on page 22) we can prove that the support graph contains exactly the same number of nodes as the BN.

Many known graph algorithms that have an exponential worst case running time on multiply connected inputs, have polynomial running times for singly connected graphs. This also holds for our support graph construction algorithm:

Property 4.8. *Given a BN graph $G = (\mathbf{V}, \mathbf{E})$ and the support graph $\mathcal{G} = (\mathbf{N}, \mathbf{L})$ constructed by Algorithm 4.1 for some variable of interest. If G is singly connected, every variable occurs exactly once in $\mathcal{G}$ and the size of the support graph is $|\mathbf{N}| = |\mathbf{V}|$.*

Proof. A variable can in theory occur multiple times in the support graph, but this only happens when the graph is loopy (multiply connected). In a singly connected graph there are no loops. This means that using the three available steps from Algorithm 4.1, the recursive construction encounters every variable exactly once after which it will be forbidden in the ancestors of the resulting support node and unreachable in the BN from any other branch of the support graph. □

More generally, the number of support nodes for a single variable $\mathtt{V_i}$ is bounded by the number of simple chains from $\mathtt{V_i}$ to $\mathtt{V}^\star$ which is smaller for less densely connected graphs. The sparser the BN graph, therefore, the more the support graph will approach size $|\mathbf{V}|$.

This shows that the support graph is a concise model to represent the inferences in a BN. We have already seen that support graphs abstract from the sometimes confusing interpretation of the directions of edges. From the bounds on the size of

the support graph a bound on the complexity of the algorithm can easily be derived. Specifically, the `expand()` function is called once for every node in the final graph and has itself a worst case complexity of $\mathcal{O}(|\mathbf{V}|)$ because it loops once over the Markov blanket of each variable which could contain all other variables in the graph in the worst case. The worst case complexity of Algorithm 4.1 is therefore bounded by $\mathcal{O}(|\mathbf{V}|^2 * 2^{|\mathbf{V}|})$ in general and $\mathcal{O}(|\mathbf{V}|^2)$ for singly connected graphs. One of the reasons why BNs are popular as a model for probability distributions is that they provide a considerable reduction in computational power when the graph is not densely connected. A similar improvement holds for our algorithm. In practice, the Markov blankets often contain only a relatively small portion of the other variables in the BN, resulting in fast execution times.

We have already proven in Theorem 4.9 that two Markov equivalent graphs share the same set of possible support graphs for a specific node of interest. We now show that for our algorithm we can prove that Markov equivalent BN graphs result in a single unique support graph.

Theorem 4.9. *Given two Markov equivalent BN graphs G and G', and a variable of interest $\mathtt{V}^\star$, the two support graphs resulting from Algorithm 4.1 ($\mathcal{G}$ and $\mathcal{G}'$) are identical.*

Proof. Consider the BN graph G and the corresponding support graph $\mathcal{G}$. In a Markov equivalent graph G' edges may be reversed but not if this creates or removes immoralities. We can prove that the support graphs for G and G' are identical by induction. First the roots have the same variable $\mathtt{V}^\star$ and the same forbidden set $\{\mathtt{V}^\star\}$ by definition. Then, in every iteration of the support graph construction algorithm the added nodes are identical if the support graphs under construction are to be identical. Following the three possible support steps we see that every supporter follows an edge from the skeleton (which stays the same) or an immorality (which also stays the same). This means that the variables that are associated with the newly added nodes must be the same. If the support graphs were to differ, this has to follow from a different forbidden set. What remains to be shown is that the forbidden sets will also be equal given that the (alrealy found) children have the same forbidden set. Let us consider the three cases of the $\mathcal{F}$ update from Algorithm 4.1 (see also Figure 4.2). Suppose that in the support graph of G, N_i with $\mathcal{V}(N_i) = \mathtt{V_i}$ is supporting N_j (i.e., N_i is a parent of N_j in $\mathcal{G}$) with $\mathcal{V}(N_j) = \mathtt{V_j}$:

- if $\mathtt{V_j}$ is a parent of $\mathtt{V_i}$ in G (case I), then
  - if the direction of the edge in G' is also from $\mathtt{V_j}$ to $\mathtt{V_i}$ the forbidden sets are trivially the same and
  - if the edge is reversed in G' (from $\mathtt{V_i}$ to $\mathtt{V_j}$) then in G' this is handled by case II. This adds any $\mathtt{V_k}$ to the forbidden set for which $(\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k})$ is an immorality. However, $(\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k})$ cannot be an immorality for any $\mathtt{V_k}$ in G' because it was not in G and the immoralities are the same.

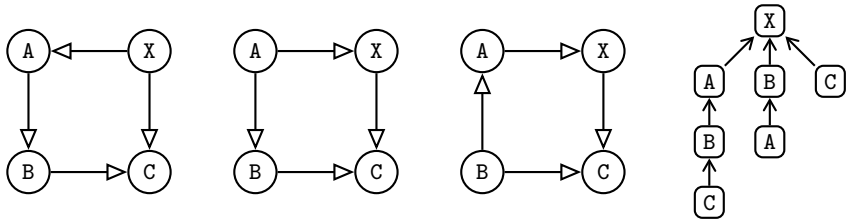


Figure 4.4: Three Markov equivalent BNs and their unique support graph for the case that $\mathtt{V}^\star = X$.

- if $\mathtt{V_j}$ is a child of $\mathtt{V_i}$ in G (case II), then
  - if the direction of the edge in G' is also from $\mathtt{V_i}$ to $\mathtt{V_j}$ the forbidden sets are trivially the same and
  - if the edge is reversed in G' (from $\mathtt{V_j}$ to $\mathtt{V_i}$) then in G' this is handled by case I. The forbidden sets are the same except that in G any $\mathtt{V_k}$ is added that constitutes an immorality $(\mathtt{V_i}, \mathtt{V_j}, \mathtt{V_k})$. Again, no such $\mathtt{V_k}$ exists because reversal of the edge would not be allowed in G'.
- if $(\mathtt{V_i}, \mathtt{V_k}, \mathtt{V_j})$ is an immorality (case III), then it must also be an immorality in G' because immoralities in G and G' are the same. Therefore the forbidden sets must also be identical.

Therefore, during the execution of the algorithm on Markov equivalent graphs, the forbidden sets are exactly identical, and therefore the constructed support graphs will be identical. □

What this theorem shows is that Markov equivalent models are mapped to the same support graph, which means that they will receive the same argumentative explanation later on. In Figure 4.4, for example, we show three different but Markov equivalent BNs and the single resulting support graph.

In Section 4.3 on argument construction the following property is helpful. It states that the support graph constructed by Algorithm 4.1 is 'minimal' in the sense that support chains have been merged as much as possible. This means that for every support node the set of supporters is 'maximal'.

Theorem 4.10. *Assume a BN with graph G and a variable of interest $\mathtt{V}^\star$. Denote the support graph constructed by Algorithm 4.1 as $\mathcal{G}$. Consider a support chain in the BN graph G ending in $\mathtt{V}^\star$. We have that there is a unique directed simple path in $\mathcal{G}$ that maps to this support chain.*

Proof. That no other simple path in $\mathcal{G}$ maps to the same support chain in G follows from the fact that the algorithm never creates multiple support nodes for the same variable with the same forbidden set, and that a support chain uniquely defines the forbidden set (through the 3 cases in the algorithm). Therefore, any support chain in G is represented by one such path in the support graph constructed by Algorithm 4.1. □

This is exactly the minimality property of Algorithm 4.1 that we hinted at earlier. It means that chains in the support graph are merged as much as possible which makes it the most concise support graph among all support graphs that are theoretically possible.

4.3 Argument construction

From a support graph, arguments can be generated that match the reasoning in the BN, since the support graph captures all possible chains of inference. In this section we show how arguments can be generated on the basis of a support graph as constructed by Algorithm 4.1. We will employ a *strength measure* to rank inferences and to prevent arguments that follow inactive paths in the BN graph.

The interpretation of an argument in this chapter is slightly different from what is common in argumentation systems and also different from our interpretation of arguments in Chapter 3. Since we try to capture the Bayesian network reasoning in arguments, the arguments extracted in this chapter will encapsulate all pro and con reasons for their conclusions. This will reflect the way in which Bayesian networks internally weigh all evidence. The resulting arguments will, therefore, attack and defeat each other in a way that is not common in argumentation. The aim of such an argumentation system is to provide an explanation of the probabilistic reasoning captured by the Bayesian network. This is in contrast to the usual modelling of argumentation, in which reasons for and against a conclusion are distributed over conflicting arguments. Consider, for example, the reasons to believe that a suspect was present at a crime scene at the time of an offence. In other argumentative models, it is usually the case that an argument in favour of a conclusion (based on a matching DNA profile that was recovered from the crime scene, for instance) and an argument against it (a witness testifying that the suspect was at another location at the time) would result in two arguments: one for the conclusion that the suspect was at the crime scene and one for the conclusion that he/she was not. In our method, however, we only find the argument for one of these conclusions that has both the DNA test and the witness testimony as premises. Which conclusion we find depends on the probabilities involved. The interpretation of such an argument is that the conclusion holds 'because or despite' the premises. In case of the example above such an argument could be: 'The suspect was at the crime scene because the DNA profiles match, despite the fact that a witness has testified otherwise'. This explains why we do not need undercutters in this chapter. Undercutters indicate exceptions that can arise by further observations. By considering all pro and con evidence together there are never further observations that can undercut the inference.

The arguments that we build will follow the structure of the support graph. As such, the support graph can be seen as a skeleton to build arguments. We present

a formal model of these explanatory arguments which instantiates the ASPIC+ framework for structured argumentation. We also discuss how the grounded extension of such a framework can be generated efficiently on the basis of the support graph.

First, we define a logical language $\mathcal{L}$ of sentences used to build arguments. For this language, we take pairs $(N, \mathtt{o})$ of a support node N and one of the outcomes $\mathtt{o}$ of the associated variable $\mathcal{V}(N)$. Elements of this language negate each other iff they assign different outcomes to the same variable.

Definition 4.11 (Language for explanatory arguments)**.** *Given a BN with graph $G = (\mathbf{V}, \mathbf{E})$ and the corresponding support graph $\mathcal{G} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$, let the logical language $\mathcal{L}$ be defined as:*

$$\mathcal{L} = \{(N, \mathtt{o}) \mid N \in \mathbf{N} \text{ and } \mathtt{o} \in \text{vals}(\mathcal{V}(N))\}$$

For which the negation is defined as

$$\overline{(N, \mathtt{o})} = (N, \mathtt{o}') \text{ such that } \mathtt{o}' \in \text{vals}(\mathcal{V}(N)) \text{ and } \mathtt{o}' \neq \mathtt{o}$$

Since the support graph captures the allowed paths of reasoning, the rules in the argumentation system should follow the edges of this support graph. When a support node has multiple parents we must consider combinations of supporting parents to form a rule for an outcome of the supported node. In particular, we should consider all parents that can themselves be supported by evidence. This means that we must first consider which chains in the support graph start with actually observed evidence. For this we create a pruned version of the support graph in which all chains start with an instantiated variable and end in the variable of interest.

Definition 4.12 (Support graph pruning)**.** *Given a support graph $\mathcal{G}$ for variable of interest $\mathtt{V}^\star$ and evidence $\mathbf{e}$ for the BN variables $\mathbf{V}_e$, the pruned support graph $\mathcal{G}_\mathbf{e}$ is obtained by repeatedly removing from $\mathcal{G}$ every node N for which either:*

- *N is an ancestor of a node N' for which $\mathcal{V}(N') \in \mathbf{V}_e$ or*
- *$\mathcal{V}(N) \notin \mathbf{V}_e$, and N has no unpruned parents.*

The second condition resembles the definition of *barren nodes* [Jensen and Nielsen, 2007] in a Bayesian network except that nodes are barren in a BN iff they are uninstantiated and their *children* are barren.

Note that when there is no evidence, the pruned support graph will be empty. This is logically explained by the fact that this method is intended to explain the inferential relation between the variable of interest and the evidence. Without evidence there is nothing to explain.

In Figure 4.5 we have depicted the support graph from the running example together with the pruned version for the evidence variables $\{\texttt{Motive}, \texttt{DNA\_match}\}$. The node for `Psych_report` has been pruned because it satisfies both conditions

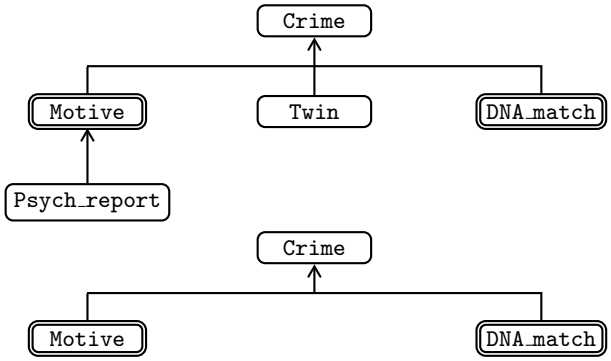


Figure 4.5: Support graph from the running example before and after pruning. Instantiated variables are depicted by double node outlines.

(the only path to $\mathtt{V}^\star$=`Crime` contains an instantiated variable and it has no unpruned ancestors) and `Twin` has been pruned by the second condition. The set of defeasible rules is defined to follow the structure of this pruned support graph.

Definition 4.13 (Defeasible rules)**.** *Given a support graph $\mathcal{G}$ as constructed by Algorithm 4.1, observations* $\mathbf{e}$ *and the pruned support graph $\mathcal{G}_\mathbf{e} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$, a rule in our argumentation system has the form $(N_1, \mathtt{o_1}), \ldots, (N_k, \mathtt{o_k}) \Rightarrow (N_c, \mathtt{o_c})$ such that*

- *$N_1, \ldots, N_k$ are all parents of N_c in $\mathcal{G}_\mathbf{e}$, and*
- *$\mathtt{o_c}$ is an outcome of the conclusion variable $\mathcal{V}(N_c)$, and*
- *$\mathtt{o_1}, \ldots, \mathtt{o_k}$ are outcomes of the associated variables $\mathcal{V}(N_1), \ldots, \mathcal{V}(N_k)$*

These rules are defeasible because they indicate a likely or probable inference rather than a strict deduction.These rules can be applied to evidence to derive conclusions about the variables in the BN. For this purpose the evidence that is entered in the BN is represented in the knowledge base:

Definition 4.14 (Knowledge base)**.** *Given a Bayesian network and evidence* $\mathbf{e}$ *for the variables $\mathbf{V}_e$. The knowledge base $\mathcal{K}_n$ contains all observations:*

$$\mathcal{K}_n = \left\{ (N_i, \mathtt{o_i}) \;\middle|\; \begin{array}{l} \mathcal{V}(N_i) \in \mathbf{V}_e, \text{ and} \\ \mathtt{o_i} \text{ is logically consistent with } \mathbf{e}, \text{ and} \\ (N_i, \mathtt{o_i}) \in \mathcal{L} \end{array} \right\}$$

Using Definitions 4.13 and 4.14 ASPIC+ specifies arguments and the attack relations between them. To resolve possible conflicts we consider how arguments can be evaluated against each other. As noted above in Definitions 2.9, arguments can attack each other on the outcome of the conclusion variable and defeat can be based on the *strength* of the arguments. To compute this strength we, again, apply one of the measures of inferential strength discussed in Chapter 3. For these,

we observed in Chapter 3 that *incremental* and *absolute* measures can be distinguished. In the examples in this chapter we will use the LR and the posterior odds measures to show how they compare. Recall that, although the support graph is not concerned with variable outcomes, the following (and in particular the likelihood ratio as a measure of strength) requires that variables are boolean-valued. Hence we assume that our input BN contains only binary-valued variables.

Inferential strength can be computed from the BN for every support graph node and depends on the evidence for variables in ancestors of that node in the support graph.

Definition 4.15 (Relevant premises to calculate strength)**.** *Consider a support graph* $\mathcal{G}_{\mathbf{e}} = ((N, L), \mathcal{V})$ *built from a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *by Algorithm 4.1 and pruned to observations* $\mathbf{e}$ *for the variables* $\mathbf{V}_e$*.*

The set of relevant premises (premises(N_i)*) of a support graph node* N_i *is an assignment to*

$$\mathbf{V}_e \cap \{\mathcal{V}(N_j) | N_j \in \text{Ancestors}(N_i)\}$$

that is logically consistent with $\mathbf{e}$*.*

As in Chapter 3, in order to correctly compute the inferential strength, it is important to take into account the correct *context*. This context is a subset of the observed evidence. Instantiations of the variable under consideration are again omitted. The evidence that overlaps with the ancestors of the node under consideration is excluded during the calculation of the strength because it occludes the potential influence between variables that we wish to detect. That is, to measure the potential influence of a DNA match on the guilt hypothesis we must (temporarily) ignore the fact that the DNA match is already observed. If we would not do that, the hypothesis would appear to be independent of the DNA match.

Definition 4.16 (Context to calculate strength)**.** *Consider support graph* $\mathcal{G}_{\mathbf{e}} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$ *built from a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *and pruned to observations* $\mathbf{e}$ *for the BN variables* $\mathbf{V}_e$*.*

The context (context(N_i)*) of a support graph node* N_i *is an assignment to*

$$\mathbf{V}_e \setminus \{\{\mathcal{V}(N_j) | N_j \in \text{Ancestors}(N_i)\} \cup \{\mathcal{V}(N_i)\}\}$$

that is logically consistent with $\mathbf{e}$*.*

As discussed, there are multiple ways to assign a numerical strength to an inference. In the following we will assume that an informed choice has been made and in our examples we will use either the LR or the posterior odds as a measure of strength:

Definition 4.17 (Likelihood ratio as measure of strength)**.** *Consider a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *and probability distribution* P*, a support graph* $\mathcal{G}_{\mathbf{e}} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$

for the variable of interest $\mathtt{V}^\star$*, and observations* $\mathbf{e}$ *for the variables* $\mathbf{V}_e$*. The LR strength of an assignment* $\mathtt{V_i} = \mathtt{o}$ *for a given support graph node* N_i *with* $\mathcal{V}(N_i) = \mathtt{V_i}$ *is*

$$strength_{LR}(\mathtt{V_i}, \mathtt{o}, N_i) = \frac{\mathrm{P}(premises(N_i) \mid (\mathtt{V_i} = \mathtt{o}) \wedge context(N_i))}{\mathrm{P}(premises(N_i) \mid (\mathtt{V_i} \neq \mathtt{o}) \wedge context(N_i))}$$

Definition 4.18 (Posterior odds as measure of strength)**.** *Consider a BN with graph* $G = (\mathbf{V}, \mathbf{E})$ *and probability distribution* P*, a support graph* $\mathcal{G}_\mathbf{e} = ((\mathbf{N}, \mathbf{L}), \mathcal{V})$ *for the variable of interest* $\mathtt{V}^\star$*, and observations* $\mathbf{e}$ *for the variables* $\mathbf{V}_e$*. The posterior odds strength of an assignment* $\mathtt{V_i} = \mathtt{o}$ *for a given support graph node* N_i *with* $\mathcal{V}(N_i) = \mathtt{V_i}$ *is*

$$strength_{odds}(\mathtt{V_i}, \mathtt{o}, N_i) = \frac{\mathrm{P}(\mathtt{V_i} = \mathtt{o} \mid premises(N_i) \wedge context(N_i))}{\mathrm{P}(\mathtt{V_i} \neq \mathtt{o} \mid premises(N_i) \wedge context(N_i))}$$

Strength as defined for assignments to support graph nodes can be lifted to argument strength directly.

Definition 4.19 (Argument strength and ordering)**.** *Let* A *be an argument with* $\mathrm{Conc}(A) = (N, \mathtt{o})$*. The strength of* A *is:*

$$\mathrm{strength}(A) = \mathrm{strength}(\mathcal{V}(N), \mathtt{o}, N)$$

From this an argument ordering follows. Recall from Chapter 2 that $A \prec B$ *is used to denote that* $A \preccurlyeq B$ *and* $A \not\succcurlyeq B$*.* $A \prec B$ *iff either:*

- B *is strict (premise argument from observation) and* A *is not, or*
- $\mathrm{strength}(A) < \mathrm{strength}(B)$

Figure 4.6 shows examples of arguments that can be constructed by ASPIC+ from the given definitions of rules and knowledge bases for the running example. Arguments A_1, A_2, A_3 and A_4 together in fact form the grounded extension of this argumentation system. This is because this argument graph uses the maximal set of premises in every inferential step and it assigns the outcomes that are probabilistically best supported. Figure 4.6 shows, in addition, a similar argument that uses the same set of premises for that conclusion variable but which draws the 'wrong' conclusion. Such an argument will always be rebutted by the similar argument for the right conclusion. If the two outcomes of the node are equally strong (which in the case of the LR measure of strength means the conclusion is independent of the premises given the evidence in the context), then arguments for both outcomes coexist but defeat each other and will therefore not be part of the grounded extension. In fact, the grounded extension in this argumentation system coincides with the set of undefeated arguments.

Theorem 4.20. *Consider an argumentation system with the above definitions for the language, rules, knowledge base and argument strength. An argument* A *is in*

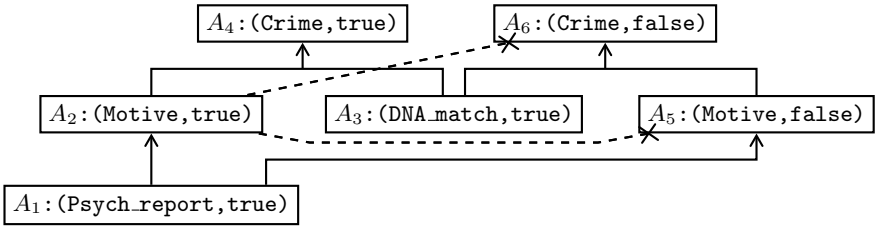


Figure 4.6: Some arguments resulting from our running example. Arrows show the immediate subargument relation. Besides the intuitively correct arguments $A_1, \ldots, A_4$ there are two additional arguments depicted that can also be made but that are successfully rebutted by A_2. The dashed arrows with crosshair tips show the defeat relation between arguments. Argument A_5 is defeated by A_2 because $(\texttt{Motive}, \texttt{true})$ is probabilistically stronger (using the likelihood ratio measure of strength in this case) than $(\texttt{Motive}, \texttt{false})$ based on this evidence. Any conclusion that builds on this second argument (such as A_6) is also defeated.

the grounded extension if and only if it is undefeated.

Proof. Undefeated arguments are by definition part of the grounded extension. For the other way around, we have to prove that any argument in the grounded extension is undefeated. We prove this by induction over subarguments.

For premise arguments, the base case, it is trivially true that they are undefeated because the argument ordering is such that by definition premise arguments are stronger than other arguments and no two premise arguments for different outcomes can exist.

Now for the induction step, we have to prove that an argument A in the grounded extension is undefeated, given the induction hypothesis which states that all immediate subarguments of A are undefeated.

By construction of our argumentation theory, an argument B with the opposite conclusion $\text{Conc}(B) = \overline{\text{Conc}(A)}$ can be constructed which has the same set of proper subarguments as A. Since A is in the grounded extension, there exists a reinstating argument C in the grounded extension that strictly defeats B. By the induction hypothesis we know that C must directly rebut B, since all the subarguments of B are undefeated. This means that $\text{strength}(C) > \text{strength}(B)$. By the definition of argument strength we have that $\text{strength}(C) = \text{strength}(A)$ and consequently $\text{strength}(A) > \text{strength}(B')$ for any B' that directly rebuts A. By the induction hypothesis we know that no subargument of A is defeated and hence, A is undefeated. □

Corollary 4.21. *For any argument A in the grounded extension with conclusion* $\text{Conc}(A) = (N, \texttt{o})$ *for variable* $\mathcal{V}(N) = \texttt{V}_\texttt{i}$*, there is no argument B in the grounded extension with* $\text{Conc}(B) = (N, \texttt{o}')$ *such that* $\text{strength}(\texttt{V}_\texttt{i}, N, \texttt{o}) < \text{strength}(\texttt{V}_\texttt{i}, N, \texttt{o}')$*. In other words, if strength is given by the posterior probability,*

then the arguments in the grounded extension are for those assignments with the highest probability in the BN.

Because our argumentation theory has no strict rules and no presumed knowledge it follows that any argument ordering is *reasonable* [Prakken, 2010]. This means that all known [Modgil and Prakken, 2014] results regarding rationality postulates [Caminada and Amgoud, 2007] on ASPIC+ also hold for our argumentation theory. In particular, consistency and strict closure of the conclusion sets of Dung extensions is guaranteed.

Important to note is that due to the nature of support graphs there may be paths in the graph that are inactive given the actual evidence and should therefore not be used to reason along. Since d-separation depends on the actual set of evidence and the support graph is meant to capture possible support independent of the actual set of evidence, these irrelevant reasoning paths are still present in the support graph. Only after evaluating the strengths of arguments will these paths explicitly become redundant.

Since the set of rules is directly based on the support graph it is possible to construct the arguments (and in particular the grounded extension) directly, simply by traversing the nodes of the support graph. For every node the 'best' supported argument can be computed using the chosen measure of strength and when both outcomes are equally well supported we immediately know that both outcomes are defeated by the other and not in the grounded extension. This means that the computation of the grounded extension, which is in general computationally hard, can be done efficiently from this argumentation system.

4.4 Skidding car case study

We will now apply our method to a more realistically sized example. For this, we use the Bayesian network as described by Huygen [2002], which is an adaptation from the causal model presented by Prakken and Renooij [2001] for a civil legal case about a car accident. The graphical structure of this network is shown in Figure 4.7. Since the probability tables described by Huygen omit two conditional probabilities we have estimated those in a similar analysis to Huygen's. For our analysis the exact values are not critical. The full specification of the conditional probability tables (CPTs) is given in Appendix B.

4.4.1 Bayesian network

The example network models the events discussed in an actual legal case about a car accident. The passenger in the car claims that the driver lost control over the vehicle. Because the driver was, supposedly, speeding in the S-curve, the passenger claims that the driver is responsible for the consequences of the accident and wants financial compensation for damages. However, according to the driver it was the

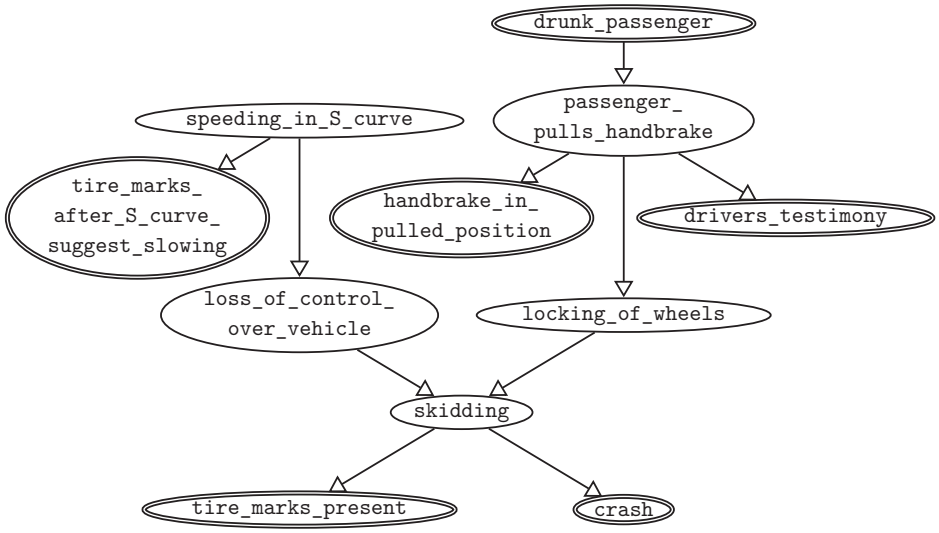


Figure 4.7: Graphical structure of the skidding car accident network [Huygen, 2002]. Observed outcomes can be distinguished by the double outlines.

passenger (who was drunk at the time of the accident) who pulled the handbrake, causing the car to skid and crash. This case is modelled in eleven variables. Six of these variables are instantiated with evidence. Most importantly, there are tire marks, indicating that the car was skidding before the accident. The nature of the tire marks beyond the S-curve indicates slowing rather than speeding. Concerning the handbrake, the police found the car with the handbrake in the pulled position. The first thing that the driver said to the police was that the passenger had pulled the handbrake. Finally, it was confirmed by the police that the passenger was drunk at the time of the accident.

4.4.2 Support graph

Based on the BN in Figure 4.7, a support graph can be constructed for any of the variables. The variable that we are interested in is `speeding in S curve` because that is what determines the liability of the driver in the accident. The support graph for this variable is shown in Figure 4.8.

What can be seen from this support graph is that the observed nature of the tire marks is direct evidence for the fact that the driver was (or was not in this case) speeding in the S-curve. Another supporter for the conclusion is the `loss of control over vehicle` variable because loss of control can occur when one is speeding and has, therefore, a strong correlation with it. The fact that the driver may have lost control over his vehicle is supported by the fact that the car was skidding, which in turn is diagnosed by the fact that the crash happened in the first place and the presence of tire marks. The locking of wheels, however,

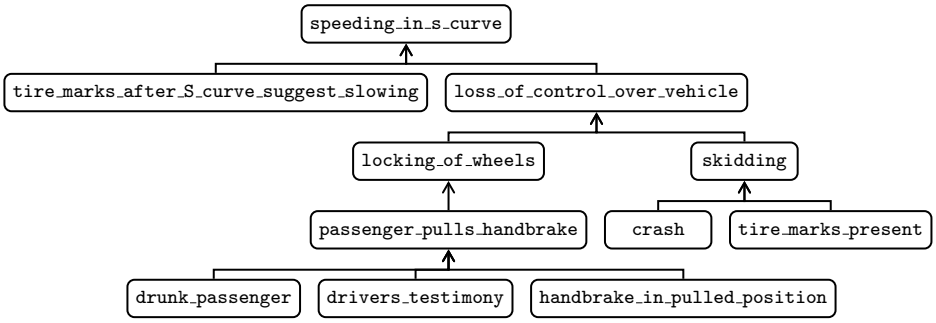


Figure 4.8: Support graph from the skidding car accident network.

can also explain the skidding and the resulting crash. This may, to some extent, explain away the `loss of control over vehicle` node. The locking of the wheels is supported by the statement that the passenger pulled the handbrake, which is supported by the three observations that the passenger was drunk, that the handbrake was in the pulled position and that the driver testified to the police about this event.

4.4.3 Arguments

The support graph does not need pruning since all (and only) leaves of the graph correspond to instantiated BN variables. This is because the BN is targeted at this specific set of evidence and no variables have been considered that are irrelevant given the current set of observations.

We first translate the support graph into arguments based on the likelihood ratio measure of inferential strength. The resulting undefeated argument tree is shown in Figure 4.9. We observe that the skidding receives an infinite LR from the evidence below it. This is the case because the probability of finding tire marks was set to 0 if the car did not skid. In other words: there is no other explanation for the tire marks than that the car must have skidded. However, this strong support does not transfer to the `loss of control over vehicle` because the `locking of wheels` (itself with a moderate amount of support) poses an alternative explanation for the skidding. We see that the final conclusion `speeding in S curve = false` is best supported by the combined evidence with a likelihood ratio of 2.797. This does, however, not mean that this conclusion has a high probability, just that it has become more likely by observing the evidence.

To see what the posterior odds are, we can relabel the support graph with posterior odds instead of likelihood ratios. We have done so in Figure 4.10. Indeed, we see that the `loss of control over vehicle` has a very low probability, even though the skidding is a certain event (probability 1.0). This shows that both the incremental and absolute reasoning patterns can be explained using this model.

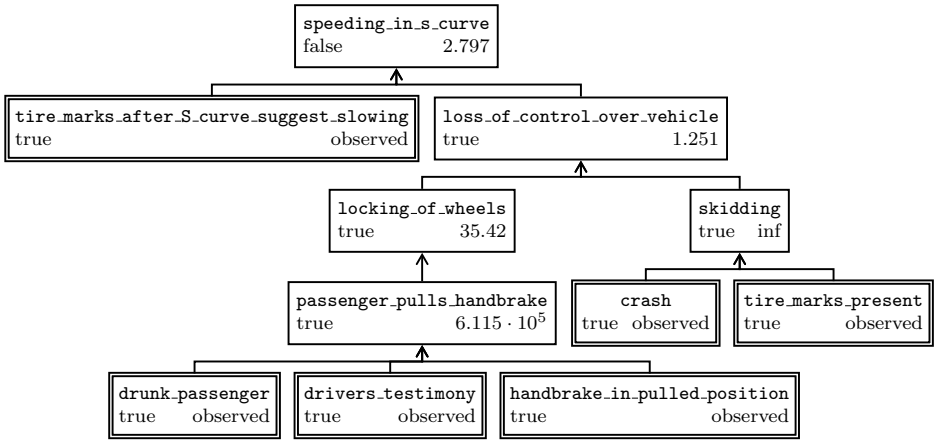


Figure 4.9: The best argument for the skidding car accident network using the LR measure of strength. The strengths have been displayed in the nodes that were not instantiated.

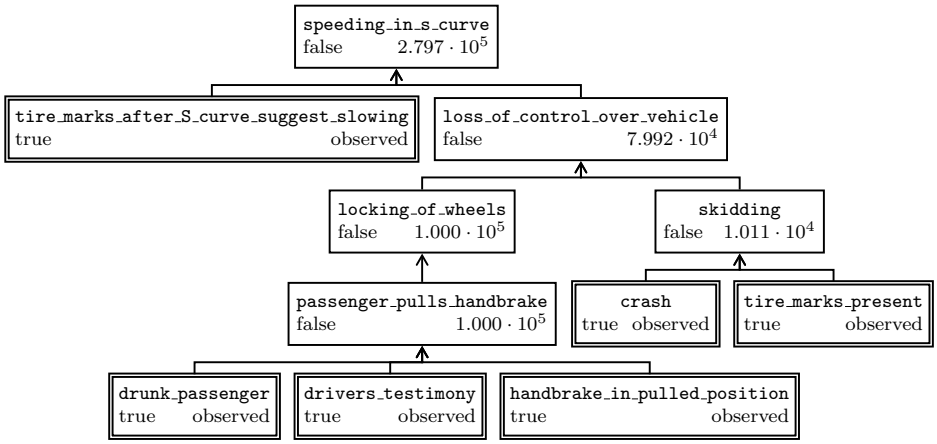


Figure 4.10: The best argument for the skidding car accident network labelled with posterior odds of the most likely outcome.

4.5 Discussion and conclusions

In this chapter we formalised a two-phase argument extraction method for explaining the reasoning in BNs. This improved on the previous chapter because we have significantly reduced the number of arguments. In particular, only rules and arguments that are directly relevant to a variable of interest are enumerated. We do inherit from the previous chapter the notion of inferential strength, but since we now present arguments as the balanced conclusions of all pro and con evidence, there is no longer a notion of undercut in the resulting argumentation.

Recall that the motivation in this study is to facilitate the correct explanation

of (Bayesian) probabilistic reasoning. Bayesian networks are notoriously hard to interpret, partly because of the fact that the directions of arrows in a BN have no intrinsic interpretation. We solve this by first extracting a support graph, which removes this confusing property, while maintaining which paths of inference are allowed by the independence relation that was captured by these directions. The support graph is used as the basis for argument construction. This results in qualitative arguments about the same case that was modelled in the BN.

We stress that the proposed method is an explanation tool and cannot be used to replace evidence propagation in Bayesian networks. It rather complements probabilistic inference by representing the same quantitative information in a qualitative, argumentative setting. The resulting argument graphs follow the inference in the BN, but are structured such that they have an argumentative interpretation.

To resolve conflicts between arguments we have applied likelihood ratios and posterior odds. These are typically expressed in terms of two complementary hypotheses. It is for this reason that we have limited the approach to binary valued variables. All other aspects of this method do not rely on this. In particular, the general case of ASPIC+ uses *contrariness*, which is a generalisation of negation.

We have shown how support graphs help in the construction of arguments because they capture the argumentative structure that is present in a BN. This provides a general purpose BN explanation method. An advantage of our method is that it offers a dynamic model of evidence: if more observations become available the first phase does not need to be repeated. When evidence is added, only paths are added that were previously pruned.

Chapter 5

Translating argumentation schemes to Bayesian network idioms

In the previous chapters we showed how arguments about legal evidence can be constructed from a Bayesian network (BN) that models this evidence and the relations between evidence and hypotheses. So far we have always presumed that a BN was given as input. One of the problems with BNs is that they can be hard to construct [Henrion, 1987; Druzdzel and Van der Gaag, 2000].

We address in this chapter the question how argumentation techniques can aid the design of Bayesian networks. In particular, we discuss how *argumentation schemes* with *critical questions* can be modelled in a BN. As such, we provide general building blocks consisting of network *fragments*, sometimes called *idioms* [Lagnado et al., 2012; Hepler et al., 2007], which can be combined and reused to create larger networks.

5.1 Introduction to BN construction

The construction of a Bayesian network model is a difficult task. Especially in domains where no data is available to automatically learn the structure and the numbers of a BN, networks are often built manually with the help of domain experts, which can sometimes be tedious and error-prone. To ease this construction, a modular construction strategy seems promising. By constructing small components for parts of the modelled case, the work can be divided in small chunks. For recurring patterns, standard modules can be defined. These standard modules can be reused in other cases and often leave room to be tuned to the particularities of the case at hand. Such a strategy has been applied to argumentation as well

as Bayesian networks. In argumentation these recurring patterns are called argumentation schemes [Walton et al., 2008] and in Bayesian networks they are called idioms [Lagnado et al., 2012; Fenton et al., 2013; Vlek et al., 2014] or network fragments [Laskey and Mahoney, 1997; Hepler et al., 2007].

Recall from Chapter 2 that schemes capture recurring patterns of argumentation. Since these patterns typically have exceptions—circumstances under which they do not apply—critical questions are usually provided with the schemes. A typical argumentation scheme from the legal setting is that "witnesses usually speak the truth". That is, if a witness W testifies X, we can conclude X. A typical critical question that can be asked is: "Is W objective in the matter X?" This scheme can be specialised for a particular witness W and a particular statement X.

To apply a modular design for BNs, they can be composed from *fragments* or *idioms*. These are partial networks that model a single aspect (critical questions in our case). Typically, these are defined on an abstract level. So they can be specialised and combined to model any specific case at hand. Note that we use the terms *fragment* and *idiom* also for approaches in which we recognise the same principle, even if the authors do not use these terms.

Recall that the aim of this work is to combine argumentative and probabilistic models of evidence. In order to ultimately enable such a combination, one possibility is to translate the schemes for arguments as idioms for a BN. For this purpose we study how argumentation schemes and critical questions can be modelled in a BN. In the literature on BN construction we observe that often structures are used that are similar to argumentation schemes and their critical questions [Fenton et al., 2013; Lagnado et al., 2012; Carofiglio, 2004; Hepler et al., 2007]. However, in this literature, several design principles are used and no clear reasons for one design pattern over the other have so far been given. The existing approaches differ in the extent to which they resemble argumentation schemes. Carofiglio et al. explicitly use argumentation schemes whereas others, such as Fenton et al. [2013]; Lagnado et al. [2012]; Hepler et al. [2007] have modelled similar cases without explicitly referring to argumentation schemes. These approaches also differ in the use of reusable constructs. Fenton et al. [2013] and Lagnado et al. [2012] use idioms to construct a BN, whereas Hepler et al. [2007] use Object-Oriented BNs, which are similar from a design perspective. An Object-Oriented BN [Koller and Pfeffer, 1997] is a BN that has been augmented with additional information that allows parts of it to be collapsed or expanded in the user-interface. Carofiglio [2004] proposes a general design method without explicitly using fragments or idioms to construct a BN. In that case the design method is general enough to be reused but they do not introduce specific fragments.

We began this chapter by presenting an overview of existing methods to capture argumentation schemes in a BN and we show that these approaches can be divided in two categories. Both of these general approaches have been shown to have their own advantages and disadvantages.

Subsequently, we introduce a number of criteria that allow us to compare these general approaches and we show that the general approaches are not necessarily contradictory. We propose a hybrid approach that has the combined advantages from both. Just like in the previous chapters, we take legal reasoning about evidence as the domain of application, but the principles also apply to other domains.

In Section 5.2, we introduce the reader to relevant background on both argumentation schemes and BN idioms. We discuss the common ideas in all methods that embed argumentation schemes in Bayesian networks (either explicitly as idioms or in a hand crafted BN in which an implicitly assumed idiom can be recognised) in Section 5.3. We introduce a number of criteria, which mainly concern the incorporation of critical questions in such idioms, in Section 5.4. We identify two general approaches that we observe throughout the literature and discuss these in Sections 5.5 and 5.6 respectively. In Section 5.7, we therefore present a third, hybrid method to model argumentation schemes and critical questions, which inherits advantages from both sides. In Section 5.8, we present conclusions and discussions as well as related research and suggestions for future research.

5.2 Background

A recurring concept in argumentation theory is that of an argumentation scheme, which was already discussed in Chapter 2. Recurring design patterns in BNs have also been identified. We discuss these two principles separately.

5.2.1 Argumentation schemes

In argumentative legal reasoning, argumentation schemes have been identified for various types of (legal) evidence [Verheij, 2003b; Bex et al., 2003]. One example is that of the argument from position-to-know. Kadane and Schum [1996], for instance, use such a scheme (although not explicitly as an argumentation scheme) for testimony evidence. On the basis of Walton et al. [2008], the following adaptation by Prakken [2014] for probabilistic reasoning will be used as an example throughout the rest of this chapter:

W is in the position to know about H
W testifies that H
Therefore, presumably, H
Critical Questions:
1. Veracity: is W sincere?
2. Objectivity: Did W's memory function properly?
3. Observational sensitivity: Did W's senses function properly?

Variations on this scheme exist for different kinds of testimonies such as experts and eye witnesses. For experts, the scheme can be specialised as follows according to Walton et al. [2008]:

E is an expert on domain D
E testifies that H
H is in the field of D
Therefore, presumably, H

Critical Questions:

1. is E's expertise on domain D acknowledged by other experts in this field?
2. is H exactly what E testified to?
3. is H backed by empirical data?
4. is H backed by other experts in domain D?
5. is H indeed in domain D?

Argumentation schemes and critical questions can guide the construction of arguments and counterarguments. Because an argumentation scheme suggests a presumptive inference, this inference can be used as defeasible inference rule in formal argumentation systems such as ASPIC+. The critical questions then directly point to possible undercutters of the inference. They are abstract models of inference in the sense that they can be specialised for a particular witness testifying on a particular topic. The argumentation scheme from position-to-know is just one of many that have been proposed for legal evidential issues. For instance, argumentation schemes can be constructed for fingerprint and DNA matching evidence. Argumentation schemes are not limited to legal reasoning since they can capture any situation in which we wish to abstract from a certain pattern of reasoning. For most of the introduced schemes, appropriate critical questions can be formulated.

Critical questions for an argumentation scheme indicate possible ways to attack application of the scheme. These questions can be divided in different kinds of attack:

- Questions that point to undermining attacks on the scheme's premises. For example, a critical question about whether an expert indeed has the required expertise to make a claim about a certain topic is of this kind;
- Questions that point to exceptions to the scheme. The veracity, objectivity and observational sensitivity questions introduced above are of this kind. Without directly attacking any of the premises or directly attacking the conclusion of the argumentation scheme, its applicability is questioned;
- Questions that rebut the conclusion. The question whether other experts agree on the conclusion is an example where the conclusion of the argumentation scheme is directly attacked.

We will focus on the second kind of critical question, since this kind point to undercutting attack, which, as explained in Chapter 3, can be modelled by explaining away in BNs.

5.2.2 Bayesian network idioms

So far, we have assumed a BN to be given as input, but the design of such a model is a challenging task by itself. One method that facilitates the construction of a BN is the use of idioms. By carefully constructing idiomatic fragments for recurring situations, the process of constructing a new BN can be speeded up significantly. Idioms for many recurring situations have already been identified. Fenton et al. [2013] have proposed a number of legal BN idioms. Vlek has extended this with BN idioms based on scenario schemes [Vlek et al., 2013, 2014]. In other (non-legal) domains similar approaches have been taken. For instance, Laskey and Mahoney [1997] introduced a number of fragments for military situation assessment. Object-Oriented BNs (OOBNs) have been introduced by Koller and Pfeffer [1997] and are similar in concept. These were later applied to legal cases by others such as Hepler et al. [2007].

Recall from Chapter 2 that a BN is a model of a probability distribution that can encode complex interactions between variables. When two adjoining edges on a path in the BN converge this is called a *head-to-head* connection. Since conditional probability tables (CPTs) are specified for all combinations of parent outcomes there is the possibility to encode intercausal interactions, such as *explaining away* or *explaining in*, in such nodes. These effects occur when the correlation between a parent and a child changes strength (or even sign) upon observation of another parent. When modelling causal systems, there often are intercausal interactions, which requires such a head-to-head connection in the BN to be modelled correctly, for instance, when two events can both cause the same (third) event. Either cause could explain the outcome of the shared effect if it is observed, but when in addition one of the causes is also observed, this has a negative effect on the other cause. BNs can model this vary naturally by means of a head-to-head connection. Head-to-head connections, therefore, have a special status, which is an important aspect in this chapter. Pearl [1988b] already argued that default reasoning exhibits such complex interactions and therefore benefits from models such as BNs, which can correctly represent these interactions.

It should be noted though that head-to-head connections increase the representational complexity of the model significantly. More complex interactions between parents can be modelled at the cost of an increase in the number of numerical parameters present in the model. Compare, for instance, the two graphical structures in Figure 5.1. The lower graphical model has fewer parameters but enables the same intercausal interactions between the variables `A`, `B`, `C` and `D`. Of course, when no such interaction exists, a graph without head-to-head connections is even more efficient in terms of the number of model parameters. In summary, head-to-head connections are sometimes necessary to model the correct probability distribution but we should aim to minimise their use when possible in order to decrease the representational complexity.

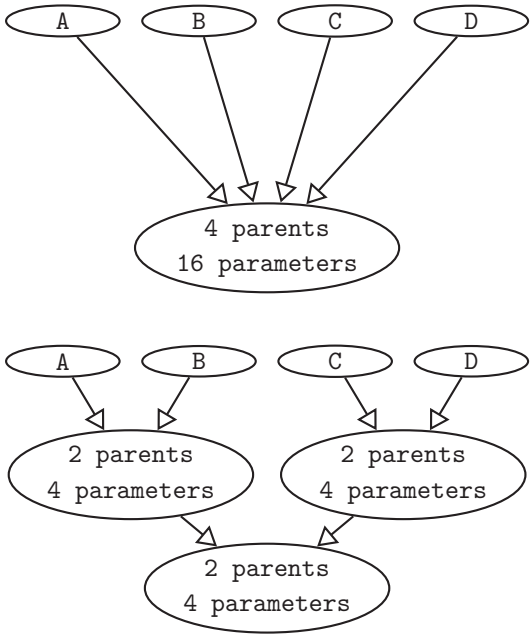
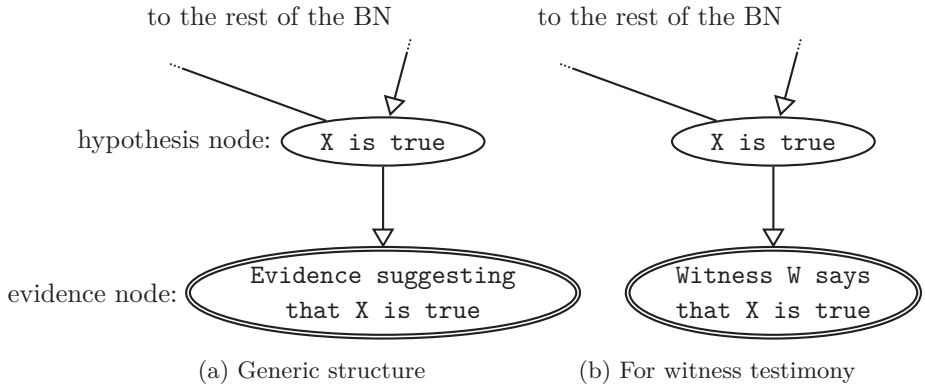


Figure 5.1: Reducing the number of head-to-head connections can lead to fewer free numerical parameters. For three of the variables the number of free parameters is shown, which depends on the number of parents and on the number of outcomes of the variables. Assuming that all variables are binary the top graph has six head-to-head connections (one between each pair of parents) and sixteen free parameters. The lower graph has three head-to-head connections and a total of twelve free parameters.

5.3 Directions of edges in BNs and arguments

Walton et al. [2008] have already identified that many argumentation schemes share similarities and that a number of general categories can be identified. Arguments from position-to-know is one such category. One of the distinguishing characteristics of an argumentation scheme that was not introduced by Walton but which turns out important for our purposes, is whether they are *diagnostic* or *predictive* in nature. A diagnostic, sometimes also referred to as *evidential*, or *explanation-invoking* [Pearl, 1988b], argument reasons from a consequence of some causal process back to a possible cause of that consequence. In the case of the argumentation scheme from testimony, the presence of causality is debatable, but it can be argued that the inference is diagnostic in the sense that the occurrence of the fact X is the supposed reason for W's testimony of X. Such a dependency shows that the testified fact X, if it did not effect the testimony, it at least affected it. In (legal) reasoning about evidence, many arguments and argumentation schemes are diagnostic in nature because there reasoning from evidence back to the facts



(a) Generic structure (b) For witness testimony

Figure 5.2: BN graph fragment representing a diagnostic argument.

that supposedly caused that evidence is most common.

Predictive arguments, on the other hand, reason from a known cause to a potential consequence of that cause. For example, having a motive may 'cause' someone to commit a crime. Not all argumentation schemes can be categorised as either predictive or diagnostic but this is not of concern for the following discussion of modelling position-to-know arguments in BNs.

In graphical representations of arguments, edges are commonly drawn from the known premise to the supposed conclusion, whereas in BNs it is customary to draw edges in the direction of the causality. That way intercausal interactions automatically appear when two or more causes of the same effect are modelled. Since diagnostic and predictive argumentation schemes traverse a causality relation in opposite directions, this implies that the associated argumentation schemes should be modelled differently in BNs. The argumentation schemes from position-to-know are all of this diagnostic kind and can therefore be modelled with similar graphical structures in BNs. In particular, we observe that edges in BNs about diagnostic evidence, such as in Fenton et al. [2013], Lagnado et al. [2012], Carofiglio [2004], Hepler et al. [2007], Kadane and Schum [1996] and Aitken et al. [2003], are always drawn from the hypothetical fact to the observed testimony, i.e. reversed with respect to the direction of the inference. Thus, we propose to model the relation between evidence and the hypothesis that it concerns as shown in Figure 5.2.

The reason why the direction of the edge is best drawn as in Figure 5.2 can also be explained in terms of the represented independence relation. If the arrow were to be drawn from the evidence to the hypothesis, the witness testimony would not necessarily be independent (which we expect it to be) of the rest of the BN (other evidence and hypotheses in the case), given that X is true. The proposed structure guarantees that the evidence is independent of the rest of the BN, no matter how the connections to those other variables are directed. That is, if (in contrast to what is normally possible) X itself would be observed to be true (or false), a testimony for (or against) X should not change any other beliefs about

the case. The evidence only pertains to the question whether X is true. If the truth of X is known, then the testimony becomes irrelevant.

The above discussion provides generally accepted reasons for how position-to-know arguments can be modelled in a BN [Fenton et al., 2013; Lagnado et al., 2012; Carofiglio, 2004; Hepler et al., 2007]. The consensus on this topic is that edges are drawn from the hypothesised cause (conclusion) to the observed effect (premise), which is the reverse of what is common in argumentation, where edges are drawn from premise to conclusion. About the incorporation of critical questions there is disagreement in the literature. To be able to discuss and compare the options to do this, we introduce in the next section a number of criteria to compare the general approaches presented in the following sections.

5.4 Criteria for embedding critical questions

In the following sections we will introduce three ways in which critical questions can be incorporated in BN idioms for argumentation schemes based on a position-to-know. To highlight the differences and to show that benefits from the first two approaches can be combined in the third approach, we first discuss a number of modelling criteria for argumentation schemes (and critical questions in particular) in BN idioms. This list of criteria is used to compare the approaches that we discuss. These are not strict requirements but advantageous property that we would like a BN idiom about critical questions to possess. The list is also not exhaustive in the sense that correctness is guaranteed for models in which all of these criteria are met. Instead, these are the distinguishing properties of the models that we discuss further on:

1. Critical questions are explicitly modelled by a node;
2. Critical questions can explain away the hypothesis (via a head-to-head connection);
3. The number of free model parameters is as small as possible;
4. No redundant edges are included;

Besides these, some further guidelines are often implicitly or explicitly used in the literature and we adhere to these in the following examples:

5. Premises and conclusions are explicitly modelled by a node;
6. There can be an active chain from the hypothesis to the evidence given the variables that can potentially be observed;
7. When applicable, edges follow the direction of temporal/causal precedence;

Criteria 5 and 6 summarise the previous section and all approaches that we find in the literature either explicitly or implicitly agree on these criteria. Although it is technically not always necessary, criterion 7 is often considered good practice because it helps to make it easier to interpret the resulting BN.

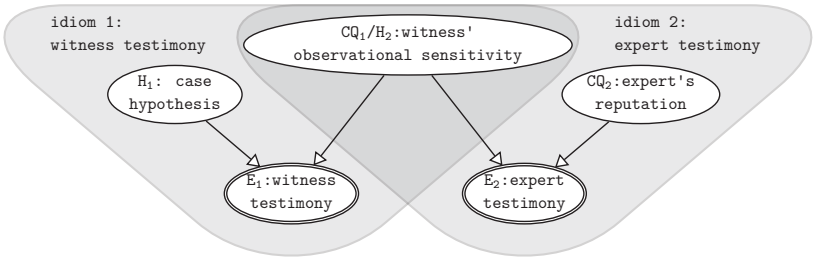


Figure 5.3: An example of two invocations of the testimony idiom.

Criterion 1, in contrast, is not followed by all approaches in the literature. We will present examples of this in Section 5.6. Although it is in some cases possible to model a critical question implicitly in the BN, we argue that it is preferable to have an explicit variable for each critical question. The main reason for this is that it improves the usability of the resulting idiom, because ancillary evidence can be connected to the critical questions. In an argumentation scheme for witness testimony, for instance, the observational sensitivity of a witness should be presented as a variable in order to be able to connect other evidence to it. Consider a case in which the observational sensitivity of a witness is disputed but a (medical) expert testifies that the witness was likely able to perceive the events that the testimony describes. In such a case it should be possible to invoke the argument from position-to-know twice, once for the witness testimony and once for the expert testimony. The critical question from the former is the hypothesis node of the latter.

An example of how this could be modelled is shown in Figure 5.3. In this example one critical question for each of the testimonies is captured by a variable in the idiom. In the following sections we will discuss how multiple critical questions can be handled. In the case of the example in Figure 5.3 an idiom is instantiated for the witness testimony in which the critical question concerning the observational sensitivity of the witness is modelled by a variable. By making this variable a parent of the evidence it is the minimal idiom that satisfies the other constraints above as well. We take this observational sensitivity variable as the hypothesis for the next idiom. For this second idiom the expert's objectivity is modelled by a node that represents the question whether the expert has a good reputation (which we suppose means that he must be objective for the sake of the example). In a more realistic setting the other critical questions for both testimonies will have to be added as well. This can be done in multiple ways, which we will discuss shortly.

Criterion 2 above states that intercausal interactions should be possible between the hypothetical conclusion (the topic of the testimony in position-to-know arguments) and the critical questions. This is because often the 'wrong' answer to a critical question (an answer that generates a counterargument) explains away the presence of the testimony. Consider, for example, the case of witness veracity.

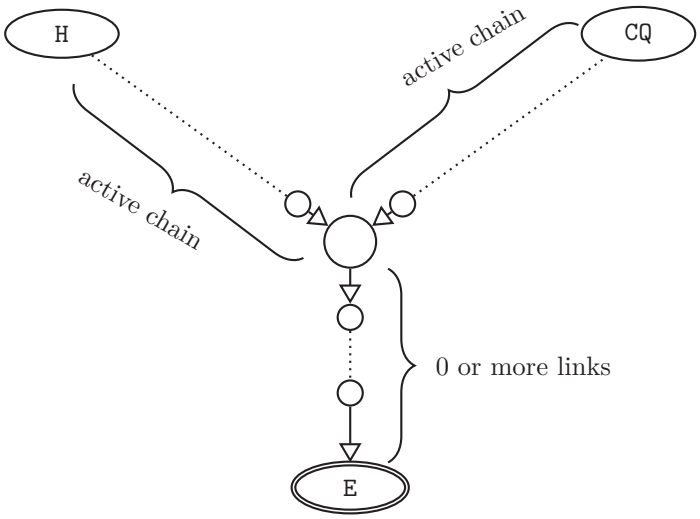


Figure 5.4: The general structure required to model intercausal interactions between the hypothesis (H) and a critical question (CQ) for evidence E. This corresponds to the description of Criterion 2.

If the witness is related to the suspect, an exculpatory statement can be explained as an attempt to help the suspect rather than as a truthful representation of the events. In a sense the veracity of the witness 'causes' the witness to make an exculpatory testimony. Such an intercausal interaction is only possible if there is a head-to-head connection from the veracity, via the testimony, to the testified fact. The explaining away effect can then emerge if there is evidence for the head-to-head variable. This evidence can be a direct observation of that variable or it can be an observation of a descendant, which enables a chain of evidential reasoning steps that increases its probability. Such a chain of inferences can represent the modelled evidence in greater detail. The same can be done for the head-to-head connection. Is does not need to be the case that the hypothesised cause and the critical question are direct parents of the head-to-head node, as long as they are connected by a chain that has such a node. The fact that the relation between the hypothesis and the critical question is represented in greater detail does not take away the possibility that one explains the other away. Figure 5.4 depicts the general structure that is required to model intercausal interactions.

All other things being equal, we prefer a model that reduces the representational complexity. In a BN conditional probabilities must be specified for each outcome of each variable given every possible combination of outcomes of the parents of that variable. These conditional probabilities are referred to as the numerical model parameters. We note that for each variable and each conditioning

set the probabilities of the outcomes must sum to one and that therefore the probability of one of these outcomes can be calculated from the others. We therefore say that all but one of these are *free parameters* and the last one is bound. We wish to prevent interactions between variables that we know do not interact and we wish to limit the number of free model parameters (criteria 3 and 4). In argumentation schemes in particular, critical questions are often independent sources of doubt. They can all explain away the evidence, but there is usually not a more complex interaction than via the common evidence that they can all influence. More complex interactions usually introduce additional model parameters. Although the method proposed in this chapter does not concern the estimation of conditional probabilities, we do take into account that models with more parameters quickly become infeasible in this aspect.

Using the criteria above as a measure for comparison we can now introduce the general approaches to modelling argumentation schemes and critical questions that have been presented in the literature.

5.5 Critical questions as additional parents

Consider again the fragment introduced in Section 5.3. We now proceed with the question of how multiple critical questions can be incorporated in an idiom for argumentation schemes from position-to-know evidence.

The key characteristic of a critical question is that it can *explain* the evidence *away*. A natural idea (and indeed the direction taken by [Fenton et al., 2013; Lagnado et al., 2012; Carofiglio, 2004]) is to model every critical question as a parent of the evidence as shown in Figure 5.5 for the running example in this chapter. This indeed creates the possibility to undercut or weaken the inference from the evidence to the hypothesis, satisfying Criterion 2. Whether this happens, of course, depends also on the numerical parameters of the model. The first option that we discuss is, therefore, to create one variable for each critical question and to connect those with edges directly to the evidence.

However, this model has a disadvantage, which is the combinatorial explosion of numerical parameters that needs to be provided to obtain a fully defined BN. In fact, such a model allows for interactions between critical questions where, for instance, one critical question can even reverse the effect of another. While this is the most flexible model in terms of interactions that can be modelled, this comes at the cost of a huge increase in representational complexity. Since one needs to specify the conditional probability of the evidence conditioned on any combination of parents of the evidence node, the number of parameters for the evidence node is doubled every time a critical question is added (presuming that variables are binary). In Figure 5.5, the total number of free parameters for the evidence node and the three critical question nodes together is nineteen because the evidence node has four parents, resulting in sixteen possible configurations, each with one

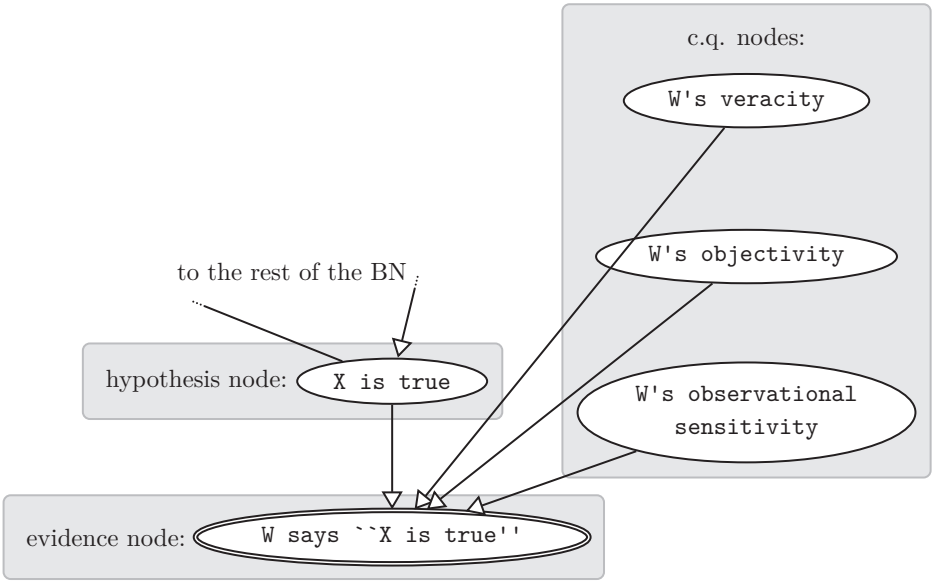


Figure 5.5: Critical questions as additional parents of the evidence.

numerical probability. The critical question nodes themselves have one free parameter as well, which is their prior. For the critical questions about arguments from position-to-know it can be argued that this degree of freedom in the complex interactions is not necessary. In terms of the above criteria this approach performs poorly on criteria 3 and 4. It is evident that the critical questions are each represented by a variable and that, therefore, Criterion 1 is satisfied. Criteria 5, 6 and 7 are also satisfied because the fact X temporally precedes W's statement about X. As discussed before, the presence of a causal connection between these events is debatable, but at least the temporal ordering is respected.

5.6 Critical questions as filters

Another method that has been used in the literature [Hepler et al., 2007; Kadane and Schum, 1996; Aitken et al., 2003] can be described as a *signal filtering* perspective in which the effect of the evidence is weakened in a number of consecutive stages. When critical questions do not interact in a more complicated sense than that they can all explain away the same evidence, the number of parameters can be reduced. This is the case when the critical questions represent independent modes of failure of the normal inference. Instead of introducing variables for the critical questions, we can split the uncertain inference from the evidence to the conclusion into a number of inferences. Each of these inferences expresses exactly the normality assumption from one of the critical questions. Every critical question is now represented by one edge in the graph instead of a variable. This violates

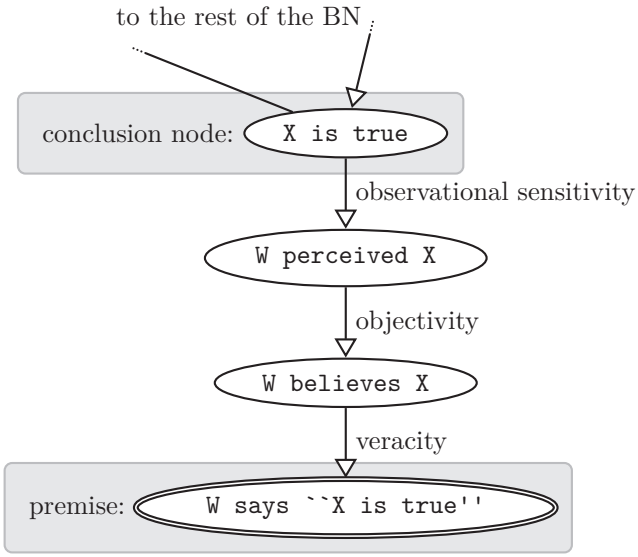


Figure 5.6: A signal filtering perspective on witness testimonies.

Criterion 1 and therefore also Criterion 2 which is a disadvantage of this approach. However, we will show that it also has some advantages.

Using the witness testimony example we can illustrate how the variables are connected, see Figure 5.6. The intuition is that the event may or may not have occurred. Through some fuzzy process this usually results in W perceiving X which again results (with the possibility of exceptions) in W believing X and finally in W testifying that X occurred. The true occurrence of the event is therefore passed through a number of "filters" that make explicit the ways in which the truth about X could result in W's testimony.

This approach is based on a long-standing view on legal argumentation where proof is built on top of evidence in a step-wise manner. From the testimony it is first deduced that W sincerely believes that X happened. From this it is deduced that W indeed perceived X, followed by the conclusion that X must have happened. Such a view on legal proof is, for instance, clearly visible in the treatment of the Sacco and Vanzetti case by Kadane and Schum [1996]. They make an almost one-to-one mapping from Wigmore-charts (which model the step-by-step process of proof as we just described) to BNs, and indeed the resulting BNs are very similar to the one in Figure 5.6.

This method to incorporate critical questions more closely follows the inferential process of reasoning from the evidence to the hypothesis. However, it requires a subtle reformulation of the variables, as can be seen from Figure 5.6. Before, we literally used the critical questions as variables in the network. Here we introduced a variable to express the modalities of belief in X after taking the critical question into account. The variable `W perceived X` implicitly models the observational

sensitivity of the witness. Similarly, the objectivity and veracity are modelled by the variables `W believes X` and `W says `X is true'` respectively.

The fact that the critical question is not represented by a variable is not really an issue in this small example. However, one can imagine that if the veracity of a witness becomes subject of debate, we may wish to add evidence for or against the veracity. In the previous model we could simply have added an extra child to the `Veracity` node to represent such evidence. In the current model this is impossible since the veracity is not represented by a variable. A second consequence of not explicitly modelling the critical questions as variables is that the prior probabilities of those questions are not easily identifiable in the model. When defending the correctness—or at least reasonableness—of such a BN it may be hard to explain how a prior belief in the witness' veracity is represented by the model.

Another disadvantage of this model is that an ordering of the critical questions is necessary. For arguments from position-to-know there is a clear intuition behind the order in which the critical questions are applied. This ordering is based on the cognitive functions of the perceiver: first the evidence must be correctly observed, then it must be remembered and then it can be reported truthfully. Each of these stages independently introduces the possibility of an error. However, such a natural, intuitive ordering need not always exist. Consider, for example, argumentation schemes from *analogy* [Walton et al., 2008]. Assume, for instance, two legal cases with very similar circumstances. For example, two robberies in the same area, both committed early in the morning, with similar escape vehicles on similar kinds of shops where similar weapons were involved. If we then learn that John committed the first, this similarity may provide a reason to think that he also committed the second robbery. The following argumentation scheme allows such an inference [Walton et al., 2008]:

Generally, case C1 is similar to case C2.
A is true (false) in case C1.
Therefore, A is true (false) in case C2.

Critical Questions:

1. Are there differences between C1 and C2 that would tend to attack the force of the similarity cited?
2. Is A true (false) in C1?
3. Is there some other case C3 that is also similar to C1, but in which A is false (true)?

Regarding the critical questions, there is no clear, natural ordering of these questions, as there is for position-to-know arguments. There is no natural way in which one of these sources of doubt precedes another. This is to be expected since they attack the argumentation scheme in different ways: the first critical question tries to attack the first premise of the argumentation scheme by searching for dissimilarities. The second critical question attempts to attack the second premise by questioning whether the hypothesis holds at all. The third critical

question attempts to attack the conclusion of the argumentation scheme because it hints at another possible argument from analogy for the opposite conclusion. This scheme is also an example of the fact that not all argumentation schemes can be categorised as either evidential or predictive.

Compared to the previous model, the number of parameters is clearly reduced, since now only two free parameters per critical question (the probabilities of finding that X is correctly represented after the filtering step given either outcome of the previous step) are required, assuming again that variables are binary. The BN fragment in Figure 5.6, for instance, requires only six free parameters (two for each variable that start with "W ...") to represent the critical questions and the evidence, whereas the previous method had nineteen of these parameters.

In summary, models that take this general approach perform poorly on Criteria 1 and 2 but better than the approach presented in the previous chapter on Criteria 3 and 4. For Criteria 5, 6 and 7 there is no difference compared to the previous model.

5.7 A hybrid approach to modelling argumentation schemes and critical questions

We have now seen two ways in which argumentation schemes and critical questions can be modelled in BNs that both perform well on two of the four distinguishing criteria that we introduced. We now propose a third method that combines aspects of both approaches, in which the chain of inferential steps is augmented with explicit exceptions in the form of alternative explanations recognisable as head-to-head connections in the graph. This allows us to explicitly model critical questions as variables, satisfying Criterion 1. As an example consider Figure 5.7. The head-to-head connection through which they are connected to the inferential path from evidence to conclusion also explicitly shows that they can have an explaining away effect on that inference as one would expect from a critical question (Crition 2). At the same time we maintain from the second method that there is an argumentative path from the evidence to the hypothesis, such that this model closely follows legal tradition of making small inferential steps starting from evidence and building towards a conclusion. Additionally, there is also no exponential growth in the number of parameters that need to be estimated to complete the model (Criteria 4 and 3). In this case, five free parameters are introduced for each critical question. One parameter is required to specify the prior probability of a critical question and four parameters for each `W perveived X`, `W believes X` and `W says ``X is true'` describe the effect of one of the critical questions. Criteria 5, 6 and 7 are again satisfied, just as in the other two approaches.

In this chapter we have identified two different approaches to modelling argumentation schemes and critical questions in BNs which are used in the literature, both with their own advantages and disadvantages. These (dis)advantages are

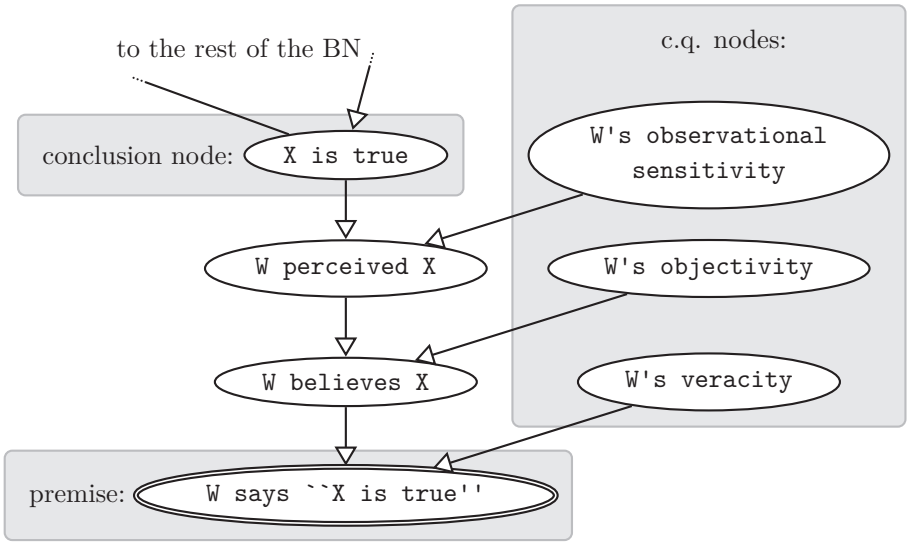


Figure 5.7: Modelling critical questions as a chain of exceptions.

relative to a number of criteria that we introduced. We then combine features of both in a way that better satisfies all of the above criteria than the individual models. Both the introduction of the criteria and the combination of positive features from the literature are new contributions. These criteria are not strict, and neither exhaustive but serve to illustrate the positive and negative features of the approaches that we identify in the literature and how our own method combines these in a unified model. We have argued that the first two approaches do not need to exclude each other and in particular that we can combine features of both to obtain a BN model of argumentation schemes and critical questions that retains the clear explaining away structure of the first and the filtering intuition of the second approach. Together, this makes a model of argumentation schemes and critical questions that satisfies all criteria that we introduced. This can be summarised by the following table:

Nr.	description	as parents	as filters	new approach
1.	explicit c.q. nodes	yes	no	yes
2.	c.q. explains away conclusion	yes	no	yes
3.	minimise free model parameters	no	yes	somewhat
4.	limit unnecessary interactions	no	yes	yes

The proposed method thus combines the advantages from both other approaches to modelling argumentation schemes and critical questions for position-to-know arguments. However, other argumentation schemes exist. See, e.g. Walton et al. [2008]. As described in Section 5.3, a distinction can be made between *diagnostic* and *predictive* schemes. In a diagnostic scheme an effect of the hypothesis has been observed and by lack of evidence of other explanations the truth of the

hypothesis is assumed. Critical questions for such schemes are usually based on alternative causes for evidence which can explain away the conclusion. Arguments from position-to-know all belong to this category. The proposed method does not necessarily work for predictive argumentation schemes or diagnostic schemes that are not a special case of the position-to-know scheme. An example of this will be discussed below.

In particular, argumentation schemes can in general have other types of critical questions, even if we only consider diagnostic argumentation schemes. The critical questions for the position-to-know schemes all attack the inference between the premise and the conclusion. However, critical questions can also be rebutted on their conclusions or undermined on their premises. This kind of attack has not been considered herein. The proposed method does not necessarily apply to these other types of critical questions.

The question that arises is what is minimally required from an argumentation scheme to be able to translate it to a BN idiom in the proposed manner. We argue there are two conditions for this:

- the argumentation scheme must be evidential/diagnostic rather than predictive, and
- there must be a natural ordering among the critical questions.

That the method does not (straightforwardly) apply to predictive argumentation schemes can be illustrated by an example. Consider the argumentation scheme *from cause to effect*, which is again a very generic scheme that can be specialised in many ways [Walton et al., 2008]:

Generally, if C occurs, then E will (might) occur.
In this case, C occurs (might occur).
Therefore, in this case, E will (might) occur.
Critical Questions:

1. How strong is the causal generalization (if it is true at all)?
2. Is the evidence cited (if there is any) strong enough to warrant the generalization as stated?
3. Are there other factors that would or will interfere with or counteract the production of the effect E in this case?

We note that in such an example the two objectives of 1) following the directions of causality when directing edges in the BN, and 2) having a structure such that the critical questions explain away the hypothesis (that E will occur) are conflicting. For both of these principles we have provided good reasons but we must note that they cannot be followed simultaneously for predictive argumentation schemes. See Figure 5.8 for an example. There is a head-to-head connection, but this connection does not allow explaining away because the conclusion is not instantiated. This means there is a conflict within the formulated criteria for predictive argumen-

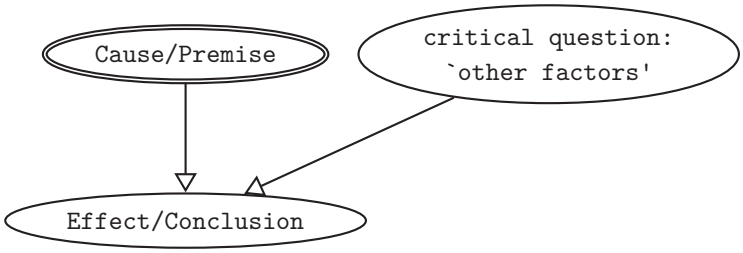


Figure 5.8: A suggestion for how a predictive argumentation scheme might be modelled in a BN idiom.

tation schemes. Although this does not conclusively prove that it is impossible to construct a similar idiom for predictive argumentation schemes (and we speculate that this should be possible), this is not straightforward with the proposed method.

The second constraint for applying this method is that there is a natural ordering in the critical questions. This ordering may be temporal or based on some other process, such as the cognitive process of forming memories from events and testimonies from memories. This order is necessary because the critical questions are applied one by one in the proposed method. We have already presented the argumentation scheme from analogy as an example in which this is not evident.

5.8 Conclusion

We have illustrated differences in nature between predictive and diagnostic argumentation schemes. We have proposed a method to embed argumentation schemes and critical questions for arguments from position-to-know in BNs because these are an important class of argumentation schemes in legal argumentation. However, we have also seen that the proposed method does not directly apply to predictive argumentation schemes because of this difference in nature. Some argumentation schemes, such as the scheme for reasoning by analogy, do not even categorise as predictive or diagnostic.

To conclude, the method applies to diagnostic argumentation schemes in which the critical questions have a natural ordering and this, most notably, includes all argumentation schemes that are based on the position-to-know scheme.

Chapter 6

Related research

This chapter discusses the relation of existing approaches in the literature to the work presented in this thesis. We start with other work on evidential reasoning and the three normative approaches on evidence that have addressed the issues that are associated to reasoning with uncertain evidence. From there we discuss how these approaches have been combined in the literature since the goal of this thesis was also to combine two of these approaches. In Section 6.2 we highlight combinations of argumentation and probabilistic reasoning that do not aim to use argumentation as an explanation method. In Section 6.3 we discuss BN explanation methods that do not use argumentation, and, finally, in Section 6.4 we discuss a number of approaches that use argumentation in BN explanations but in a different way than proposed in this thesis. The research related to Chapter 5 on BN construction using argumentation schemes is discussed in Section 6.5.

6.1 Three normative perspectives on evidence

In Chapter 1 we noted that different rational models of proof exist. These are *argumentative*, *probabilistic* and *narrative* approaches. Each of these normative perspectives has its own advantages and disadvantages.

Argumentation approaches, such as for instance [Bex et al., 2003] or [Anderson and Twining, 1991], have the advantage that argumentative structures are usually intuitive to understand, since they follow patterns common to everyday reasoning such as chaining inferences to derive complex conclusions and reinstating a conclusion by counterattacking an attacker. The main strength of the argumentative approach is the emphasis on the adversarial setting. Arguments against and in favour of a conclusion can be modelled simultaneously and the justification status of each argument can be computed. A weakness of argumentation is that it has difficulty dealing with numerical probabilities. Below we will further discuss some work in this direction.

Probabilistic models, such as BNs, on the other hand, can represent and reason with numerical information in a theoretically grounded way. The main strength of the probabilistic approach is the possibility to do exact computations of probabilities. However, the required input probabilities are in practice often not easily available [Druzdzel and Van der Gaag, 2000].

Research has been conducted to facilitate the elicitation of the required parameters from experts [Van der Gaag et al., 1999]. Alternatively, one can use qualitative BNs [Wellman, 1990], which reduces the issue of estimating numbers to estimating qualitative influences.

In the literature about legal evidential BNs, most research is focussed on the graphical structure instead of the numerical probabilities, such as in e.g. [Lagnado et al., 2012; Vlek et al., 2014; Fenton et al., 2013; Laskey and Mahoney, 1997; Hepler et al., 2007]. This can at least partially be contributed to the facts that networks are very case specific and data to learn probabilities from is not readily available.

Narrative methods use scenarios [Pennington and Hastie, 1993] and more easily present an overview of the evidence that is available in a particular (legal, medical or other) case than argumentation and probabilistic reasoning. The narrative approach emphasises the global coherence of events, which is an important aspect in (legal) reasoning about evidence. However, narrative models of evidence have a similar difficulty with probabilities as argumentation, which is that incorporating numerical evidence can be hard. Furthermore, scenario models suffer from the fact that the truth is not always the best, or most convincing, scenario [Wagenaar et al., 1993].

The three normative approaches also differ in the extent to which they have led to formalisms and algorithms. The strongest in this aspect is the probabilistic approach, which has a solid formal underpinning in probability calculus. BNs are widely accepted models of probability distributions for given independence relations. In argumentation, several formalisms exist. Most of these build on the work by Dung on abstract argumentation [Dung, 1995]. Several approaches that build on this framework coexist, with differences in how they extend Dung's abstract framework. ASPIC+, which we used in this thesis, is one such framework, but alternatives exist, such as Carneades [Gordon et al., 2007] and DefLog [Verheij, 2003a] for evidential reasoning. More general is the work on argumentation using classical logic or on assumption based argumentation; see for instance [Hunter, 2014]. The least formally developed approach is that of scenario models. In this approach scenarios are often depicted as boxes connected by arrows, but there is no standard model of how these are to be interpreted. One way of formalising narrative models of evidence is through *inference to the best explanation* [Thagard, 2004].

Recently, a scientific interest in combining models of proof has arisen [Bex, 2011; Keppens, 2012; Verheij, 2014; Vlek et al., 2015; Verheij et al., 2016]. One possible combination is the use of argumentation to explain and support proba-

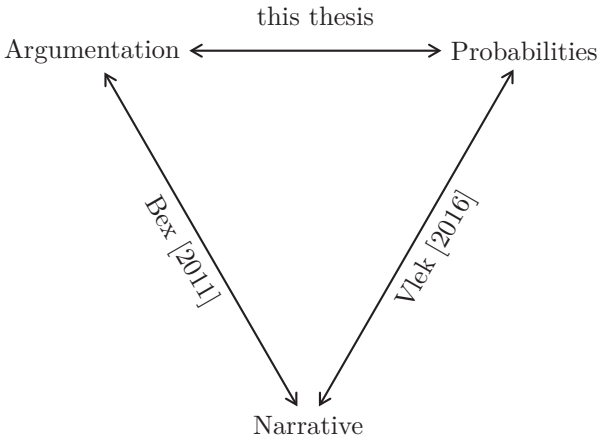


Figure 6.1: Probabilistic, argumentative and narrative models of evidence have been combined in other research as in the depicted triangle [Verheij et al., 2016].

bilistic reasoning, which we have explored in this thesis. This research is part of a bigger line of research that tries to combine argumentative, probabilistic and narrative approaches to modelling evidence, see Figure 6.1.

In an earlier project, Bex et al. [2010], Bex [2011] and Bex and Verheij [2013] have combined narrative and argumentation models. This is done by relating elements of a scenario to evidence using argumentation structures. The coherent story is modelled as a scenario and the individual events can be argued about using the argumentation. Figure 6.2 shows this for a legal case where two suspects, John and an anonymous asylum seeker (AS), are considered. Arguments A_1, A_2 and A_3 support differents parts of the two scenarios (S_1 and S_2) each.

More recently, Vlek [Vlek et al., 2014, 2015; Vlek, 2016] has combined probabilistic models in the form of BNs with scenario models. This takes the form of a design protocol and an explanation method for BNs in which scenarios play an important role. When designing BNs, each scenario, consisting of a number of events, is at least connected by links that follow the temporal order of these events in the scenario and by links between each event and a special node corresponding to the scenario as a whole. In this way, the scenario as a whole is more likely when its events are more likely, but also vice versa: individual components of a believed story become highly likely when no directly contradicting evidence for that particular event is present. The BNs that are constructed in this way have a well-defined structure that allows for an intuitive explanation, in terms of scenarios. Figure 6.3 shows an abstract example of this. Scenarios are hierarchically ordered and events P_i are connected as sequences but also to the scenarios in which events happen.

An integrated approach that has all three components has been proposed [Verheij, 2014]. The resulting method offers an argumentative perspective on classical probability theory (in which no BNs are used).

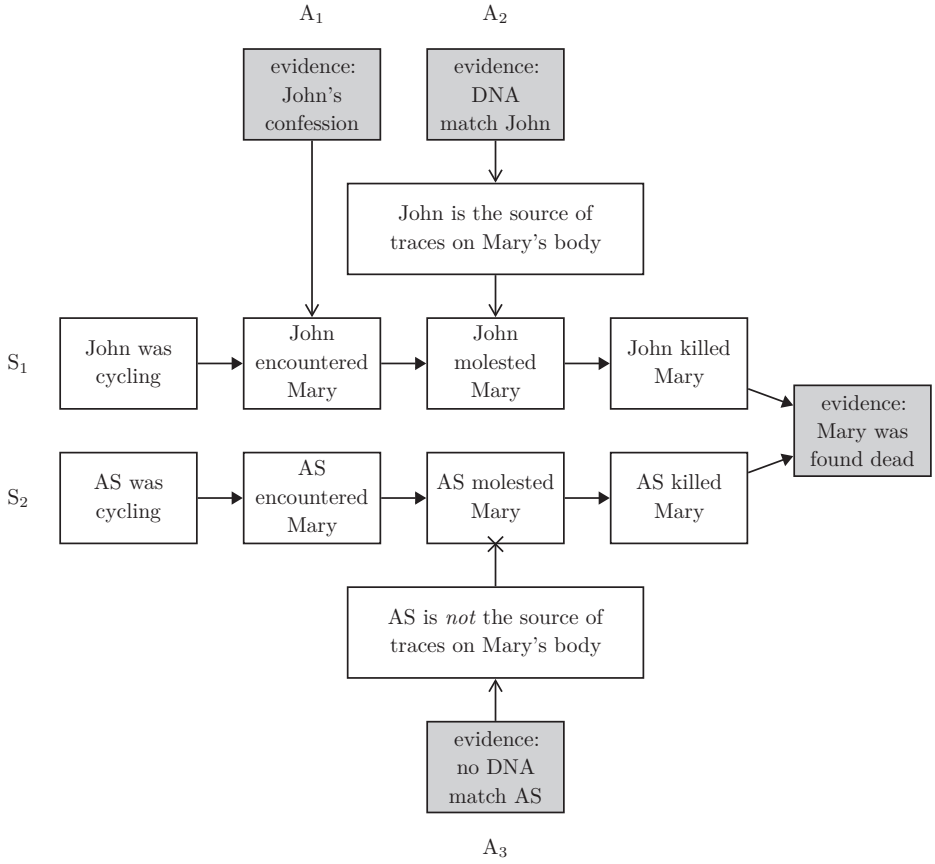


Figure 6.2: Combining scenarios and arguments (from Bex [2011]). Arrows with open arrowheads stand for evidential inferences, and arrows with closed arrowheads stand for causal/temporal relations in the scenarios.

6.2 Probabilistic argumentation

We have now seen that three normative perspectives exist and that several approaches have tried to combine these perspectives. For instance, Schum [1994] discussed argumentative interpretations of many types of probabilistic inferences. In this thesis we have combined formal models of argumentation and probabilities. We note that the models discussed in this section differ from our approach in that the aim of such systems is not to explain or design Bayesian networks but to embed numerical uncertainty in argumentation. In Section 6.3 we will return to explanation methods for BNs.

Several approaches exist to incorporate probabilistic reasoning in argumentation [Pollock, 2001; Hahn and Oaksford, 2007; Hahn and Hornikx, 2016; Dung and Thang, 2010; Li et al., 2012; Hunter, 2013; Hunter and Thimm, 2014, 2016]. Many of these approaches are described as *probabilistic argumentation*. However,

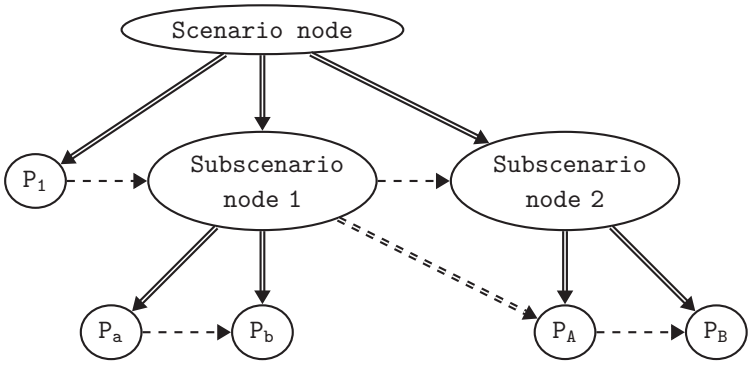


Figure 6.3: Nested scenarios from Vlek [2016].

no universally accepted definition of that term exists and the approaches are often fundamentally different. The method proposed in Chapters 3 and 4 can also be called probabilistic argumentation, but it also differs from existing work in several ways, as we will demonstrate. In these methods, probabilities are used to express grades of uncertainty about the arguments. These numerical uncertainties are built into the logic of the argumentation system, such that conclusions about probabilities can be drawn from them. Much work has been done in overcomming the difficulties that arise when combining arguments and probabilities in this way (designing a logic that deals with probabilities in the statements such that both logical and probabilistic axioms are satisfied). For instance, Pfeifer [2013] discussed how arguments with conditionals als premises should be treated.

The combination of argumentation and probabilistic reasoning has also been studied from a psychological perspective [Hahn and Hornikx, 2016; Hahn et al., 2013; Hahn and Oaksford, 2007]. This work provides psychological evaluation of reasoning by human subjects. For instance, in [Hahn et al., 2013] the assessment of circular reasoning in test subjects in studied. This leads to a model of argumentation in which argumentation schemes are enhanced with probabilistic statements [Hahn and Hornikx, 2016]. This work is, however, not a formal argumentation system.

Pollock [2001] proposed to allow defeasible arguments to be compared on a numerical strength. He turns to the likelihood ratio and the weakest link principle to do so. First, *reasons* (or arguments) for a conclusion are given a strength. This is done by applying the likelihood ratio of the antecedents (i.e., what would be immediate subarguments in ASPIC+ terminology) for the conclusion. This results in a degree of justification for inferences. These are then combined into graphs that feature both inference and defeat links. The premises and inferences have varying degrees of justification but the defeat links do not have a numerical strength. That is, the effect of an inference depends on the justification of the premises and on the justification of the inference rule, but the effect of an attack solely depends on the justification of the attacker, much like undercutting in ASPIC+. The degree

of justification s of an inference is computed as

$$s = \log_2(0.5) - \log_2(1 - r)$$

where r is the probability used in the statistical syllogism on which the inference is based. For instance, if the generalisation "witnesses usually tell the truth" is used and we estimate that this is true in 90% of the cases, then the degree of justification of the inference from 'John says X' to 'X' equals

$$\log_2(0.5) - \log_2(1 - 0.9) = 2.323$$

This formula poses a measure of strength that maps inferences to a domain for which 0 means no positive or negative influence and infinite justification is assigned to certain inferences. Pollock then applies the weakest link principle to determine the degree of justification of the conclusions that are derived from a sequence of these inferences. In the simplest case, the justification of an argument with a top rule with justification s, immediate subarguments with justification $s_1, \ldots, s_n$, and defeaters whose conclusions have degrees of justification $d_1, \ldots, d_m$ is:

$$\text{justification} = \min(s,\ s_1, \ldots, s_n) - \max(d_1, \ldots, d_m)$$

This means that an argument is at most as strong as its weakest subargument and its weakest rule. For cyclical graphs the computation is more complicated but follows the same principle. This is a variation on the weakest link principle that we used in Chapter 3. Attacking arguments have a diminishing effect on the justification of an argument so defeat is not resolved using Dung style semantics but by the system itself. We applied ASPIC+ and standard Dung semantics in the work presented in this thesis—rather than Pollocks variable degrees of justification—because Pollock's method for recombining argument strength is not appropriate for probabilities, since the probability of a conclusion can be higher than the minimum of the probabilities of the supporters (for instance if the conclusion has a high prior probability or if the premises collectively bring about the conclusion). Furthermore, by conforming to ASPIC+ and Dung semantics we build on well studied and widely accepted models argumentation. In particular we have discussed the rationality postulates that were proven to hold for this approach.

The methods proposed in this thesis differ from the work by Pollock in a number of ways. Our methods extract arguments from a BN model which represents a probability distribution, whereas Pollock's work assigns quantitative degrees of belief to arguments. The method proposed by Pollock integrates numerical justification in the argumentative process which is not what we aimed to do. This implies that the methods proposed herein and those proposed by Pollock are of a different nature and have different aims. Consequently, the approaches are quite different. For instance, Pollock does not use Bayesian networks and is not concerned with independence information or intercausal interactions. In fact, Pollock is explicitly opposed to a Bayesian interpretation of degrees of justification and hence introduces his own notion of degrees of justification instead.

Another approach to probabilistic argumentation was proposed by Dung and Thang [2010]. In this method a number of jurors is assumed with different internal models of the world. The possible worlds are shared but each juror has individual probabilities assigned to these worlds and an individual way of determining what arguments are consistent with that world. The probability of an argument is then defined (with respect to one juror) as the sum of the probabilities (according to that juror) of the worlds with which this argument is consistent and in which it is in the grounded extension. Such a framework is an abstract probabilistic argumentation framework because the notion of how arguments can be consistent with a possible world is left unspecified. This is then instantiated with assumption based argumentation, in which inference rules are chained to form arguments. An argument in that framework is consistent with a world if its conclusion follows by rule applications from premises in that world. A problem with such an approach is that possible worlds are assumed to warrant some arguments with a probabilistic weight. There is no clear intuition of how these weights should be assigned and how the weight of an argument should be interpreted. Hence we applied Dung's abstract argumentation framework in its original form in Chapters 3 and 4.

A similar approach is taken by Li et al. [2012], who build on Dung's abstract argumentation by adding a probability in the $(0, 1]$ interval to each argument and to each defeat link. This results in a probability distribution over possible Dung frameworks. Each of these Dung frameworks has a subset of the arguments and defeat links of the probabilistic Dung framework. That is, for each combination of arguments and defeat links a probability is obtained that it is the actual Dung framework. For a set of arguments it can then be computed with what probability it is a subset of a certain Dung semantics. This probability is used as the degree of justification of those arguments. The paper by Li et al. [2012] proceeds by introducing an efficient approximation algorithm but this is of less interest for the comparison to our own contribution. Similar to the above, in this approach probabilities are assigned to arguments. Besides the fact that the *probability of an argument* is not an intuitive notion, this approach has no notion of defeasible inference. It remains purely at the abstract level. In our work we used ASPIC+ to express uncertainty about the inference rules and this results in arguments and attack between arguments.

Similarly, Hunter [2013] has proposed to add probability assignments to abstract argument graphs. In his approach a probability distribution is given over the possible spanning subgraphs of the Dung argument graph. A spanning subgraph is a graph that has the same arguments and a subset of the defeat links. Each of these graphs is assigned a probability and these probabilities must sum to one. For each individual link a probability of defeat can then be calculated by simply summing over the possible graphs that include this link. It is shown that, under sufficient independence assumptions, the reverse can also be done. That is, by specifying the probability of each link, the probability of each subgraph can be calculated and hence sets of arguments—like extensions—can be be numerically

evaluated. Additionally, probabilities assigned to the premises of arguments can be used to derive probabilities for argument to be used in the discussed probabilistic argumentation. In contrast to ASPIC+ the proposed argumentation does not use defeasible inference rules but only defeasible premises. A strong point of this approach is that, building on this, probabilistic models of argumentation have been introduced that can also deal with incomplete information [Hunter and Thimm, 2014, 2016].

The latter three approaches (Dung and Thang [2010], Li et al. [2012] and Hunter [2013]), although different in the details, feature similar differences with our approach. Most importantly, as was the case for [Pollock, 2001], is that these frameworks evaluate arguments using probabilities—either assigned to edges, to subgraphs of the defeat graph or to arguments—on an abstract level. In comparison, our methods take an explanation perspective and construct structured arguments on the basis of a Bayesian network. The probabilities are not primarily used to evaluate arguments but arguments are constructed that explain the probabilities modelled by the BN. From this, further differences follow. Firstly, we put a strong emphasis on methods to enforce correct interpretation of independence information that is modelled in the BN. Independence (besides being used as an assumption in [Hunter, 2013]), and conditional independence, which allows intercausal interactions to be modelled, does not play a role in other systems. Secondly, these systems assume uncertainty in the defeat relation but not in the support relation by which arguments are constructed. Our system takes into account that reasons are defeasible as well as the attacks between them. Thirdly, in all of the above systems, the probabilities are used to assign uncertainty to arguments. Li et al. [2012] state that

> *"These probabilities represent the likelihood of existence of a specific argument or defeat."*

In our argumentation systems the arguments exists with certainty but they express uncertain inferences. As such the uncertainty in our argumentation system is at the level of the defeasibility of rules and not at the level of arguments or the attack relation between arguments. Hunter [2013] seems to have a similar interpretation of the probability of an argument:

> *"We can qualify each argument in an argument graph by a probability value that indicates the belief that the argument is true."*

How such a probability follows from the strengths of defeasible inferences and exceptions is not defined.

In the following section we will discuss a number of approach from the literature that do take the explanation perspective of argumentation.

6.3 Explanation methods for Bayesian networks

We distinguish related work on BN explanation methods that use argumentation from those that do not. We discuss the former in Section 6.4 and first focus on other explanation methods. Existing explanation methods for BNs can broadly be divided in three categories. First, the elements of the model can be explained. See, for instance, the work of Lacave et al. [2007] or Koiter [2006] in which the links in the nodes and the edges in the BN are explained. Secondly, the evidence that is instantiated in the BN can be explained by calculating, for instance, the so-called *most probable explanation* (MPE) or the *maximum a-posteriori probability* (MAP) assignment, which is the most likely configuration of a (sub)set of non-evidence variables [Pearl, 1988a]. Thirdly, and this is also the approach to explaining that we have taken in this thesis, the reasoning chains in the Bayesian network can be explained. We have done so by giving argumentative interpretations of these reasoning chains, but other explanations of reasoning chains exist.

For example, Suermondt [1992] explains the reasoning chains in a BN based on identifying important nodes and important edges. The importance of nodes and edges is determined on the basis of a probabilistic measure and compared against a threshold. Verbal explanations for the outcome of a particular node of interest are then generated on the basis of different metrics. For instance, positively and negatively contributing nodes and links are identified and chains of sufficiently important links are presented to the user as part of the explanation. Also the relations between links are considered, such that, for instance, explaining away is identified when two edges converge.

A similar approach has been taken by Yap et al. [2008], who extract from BNs verbal explanations that are structured in a tree. To do so a graphical transformation is performed on the input BN. Unlike our transformation to support graphs, the inferential directions of edges are obtained by reversing edges where necessary, and recreating any dependency that is violated in that process by adding extra edges to the graph.

Non-verbal explanation methods that are similar to MAP/MPE have been proposed, such as the *most inforbable explanation* [Kwisthout, 2013], which is an assignment that is both *informative* and *probable*, or the *most relevant explanation* [Yuan and Lu, 2011], which is an assignment to evidence variables that maximises a generalised version of the likelihood ratio. A MPE (or any of the named alternatives) helps to explain the evidence, but does not explain why the posterior probabilities of variables of interest are high or low nor do they explain the reasoning steps between evidence and hypotheses. For other work in this direction see Kwisthout [2015].

It has also been proposed to explain probabilistic inferences visually such that they become more intuitive for non-mathematical experts. See for instance the work of Lacave et al. [2007], Lacave and Díez [2002], Koiter [2006], Madigan et al. [1997] and Druzdzel [1996]. All these methods explain properties of the represented

probability distribution by adding visual cues to the edges, such as colour and thickness. Koiter [2006], for instance, uses colours to show the qualitative sign of the influence between variables and thickness to show the magnitude of that influence. Koiter [2006], for instance, presents the BN with edges that have a distinct color and thickness. These are used to indicate a strength and a sign. In contrast to the strength that we used in both Chapters 3 and 4, these strengths are calculated per edge and cannot capture interactions between variables. In our approach the strength is calculated in the context of a particular set of evidence and weighing all factors that influence the outcome. A similarity that is shared with our work is that a distinction is made between absolute and incremental strength assignments, although Koiter [2006] calls these *static* and *dynamic*.

Madigan et al. [1997] construct visual explanations of a Bayesian network based on the graphical structure of the BN and a measure similar to the likelihood ratio but they discard the directions of edges, thereby disregarding information about possible dynamic interactions such as explaining away that is present in the graphical structure of the BN. In contrast to the above visual explanation methods, Madigan et al. present visual explanations of the *flow of information* during evidence propagation, which better resembles the argumentative chains that we identified in this thesis than the other visual explanations methods.

Another approach that explains BNs by their reasoning chains is taken by Van Leersum [2015], who identifies important nodes on the basis of their graphical properties in the BN graph. That is, nodes that form bottlenecks in the propagation or that aggregate information from many neighbours are deemed more important. Based on these important intermediate explanation nodes, clusters of evidence variables are identified. These can also be reported verbally and visually.

Vlek [2016] has introduced yet another approach to explaining BNs, which presents relevant scenarios bases of a BN. A requirement of this method is that the BN is constructed with some metadata about scenarios already present. This method constructs textual explanations that feature coherent stories about the modelled (legal) case. This means that the explanation focusses on the global coherence of the evidence, rather than on the local correlations between variables, such as most other explanations described herein.

We have now discussed several approaches that either verbally, visually, or otherwise explain the BN model. In order to explain the probabilistic reasoning chains that are possible in the BN we applied structured argumentation.

6.4 Explanations using argumentation

All of the previously described work is non-argumentative. More argumentation inspired explanation methods for BNs have also been proposed.

Vreeswijk [2005] proposed a simple method to construct rules from BNs to form arguments in an argumentation system similar to assumption based argumentation.

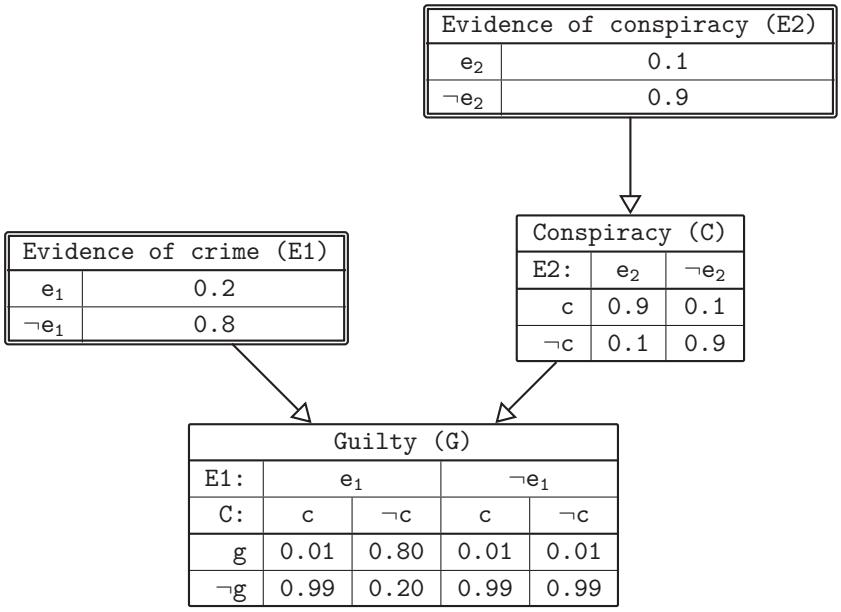


Figure 6.4: Example BN that fits the input requirements for Vreesewijk's method.

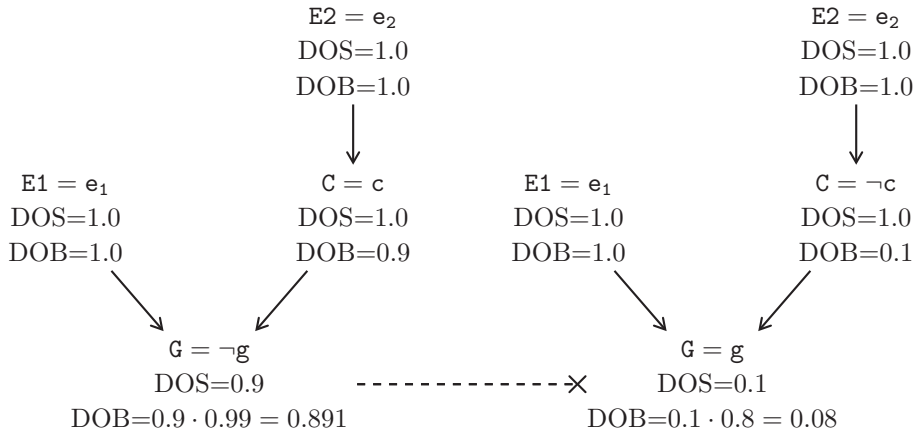


Figure 6.5: The two arguments with highest degree of belief (DOB) resulting from Vreeswijk's algorithm for argument extraction as applied to the example in Figure 6.4.

This approach only works under some limiting assumptions on the BN models that can be used as input. In particular, Vreeswijk assumes that the variable of interest is a sink in the BN graph and that all other variables are ancestors of this sink. Although head-to-head connections can exist in the graph, there are in his model no intercausal interactions effective between the evidence and the variable of interest.

The algorithm proposed by Vreeswijk works by recursively multiplying values from the conditional probability tables (CPTs) from the evidence down. For a value assignment to a node, first the degree of support (DOS) is calculated by multiplying the degrees of belief (DOB) for its parents. The degree of belief for that outcome is simply the degree of support from parents multiplied by the probability of that value given the values of the parents in the subarguments (each parent constitutes a subargument).

Example 6.1. *Consider the BN graph as shown in Figure 6.4, which resembles our earlier example but transformed to fit the input requirements of Vreeswijk's method. We first note that the priors of the evidence are arbitrary because they are not used in the algorithm. The CPT for the* `Conspiracy` *node encodes a strong but not certain correlation between the corresponding evidence. The* `Guilty` *CPT encodes that the suspect is likely guilty if there is evidence for the crime and not for conspiracy, and is very likely not guilty otherwise. Figure 6.5 shows the two arguments with the highest degrees of belief as extracted by Vreeswijk's algorithm. For each argument the degree of belief and degree of support is shown. Evidence* `E1` *and* `E2` *result in arguments with degree of support (DOS) and degree of belief (DOB) both* 1.0*. Further arguments are developed by following the edges of the BN graph. Two arguments are then possible for* `C`*, one where* $\texttt{C} = \texttt{c}$ *(left) and one where* $\texttt{C} = \neg\texttt{c}$ *(right). The degree of support for such an argument is the product of the degrees of belief in the immediate subarguments, which is equal for both outcomes (*1.0 *in this case of* `C` *because the only subargument has DOB=*1.0*). The degree of belief is then calculated by multiplying the degree of support with the conditional probability of the argument's outcome given the outcomes of the immediate subarguments. In the case of the argument for* $\texttt{G} = \neg\texttt{g}$*, for instance, the degree of support is* 0.9 *and the conditional probability of the outcome* $\texttt{G} = \neg\texttt{g}$ *given that* $\texttt{E1} = \texttt{e}_1$ *and* `C=c` *is* 0.99*. The argument on the left in Figure 6.5 defeats that on the right because it has a higher degree of belief.*

One shortcoming of this method is that arguments for some variable outcome are only based on the parents (and ancestors) of the respective node. This means that only reasoning along the edges in the BN is possible. In the setting of legal reasoning this is a major issue, since it is common for the evidence nodes to be the children of a (possibly intermediate) hypothesised node. Without this possibility, it is impossible to model induced intercausal interactions. The *support graph* method that we introduced in Chapter 4 arguably improves on Vreeswijk's method, because our method respects the independence properties of a Bayesian network (BN) without the limiting conditions required by Vreeswijk. In particular, he assumes that the node of interest is a sink in the BN graph and that inference is only allowed in the direction of the BN edges. This means, among other things, that no induced intercausal interaction allowed by the BN will be represented in the argumentation.

Bayesian Networks can be learned from data. It is, however, more difficult to learn arguments from a collection of data. Williams and Williamson [2006] have, therefore, proposed a method to extract arguments from Bayesian Networks that were learned from a database. Their particular field of application is medical diagnosis. A diagnosis can in fact be seen as an argument for a certain treatment.

The construction of the network consists of the following steps; The data is split into a large learning set and a smaller set of test cases. From the learning set a Bayesian Network is learned via some pre-existing learning algorithm. A set of BN constraints is taken into account during the construction. These constraints are derived from domain knowledge of non-causation between variables.

Once a Bayesian Network has been constructed it is applied to the test cases. This already yields a method to calculate the probability of the nodes of interest for these particular cases. That means that in a medical case the probability of different treatments can be calculated.

Williams and Williamson take the system one step further; from the BN they extract arguments about the test cases. The way to extract defeasible inference rules from a probability distribution is defined by a heuristic that strongly resembles influence in Qualitative probabilistic networks [Wellman, 1990]:

- if $P(\mathtt{y} \mid \mathtt{x}, Z) \geq P(\mathtt{y} \mid \neg\mathtt{x}, Z)$ for any Z (either $\mathtt{z}$ or $\neg\mathtt{z}$) which already influences $\mathtt{y}$ through some rule, then $\mathtt{x} \Rightarrow \mathtt{y}$
- if $P(\mathtt{y} \mid \mathtt{x}, Z) \leq P(\mathtt{y} \mid \neg\mathtt{x}, Z)$ for any Z which already influences $\mathtt{y}$ through some rule, then $\mathtt{x} \Rightarrow \neg\mathtt{y}$

They basically compare the likelihood ratio of the cause on the effect but with the added restriction that the likelihood must favour the conclusion independent of possible other rules influencing the conclusion.

This yields inferences that are applicable to the case at hand. In the light of our research this approach might not be very useful since Bayesian Networks for legal cases can usually not be learnt from data. The cases in which probabilistic analyses are used are rare and often so much dissimilar that every case has different variables which makes it impossible to learn one network that can be used for multiple cases. It seems that the rule extraction method proposed by Williams and Williamson [2006] can also be applied to hand-crafted BNs. A problem with this rule extraction heuristic is posed by the fact that extracted rules are such that an influence holds if and only if the inequality stands for any $\mathtt{z}$ that already affect the consequent. This disregards many interesting inference rules. The arguments that are extracted in this way are all strict, due to the fact that all inferences are based on a positive likelihood that must hold irrespective of any further observations.

A similar treatments of arguments on the basis of probabilistic inequalities has been proposed by Parsons [1998, 2004], who uses Qualitative probabilistic networks (a construct similar to BNs but with qualitative constraints rather than numerical parameters) as input to construct arguments. This branch of research applies a filter to discard arguments that use inactive chains in the graph as arguments.

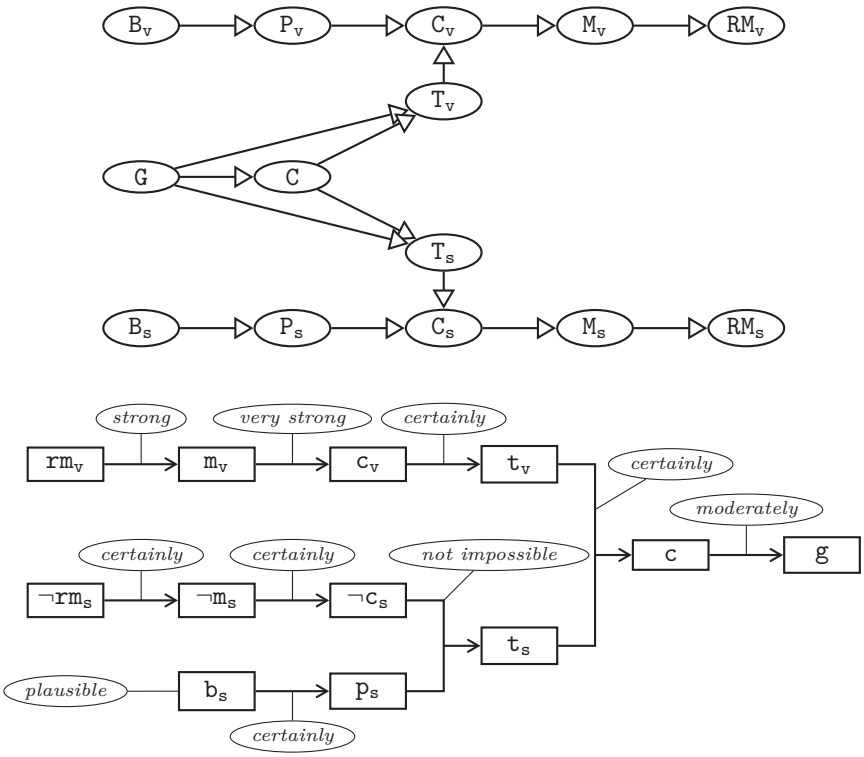


Figure 6.6: Example of a BN (top) and the corresponding extracted Argumentation Diagram (bottom) from [Keppens, 2012]. CPTs are used to compute the outcomes of argument and the verbal strength expressions but we have omitted them here for readability. The $\mathtt{RM}_v$ and $\mathtt{RM}_s$ represent the reported match of the stains on the victim (matching with the suspect) and the suspect (matching with the victim) respectively. From these findings it is derived that the suspect is guilty (variable G).

Another approach to extract arguments from BNs was taken by Keppens [2012], who extracts Argumentation Diagrams (ADs) from BNs. An Argumentation Diagram is a graphical structure that informally represents support between statements. Such a structure can be constructed by first including all directed paths between evidence and the variable of interest and expanding this with variables that have a *second-order influence* on those paths. Second-order influence means that these variables do not directly influence the variable of interest, but can modify the strength of another influence. This resembles the complex interactions modelled in Chapter 4. Like in our proposal, assignments are added after identifying the structure of the inference graph. This is done by calculating the Most Probable Explanation (MPE), which is the set of assignments that is *collectively* the most likely outcome of the nodes. For each inference step in the AD, a likelihood ratio is computed which is translated to a verbal expression such as

'plausible' or 'certainly' added as a label to the arrow in the AD. An example of such an automatically extracted AD is shown in Figure 6.6.

Argumentation Diagrams—much like structured argumentation—represent inference from premises to conclusions. In formal argumentation frameworks, such as ASPIC+, the justification status of arguments can be computed, which is something that cannot be done with Argumentation Diagrams. In the argumentation approach presented in Chapter 3 we chose to use ASPIC+ arguments rather than Argumentation diagrams because of the inherent adversarial setting and the strong formal underpinning present with such a system. In Chapter 4 we considered pro and con arguments collectively and weighted them in accordance with the Bayesian network. The latter is more similar to AD extraction. However, by looking at the support graph before the pruning step, these possible counterarguments can be identified. That is, the adversarial perspective only disappears after pruning the support graph to a specific set of observations.

In summary, the argument extraction methods presented in Chapters 3 introduces the idea to extract inference rules from a BN, a way to combine those into formal, structured argumentation and a number of considerations for such an approach. Chapter 4 focusses on extracting the reasoning chains from the BN first. This helps to identify which variables should be taken into account in 'good' arguments. The use of argumentation, rather than visual or textual explanation methods, is arguably more useful because of its inherent adversarial setting. Furthermore, our proposed method takes into account the independence information that is present in the graphical structure of the BN, which is something that has not received much attention in the literature on explanation methods for BNs.

6.5 Construction of Bayesian networks

In Chapter 5 we proposed a method for BN construction that captures argumentative reasoning with position-to-know argumentation schemes. For designing BNs several methods have been proposed. The use of *fragments* or *idioms* is a recurring method to construct BNs. In Chapter 5 we have already mentioned several different approaches to construct these idioms. In particular, Fenton et al. [2013] have proposed a number of BN idioms for handling evidence in legal cases. Vlek has extended this work with BN idioms based on scenario schemes [Vlek et al., 2015]. In other (non-legal) domains related approaches have been taken. For instance, Laskey and Mahoney [1997] introduced a number of fragments for military situation assessment. Object-Oriented BNs (OOBNs) have been introduced by Koller and Pfeffer [1997] and are conceptually similar. These were later applied to legal cases by others such as Hepler et al. [2007]. We recognise a similar design in work by others, such as Kadane and Schum [1996] and Aitken et al. [2003].

Many methods focus on the structural part of the BN. This leaves the numerical part (the CPTs) to other sources. These numerical parameters are usually consid-

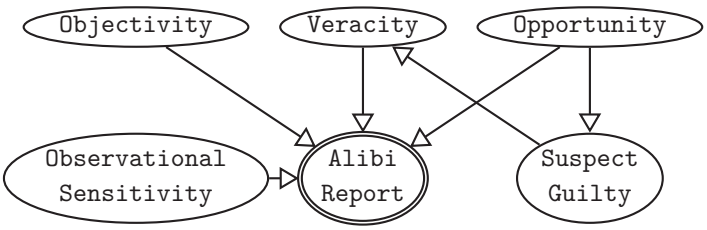


Figure 6.7: Alibi Idiom from [Lagnado et al., 2012].

ered harder to obtain. Sometimes a sensitivity analysis can help in this respect by showing for which variables the probabilities are critical for some variables of interest. Kadane and Schum [1996] build BNs including the CPTs but they do not present a structural method to define those CPTs. Instead, they make informal arguments to derive reasonable, conservative estimates based on the available evidence for one case. Druzdzel and Van der Gaag [1995] take a more formal approach to parameter estimation in their method that applies inequality constraints based on qualitative domain information to choose appropriate conditional probabilities.

Methods have also been proposed to construct BNs by means of a compositional approach where BNs are made by combining other BNs [Nielsen and Parsons, 2007]. In this work a BN is viewed as a cognitive representation of the world and agents with different models use argumentation to agree on a shared model (BN).

We have already mentioned that the way in which in the literature argumentation schemes and critical questions are modelled in BNs is dichotomised. Some of the approaches explicitly use argumentation schemes, others model position-to-know arguments in BNs without (at least explicitly) referring to argumentation schemes. Some approaches construct fragments for reuse while others directly construct a BN about a particular case. In all of these we have recognised two general approaches to modelling critical questions in BNs which we have taken as inspiration for our third, hybrid approach.

The first method that we discussed is followed by Lagnado et al. [2012], Fenton et al. [2013] and Carofiglio [2004], who introduced network fragments in which critical questions are modelled as parents of the evidence node. Let us consider [Lagnado et al., 2012] in more detail. This work applies the idioms that were introduced earlier by Fenton et al. [2013] to an example case. They emphasise the use of Bayesian models as a normative tool to compare human reasoning with statistical results. The alibi idiom in particular is taken as a prominent example. This idiom is shown in Figure 6.7.

The question that is discussed by Lagnado et al. [2012] is whether or not the veracity pertaining to an alibi should be linked to the guilt of the suspect. This boils down to the question whether or not an alibi that turns out to be false can be used as evidence for the suspect's guilt. Independent of this question we see that *objectivity*, *veracity* and *observational sensitivity* are taken to be parents of the evidence variable.

Figure 6.8 also shows an example from [Fenton et al., 2013] in which eye witness testimonies are modelled. Note that also here the critical questions (assuming competence is intended to mean competent to observe as in observational sensitivity) are modelled as three parents to one node, allowing complex interactions between them. Note that in the latter case an intermediate variable is introduced for the accuracy of the witness.

The second general approach that we discussed was followed by Hepler et al. [2007], Kadane and Schum [1996] and Aitken et al. [2003], who presented a number of BNs (not necessarily fragments) in which an evidential sequence of abductions is followed, resulting in a BN in which the evidence is step-wise weakened. We

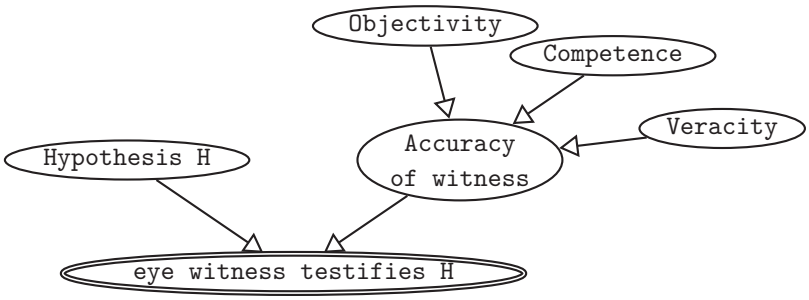


Figure 6.8: Eye witness idiom from [Fenton et al., 2013] (node names slightly abbreviated).

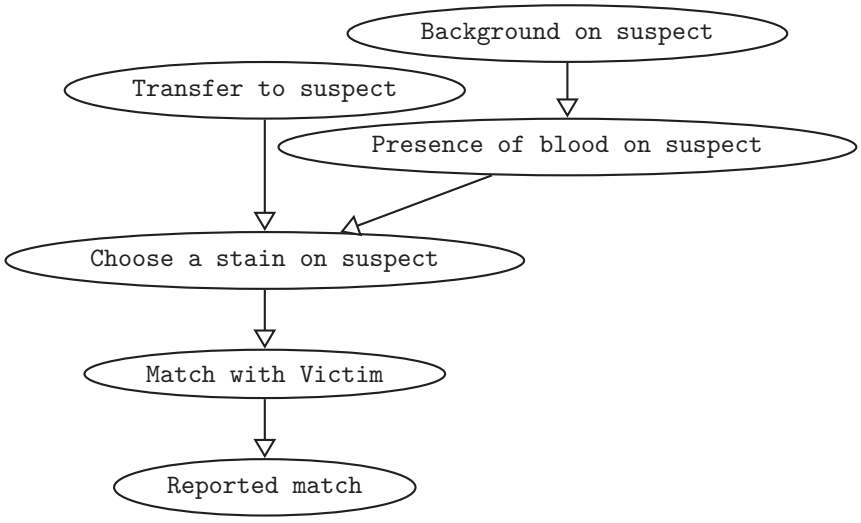


Figure 6.9: Chain of nodes relating to DNA transfer from [Aitken et al., 2003]. The primary chain of reasoning goes from the reported match to the fact that DNA was transferred, but a second chain (background noise) could also explain the evidence. Their full network contains two of these structures and shows how they are interconnected but that is not relevant for our purposes.

dubbed this approach a signal-filtering method because of its resemblance to a signal that passes through a number of noisy filters. Take, for example, the treatment of DNA stain evidence matches in [Aitken et al., 2003]. The fragment that is used (although Aitken et al. do not call this a fragment) is shown in Figure 6.9. What can be seen is that there is a chain of weakening steps, all of which can introduce an error in the inferential process. This process starts with the (hypothesised) transfer of blood and ends in the reported DNA match (evidence).

Very similarly, chains of inference have been modelled by Kadane and Schum [1996]. These are not based on a position-to-know argument, but they are constructed on the basis of Wigmore charts, which are a precursor to structured argumentation. Figure 6.10 shows one of their main examples about a robbery. The inferential structure starts at the bottom with three testimonies regarding Sacco's hand movement. From this it is derived that he was reaching for his weapon, and consequently that he was involved in the robbery (since innocent bystanders would presumably not draw a weapon).

A BN design method that uses idioms to construct BNs and which models witness testimonies is proposed by Hepler et al. [2007]. They use the notion of Object-Oriented Bayesian Networks (OOBNs) and apply it to a legal setting. An OOBN [Koller and Pfeffer, 1997] consists of nested modules that collectively constitute a normal BN. These modules are similar to the fragments or idioms that we discussed before. They can be specified at an abstract level, in which case they are referred to as 'generic' networks, and they can be instantiated for a real situation. An OOBN is computationally equivalent to a regular BN, but it differs in the presentation because certain subsets of nodes can be 'hidden' inside other nodes. So if, for instance, the model describes a police investigation, then at first the alibi of the suspect might be visible as a single node, but on closer inspection it might contain nodes for testimonies and other evidence that supports or attacks the alibi. The 'hidden' variables are typically the evidence that pertains to the alibi only, and are connected to the model only through that alibi. In this way the user is first confronted with a global, easy to oversee representation of the case and can then interact with the model to see all the details.

From the cited approaches, Hepler et al. [2007] most closely approximate the model of testimony evidence that we proposed in Chapter 5 because they mention the possibility to extend the objectivity and veracity nodes into a *noisy filter*. This is depicted in Figure 6.11, where a credibility subnetwork is expanded to accommodate exactly the three sources of doubt that we have also discussed. They provide no detail, nor argue why such a model was chosen or where it is or is not appropriate. This can be explained by the fact that the aim of their work is not to model testimony evidence but to introduce a method to structure large scale networks hierarchically. The proposed way to model evidence follows the more general method that we proposed in Chapter 5.

To summarise, our work combines advantages from design principles that we have identified in the literature. These principles are sometimes explicitly named

and sometimes implicitly followed. In terms of representational complexity the newly introduced hybrid approach ranks between the existing methods, which is to be expected based on the freedom that these models have in the complexity of interactions that they can model.

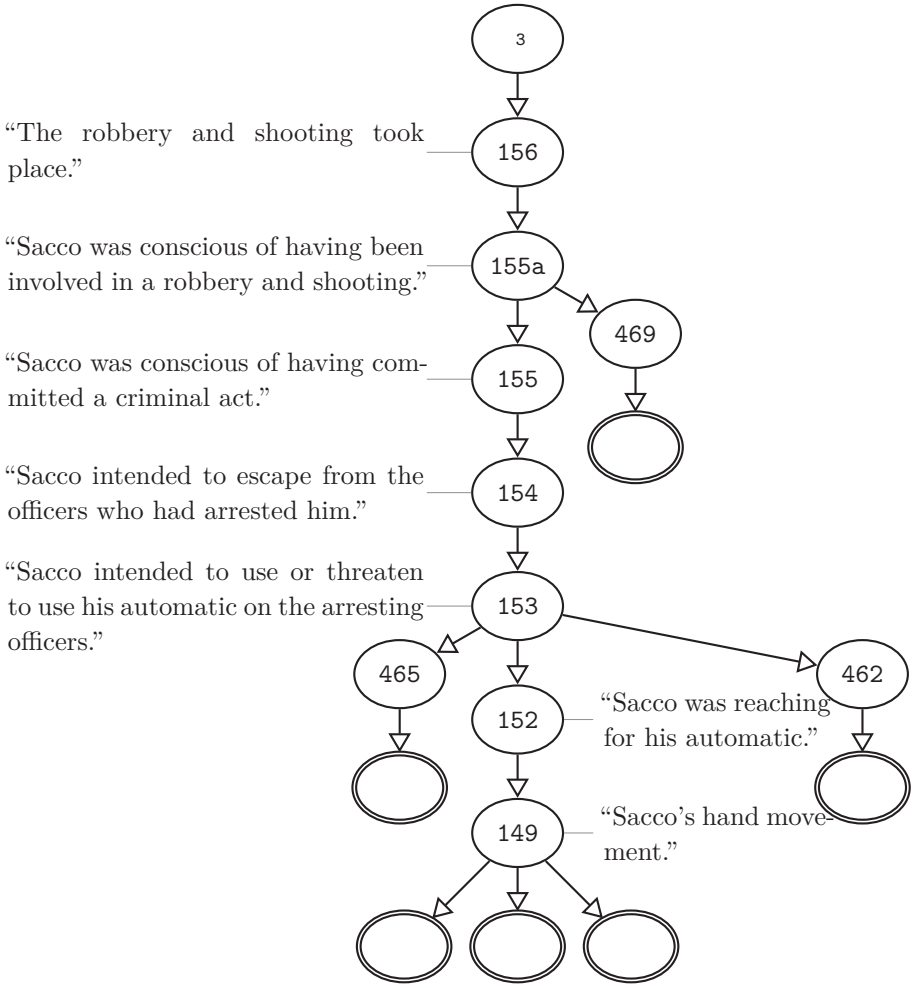


Figure 6.10: BN by Kadane and Schum [1996] about a robbery taking place (Π_3) supported through a series of inferences by evidence for a hand movement. Labels show what the nodes on the main path from 149 to (the hypothesis) Π_3 represent.

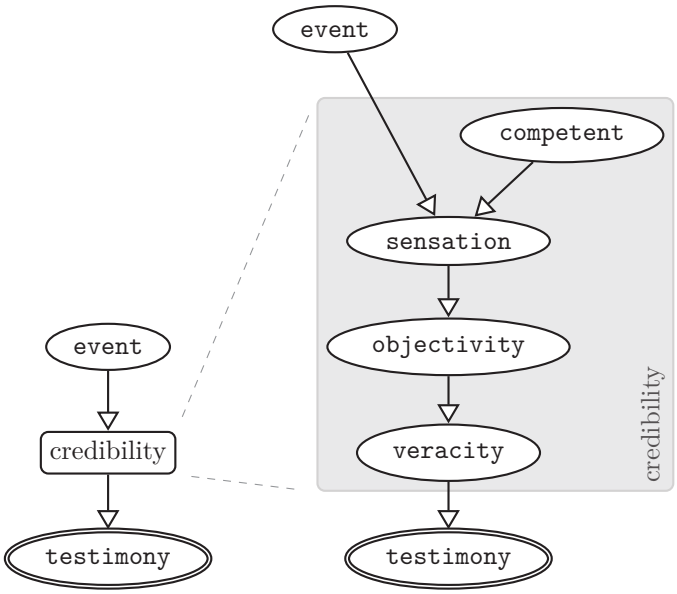


Figure 6.11: Part of OOBN from [Hepler et al., 2007] before and after expanding the credibility fragment.

Chapter 7

Conclusions

One of the motivations for this research has been the increasing use of probabilities in court and in particular the communication problems between experts with different backgrounds to which this has led. In order to better understand and possibly alleviate this problem we addressed the following main research topics:

Topic 1 How can argumentation techniques aid the understanding of Bayesian networks that model legal evidence?

Topic 2 How can argumentation techniques aid the design of Bayesian network models of legal evidence?

Since the motivation for this research indicates that there is a communication problem between experts with different backgrounds we chose to attack the problem from a translation perspective. Automated translation methods between different formal models of evidence may prove helpful to bridge such a communication gap. With this perspective in mind we address our main topics by answering the following research questions:

Research question 1 Can explanatory arguments be extracted from a given Bayesian network?

Research question 2 Can argumentation schemes for evidential reasoning be modelled in a Bayesian network?

In the following sections we summarise our answers to both questions separately and we provide some suggestions for future work.

7.1 Extracting arguments from Bayesian networks

Before we proceed to answer the first question we first summarise our contributions concerning the extraction of explanatory arguments. This summarises Chapters 3 and 4.

7.1.1 Summary

In Chapter 3 we introduced a formalisation of argumentation in a Bayesian network (BN) setting. This is an initial step towards a formal connection between concepts from Bayesian reasoning and argumentative reasoning. The intuitive interpretation of argumentation structures may facilitate the interpretation of BNs by non-mathematical experts, making such a translation arguably valuable in the field of legal reasoning about evidence.

An important aspect of such a translation is the use of a strength measure, the choice of which has a strong influence on the interpretation of the resulting arguments. When using an incremental measure of strength, the resulting arguments have an incremental interpretation. For example, conclusions like "the evidence supports hypothesis A better than hypothesis B" can be drawn using incremental measures of strength. When using an absolute measure of strength, the resulting arguments have an absolute interpretation. Conclusions like "given the evidence, hypothesis A is more likely than hypothesis B" can be drawn using such a measure. A second important aspect of the use of strength measures is that we introduced the notion of a *context* of a rule, which is the subset of evidence not used in the rule itself. This ensures that rules are compared on their probabilistic strength in the context of all available evidence that is relevant to the rule.

Using a strength measure, rules and undercutters to those rules can be enumerated automatically. Subsequently, using the rules and their undercutters, arguments can be constructed. We applied the ASPIC+ argumentation framework to do this because it is one of the major modern argumentation frameworks for structured argumentation. A problem that arises when applying such a method naively is that, without further constraints, the number of possible inference rules is already large, but the number of ways in which they can be chained into arguments is even larger and becomes quickly infeasible for larger input BNs. For that reason we identified a number of constraints that can be put on this chaining process that reduces spurious, redundant results. First, during the enumeration of candidate rules, only premises from the Markov blanket of the conclusion are considered. This limits the number of possible rules considerably while maintaining the possibility to reason along chains in the graph and to find the complex interactions between parents of a common child. We argued that this is in fact a more natural representation of the inference in the BN. Recall that Pearl's evidence propagation algorithm also follows these paths [Pearl, 1988a] and that this provides an intuitive interpretation of how observations influence other variables. Secondly, we introduced a labelling scheme inspired by Pearl's CE-system for evidential reasoning which ensures that no inference chains are created that are not valid in the BN.

Although these constraints on the production of rules and arguments makes the method more feasible, we improved on this in Chapter 4 with a new intermediate representation of evidence, and its relation to the hypothesis, called a *support graph*.

This allowed us to more efficiently capture the potential reasoning chains that at the same time provide a skeleton for fast construction of the best argument that can be constructed on the basis of the given BN and evidence. This intermediate model leads to a two-phase argument extraction method for explaining the reasoning in BNs. In the first phase, only the structure of arguments that are allowed by the independence relation modelled in the BN is determined and represented as a support graph. Reusing the idea of strength measures to calculate inferential strength from Chapter 3, arguments can be constructed from such a support graph.

7.1.2 Answering research question 1

Let us now reconsider the first subquestion that we posed:

> **Question 1.1:** Can Bayesian inference be translated to argumentative inference rules?

To answer this question we introduced in Chapter 3 a method to extract inference rules from a BN. Recall our intuitively appealing interpretation of BN inference following active chains in the BN graph. This means that BN inference can be viewed as a sequence of small inference steps, similar to how argumentation organises complex proof by combining smaller inference steps. This idea was further developed in Chapter 4, where we bundled all allowed chains of inference in a BN together in a support graph. By using a probabilistic measure of strength, not only the structure of the BN is mirrored by the argumentative inference, but also its quantitative properties. Together, this means that we can positively answer the question whether Bayesian inference can be translated to argumentative inference rules. In particular, we have introduced two ways to identify these rules that mirror the inferences possible in the BN.

Special care needs to be taken where edges in the BN converge because intercausal interactions introduce complex behaviour that is different from the naively combined effect along the followed edges. This brings us to the next subquestion:

> **Question 1.2:** Can probabilistic arguments be constructed that respect the dynamic interactions present in Bayesian networks?

To answer this question we have looked at how induced intercausal dependencies—and most notably explaining away—can be translated to argument undercutting. Since explaining away captures a dynamic aspect in Bayesian inference which is similar to the dynamics of undercutting, we identified a method to use the same strength measure used to extract rules, to extract undercutters to those rules in Chapter 3. Furthermore, we formally showed that only active chains in the BN given the observed evidence can result in argument chains in the extracted arguments. Observing additional evidence can unblock new chains of inference, which we mirrored in the argumentation by introducing a labelling scheme. Additional evidence can also block an existing inference, which happens when this evidence (or something that can be derived from it) poses an undercutter. In Chapter 4

we further focussed on the intercausal interactions that can be modelled by a BN and the resulting support graph such that exactly and only the potentially allowed paths of inference were represented by such a graph. This demonstrates that it is indeed possible to capture the dynamic aspects of Bayesian inference in argumentation.

The argument generation method in Chapter 3 is not efficient. This is a computational issue for larger networks where the size of the resulting argument graph grows exponentially. More importantly, large argument graphs are hard to interpret. Many of the identified arguments were however redundant or not important for the variables of interest, indicating room for improvement. Therefore, another motivation for the work in Chapter 4 was to both reduce computation time and increase readability of the resulting argumentation. This brings us to question 1.3 about the computational efficiency of argument extraction.

> **Question 1.3:** What is the computational complexity of a translation from Bayesian network to arguments via rules?

We note that to compute the strength of even a single inference rule, we have to calculate posterior probabilities from the BN, which is by itself computationally expensive and worst case exponential in the size of the network. However, algorithms for evidence propagation in BNs have been introduced that can do this fast in most practical situations. Leaving aside for now the fact that calculating rule strengths and enumerating rules, such as in Chapter 3, is computationally hard, we observed that the real bottleneck arises when we try to combine these rules in all possible ways to identify arguments. We have also seen that, even for an input network of four variables, the size of the argument graphs is significantly larger than the size of the input network. For larger networks the size of the argument graph quickly becomes unmanageable. This is not just a computational issue, but also one that hinders the explanatory value of the constructed arguments. That is, automatically extracted arguments are less valuable when surrounded by too many irrelevant arguments. Only the arguments that are in some chosen extension have to be reported, but we noted that a lot of the extracted rules and arguments are not defeated but merely irrelevant or redundant for the variable of interest. The method proposed in Chapter 4, therefore, does not enumerate all possible rules, but first constructs a graphical representation of the allowed structure of the extracted arguments. The available evidence is then used to prune all parts of this graph that cannot contribute to a conclusion about the variable of interest. Although the worst case size of such a support graph is in general exponential in the number of nodes in the input BN, we showed that such a support graph will benefit from the sparseness of the input BN, similarly to how BN propagation algorithms are typically faster for sparser graphs. In particular we showed that both the size and computation time of a support graph grow linearly with the size of the input network for singly connected graphs.

Throughout Chapters 3 and 4 we have made only one assumption regarding

the variables, which was that these were boolean valued. Technically, any other variable can be turned into a set of boolean variables but the result may not be elegant for explanation purposes. The only place where we relied on the boolean values of nodes is for the computation of odds and likelihood ratios in the measures of strength. We made no assumption about the structure of the BN graph or the probabilistic parameters of the BN model. This answers the last subquestion:

> **Question 1.4:** Does a translation method impose limiting constraints on the Bayesian networks that are applicable as inputs to such a system?

Moreover, we have left the choice for a measure of strength open so that another measure can be plugged in without changing the formalism.

7.2 Modelling argumentation schemes in Bayesian networks

In Chapter 5 we addressed the second topic of this thesis:

> **Topic 2:** How can argumentation techniques aid the design of Bayesian network models of legal evidence?

Again, we approached this from a translation perspective. We addressed the topic by combining the best features of two general approaches from the literature for modelling argumentation schemes for position-to-know arguments in BNs. The research question that we tried to answer was:

> **Question 2:** Can argumentation schemes for evidential reasoning be modelled in a Bayesian network?

7.2.1 Summary

Argumentation and Bayesian networks can both be constructed by combining reusable templates. In argumentation these are called argumentation schemes and in BNs they are called idioms or fragments. We proposed a translation from one type of argumentation scheme to BN idioms. The proposed method is a hybrid between two general approaches that we recognised in the literature. These two general approaches have their own advantages. In essence we can say that one is more concise while the other is more versatile. Consequently they differ in a number of ways. We identified criteria on the basis of which these approaches can be compared. As we showed, the features from the two approaches that we discuss are not necessarily exclusive and we introduced a new approach that performs well on all of the introduced criteria simultaneously. We have seen that most approaches to modelling argumentation schemes agree on the way in which argumentative inference should be modelled in a BN (or indeed in a BN idiom to

maintain the possibility to reuse the fragment in other cases). In particular we have identified that in all of the discussed approaches, argumentative inference is modelled as a possibly active chain in a BN conditioned on the variables that can be observed. We have seen that, consequently, it is always the case that the premise and the conclusion are modelled by a node in the BN.

The discussed methods diverge on the topic of modelling critical questions. We have seen that two general approaches for modelling critical questions can be identified. First, a number of authors have proposed methods in which the critical questions are modelled by adding parents to the conclusion node. This creates head-to-head connections between the critical questions and the premise of the scheme. Such a connection is necessary to express complex interactions such as explaining away. Since critical questions need this type of interaction with the premise of the argumentation scheme, the use of these head-to-head connections is inevitable in this approach. There is, however, another general approach to modelling critical questions in BNs, which is to encode the uncertainty introduced by these questions in the edges of the path from the premise to the conclusion. This has some advantages with regard to the complexity of the model, but it means that the critical questions are not represented directly by nodes in the BN graph. We have argued that this is a disadvantage of such a model of critical questions. To alleviate this, we introduced a mixed model in which the critical questions are applied sequentially in a chain of head-to-head connections. This combines most of the advantages from both other approaches. The only aspect in which this hybrid approach is outperformed by one of the existing general approaches is in its representation complexity.

7.2.2 Answering research question 2

Now, consider the first subquestion that we posed:

> **Question 2.1:** Can argumentation schemes be captured by Bayesian networks?

In the literature a number of design methods for incorporating argumentation schemes in BNs have been proposed, which implies that the respective authors believe that this question should be answered positively. We agree with this conclusion and, as already noted, we introduced a number of desirable features of methods that do this. Our own method performs well on all of these features.

About the modelling of critical questions in a BN there is no such agreement in the literature. Hence we proceed to address the next question:

> **Question 2.2:** How are critical questions to be incorporated in such a BN (fragment)?

As discussed above, we identified two general approaches to do this and we introduced a hybrid approach that outperforms both on most of the differentiating criteria. One of these criteria was that the representational complexity should be

minimised. That is, the number of probabilities that needs to be entered in the conditional probability tables should be reduced when possible. We saw that in this respect the proposed hybrid model falls between the two other approaches. The fact that the two general approaches and the newly introduced hybrid approach differ in the number of model parameters also hints at the answer to the next question:

> **Question 2.3:** What are the representational complexity implications of the answers to the above?

When critical questions are embedded in a BN by adding for each critical question a parent to the conclusion, the possibility is created that all critical questions feature complex interactions with not only the premise but also each other. Although this may be desirable in some occasions, it is generally not necessary and introduces many numerical parameters to the model. Since each parameter needs to be estimated and (at least for legal applications) data is rarely available to learn parameters from, this leads to overly complex models of the evidence. The other approach (modelling critical questions in the probability tables of the nodes on the chain from premise to conclusion) is much more efficient. However, as we have discussed, this has other disadvantages. We have shown that the hybrid approach introduced in Chapter 5 results in a compromise between the other two approaches. On a scale from least generic (but with few parameters) to very generic models with many parameters, our method ranks between the existing approaches in the literature. If the number of parameters was the only criterion of interest, then the best approach would be the one with just enough freedom to represent the required interactions between the variables.

Chapter 5 introduced a method to model argumentation schemes for position-to-know arguments in a BN. Although this covers a large portion of legal arguments, many other argumentation schemes have been introduced in the literature. In this context we have asked the following question:

> **Question 2.4:** Are all argumentation schemes sufficiently similar to be captured by one BN model?

We have studied the possibility of creating BN models of argumentation schemes. The approach introduced in Chapter 5 is not limited to eye witness testimonies nor to witness testimonies in general. Instead, it applies to the more general category of argumentation schemes from position-to-know, which includes the former two types of evidence. However, other types of argumentation schemes exist and we discussed examples where the approach does not apply intuitively. To make this more specific we have concluded that our method applies to evidential argumentation schemes with a natural ordering on the critical questions.

7.3 Future research

The research as described in the previous chapters has led to the development of new theories and insights, but has also raised new questions. We discuss a number of potentially interesting future extensions and variations of our work.

7.3.1 Verbal explanations

The method to extract arguments as presented in Chapter 3 generates a large number of irrelevant arguments. This results in cluttered argument graphs. By using support graphs the number of arguments can be reduced to those directly relevant to the conclusion and hence the resulting argumentation is easier to oversee for a human interpreter, such as a legal expert. One question that arises is whether the arguments constructed by the support graph method can be linearised into verbal explanations. We conjecture that this is relatively straightforward. This may benefit the practical applicability of such a method.

Consider again the resulting argument that was found for the case study in Section 4.4. The same argument could be expressed in natural language as follows:

> The evidence weakly suggests that the driver was not speeding because the tire marks suggest that he was in fact slowing, even though he arguably also lost control over the vehicle. The former fact is observed by the police and the latter is not uncommon to deduce from the fact that the car was skidding as evidenced by the very occurrence of the crash and the presence of tire marks. However, the skidding may also have been caused by locking of the wheels, in turn caused by the passenger pulling the handbrake, which is strongly supported by the driver's testimony, the pulled state of the handbrake after the crash and the fact that the passenger was drunk.

We note that to automate such a verbal explanation two necessary components will have to be developed:

- Verbal classifications for both positive and negative influence in order to distinguish the different strengths of inferences. Inspiration for such a classification could, for instance be taken from verbal expressions used in probability elicitation methods [Renooij, 2001]. See Druzdzel [1996] or Vlek [2016] for other work in this direction.
- A better understanding of when to use the terms 'even though' and 'despite' as in the example above. This does not follow from the strengths of the inference because those only report the total strength of all subarguments after weighing. A subargument negatively contributes to the conclusion if it would support the other outcome on its own. For this, inspiration could be taken from Kadane and Schum [1996] and Schum [1994].

7.3.2 More constraints on the support graph

The support graph as introduced in Chapter 4 is constructed by iteratively expanding the roots of the graph under construction. During this construction a set of forbidden variables is maintained to ensure that certain patterns are not present in the resulting argumentation. In particular we have used this to prevent two types of reasoning:

- Cyclical arguments, where a statement about variable $\mathtt{V}$ is used to support another statement about that same variable $\mathtt{V}$;
- Arguments that follow an immorality in the BN because in those cases the explaining away step, if any, should be preferred over the indirect reasoning.

This technique could be extended to blacklist other variables or structures. Reasons for this will typically be domain specific. Consider, for example, the case when a particular variable should be used in weighing the strengths of other inferences (i.e. when it is in the *context* of the evaluated rule), but should not itself be considered a reason. We also foresee one possible technical application of such an additional filter. In theory, when one observed variable has more than two parents, it is possible to make two consecutive explaining away steps. See Figure 7.1. In this case it is possible to reason from $\mathtt{V_1}$ to either of $\mathtt{V_2}$ or $\mathtt{V_3}$. From $\mathtt{V_2}$ it is, however, also allowed to reason to $\mathtt{V_3}$. This means that in the support graph two paths will be found between $\mathtt{V_1}$ and $\mathtt{V_3}$, one directly, and one via $\mathtt{V_2}$. Although this is technically not an issue (making two such intercausal steps is not a fallacy or mistake) it may be counterintuitive to present this redundant step as an option in the resulting argument. Filtering this extra chain can be done by adding other parents of a node to the forbidden set when making an explaining away step.

7.3.3 Experimental validation of our methods

We have proposed a number of models and methods to design and explain BNs. We have evaluated these by means of formal analyses and conceptual comparisons with existing work. However, this does not prove that these methods are useful for legal or forensic professionals in practice. Recall again that the motivating problem

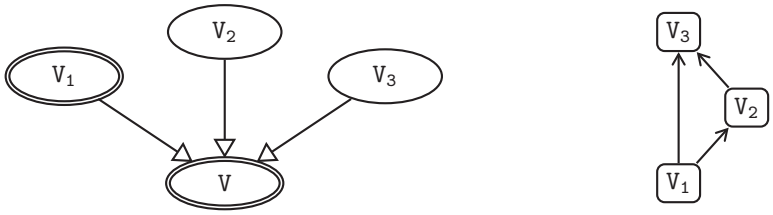


Figure 7.1: Example of how multiple parents can create a redundant explaining away step.

of our research was the communication barrier between these fields. Determining whether the proposed methods indeed alleviate the described problem requires experimental validation. This could, for instance, be done by investigating if judgements concerning probabilities by legal expert are improved when presented with the extracted explanatory arguments. The argument extraction methods proposed in Chapter 3 and 4 are at this point algorithmic tools. To lawyers, judges and other legal experts this may not be useful in that form. How these arguments are to be presented (for instance using verbal explanations as described above) and whether their interpretation by legal experts is indeed the interpretation intended by the person constructing the BN has to be studied.

Similarly, for the BN construction method proposed in Chapter 5, It could be studied if and how the proposed method helps experts to construct Bayesian models of legal (testimony) evidence and if the constructed models are useful in the cases for which they are built.

7.3.4 More generalised argument extraction

We have used ASPIC+ and Dung-style argumentation frameworks as the target for our BN translation method. Reasons for this are the fact that ASPIC+ is one of the major modern argumentation frameworks and that rationality postulates follow straightforwardly from our definitions and known results from the literature. However, other argumentation approaches exist, such as Carneades [Gordon et al., 2007] and ADFs [Brewka and Woltran, 2010]. Such a system might align better with the intuition of summarizing pro and con arguments which was an important aspect in Chapter 4. In both Carneades and ADFs the attacks and supports are modelled at the same conceptual level.

Another question that could be investigated is how uncertainty about inferences results in uncertainty about arguments. We already discussed a number of probabilistic generalisations of Dung's abstract argumentation [Hunter, 2013; Li et al., 2012]. ADFs pose another such generalisation. However, there is presently no formal connection between the uncertainty in the applied rules and the uncertainty about the whole argument. In particular, because these generalised Dung frameworks allow one to model uncertain arguments and our argument extraction methods resulted in uncertain inferences (where the existence of the arguments was not disputed) the connection between these two levels of uncertainty could result in a better understanding of the relation between probabilistic and argumentative reasoning.

7.3.5 More generalised argumentation scheme models

We have proposed in Chapter 5 a BN model of argumentation schemes for position-to-know arguments. This includes different kinds of testimony evidence and hence a large portion of evidence that is common in legal cases. However, we have not looked at other domains where other types of arguments may be more common. We did identify two critical properties of argumentation schemes for our method

to apply. These were the presence of an ordering on the critical questions and an evidential (rather than predictive) nature. In future research it could be investigated how common these properties are or if there are variations on our method that eliminate or alleviate these requirements.

7.4 Final remarks

To summarize, we have addressed the combination of two of the three normative perspectives on modelling evidence, which are argumentation, probabilities and scenarios. Of these, we have investigated how argumentation and probabilistic reasoning can be combined. The aim of this was to ultimately facilitate the communication between forensic and legal experts. The different backgrounds of these experts are better aligned with the two normative perspectives respectively. That is, legal experts are better accustomed to argumentative representations of evidence, whereas forensic experts are more familiar with probabilistic models. The work presented here is a step towards that goal of integrating the two approaches. The angle from which this problem was attacked was to introduce conversions or translation methods (although the latter may suggest that the translation is reversible, which is at least in the first proposed method not true) between the representations of evidence. To translate BNs to arguments we have developed an argument extraction method from BNs which pays special attention to the independence information which is embedded in the graphical structure of the BN. To translate arguments to BNs, we have introduced a method to embed argumentation schemes in BN idioms. Where argumentation schemes are common building blocks for arguments, idioms fulfil a similar role in BN design and therefore a translation perspective offers a natural approach to the problem.

We have already stressed that the proposed support graph method does not replace evidence propagation in Bayesian networks but rather complements it by adding argumentative explanations of the probabilistic reasoning. Necessarily, any method that translates quantitative models of evidence, such as BNs, to qualitative models of evidence, such as arguments, discards some information. The BN construction method is also not a replacement for other BN design methods, but it provides a guideline or starting point for modelling one common type of evidence.

Regarding the motivating problem for this research, which is the communication gap between experts with different backgrounds, we argued that the support graph notion has a beneficial effect in terms of explaining the reasoning chains in BNs. The fact that the directions of arrows in a BN have no intrinsic interpretation by themselves is one of the often confusing aspects for non-probabilistic experts. We address this problem by first extracting a support graph, which removes this confusing property, while maintaining conditional independence information. Similarly, the argument construction method results in BNs that have an argumentative interpretation making them more familiar to legal experts. By

investigating the translation of probabilistically modelled information to an argumentative representation and vice versa we advanced our understanding of how the two fields relate. Such an understanding is, in our opinion, crucial to seal the communication gap between probabilistic and non-probabilistic experts, for instance in legal debates about forensic, probabilistic evidence.

Appendices

Appendix A

Arguments extracted in Chapter 3

Scenario 1

In order to reduce line breaks we have abbreviated `Guilty` = `true` to `g` and `Guilty` = `false` to ¬`g`. Similarly we have abbreviated assignments to `Conspiracy` to `c` and ¬`c`, assignments to `Evidence of Crime` to `e1` and ¬`e1`, and assignments to `Evidence of Conspiracy` to `e2` and ¬`e2`. These are all labelled and extracted rules:

$r_1 : C(\texttt{c})$	$\Rightarrow C(\texttt{e2})$	$(strength = 4.977)$
$r_2 : E(\texttt{c})$	$\Rightarrow C(\texttt{e2})$	$(strength = 4.977)$
$r_6 : E(\neg\texttt{c})$	$\Rightarrow C(\neg\texttt{e2})$	$(strength = 1.098)$
$r_7 : C(\neg\texttt{c})$	$\Rightarrow C(\neg\texttt{e2})$	$(strength = 1.098)$
$r_{12} : E(\neg\texttt{e1})$	$\Rightarrow E(\neg\texttt{g})$	$(strength = 1.086)$
$r_{14} : E(\texttt{e1})$	$\Rightarrow E(\texttt{g})$	$(strength = 8.990)$
$r_{22} : E(\neg\texttt{e1}), E(\neg\texttt{c})$	$\Rightarrow E(\neg\texttt{g})$	$(strength = 1.086)$
$r_{23} : C(\neg\texttt{c}), E(\neg\texttt{e1})$	$\Rightarrow E(\neg\texttt{g})$	$(strength = 1.086)$
$r_{29} : E(\texttt{e1}), E(\neg\texttt{c})$	$\Rightarrow E(\texttt{g})$	$(strength = 9.888)$
$r_{30} : C(\neg\texttt{c}), E(\texttt{e1})$	$\Rightarrow E(\texttt{g})$	$(strength = 9.888)$
$r_{36} : E(\neg\texttt{e2})$	$\Rightarrow E(\neg\texttt{c})$	$(strength = 1.098)$
$r_{38} : E(\texttt{e2})$	$\Rightarrow E(\texttt{c})$	$(strength = 4.977)$
$r_{42} : E(\neg\texttt{e1})$	$\Rightarrow E(\neg\texttt{c})$	$(strength = 1.008)$

$r_{44} : E(\mathtt{e1}) \Rightarrow E(\mathtt{c}) \quad (strength = 10.102)$

$r_{60} : E(\neg\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{62} : E(\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 50.282)$

$r_{65} : E(\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 1.233)$

$r_{69} : E(\neg\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.000)$

$r_{73} : C(\neg\mathtt{g}), E(\neg\mathtt{e1}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{72} : E(\neg\mathtt{e1}), E(\neg\mathtt{g}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{75} : C(\mathtt{g}), E(\mathtt{e1}) \Rightarrow E(\mathtt{c}) \quad (strength = 1.123)$

$r_{76} : E(\mathtt{e1}), E(\mathtt{g}) \Rightarrow E(\mathtt{c}) \quad (strength = 1.123)$

$r_{80} : C(\mathtt{g}), E(\neg\mathtt{e1}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.005)$

$r_{81} : E(\neg\mathtt{e1}), E(\mathtt{g}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.005)$

$r_{83} : E(\mathtt{e1}), E(\neg\mathtt{g}) \Rightarrow E(\mathtt{c}) \quad (strength = 90.090)$

$r_{84} : C(\neg\mathtt{g}), E(\mathtt{e1}) \Rightarrow E(\mathtt{c}) \quad (strength = 90.090)$

$r_{88} : C(\neg\mathtt{g}), E(\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 98.792)$

$r_{87} : E(\mathtt{e1}), E(\mathtt{e2}), E(\neg\mathtt{g}) \Rightarrow E(\mathtt{c}) \quad (strength = 98.792)$

$r_{93} : C(\mathtt{g}), E(\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.008)$

$r_{92} : E(\mathtt{e1}), E(\mathtt{g}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.008)$

$r_{96} : C(\mathtt{g}), E(\neg\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 4.347)$

$r_{95} : E(\neg\mathtt{e1}), E(\mathtt{g}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 4.347)$

$r_{99} : C(\mathtt{g}), E(\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 9.278)$

$r_{100} : E(\mathtt{e1}), E(\mathtt{g}), E(\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 9.278)$

$r_{104} : E(\neg\mathtt{e1}), E(\neg\mathtt{g}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{105} : C(\neg\mathtt{g}), E(\neg\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{109} : C(\mathtt{g}), E(\neg\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{108} : E(\neg\mathtt{e1}), E(\mathtt{g}), E(\neg\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.009)$

$r_{112} : C(\neg\mathtt{g}), E(\mathtt{e1}), E(\neg\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 50.251)$

$r_{111} : E(\mathtt{e1}), E(\neg\mathtt{g}), E(\neg\mathtt{e2}) \Rightarrow E(\mathtt{c}) \quad (strength = 50.251)$

$r_{117} : C(\neg\mathtt{g}), E(\neg\mathtt{e1}), E(\mathtt{e2}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.000)$

$r_{116} : E(\neg\mathtt{e1}), E(\mathtt{e2}), E(\neg\mathtt{g}) \Rightarrow E(\neg\mathtt{c}) \quad (strength = 1.000)$

$r_{121} : E(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{122} : C(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{126} : E(\neg\mathtt{c}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.008)$

$r_{127} : C(\neg\mathtt{c}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.008)$

$r_{130} : C(\neg\mathtt{g}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.086)$

$r_{131} : E(\neg\mathtt{g}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.086)$

$r_{133} : C(\mathtt{g}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.990)$

$r_{134} : E(\mathtt{g}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.990)$

$r_{140} : E(\neg\mathtt{g}), E(\neg\mathtt{c}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.096)$

$r_{139} : C(\neg\mathtt{g}), E(\neg\mathtt{c}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.096)$

$r_{141} : C(\neg\mathtt{g}), C(\neg\mathtt{c}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.096)$

$r_{138} : C(\neg\mathtt{c}), E(\neg\mathtt{g}) \Rightarrow C(\neg\mathtt{e1}) \quad (strength = 1.096)$

$r_{144} : C(\mathtt{g}), E(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{143} : E(\mathtt{c}), E(\mathtt{g}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{146} : C(\mathtt{g}), C(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{145} : E(\mathtt{g}), C(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{150} : E(\neg\mathtt{g}), C(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{151} : C(\neg\mathtt{g}), C(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{149} : C(\neg\mathtt{g}), E(\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{152} : E(\mathtt{c}), E(\neg\mathtt{g}) \Rightarrow C(\mathtt{e1}) \quad (strength = 10.102)$

$r_{157} : C(\mathtt{g}), C(\neg\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.979)$

$r_{156} : C(\mathtt{g}), E(\neg\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.979)$

$r_{155} : C(\neg\mathtt{c}), E(\mathtt{g}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.979)$

$r_{158} : E(\mathtt{g}), E(\neg\mathtt{c}) \Rightarrow C(\mathtt{e1}) \quad (strength = 8.979)$

Scenario 2

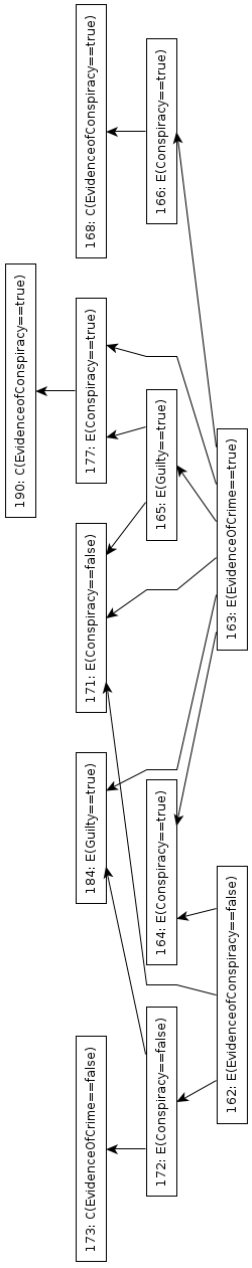


Figure A.1: Arguments built from the extracted rule base, with observations for `EvidenceofCrime` = `true` and `EvidenceofConspiracy` = `false`.

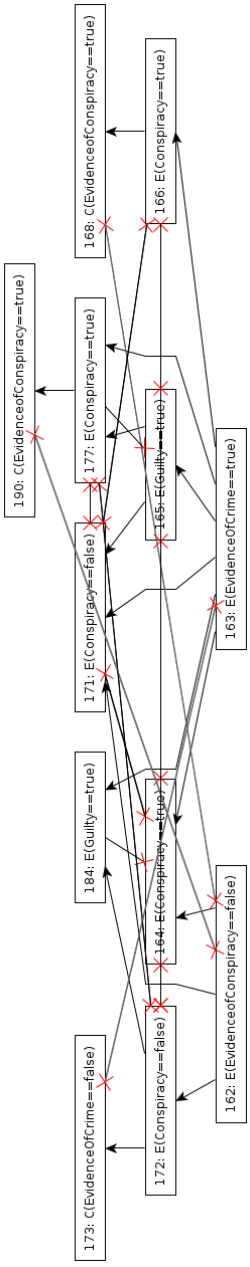


Figure A.2: The same arguments as in Figure A.1, but with the attack relation. Undercutting, rebutting and undermining are not visually distinguishable, but from the conclusions of the arguments it can be deduced what the type of attack is. Again, only direct attack is shown.

Scenario 3

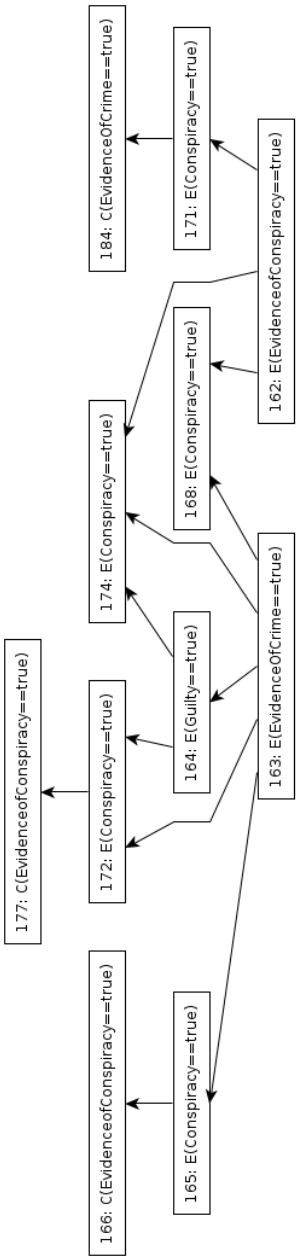


Figure A.3: Arguments built from the extracted rule base, with observations for `EvidenceofCrime` = `true` and `EvidenceofConspiracy` = `true`.

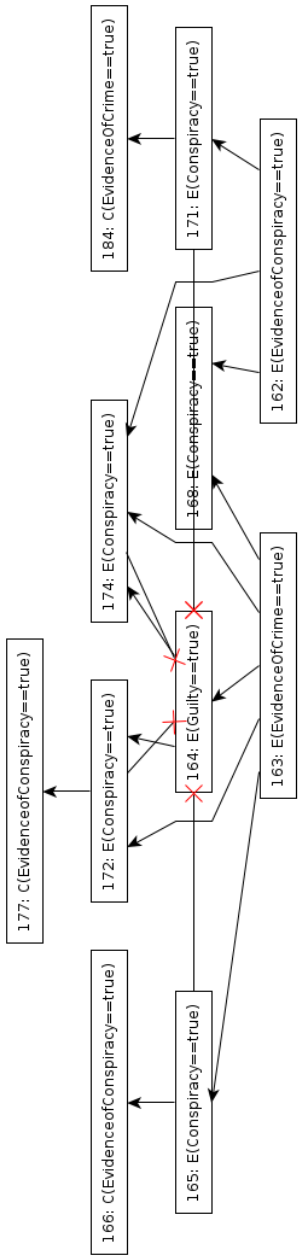


Figure A.4: The same arguments as in Figure A.3, but with the attack relation shown. Compared to Figure A.2, there are far fewer attacks, which is the result of the fact that the evidence does not contradict itself in any way. Again, only direct attack is shown and all attackers are undercutters.

Appendix B

Conditional probabilities for the case in Chapter 4

drunk_passenger=true	0.05
drunk_passenger=false	0.95

speeding_in_S_curve=true	0.00001
speeding_in_S_curve=false	0.99999

passenger_pulls_handbrake	true	false
locking_of_wheels=true	0.8	0.1
locking_of_wheels=false	0.2	0.9

speeding_in_S_curve	true	false
loss_of_control_over_vehicle=true	1.0	0.00001
loss_of_control_over_vehicle=false	0.0	0.99999

drunk_passenger	true	false
passenger_pulls_handbrake=true	0.03	0.0
passenger_pulls_handbrake=false	0.97	1.0

loss_of_control_over_vehicle	true		false	
locking_of_wheels	true	false	true	false
skidding=true	1.0	1.0	1.0	0.0
skidding=false	0.0	0.0	0.0	1.0

skidding	true	false
crash=true	0.2	0.0001
crash=false	0.8	0.9999

passenger_pulls_handbrake	true	false
drivers_testimony=true	0.9	0.03
drivers_testimony=false	0.1	0.97

passenger_pulls_handbrake	true	false
handbrake_in_pulled_position=true	0.99	0.001
handbrake_in_pulled_position=false	0.01	0.999

speeding_in_S_curve	true	false
tire_marks_after_S_curve_suggest_slowing=true	0.2	0.7
tire_marks_after_S_curve_suggest_slowing=false	0.8	0.3

skidding	true	false
tire_marks_present=true	1.0	0.0
tire_marks_present=false	0.0	1.0

Samenvatting

Sluitend bewijs is essentieel voor een goede rechtsgang. In het juridische domein wordt bewijsmateriaal aan conclusies gekoppeld met één of meer redeneerstappen. Argumentatieleer bestudeert deze redeneerstappen en hoe zij gecombineerd kunnen worden tot een sluitend bewijs. Deze studie is gebaseerd op ideeën uit de formele logica. Een bewijsvoering moet immers aan de wetten van de logica voldoen om overtuigend te zijn. Recente ontwikkelingen op het gebied van forensische wetenschap hebben echter voor problemen gezorgd met deze aanpak. Bij het vergelijken van DNA of vingerafdrukken maakt men vaak gebruik van een kansanalyse om de de bewijskracht van die sporen te beschrijven. Die kansanalyse beschrijft getalsmatig wat de zeldzaamheid van het gevonden bewijs is. Het interpreteren van die resultaten lijkt simpel: hoe zeldzamer de eigenschappen van het bewijs, hoe groter de bewijskracht van een gevonden match.

De praktijk is weerbarstiger. Vaak is er niet één bewijsstuk maar zijn er meerdere. Er kan dan discussie ontstaan over de interacties tussen de sporen onderling en met de conclusie van het onderzoek. Ook is het vaak niet duidelijk hoe kwantificeerbare sporen (zoals vingerafdrukken) gecombineerd moeten worden met niet-kwantificeerbare sporen (zoals getuigenverklaringen). De hierdoor ontstane verwarring heeft al meermaals geleid tot juridisch dwalen. Bekende voorbeelden hiervan zijn de zaak tegen Lucia de B. in Nederland of Sally Clark in Groot-Brittannië.

Het lijkt alsof er een communicatiekloof is tussen juridische en forensische experts. Waar juristen gewend zijn aan argumentatief redeneren, kwantificeren forensische experts onzekerheden in kansen. De rol van numerieke analyse als bewijs in de rechtszaal is daardoor onder vuur komen te liggen. Er is echter een onderliggend probleem: experts (zowel juridisch als forensisch) gaan op verschillende manieren om met de onzekerheid die bewijsstukken met zich meebrengen. Argumentatieleer en kansanalyses bieden twee verschillende perspectieven op onzekerheid in bewijs. In dit proefschrift combineer ik deze twee aspecten van bewijs in een poging de werelden van juridische en forensische experts dichter bij elkaar te brengen.

Voor zowel argumentatief als kansmatig redeneren bestaan computationele modellen. Als input hebben dergelijk modellen over het algemeen bewijsstukken en redeneerregels (in het geval van argumentatiesystemen) of tabellen met numerieke afhankelijkheden (in het geval van kansredeneren). Op basis van die input worden

dan conclusies gegenereerd die op basis van dat bewijs getrokken kunnen worden, al dan niet voorzien van een numerieke betrouwbaarheid.

De aanpak waar ik in dit proefschrift voor kies is om vertaalmethoden te onderzoeken die als brug kunnen dienen tussen deze twee modellen. Daarmee kan dus bewijs dat argumentatief gepresenteerd wordt omgezet worden naar iets dat bruikbaar is in een kansmatige analyse. Eveneens kan kansmatig gemodelleerd bewijs omgezet worden naar argumentatief bewijs. Deze tweezijdige vertaalmethode helpt dus enerzijds om kansmodellen op te stellen en anderzijds om deze uit te leggen in argumentatieve termen. Een dergelijke aanpak kan in de toekomst hopelijk bijdragen aan een beter begrip van kansbewijs in rechtzaken.

Acknowledgement

I would like to express my gratitude towards a number of people without whom I would not have been able to produce this thesis in its current form or at all.

First and foremost I should thank my supervisors. Without their dedication to keep me on the track towards our common goal—the completion of this thesis and the underlying research—I would have been lost. John-Jules, I am thankful for the fact that you were always available. Not just for feedback on my progress, but also to discuss the difficulties (and sometimes pleasures) of the scientific discourse and life in general. Your presence lightened up even the sometimes difficult meetings. Henry, you have been a real mentor for the last four years. I thank you for your time and dedication and the patience with which you kept steering me in the right direction, while allowing me to choose my own course as well. It must not have been easy at times. Silja, your contribution to my work is best described as *meticulous*. When you were done there was always more red than black ink on the page. I promise that if I ever get to supervise students, I will remember you and try to be just as thorough, knowing that if they curse me for it at first, they will ultimately be thankful. Bart, besides all the research feedback, I fondly remember our off-topic discussions. Besides a good colleague you have positioned yourself as a good friend as well. I am grateful for your personal guidance.

I also owe an acknowledgement to those colleagues that provided their scientific input to my work. First, there are Floris and Charlotte, working in our shared models of evidence project. It's good to have people around with fresh recollections of what PhD research is like (or are still undergoing it). Roland and Rutger, thanks for proof reading (parts of) this manuscript. Thanks to the numerous colleagues that I met during all the conferences and workshops over the last years. I cannot begin to list them all here but I thank you all for the engaging talks and the inspiration you have been. Finally, I thank the members of the reading committee—Anthony, Marjan, Peter, Rosalie and Simon—for their time and feedback.

I would also like to thank the (former) colleagues from inside and outside our group for the entertaining discussions we had over lunch and during coffee breaks. Some of them became friends with whom I share much more than small talk. Allard, Bas, Cheah, Chide, Chris, Eric, Estelle, Frank, Gerard, Hein, Jan,

Janneke, Jesse, Jeroen, Joost, Loïs, Luora, Marieke, Marlo, Max, Mehdi, Rogier, Ruud, Samaneh, Sjur, And Tom. A few stand out in particular. Bas, my office mate, I could always rely on you when I was in need of a conversation. I'm going to miss that. Cheah and Luora, from you I learnt so much about myself from your inquisitive attitude. Your questions have opened my eyes so many times that I will always remember to keep looking at the world and myself from as many angles as possible. Thank you. Chide, your fascination for technology is truly contagious. Although we did not always agree I enjoyed every last bit of our discussions. Loïs, you were the mastermind behind many social events ranging from mental (board game night) to physical (laser gaming) to slightly crazy (pirate themed murder dinner party). Your enthusiasm, I believe, sparked of onto others making all of those other great adventures possible.

Outside the university there is one group of people that I would like to acknowledge in particular. These are the 'hackers' from RandomData, the hackerspace in Utrecht, a lively community of makers, builders, and software security experts. There were times when our weekly chill-out on Tuesday night was the only thing that could take my mind of work for a couple of hours. 1sand0s, [com]buster, Ardillo, fish_, harmless, potatomas, RSpliet, synnack, TeaJay, and zkyp. I thank you all very much for that.

Finally, I thank my family. Eric, Sietske, Simba and Els. You were there, watching from the sideline, for the last four year. Always supportive and understanding. Most of all I am thankful towards Inge. You were patient when I was absent minded, understanding when I was hard to live with and most of all supportive when I most needed it. My endeavour—that you had not asked for—had just as big an impact on your life as it had on mine and I thank you for supporting me from the start to the end.

Many thanks to all of you,
Sjoerd

Curriculum vitae

2001 – 2006
Piter Jelles Gymnasium, Leeuwarden: VWO

2006 – 2010
Utrecht University: BSc, Computing science

2010 – 2012
Utrecht University: MSc Computing science, cum laude, direction: Algorithm Design and Complexity

2012 – 2016
Utrecht University: PhD, direction: Artificial intelligence

Bibliography

C. Aitken, F. Taroni, and P. Garbolino. A graphical model for the evaluation of cross-transfer evidence in DNA profiles. *Theoretical Population Biology*, 63(3): 179–190, 2003.

T. Anderson and W. Twining. *Analysis of Evidence. How to Do Things with Facts Based on Wigmore's Science of Judicial Proof*. Little, Brown and Company, Boston, MA, 1991.

S. A. Andersson, D. Madigan, and M. D. Perlman. A characterization of markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2): 505–541, 1997.

C. Berger, J. Buckleton, C. Champod, I. Evett, and G. Jackson. Evidence evaluation: A response to the appeal court judgment R v T. *Science and Justice*, 51(2):43–49, 2011.

P. Besnard and A. Hunter. Argumentation based on classical logic. In G. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 133–152. Springer, 2009.

F. J. Bex. *Arguments, Stories and Criminal Evidence*, volume 92 of *Law and Philosophy Library*. Spinger, 2011.

F. J. Bex and B. Verheij. Legal stories and the process of proof. *Artificial Intelligence and Law*, 21(3):253–278, 2013.

F. J. Bex, H. Prakken, C. Reed, and D. N. Walton. Towards a formal account of reasoning about evidence: Argumentation schemes and generalisations. *Artificial Intelligence and Law*, 11(2):125–165, 2003.

F. J. 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.

G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.

M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

V. Carofiglio. Modelling argumentation with belief networks. In F. Grasso, C. Reed, and G. Gareninim, editors, *Proceedings of the 4th Workshop on Computational Models of Natural Argument*, 2004.

G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.

V. Crupi, K. Tentori, and M. Gonzales. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74(2):229–252, 2007.

A. P. Dawid. Beware of the DAG! In I. Guyon, D. Janzing, and B. Schölkopf, editors, *NIPS Causality: Objectives and Assessment*, pages 59–86, 2010.

T. Derksen and M. Meijsing. The fabrication of facts: the lure of the credible coincidence. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, chapter 2, pages 39–70. Ashgate, Farham, 2009.

P. Diaconis and D. Freedman. The persistence of cognitive illusions. *Behavioral and Brain Sciences*, 4(3):333–334, 1981.

M. J. Druzdzel. Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly*, 94:43–54, 1996.

M. J. Druzdzel and L. C. Van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In P. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 141–148. Morgan Kaufmann, 1995.

M. J. Druzdzel and L. C. Van der Gaag. Building probabilistic networks: Where do the numbers come from? *IEEE Transactions on knowledge and data engineering*, 12(4):481–486, 2000.

P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

P. M. Dung and P. M. Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In *Computational Models of Argument. Proceedings of COMMA 2010*, pages 171–182, Amsterdam, 2010. IOS Press.

P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-Based Argumentation. In G. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009.

I. Evett. Avoiding the transposed conditional. *Science & Justice*, 35(2):127–131, 1995.

N. E. Fenton. Science and law: Improve statistics in court. *Nature*, 479:36–37, 2011.

N. E. Fenton, M. Neil, and D. A. Lagnado. A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science*, 37(1):61–102, 2013.

T. F. Gordon and D. Walton. Proof burdens and standards. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, 2009.

T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

U. Hahn and J. Hornikx. A normative framework for argument quality: argumentation schemes with a Bayesian foundation. *Synthese*, 193(6):1833–1873, 2016.

U. Hahn and M. Oaksford. The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3):704–732, 2007.

U. Hahn, A. J. Harris, and M. Oaksford. Rational argument, rational inference. *Argument & Computation*, 4(1):21–35, 2013.

M. Henrion. Practical issues in constructing a bayes' belief network. In J. Lemmer, T. Levitt, and L. Kanal, editors, *Proceedings of the 3rd Conference Annual Conference on Uncertainty in Artificial Intelligence*, pages 132–139. AUAI Press, 1987.

A. B. Hepler, A. P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability & Risk*, 6(1-4): 275–293, 2007.

A. J. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.

A. J. Hunter, editor. *Argument and Computation*, volume 5. 2014. Special issue with Tutorials on Structured Argumentation.

A. J. Hunter and M. Thimm. Probabilistic argumentation with incomplete information. In *Proceedings of the European Conference on Artificial Intelligence*, pages 1033–1034. IOS Press, 2014.

A. J. Hunter and M. Thimm. On partial information and contradictions in probabilistic abstract argumentation. In *Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning*, pages 53–62. AAAI Press, 2016.

P. E. M. Huygen. Use of Bayesian belief networks in legal reasoning. In *17th BILETA Annual Conference*. John Wiley & Sons, Inc, 2002.

F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science & Statistics. Springer Verlag, 2nd edition, 2007.

J. B. Kadane and D. A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. Wiley, New York, 1996.

D. Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

H. Kaptein, H. Prakken, and B. Verheij, editors. *Legal Evidence and Proof: Statistics, Stories, Logic*. Applied Legal Philosophy. Ashgate Publishing, 2009.

J. Keppens. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law*, 20(2):109–143, 2012.

J. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

J. R. Koiter. Visualizing inference in Bayesian networks. Master's thesis, Delft University of Technology, 2006.

D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 302–313. Morgan Kaufmann Publishers Inc., 1997.

J. Kwisthout. Most inforbable explanations: Finding explanations in Bayesian networks that are both probable and informative. In L. Van der Gaag, editor, *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 7958 of *Lecture Notes in Artificial Intelligence*, pages 328–339. Springer-Verlag, 2013.

J. Kwisthout. Most frugal explanations in Bayesian networks. *Artificial Intelligence*, 218(C):56–73, 2015.

C. Lacave and F. J. Díez. A review of explanation methods for Bayesian networks. *Knowledge Engineering Review*, 17(2):107–127, 2002.

C. Lacave, M. Luque, and F. J. Díez. Explanation of Bayesian networks and influence diagrams in Elvira. *Systems, Man, and Cybernetics, Part B*, 37(4):952–965, 2007.

D. A. Lagnado, N. E. Fenton, and M. Neil. Legal idioms: a framework for evidential reasoning. *Argument and Computation*, 4(1):1–18, 2012.

K. B. Laskey and S. M. Mahoney. Network fragments: Representing knowledge for constructing probabilistic models. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 334–341. Morgan Kaufmann, 1997.

S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50(2):157–224, 1988.

R. Lempert. The new evidence scholarship: Analyzing the process of proof. *Boston University Law Review*, 66(3):439–478, 1986.

H. Li, N. Oren, and T. J. Norman. Probabilistic argumentation frameworks. In S. Modgil, N. Oren, and F. Toni, editors, *Theory and Applications of Formal Argumentation*, volume 7132 of *Lecture Notes in Computer Science*, pages 1–16, Berlin, 2012. Springer.

D. Madigan, K. Mosurski, and R. G. Almond. Graphical explanation in belief networks. *In Journal of Computational and Graphical Statistics*, 6(2):160–181, 1997.

R. Meester, M. Collins, R. Gill, and M. van Lambalgen. On the (ab)use of statistics in the legal case against the nurse Lucia de B. *Law, Probability & Risk*, 5(3-4):233–250, 2007.

S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5(1):31–62, 2014.

S. H. Nielsen and S. Parsons. Argumentation in artificial intelligence an application of formal argumentation: Fusing bayesian networks in multi-agent systems. *Artificial Intelligence*, 171(10):754 – 775, 2007.

S. Parsons. A proof theoretic approach to qualitative probabilistic reasoning. *International Journal of Approximate Reasoning*, 19(3):265 – 297, 1998.

S. Parsons. On precise and correct qualitative probabilistic inference. *International Journal of Approximate Reasoning*, 35(2):111 – 135, 2004.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, 1988a.

J. Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35: 259–271, 1988b.

N. Pennington and R. Hastie. Reasoning in explanation-based decision making. *Cognition*, 49(1–2):123–163, 1993.

N. Pfeifer. On argument strength. In F. Zenker, editor, *Bayesian Argumentation: The practical side of probability*, pages 185–193. Springer Netherlands, 2013.

J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

J. L. Pollock. Justification and defeat. *Artificial Intelligence*, 67(2):377–407, 1994.

J. L. Pollock. *Cognitive Carpentry: A Blueprint for how to Build a Person*. MIT Press Cambridge, 1995.

J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence Journal*, 133(1):233–282, 2001.

H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

H. Prakken. Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and dungean semantics. *Artificial Intelligence and Law*, 20(1):57–82, 2012.

H. Prakken. On direct and indirect probabilistic reasoning in legal proof. *Law, Probability & Risk*, 13(3-4):327–337, 2014.

H. Prakken and S. Renooij. Reconstructing causal reasoning about evidence: a case study. In B. Verheij, A. R. Lodder, R. P. Loui, and A. J. Muntjewerff, editors, *Legal Knowledge and Information Systems. JURIX 2001: The 14th Annual Conference*, pages 131–137. IOS Press, 2001.

V. Reddy, A. C. Farr, P. Wu, K. Mengersen, and P. K. D. V. Yarlagadda. An intuitive dashboard for Bayesian network inference. *Journal of Physics*, 490: 12023–12026, 2014.

S. Renooij. *Qualitative Approaches to Quantifying Probabilistic Networks*. PhD thesis, Utrecht University, 2001.

L. Schneps and C. Colmez. *Math on Trial: How Numbers Get Used and Abused in the Courtroom*. Basic Books, 2013.

D. Schum and P. Tillers. Marshaling evidence for adversary litigation. *Cardozo Law Review*, 13:657–704, 1991.

D. A. Schum. *The evidential foundation of probabilistic reasoning*. Northwestern University Press, 1994.

G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157, 1992.

M. Sjerps. Bewijskracht 10, volle vaart recht vooruit, 2011. inaugural lecture.

H. J. Suermondt. *Explanation in Bayesian Belief Networks*. PhD thesis, Departments of Computer Science and Medicine, Stanford University, Stanford, 1992.

P. Thagard. Causal inference in legal decision making: explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, 18(3-4):231–249, 2004.

W. C. Thompson and E. L. Schumann. Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, 11(3):167–187, 1987.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Inference and attack in Bayesian networks. In K. Hindriks, M. de Weerdt, B. van Riemsdijk, and M. Warnier, editors, *Proceedings of the 25th Benelux Conference on Artificial Intelligence*, pages 199–206, 2013.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Extracting legal arguments from forensic Bayesian networks. In R. Hoekstra, editor, *Legal Knowledge and Information Systems. JURIX 2014: The 27th Annual Conference*, volume 217, pages 71–80. IOS Press, 2014.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Explaining Bayesian networks using argumentation. In S. Destercke and T. Denoeux, editors, *Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 9161 of *Lecture Notes in Artificial Intelligence*, pages 83–92. Springer, 2015a.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 109–118. ACM, 2015b.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Capturing critical questions in Bayesian network fragments. In A. Rotolo, editor, *Legal Knowledge and Information Systems. JURIX 2015: The Twenty-eighth Annual Conference*, pages 173–176. IOS Press, 2015c.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Explaining legal Bayesian networks using support graphs. In A. Rotolo, editor, *Legal Knowledge and Information Systems. JURIX 2015: The Twenty-eighth Annual Conference*, pages 121–130. IOS Press, 2015d.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. Demonstration of a structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law, ICAIL 2015, San Diego, CA, USA, June 8-12, 2015*, pages 233–234. ACM, 2015e.

S. T. Timmer, J.-J. Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, 2016. accepted for publication.

L. C. Van der Gaag, S. Renooij, C. Witteman, B. M. P. Aleman, and B. G. Taal. How to elicit many probabilities. In K. B. Laskey and H. Prade, editors, *UAI '99: Proceedings of the Fifteenth Conference on Uncertaintyin Artificial Intelligence*, pages 647–654. Morgan Kaufmann, 1999.

F. H. Van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. S. Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of Argumentation Theory*. Springer, Dordrecht, 2014.

J. Van Leersum. Explaining the reasoning of Bayesian networks, 2015. master thesis, Utrecht University.

B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003a.

B. Verheij. Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artificial Intelligence and Law*, 11(1–2):167–195, 2003b.

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

B. Verheij, F. Bex, S. T. Timmer, C. S. Vlek, J.-J. Ch. Meyer, S. Renooij, and H. Prakken. Arguments, scenarios and probabilities: Connections between three normative frameworks for evidential reasoning. *Law, Probability & Risk*, 15 (1):35–70, 2016.

T. Verma and J. Pearl. Equivalence and synthesis of causal models. In R. Shachter, T. Levitt, and L. Kanal, editors, *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, pages 255–270, New York, NY, USA, 1991. Elsevier Science Inc.

C. S. Vlek. *When Stories and Numbers Meet in Court: Constructing and Explaining Bayesian Networks for Criminal Cases with Scenarios*. PhD thesis, University of Groningen, Faculty of Mathematics and Natural Sciences, 2016.

C. S. 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*, pages 150–159. ACM Press, 2013.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Building Bayesian networks for legal evidence with narratives: a case study evaluation. *Artificial Intelligence and Law*, 22(4):375–421, 2014.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Constructing and understanding Bayesian networks for legal evidence with scenario schemes. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 128–137. ACM Press, 2015.

G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90 (1-2):225–279, 1997.

G. A. W. Vreeswijk. Argumentation in Bayesian belief networks. In I. Rahwan, P. Moraïtis, and C. Reed, editors, *Argumentation in Multi-Agent Systems*, volume 3366 of *Lecture Notes in Computer Science*, pages 111–129. Springer Berlin / Heidelberg, 2005.

W. A. Wagenaar, P. J. van Koppen, and H. F. M. Crombag. *Anchored Narratives. The Psychology of Criminal Evidence*. Harvester Wheatsheaf, London, 1993.

D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.

M. P. Wellman. Fundamental concepts in qualitative probabilistic networks. *Artificial Intelligence*, 44(3):257–303, 1990.

J. H. Wigmore. *The Principles of Judicial Proof*. Little, Brown and Company, Boston, 1913.

M. Williams and J. Williamson. Combining argumentation and Bayesian Nets for breast cancer prognosis. *Journal of Logic, Language and Information*, 15(1–2): 155–178, 2006.

G.-E. Yap, A.-H. Tan, and H.-H. Pang. Explaining inferences in Bayesian networks. *Applied Intelligence*, 29(3):263–278, 2008.

H. L. Yuan and T. Lu. Most relevant explanation in Bayesian networks. *Journal of Artificial Intelligence Research*, 42(1):309–352, 2011.

SIKS Dissertation Series

1998

1 Johan van den Akker (CWI) *DEGAS: An Active, Temporal Database of Autonomous Objects*

2 Floris Wiesman (UM) *Information Retrieval by Graphically Browsing Meta-Information*

3 Ans Steuten (TUD) *A Contribution to the Linguistic Analysis of Business Conversations*

4 Dennis Breuker (UM) *Memory versus Search in Games*

5 E. W. Oskamp (RUL) *Computerondersteuning bij Straftoemeting*

1999

1 Mark Sloof (VUA) *Physiology of Quality Change Modelling: Automated modelling of*

2 Rob Potharst (EUR) *Classification using decision trees and neural nets*

3 Don Beal (UM) *The Nature of Minimax Search*

4 Jacques Penders (UM) *The practical Art of Moving Physical Objects*

5 Aldo de Moor (KUB) *Empowering Communities: A Method for the Legitimate User-Driven*

6 Niek J. E. Wijngaards (VUA) *Re-design of compositional systems*

7 David Spelt (UT) *Verification support for object database design*

8 Jacques H. J. Lenting (UM) *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism*

2000

1 Frank Niessink (VUA) *Perspectives on Improving Software Maintenance*

2 Koen Holtman (TUe) *Prototyping of CMS Storage Management*

3 Carolien M. T. Metselaar (UvA) *Sociaal-organisatorische gevolgen van kennistechnologie*

4 Geert de Haan (VUA) *ETAG, A Formal Model of Competence Knowledge for User Interface*

5 Ruud van der Pol (UM) *Knowledge-based Query Formulation in Information Retrieval*

6 Rogier van Eijk (UU) *Programming Languages for Agent Communication*

7 Niels Peek (UU) *Decision-theoretic Planning of Clinical Patient Management*

8 Veerle Coupé (EUR) *Sensitivity Analyis of Decision-Theoretic Networks*

9 Florian Waas (CWI) *Principles of Probabilistic Query Optimization*

10 Niels Nes (CWI) *Image Database Management System Design Considerations, Algorithms and Architecture*

11 Jonas Karlsson (CWI) *Scalable Distributed Data Structures for Database Management*

2001

1 Silja Renooij (UU) *Qualitative Approaches to Quantifying Probabilistic Networks*

2 Koen Hindriks (UU) *Agent Programming Languages: Programming with Mental Models*

3 Maarten van Someren (UvA) *Learning as problem solving*

4 Evgueni Smirnov (UM) *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*

5 Jacco van Ossenbruggen (VUA) *Processing Structured Hypermedia: A Matter of Style*

6 Martijn van Welie (VUA) *Task-based User Interface Design*

7 Bastiaan Schonhage (VUA) *Diva: Architectural Perspectives on Information Visualization*

8 Pascal van Eck (VUA) *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*

9 Pieter Jan 't Hoen (RUL) *Towards Distributed Development of Large Object-Oriented Models*

10 Maarten Sierhuis (UvA) *Modeling and Simulating Work Practice*

11 Tom M. van Engers (VUA) *Knowledge Management*

2002

1 Nico Lassing (VUA) *Architecture-Level Modifiability Analysis*

2 Roelof van Zwol (UT) *Modelling and searching web-based document collections*

3 Henk Ernst Blok (UT) *Database Optimization Aspects for Information Retrieval*

4 Juan Roberto Castelo Valdueza (UU) *The Discrete Acyclic Digraph Markov Model in Data Mining*

5 Radu Serban (VUA) *The Private Cyberspace Modeling Electronic*

6 Laurens Mommers (UL) *Applied legal epistemology: Building a knowledge-based ontology of*

7 Peter Boncz (CWI) *Monet: A Next-Generation DBMS Kernel For Query-Intensive*

8 Jaap Gordijn (VUA) *Value Based Requirements Engineering: Exploring Innovative*

9 Willem-Jan van den Heuvel (KUB) *Integrating Modern Business Applications with Objectified Legacy*

10 Brian Sheppard (UM) *Towards Perfect Play of Scrabble*

11 Wouter C. A. Wijngaards (VUA) *Agent Based Modelling of Dynamics: Biological and Organisational Applications*

12 Albrecht Schmidt (UvA) *Processing XML in Database Systems*

13 Hongjing Wu (TUe) *A Reference Architecture for Adaptive Hypermedia Applications*

14 Wieke de Vries (UU) *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*

15 Rik Eshuis (UT) *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*

16 Pieter van Langen (VUA) *The Anatomy of Design: Foundations, Models and Applications*

17 Stefan Manegold (UvA) *Understanding, Modeling, and Improving Main-Memory Database Performance*

2003

1 Heiner Stuckenschmidt (VUA) *Ontology-Based Information Sharing in Weakly Structured Environments*

2 Jan Broersen (VUA) *Modal Action Logics for Reasoning About Reactive Systems*

3 Martijn Schuemie (TUD) *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*

4 Milan Petkovic (UT) *Content-Based Video Retrieval Supported by Database Technology*

5 Jos Lehmann (UvA) *Causation in Artificial Intelligence and Law: A modelling approach*

6 Boris van Schooten (UT) *Development and specification of virtual environments*

7 Machiel Jansen (UvA) *Formal Explorations of Knowledge Intensive Tasks*

8 Yongping Ran (UM) *Repair Based Scheduling*

9 Rens Kortmann (UM) *The resolution of visually guided behaviour*

10 Andreas Lincke (UvT) *Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture*

11 Simon Keizer (UT) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*

12 Roeland Ordelman (UT) *Dutch speech recognition in multimedia information retrieval*

13 Jeroen Donkers (UM) *Nosce Hostem: Searching with Opponent Models*

14 Stijn Hoppenbrouwers (KUN) *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*

15 Mathijs de Weerdt (TUD) *Plan Merging in Multi-Agent Systems*

16 Menzo Windhouwer (CWI) *Feature Grammar Systems: Incremental Maintenance of Indexes to Digital Media Warehouses*

17 David Jansen (UT) *Extensions of Statecharts with Probability, Time, and Stochastic Timing*

18 Levente Kocsis (UM) *Learning Search Decisions*

2004

1 Virginia Dignum (UU) *A Model for Organizational Interaction: Based on Agents, Founded in Logic*

2 Lai Xu (UvT) *Monitoring Multi-party Contracts for E-business*

3 Perry Groot (VUA) *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*

4 Chris van Aart (UvA) *Organizational Principles for Multi-Agent Architectures*

5 Viara Popova (EUR) *Knowledge discovery and monotonicity*

6 Bart-Jan Hommes (TUD) *The Evaluation of Business Process Modeling Techniques*

7 Elise Boltjes (UM) *Voorbeeldig onderwijs: voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes*

8 Joop Verbeek (UM) *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise*

9 Martin Caminada (VUA) *For the Sake of the Argument: explorations into argument-based reasoning*

10 Suzanne Kabel (UvA) *Knowledge-rich indexing of learning-objects*

11 Michel Klein (VUA) *Change Management for Distributed Ontologies*

12 The Duy Bui (UT) *Creating emotions and facial expressions for embodied agents*

13 Wojciech Jamroga (UT) *Using Multiple Models of Reality: On Agents who Know how to Play*

14 Paul Harrenstein (UU) *Logic in Conflict. Logical Explorations in Strategic Equilibrium*

15 Arno Knobbe (UU) *Multi-Relational Data Mining*

16 Federico Divina (VUA) *Hybrid Genetic Relational Search for Inductive Learning*

17 Mark Winands (UM) *Informed Search in Complex Games*

18 Vania Bessa Machado (UvA) *Supporting the Construction of Qualitative Knowledge Models*

19 Thijs Westerveld (UT) *Using generative probabilistic models for multimedia retrieval*

20 Madelon Evers (Nyenrode) *Learning from Design: facilitating multidisciplinary design teams*

2005

1 Floor Verdenius (UvA) *Methodological Aspects of Designing Induction-Based Applications*

2 Erik van der Werf (UM) *AI techniques for the game of Go*

3 Franc Grootjen (RUN) *A Pragmatic Approach to the Conceptualisation of Language*

4 Nirvana Meratnia (UT) *Towards Database Support for Moving Object data*

5 Gabriel Infante-Lopez (UvA) *Two-Level Probabilistic Grammars for Natural Language Parsing*

6 Pieter Spronck (UM) *Adaptive Game AI*

7 Flavius Frasincar (TUe) *Hypermedia Presentation Generation for Semantic Web Information Systems*

8 Richard Vdovjak (TUe) *A Model-driven Approach for Building Distributed Ontology-based Web Applications*

9 Jeen Broekstra (VUA) *Storage, Querying and Inferencing for Semantic Web Languages*

10 Anders Bouwer (UvA) *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*

11 Elth Ogston (VUA) *Agent Based Matchmaking and Clustering: A Decentralized Approach to Search*

12 Csaba Boer (EUR) *Distributed Simulation in Industry*

13 Fred Hamburg (UL) *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*

14 Borys Omelayenko (VUA) *Web-Service configuration on the Semantic Web: Exploring how semantics meets pragmatics*

15 Tibor Bosse (VUA) *Analysis of the Dynamics of Cognitive Processes*

16 Joris Graaumans (UU) *Usability of XML Query Languages*

17 Boris Shishkov (TUD) *Software Specification Based on Re-usable Business Components*

18 Danielle Sent (UU) *Test-selection strategies for probabilistic networks*

19 Michel van Dartel (UM) *Situated Representation*

20 Cristina Coteanu (UL) *Cyber Consumer Law, State of the Art and Perspectives*

21 Wijnand Derks (UT) *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*

2006

1 Samuil Angelov (TUe) *Foundations of B2B Electronic Contracting*

2 Cristina Chisalita (VUA) *Contextual issues in the design and use of information technology in organizations*

3 Noor Christoph (UvA) *The role of metacognitive skills in learning to solve problems*

4 Marta Sabou (VUA) *Building Web Service Ontologies*

5 Cees Pierik (UU) *Validation Techniques for Object-Oriented Proof Outlines*

6 Ziv Baida (VUA) *Software-aided Service Bundling: Intelligent Methods & Tools for Graphical Service Modeling*

7 Marko Smiljanic (UT) *XML schema matching: balancing efficiency and effectiveness by means of clustering*

8 Eelco Herder (UT) *Forward, Back and Home Again: Analyzing User Behavior on the Web*

9 Mohamed Wahdan (UM) *Automatic Formulation of the Auditor's Opinion*

10 Ronny Siebes (VUA) *Semantic Routing in Peer-to-Peer Systems*

11 Joeri van Ruth (UT) *Flattening Queries over Nested Data Types*

12 Bert Bongers (VUA) *Interactivation: Towards an e-cology of people, our technological environment, and the arts*

13 Henk-Jan Lebbink (UU) *Dialogue and Decision Games for Information Exchanging Agents*

14 Johan Hoorn (VUA) *Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change*

15 Rainer Malik (UU) *CONAN: Text Mining in the Biomedical Domain*

16 Carsten Riggelsen (UU) *Approximation Methods for Efficient Learning of Bayesian Networks*

17 Stacey Nagata (UU) *User Assistance for Multitasking with Interruptions on a Mobile Device*

18 Valentin Zhizhkun (UvA) *Graph transformation for Natural Language Processing*

19 Birna van Riemsdijk (UU) *Cognitive Agent Programming: A Semantic Approach*

20 Marina Velikova (UvT) *Monotone models for prediction in data mining*

21 Bas van Gils (RUN) *Aptness on the Web*

22 Paul de Vrieze (RUN) *Fundaments of Adaptive Personalisation*

23 Ion Juvina (UU) *Development of Cognitive Model for Navigating on the Web*

24 Laura Hollink (VUA) *Semantic Annotation for Retrieval of Visual Resources*

25 Madalina Drugan (UU) *Conditional log-likelihood MDL and Evolutionary MCMC*

26 Vojkan Mihajlovic (UT) *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*

27 Stefano Bocconi (CWI) *Vox Populi: generating video documentaries from semantically annotated media repositories*

28 Borkur Sigurbjornsson (UvA) *Focused Information Access using XML Element Retrieval*

2007

1 Kees Leune (UvT) *Access Control and Service-Oriented Architectures*

2 Wouter Teepe (RUG) *Reconciling Information Exchange and Confidentiality: A Formal Approach*

3 Peter Mika (VUA) *Social Networks and the Semantic Web*

4 Jurriaan van Diggelen (UU) *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*

5 Bart Schermer (UL) *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*

6 Gilad Mishne (UvA) *Applied Text Analytics for Blogs*

7 Nataša Jovanović (UT) *To Whom It May Concern: Addressee Identification in Face-to-Face Meetings*

8 Mark Hoogendoorn (VUA) *Modeling of Change in Multi-Agent Organizations*

9 David Mobach (VUA) *Agent-Based Mediated Service Negotiation*

10 Huib Aldewereld (UU) *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*
11 Natalia Stash (TUe) *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*
12 Marcel van Gerven (RUN) *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*
13 Rutger Rienks (UT) *Meetings in Smart Environments: Implications of Progressing Technology*
14 Niek Bergboer (UM) *Context-Based Image Analysis*
15 Joyca Lacroix (UM) *NIM: a Situated Computational Memory Model*
16 Davide Grossi (UU) *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*
17 Theodore Charitos (UU) *Reasoning with Dynamic Networks in Practice*
18 Bart Orriens (UvT) *On the development an management of adaptive business collaborations*
19 David Levy (UM) *Intimate relationships with artificial partners*
20 Slinger Jansen (UU) *Customer Configuration Updating in a Software Supply Network*
21 Karianne Vermaas (UU) *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*
22 Zlatko Zlatev (UT) *Goal-oriented design of value and process models from patterns*
23 Peter Barna (TUe) *Specification of Application Logic in Web Information Systems*
24 Georgina Ramírez Camps (CWI) *Structural Features in XML Retrieval*
25 Joost Schalken (VUA) *Empirical Investigations in Software Process Improvement*

2008

1 Katalin Boer-Sorbán (EUR) *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*
2 Alexei Sharpanskykh (VUA) *On Computer-Aided Methods for Modeling and Analysis of Organizations*
3 Vera Hollink (UvA) *Optimizing hierarchical menus: a usage-based approach*
4 Ander de Keijzer (UT) *Management of Uncertain Data: towards unattended integration*
5 Bela Mutschler (UT) *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*
6 Arjen Hommersom (RUN) *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*
7 Peter van Rosmalen (OU) *Supporting the tutor in the design and support of adaptive e-learning*
8 Janneke Bolt (UU) *Bayesian Networks: Aspects of Approximate Inference*
9 Christof van Nimwegen (UU) *The paradox of the guided user: assistance can be counter-effective*
10 Wauter Bosma (UT) *Discourse oriented summarization*
11 Vera Kartseva (VUA) *Designing Controls for Network Organizations: A Value-Based Approach*
12 Jozsef Farkas (RUN) *A Semiotically Oriented Cognitive Model of Knowledge Representation*
13 Caterina Carraciolo (UvA) *Topic Driven Access to Scientific Handbooks*
14 Arthur van Bunningen (UT) *Context-Aware Querying: Better Answers with Less Effort*
15 Martijn van Otterlo (UT) *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*
16 Henriette van Vugt (VUA) *Embodied agents from a user's perspective*
17 Martin Op 't Land (TUD) *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*
18 Guido de Croon (UM) *Adaptive Active Vision*
19 Henning Rode (UT) *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*
20 Rex Arendsen (UvA) *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven*
21 Krisztian Balog (UvA) *People Search in the Enterprise*
22 Henk Koning (UU) *Communication of IT-Architecture*

23 Stefan Visscher (UU) *Bayesian network models for the management of ventilator-associated pneumonia*

24 Zharko Aleksovski (VUA) *Using background knowledge in ontology matching*

25 Geert Jonker (UU) *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*

26 Marijn Huijbregts (UT) *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*

27 Hubert Vogten (OU) *Design and Implementation Strategies for IMS Learning Design*

28 Ildiko Flesch (RUN) *On the Use of Independence Relations in Bayesian Networks*

29 Dennis Reidsma (UT) *Annotations and Subjective Machines: Of Annotators, Embodied Agents, Users, and Other Humans*

30 Wouter van Atteveldt (VUA) *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*

31 Loes Braun (UM) *Pro-Active Medical Information Retrieval*

32 Trung H. Bui (UT) *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*

33 Frank Terpstra (UvA) *Scientific Workflow Design: theoretical and practical issues*

34 Jeroen de Knijf (UU) *Studies in Frequent Tree Mining*

35 Ben Torben Nielsen (UvT) *Dendritic morphologies: function shapes structure*

2009

1 Rasa Jurgelenaite (RUN) *Symmetric Causal Independence Models*

2 Willem Robert van Hage (VUA) *Evaluating Ontology-Alignment Techniques*

3 Hans Stol (UvT) *A Framework for Evidence-based Policy Making Using IT*

4 Josephine Nabukenya (RUN) *Improving the Quality of Organisational Policy Making using Collaboration Engineering*

5 Sietse Overbeek (RUN) *Bridging Supply and Demand for Knowledge Intensive Tasks: Based on Knowledge, Cognition, and Quality*

6 Muhammad Subianto (UU) *Understanding Classification*

7 Ronald Poppe (UT) *Discriminative Vision-Based Recovery and Recognition of Human Motion*

8 Volker Nannen (VUA) *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*

9 Benjamin Kanagwa (RUN) *Design, Discovery and Construction of Service-oriented Systems*

10 Jan Wielemaker (UvA) *Logic programming for knowledge-intensive interactive applications*

11 Alexander Boer (UvA) *Legal Theory, Sources of Law & the Semantic Web*

12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin) *Operating Guidelines for Services*

13 Steven de Jong (UM) *Fairness in Multi-Agent Systems*

14 Maksym Korotkiy (VUA) *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*

15 Rinke Hoekstra (UvA) *Ontology Representation: Design Patterns and Ontologies that Make Sense*

16 Fritz Reul (UvT) *New Architectures in Computer Chess*

17 Laurens van der Maaten (UvT) *Feature Extraction from Visual Data*

18 Fabian Groffen (CWI) *Armada, An Evolving Database System*

19 Valentin Robu (CWI) *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*

20 Bob van der Vecht (UU) *Adjustable Autonomy: Controling Influences on Decision Making*

21 Stijn Vanderlooy (UM) *Ranking and Reliable Classification*

22 Pavel Serdyukov (UT) *Search For Expertise: Going beyond direct evidence*

23 Peter Hofgesang (VUA) *Modelling Web Usage in a Changing Environment*

24 Annerieke Heuvelink (VUA) *Cognitive Models for Training Simulations*

25 Alex van Ballegooij (CWI) *RAM: Array Database Management through Relational Mapping*

26 Fernando Koch (UU) *An Agent-Based Model for the Development of Intelligent Mobile Services*

27 Christian Glahn (OU) *Contextual Support of social Engagement and Reflection on the Web*

28 Sander Evers (UT) *Sensor Data Management with Probabilistic Models*

29 Stanislav Pokraev (UT) *Model-Driven Semantic Integration of Service-Oriented Applications*

30 Marcin Zukowski (CWI) *Balancing vectorized query execution with bandwidth-optimized storage*

31 Sofiya Katrenko (UvA) *A Closer Look at Learning Relations from Text*

32 Rik Farenhorst (VUA) *Architectural Knowledge Management: Supporting Architects and Auditors*

33 Khiet Truong (UT) *How Does Real Affect Affect Affect Recognition In Speech?*

34 Inge van de Weerd (UU) *Advancing in Software Product Management: An Incremental Method Engineering Approach*

35 Wouter Koelewijn (UL) *Privacy en Politiegegevens: Over geautomatiseerde normatieve informatie-uitwisseling*

36 Marco Kalz (OUN) *Placement Support for Learners in Learning Networks*

37 Hendrik Drachsler (OUN) *Navigation Support for Learners in Informal Learning Networks*

38 Riina Vuorikari (OU) *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*

39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin) *Service Substitution: A Behavioral Approach Based on Petri Nets*

40 Stephan Raaijmakers (UvT) *Multinomial Language Learning: Investigations into the Geometry of Language*

41 Igor Berezhnyy (UvT) *Digital Analysis of Paintings*

42 Toine Bogers (UvT) *Recommender Systems for Social Bookmarking*

43 Virginia Nunes Leal Franqueira (UT) *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*

44 Roberto Santana Tapia (UT) *Assessing Business-IT Alignment in Networked Organizations*

45 Jilles Vreeken (UU) *Making Pattern Mining Useful*

46 Loredana Afanasiev (UvA) *Querying XML: Benchmarks and Recursion*

2010

1 Matthijs van Leeuwen (UU) *Patterns that Matter*

2 Ingo Wassink (UT) *Work flows in Life Science*

3 Joost Geurts (CWI) *A Document Engineering Model and Processing Framework for Multimedia documents*

4 Olga Kulyk (UT) *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*

5 Claudia Hauff (UT) *Predicting the Effectiveness of Queries and Retrieval Systems*

6 Sander Bakkes (UvT) *Rapid Adaptation of Video Game AI*

7 Wim Fikkert (UT) *Gesture interaction at a Distance*

8 Krzysztof Siewicz (UL) *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*

9 Hugo Kielman (UL) *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*

10 Rebecca Ong (UL) *Mobile Communication and Protection of Children*

11 Adriaan Ter Mors (TUD) *The world according to MARP: Multi-Agent Route Planning*

12 Susan van den Braak (UU) *Sensemaking software for crime analysis*

13 Gianluigi Folino (RUN) *High Performance Data Mining using Bio-inspired techniques*

14 Sander van Splunter (VUA) *Automated Web Service Reconfiguration*

15 Lianne Bodenstaff (UT) *Managing Dependency Relations in Inter-Organizational Models*

16 Sicco Verwer (TUD) *Efficient Identification of Timed Automata, theory and practice*

17 Spyros Kotoulas (VUA) *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*

18 Charlotte Gerritsen (VUA) *Caught in the Act: Investigating Crime by Agent-Based Simulation*

19 Henriette Cramer (UvA) *People's Responses to Autonomous and Adaptive Systems*

20 Ivo Swartjes (UT) *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*

21 Harold van Heerde (UT) *Privacy-aware data management by means of data degradation*

22 Michiel Hildebrand (CWI) *End-user Support for Access to Heterogeneous Linked Data*

23 Bas Steunebrink (UU) *The Logical Structure of Emotions*

24 Zulfiqar Ali Memon (VUA) *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*

25 Ying Zhang (CWI) *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*

26 Marten Voulon (UL) *Automatisch contracteren*

27 Arne Koopman (UU) *Characteristic Relational Patterns*

28 Stratos Idreos (CWI) *Database Cracking: Towards Auto-tuning Database Kernels*

29 Marieke van Erp (UvT) *Accessing Natural History: Discoveries in data cleaning, structuring, and retrieval*

30 Victor de Boer (UvA) *Ontology Enrichment from Heterogeneous Sources on the Web*

31 Marcel Hiel (UvT) *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*

32 Robin Aly (UT) *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*

33 Teduh Dirgahayu (UT) *Interaction Design in Service Compositions*

34 Dolf Trieschnigg (UT) *Proof of Concept: Concept-based Biomedical Information Retrieval*

35 Jose Janssen (OU) *Paving the Way for Lifelong Learning: Facilitating competence development through a learning path specification*

36 Niels Lohmann (TUe) *Correctness of services and their composition*

37 Dirk Fahland (TUe) *From Scenarios to components*

38 Ghazanfar Farooq Siddiqui (VUA) *Integrative modeling of emotions in virtual agents*

39 Mark van Assem (VUA) *Converting and Integrating Vocabularies for the Semantic Web*

40 Guillaume Chaslot (UM) *Monte-Carlo Tree Search*

41 Sybren de Kinderen (VUA) *Needs-driven service bundling in a multi-supplier setting: the computational e3-service approach*

42 Peter van Kranenburg (UU) *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*

43 Pieter Bellekens (TUe) *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*

44 Vasilios Andrikopoulos (UvT) *A theory and model for the evolution of software services*

45 Vincent Pijpers (VUA) *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*

46 Chen Li (UT) *Mining Process Model Variants: Challenges, Techniques, Examples*

47 Jahn-Takeshi Saito (UM) *Solving difficult game positions*

48 Bouke Huurnink (UvA) *Search in Audiovisual Broadcast Archives*

49 Alia Khairia Amin (CWI) *Understanding and supporting information seeking tasks in multiple sources*

50 Peter-Paul van Maanen (VUA) *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*

51 Edgar Meij (UvA) *Combining Concepts and Language Models for Information Access*

2011

1 Botond Cseke (RUN) *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*

2 Nick Tinnemeier (UU) *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*

3 Jan Martijn van der Werf (TUe) *Compositional Design and Verification of Component-Based Information Systems*

4 Hado van Hasselt (UU) *Insights in Reinforcement Learning: Formal analysis and empirical evaluation of temporal-difference*

5 Base van der Raadt (VUA) *Enterprise Architecture Coming of Age: Increasing the Performance of an Emerging Discipline*

6 Yiwen Wang (TUe) *Semantically-Enhanced Recommendations in Cultural Heritage*

7 Yujia Cao (UT) *Multimodal Information Presentation for High Load Human Computer Interaction*

8 Nieske Vergunst (UU) *BDI-based Generation of Robust Task-Oriented Dialogues*

9 Tim de Jong (OU) *Contextualised Mobile Media for Learning*

10 Bart Bogaert (UvT) *Cloud Content Contention*

11 Dhaval Vyas (UT) *Designing for Awareness: An Experience-focused HCI Perspective*

12 Carmen Bratosin (TUe) *Grid Architecture for Distributed Process Mining*

13 Xiaoyu Mao (UvT) *Airport under Control. Multiagent Scheduling for Airport Ground Handling*

14 Milan Lovric (EUR) *Behavioral Finance and Agent-Based Artificial Markets*

15 Marijn Koolen (UvA) *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*

16 Maarten Schadd (UM) *Selective Search in Games of Different Complexity*

17 Jiyin He (UvA) *Exploring Topic Structure: Coherence, Diversity and Relatedness*

18 Mark Ponsen (UM) *Strategic Decision-Making in complex games*

19 Ellen Rusman (OU) *The Mind's Eye on Personal Profiles*

20 Qing Gu (VUA) *Guiding service-oriented software engineering: A view-based approach*

21 Linda Terlouw (TUD) *Modularization and Specification of Service-Oriented Systems*

22 Junte Zhang (UvA) *System Evaluation of Archival Description and Access*

23 Wouter Weerkamp (UvA) *Finding People and their Utterances in Social Media*

24 Herwin van Welbergen (UT) *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*

25 Syed Waqar ul Qounain Jaffry (VUA) *Analysis and Validation of Models for Trust Dynamics*

26 Matthijs Aart Pontier (VUA) *Virtual Agents for Human Communication: Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*

27 Aniel Bhulai (VUA) *Dynamic website optimization through autonomous management of design patterns*

28 Rianne Kaptein (UvA) *Effective Focused Retrieval by Exploiting Query Context and Document Structure*

29 Faisal Kamiran (TUe) *Discrimination-aware Classification*

30 Egon van den Broek (UT) *Affective Signal Processing (ASP): Unraveling the mystery of emotions*

31 Ludo Waltman (EUR) *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*

32 Nees-Jan van Eck (EUR) *Methodological Advances in Bibliometric Mapping of Science*

33 Tom van der Weide (UU) *Arguing to Motivate Decisions*

34 Paolo Turrini (UU) *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations*

35 Maaike Harbers (UU) *Explaining Agent Behavior in Virtual Training*

36 Erik van der Spek (UU) *Experiments in serious game design: a cognitive approach*

37 Adriana Burlutiu (RUN) *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference*

38 Nyree Lemmens (UM) *Bee-inspired Distributed Optimization*

39 Joost Westra (UU) *Organizing Adaptation using Agents in Serious Games*

40 Viktor Clerc (VUA) *Architectural Knowledge Management in Global Software Development*

41 Luan Ibraimi (UT) *Cryptographically Enforced Distributed Data Access Control*

42 Michal Sindlar (UU) *Explaining Behavior through Mental State Attribution*

43 Henk van der Schuur (UU) *Process Improvement through Software Operation Knowledge*

44 Boris Reuderink (UT) *Robust Brain-Computer Interfaces*

45 Herman Stehouwer (UvT) *Statistical Language Models for Alternative Sequence Selection*

46 Beibei Hu (TUD) *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work*

47 Azizi Bin Ab Aziz (VUA) *Exploring Computational Models for Intelligent Support of Persons with Depression*

48 Mark Ter Maat (UT) *Response Selection and Turn-taking for a Sensitive Artificial Listening Agent*

49 Andreea Niculescu (UT) *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality*

2012

1 Terry Kakeeto (UvT) *Relationship Marketing for SMEs in Uganda*

2 Muhammad Umair (VUA) *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models*

3 Adam Vanya (VUA) *Supporting Architecture Evolution by Mining Software Repositories*

4 Jurriaan Souer (UU) *Development of Content Management System-based Web Applications*

5 Marijn Plomp (UU) *Maturing Interorganisational Information Systems*

6 Wolfgang Reinhardt (OU) *Awareness Support for Knowledge Workers in Research Networks*

7 Rianne van Lambalgen (VUA) *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*

8 Gerben de Vries (UvA) *Kernel Methods for Vessel Trajectories*

9 Ricardo Neisse (UT) *Trust and Privacy Management Support for Context-Aware Service Platforms*

10 David Smits (TUe) *Towards a Generic Distributed Adaptive Hypermedia Environment*

11 J. C. B. Rantham Prabhakara (TUe) *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*

12 Kees van der Sluijs (TUe) *Model Driven Design and Data Integration in Semantic Web Information Systems*

13 Suleman Shahid (UvT) *Fun and Face: Exploring non-verbal expressions of emotion during playful interactions*

14 Evgeny Knutov (TUe) *Generic Adaptation Framework for Unifying Adaptive Web-based Systems*

15 Natalie van der Wal (VUA) *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes*

16 Fiemke Both (VUA) *Helping people by understanding them: Ambient Agents supporting task execution and depression treatment*

17 Amal Elgammal (UvT) *Towards a Comprehensive Framework for Business Process Compliance*

18 Eltjo Poort (VUA) *Improving Solution Architecting Practices*

19 Helen Schonenberg (TUe) *What's Next? Operational Support for Business Process Execution*

20 Ali Bahramisharif (RUN) *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing*

21 Roberto Cornacchia (TUD) *Querying Sparse Matrices for Information Retrieval*

22 Thijs Vis (UvT) *Intelligence, politie en veiligheidsdienst: verenigbare grootheden?*

23 Christian Muehl (UT) *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction*

24 Laurens van der Werff (UT) *Evaluation of Noisy Transcripts for Spoken Document Retrieval*

25 Silja Eckartz (UT) *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application*

26 Emile de Maat (UvA) *Making Sense of Legal Text*

27 Hayrettin Gurkok (UT) *Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games*

28 Nancy Pascall (UvT) *Engendering Technology Empowering Women*

29 Almer Tigelaar (UT) *Peer-to-Peer Information Retrieval*

30 Alina Pommeranz (TUD) *Designing Human-Centered Systems for Reflective Decision Making*

31 Emily Bagarukayo (RUN) *A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure*

32 Wietske Visser (TUD) *Qualitative multi-criteria preference representation and reasoning*

33 Rory Sie (OUN) *Coalitions in Cooperation Networks (COCOON)*

34 Pavol Jancura (RUN) *Evolutionary analysis in PPI networks and applications*

35 Evert Haasdijk (VUA) *Never Too Old To Learn: On-line Evolution of Controllers in Swarm- and Modular Robotics*

36 Denis Ssebugwawo (RUN) *Analysis and Evaluation of Collaborative Modeling Processes*

37 Agnes Nakakawa (RUN) *A Collaboration Process for Enterprise Architecture Creation*

38 Selmar Smit (VUA) *Parameter Tuning and Scientific Testing in Evolutionary Algorithms*

39 Hassan Fatemi (UT) *Risk-aware design of value and coordination networks*

40 Agus Gunawan (UvT) *Information Access for SMEs in Indonesia*

41 Sebastian Kelle (OU) *Game Design Patterns for Learning*

42 Dominique Verpoorten (OU) *Reflection Amplifiers in self-regulated Learning*

43 Anna Tordai (VUA) *On Combining Alignment Techniques*

44 Benedikt Kratz (UvT) *A Model and Language for Business-aware Transactions*

45 Simon Carter (UvA) *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation*

46 Manos Tsagkias (UvA) *Mining Social Media: Tracking Content and Predicting Behavior*

47 Jorn Bakker (TUe) *Handling Abrupt Changes in Evolving Time-series Data*

48 Michael Kaisers (UM) *Learning against Learning: Evolutionary dynamics of reinforcement learning algorithms in strategic interactions*

49 Steven van Kervel (TUD) *Ontologogy driven Enterprise Information Systems Engineering*

50 Jeroen de Jong (TUD) *Heuristics in Dynamic Sceduling: a practical framework with a case study in elevator dispatching*

2013

1 Viorel Milea (EUR) *News Analytics for Financial Decision Support*

2 Erietta Liarou (CWI) *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing*

3 Szymon Klarman (VUA) *Reasoning with Contexts in Description Logics*

4 Chetan Yadati (TUD) *Coordinating autonomous planning and scheduling*

5 Dulce Pumareja (UT) *Groupware Requirements Evolutions Patterns*

6 Romulo Goncalves (CWI) *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience*

7 Giel van Lankveld (UvT) *Quantifying Individual Player Differences*

8 Robbert-Jan Merk (VUA) *Making enemies: cognitive modeling for opponent agents in fighter pilot simulators*

9 Fabio Gori (RUN) *Metagenomic Data Analysis: Computational Methods and Applications*

10 Jeewanie Jayasinghe Arachchige (UvT) *A Unified Modeling Framework for Service Design*

11 Evangelos Pournaras (TUD) *Multi-level Reconfigurable Self-organization in Overlay Services*

12 Marian Razavian (VUA) *Knowledge-driven Migration to Services*

13 Mohammad Safiri (UT) *Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly*

14 Jafar Tanha (UvA) *Ensemble Approaches to Semi-Supervised Learning Learning*

15 Daniel Hennes (UM) *Multiagent Learning: Dynamic Games and Applications*

16 Eric Kok (UU) *Exploring the practical benefits of argumentation in multi-agent deliberation*

17 Koen Kok (VUA) *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*

18 Jeroen Janssens (UvT) *Outlier Selection and One-Class Classification*

19 Renze Steenhuizen (TUD) *Coordinated Multi-Agent Planning and Scheduling*

20 Katja Hofmann (UvA) *Fast and Reliable Online Learning to Rank for Information Retrieval*

21 Sander Wubben (UvT) *Text-to-text generation by monolingual machine translation*

22 Tom Claassen (RUN) *Causal Discovery and Logic*

23 Patricio de Alencar Silva (UvT) *Value Activity Monitoring*

24 Haitham Bou Ammar (UM) *Automated Transfer in Reinforcement Learning*

25 Agnieszka Anna Latoszek-Berendsen (UM) *Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System*

26 Alireza Zarghami (UT) *Architectural Support for Dynamic Homecare Service Provisioning*

27 Mohammad Huq (UT) *Inference-based Framework Managing Data Provenance*

28 Frans van der Sluis (UT) *When Complexity becomes Interesting: An Inquiry into the Information eXperience*

29 Iwan de Kok (UT) *Listening Heads*

30 Joyce Nakatumba (TUe) *Resource-Aware Business Process Management: Analysis and Support*

31 Dinh Khoa Nguyen (UvT) *Blueprint Model and Language for Engineering Cloud Applications*

32 Kamakshi Rajagopal (OUN) *Networking For Learning: The role of Networking in a Lifelong Learner's Professional Development*

33 Qi Gao (TUD) *User Modeling and Personalization in the Microblogging Sphere*

34 Kien Tjin-Kam-Jet (UT) *Distributed Deep Web Search*

35 Abdallah El Ali (UvA) *Minimal Mobile Human Computer Interaction*

36 Than Lam Hoang (TUe) *Pattern Mining in Data Streams*

37 Dirk Börner (OUN) *Ambient Learning Displays*

38 Eelco den Heijer (VUA) *Autonomous Evolutionary Art*

39 Joop de Jong (TUD) *A Method for Enterprise Ontology based Design of Enterprise Information Systems*

40 Pim Nijssen (UM) *Monte-Carlo Tree Search for Multi-Player Games*

41 Jochem Liem (UvA) *Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning*

42 Léon Planken (TUD) *Algorithms for Simple Temporal Reasoning*

43 Marc Bron (UvA) *Exploration and Contextualization through Interaction and Concepts*

2014

1 Nicola Barile (UU) *Studies in Learning Monotone Models from Data*

2 Fiona Tuliyano (RUN) *Combining System Dynamics with a Domain Modeling Method*

3 Sergio Raul Duarte Torres (UT) *Information Retrieval for Children: Search Behavior and Solutions*

4 Hanna Jochmann-Mannak (UT) *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation*

5 Jurriaan van Reijsen (UU) *Knowledge Perspectives on Advancing Dynamic Capability*

6 Damian Tamburri (VUA) *Supporting Networked Software Development*

7 Arya Adriansyah (TUe) *Aligning Observed and Modeled Behavior*

8 Samur Araujo (TUD) *Data Integration over Distributed and Heterogeneous Data Endpoints*

9 Philip Jackson (UvT) *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*

10 Ivan Salvador Razo Zapata (VUA) *Service Value Networks*

11 Janneke van der Zwaan (TUD) *An Empathic Virtual Buddy for Social Support*

12 Willem van Willigen (VUA) *Look Ma, No Hands: Aspects of Autonomous Vehicle Control*

13 Arlette van Wissen (VUA) *Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains*

14 Yangyang Shi (TUD) *Language Models With Meta-information*

15 Natalya Mogles (VUA) *Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare*

16 Krystyna Milian (VUA) *Supporting trial recruitment and design by automatically interpreting eligibility criteria*

17 Kathrin Dentler (VUA) *Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability*

18 Mattijs Ghijsen (UvA) *Methods and Models for the Design and Study of Dynamic Agent Organizations*

19 Vinicius Ramos (TUe) *Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support*

20 Mena Habib (UT) *Named Entity Extraction and Disambiguation for Informal Text: The Missing Link*

21 Kassidy Clark (TUD) *Negotiation and Monitoring in Open Environments*

22 Marieke Peeters (UU) *Personalized Educational Games: Developing agent-supported scenario-based training*

23 Eleftherios Sidirourgos (UvA/CWI) *Space Efficient Indexes for the Big Data Era*

24 Davide Ceolin (VUA) *Trusting Semi-structured Web Data*

25 Martijn Lappenschaar (RUN) *New network models for the analysis of disease interaction*

26 Tim Baarslag (TUD) *What to Bid and When to Stop*

27 Rui Jorge Almeida (EUR) *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*

28 Anna Chmielowiec (VUA) *Decentralized k-Clique Matching*

29 Jaap Kabbedijk (UU) *Variability in Multi-Tenant Enterprise Software*

30 Peter de Cock (UvT) *Anticipating Criminal Behaviour*

31 Leo van Moergestel (UU) *Agent Technology in Agile Multiparallel Manufacturing and Product Support*

32 Naser Ayat (UvA) *On Entity Resolution in Probabilistic Data*

33 Tesfa Tegegne (RUN) *Service Discovery in eHealth*

34 Christina Manteli (VUA) *The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems*

35 Joost van Ooijen (UU) *Cognitive Agents in Virtual Worlds: A Middleware Design Approach*

36 Joos Buijs (TUe) *Flexible Evolutionary Algorithms for Mining Structured Process Models*

37 Maral Dadvar (UT) *Experts and Machines United Against Cyberbullying*

38 Danny Plass-Oude Bos (UT) *Making brain-computer interfaces better: improving usability through post-processing*

39 Jasmina Maric (UvT) *Web Communities, Immigration, and Social Capital*

40 Walter Omona (RUN) *A Framework for Knowledge Management Using ICT in Higher Education*

41 Frederic Hogenboom (EUR) *Automated Detection of Financial Events in News Text*

42 Carsten Eijckhof (CWI/TUD) *Contextual Multidimensional Relevance Models*

43 Kevin Vlaanderen (UU) *Supporting Process Improvement using Method Increments*

44 Paulien Meesters (UvT) *Intelligent Blauw: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden*

45 Birgit Schmitz (OUN) *Mobile Games for Learning: A Pattern-Based Approach*

46 Ke Tao (TUD) *Social Web Data Analytics: Relevance, Redundancy, Diversity*

47 Shangsong Liang (UvA) *Fusion and Diversification in Information Retrieval*

2015

1 Niels Netten (UvA) *Machine Learning for Relevance of Information in Crisis Response*

2 Faiza Bukhsh (UvT) *Smart auditing: Innovative Compliance Checking in Customs Controls*

3 Twan van Laarhoven (RUN) *Machine learning for network data*

4 Howard Spoelstra (OUN) *Collaborations in Open Learning Environments*

5 Christoph Bösch (UT) *Cryptographically Enforced Search Pattern Hiding*

6 Farideh Heidari (TUD) *Business Process Quality Computation: Computing Non-Functional Requirements to Improve Business Processes*

7 Maria-Hendrike Peetz (UvA) *Time-Aware Online Reputation Analysis*

8 Jie Jiang (TUD) *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions*

9 Randy Klaassen (UT) *HCI Perspectives on Behavior Change Support Systems*

10 Henry Hermans (OUN) *OpenU: design of an integrated system to support lifelong learning*

11 Yongming Luo (TUe) *Designing algorithms for big graph datasets: A study of computing bisimulation and joins*

12 Julie M. Birkholz (VUA) *Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks*

13 Giuseppe Procaccianti (VUA) *Energy-Efficient Software*

14 Bart van Straalen (UT) *A cognitive approach to modeling bad news conversations*

15 Klaas Andries de Graaf (VUA) *Ontology-based Software Architecture Documentation*

16 Changyun Wei (UT) *Cognitive Coordination for Cooperative Multi-Robot Teamwork*

17 André van Cleeff (UT) *Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs*

18 Holger Pirk (CWI) *Waste Not, Want Not!: Managing Relational Data in Asymmetric Memories*

19 Bernardo Tabuenca (OUN) *Ubiquitous Technology for Lifelong Learners*

20 Loïs Vanhée (UU) *Using Culture and Values to Support Flexible Coordination*

21 Sibren Fetter (OUN) *Using Peer-Support to Expand and Stabilize Online Learning*

22 Zhemin Zhu (UT) *Co-occurrence Rate Networks*

23 Luit Gazendam (VUA) *Cataloguer Support in Cultural Heritage*

24 Richard Berendsen (UvA) *Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation*

25 Steven Woudenberg (UU) *Bayesian Tools for Early Disease Detection*

26 Alexander Hogenboom (EUR) *Sentiment Analysis of Text Guided by Semantics and Structure*

27 Sándor Héman (CWI) *Updating compressed colomn stores*

28 Janet Bagorogoza (TiU) *Knowledge Management and High Performance: The Uganda Financial Institutions Model for HPO*

29 Hendrik Baier (UM) *Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains*

30 Kiavash Bahreini (OU) *Real-time Multimodal Emotion Recognition in E-Learning*

31 Yakup Koç (TUD) *On the robustness of Power Grids*

32 Jerome Gard (UL) *Corporate Venture Management in SMEs*

33 Frederik Schadd (TUD) *Ontology Mapping with Auxiliary Resources*

34 Victor de Graaf (UT) *Gesocial Recommender Systems*

35 Jungxao Xu (TUD) *Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction*

2016

1 Syed Saiden Abbas (RUN) *Recognition of Shapes by Humans and Machines*

2 Michiel Christiaan Meulendijk (UU) *Optimizing medication reviews through decision support: prescribing a better pill to swallow*

3 Maya Sappelli (RUN) *Knowledge Work in Context: User Centered Knowledge Worker Support*

4 Laurens Rietveld (VU) *Publishing and Consuming Linked Data*

5 Evgeny Sherkhonov (UVA) *Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers*

6 Michel Wilson (TUD) *Robust scheduling in an uncertain environment*

7 Jeroen de Man (VU) *Measuring and modeling negative emotions for virtual training*

8 Matje van de Camp (TiU) *A Link to the Past: Constructing Historical Social Networks from Unstructured Data*

9 Archana Nottamkandath (VU) *Trusting Crowdsourced Information on Cultural Artefacts*

10 George Karafotias (VUA) *Parameter Control for Evolutionary Algorithms*

11 Anne Schuth (UVA) *Search Engines that Learn from Their Users*

12 Max Knobbout (UU) *Logics for Modelling and Verifying Normative Multi-Agent Systems*

13 Nana Baah Gyan (VU) *The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach*

14 Ravi Khadka (UU) *Revisiting Legacy Software System Modernization*

15 Steffen Michels (RUN) *Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments*

16 Guangliang Li (UVA) *Socially Intelligent Autonomous Agents that Learn from Human Reward*

17 Berend Weel (VU) *Towards Embodied Evolution of Robot Organisms*

18 Albert Meroño Peñuela (VU) *Refining Statistical Data on the Web*

19 Julia Efremova (Tu) *e)Mining Social Structures from Genealogical Data*

20 Daan Odijk (UVA) *Context & Semantics in News & Web Search*

21 Alejandro Moreno Célleri (UT) *From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground*

22 Grace Lewis (VU) *Software Architecture Strategies for Cyber-Foraging Systems*

23 Fei Cai (UVA) *Query Auto Completion in Information Retrieval*

24 Brend Wanders (UT) *Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach*

25 Julia Kiseleva (TU) *e)Using Contextual Information to Understand Searching and Browsing Behavior*

26 Dilhan Thilakarathne (VU) *In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains*

27 Wen Li (TUD) *Understanding Geo-spatial Information on Social Media*

28 Mingxin Zhang (TUD) *Large-scale Agent-based Social Simulation - A study on epidemic prediction and control*

29 Nicolas Höning (TUD) *Peak reduction in decentralised electricity systems -Markets and prices for flexible planning*

30 Ruud Mattheij (UvT) *The Eyes Have It*

31 Mohammad Khelghati (UT) *Deep web content monitoring*

32 Eelco Vriezekolk (UT) *Assessing Telecommunication Service Availability Risks for Crisis Organisations*

33 Peter Bloem (UVA) *Single Sample Statistics, exercises in learning from just one example*

34 Dennis Schunselaar (TUE) *Title: Configurable Process Trees: Elicitation, Analysis, and Enactment*

35 Zhaochun Ren (UVA) *Monitoring Social Media: Summarization, Classification and Recommendation*

36 Daphne Karreman (UT) *Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies*

37 Giovanni Sileno (UvA) *Aligning Law and Action - a conceptual and computational inquiry*

38 Andrea Minuto (UT) *Materials that Matter - Smart Materials meet Art & Interaction Design*

39 Merijn Bruijnes (UT) *Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect*

40 Christian Detweiler (TUD) *Accounting for Values in Design*

41 Thomas King (TUD) *Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance*

42 Spyros Martzoukos (UVA) *Combinatorial and Compositional Aspects of Bilingual Aligned Corpora*

43 Saskia Koldijk (RUN) *Context-Aware Support for Stress Self-Management: From Theory to Practice*

44 Thibault Sellam (UVA) *Automatic Assistants for Database Exploration*

45 Bram van de Laar (UT) *Experiencing Brain-Computer Interface Control*

46 Jorge Gallego Perez (UT) *Robots to Make you Happy*

47 Christina Weber (UL) *Real-time foresight - Preparedness for dynamic innovation networks*

48 Tanja Buttler (TUD) *Collecting Lessons Learned*

49 Gleb Polevoy (TUD) *The title: Participation and Interaction in Projects. A Game-Theoretic Analysis*

50 Yan Wang (UVT) *The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains*

2017

1 Jan-Jaap Oerlemans (UL) *Investigating Cybercrime*

2 Sjoerd T. Timmer (UU) *Designing and Understanding Forensic Bayesian Networks using Argumentation*