Bridging the Gap Between Informal Sense-Making Tools and Formal Systems

Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

Remi Wieten

Bridging the Gap Between Informal Sense-Making Tools and Formal Systems

Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

Remi Wieten

Printing: Ridderprint | www.ridderprint.nl
ISBN: 978-90-393-7389-7
doi:10.33540/662

SIKS Dissertation Series No. 2021-18
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Bridging the Gap Between Informal Sense-Making Tools and Formal Systems

Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

Het Overbruggen van de Kloof tussen Informele Sense-Making Tools en Formele Systemen - Ondersteuning van de Constructie van Bayesiaanse Netwerken en Argumentatieraamwerken

(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. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 8 september 2021 des ochtends te 10.15 uur

door

Gerhardus Marinus Wieten

geboren op 7 januari 1990
te IJsselmuiden

Promotoren:	Prof.dr. F.J. Bex
	Prof.dr.mr. H. Prakken
Copromotor:	Dr. S. Renooij

Contents

Chapter 1

Introduction

Reasoning with uncertainty and evidence plays an important role in decision-making and problem solving in many domains, including medicine, engineering, forensics, intelligence and law. To aid domain experts in performing their tasks, various tools and techniques exist that allow them to make sense of a problem, including informal graph-based sense-making tools such as mind maps [Okada et al., 2014][1], argument diagrams [Bex et al., 2003, 2013; Okada et al., 2014], and Wigmore charts [Wigmore, 1913], which allow for structuring and visualising the problem and the user's reasoning involved in solving it. Formal systems for reasoning about evidence have been proposed in artificial intelligence (AI), including argumentation formalisms (see e.g. Prakken [2018a] for an overview) and probabilistic models such as Bayesian networks (BNs) [Pearl, 1988b; Jensen and Nielsen, 2007]. These systems allow for automated reasoning and computation, and thus support experts in formally evaluating their problem. In practice (e.g. in law [Prakken et al., 2020]) both sense-making tools and instantiations of formal AI systems are used by domain experts, as both have their merits and limitations.

Graph-based *sense-making tools* such as mind maps, Wigmore charts, and argument diagrams support domain experts in structuring and analysing a mass of evidence, thereby allowing them to obtain an overview of the problem under consideration [Okada et al., 2014; van den Braak, 2010]. In analyses performed using these tools, inferences, or reasoning steps, made between claims are captured and visualised. A limitation of these tools is that they are only intended for visualising the user's reasoning and thinking: they do not allow for automated reasoning or computations. Hence, while these tools are suited for creating an initial sketch of a problem, they do not support experts in formally evaluating the problem.

Formal systems for reasoning about evidence have been proposed in the field of AI (see e.g. Verheij et al. [2016] for an overview), which in contrast with aforementioned sense-making tools are precisely defined in terms of their notation and semantics and allow for automated inference, formal evaluation and computation.

[1] See also Master's research [Timmers, 2017].

The inner workings of formal systems are well-known, and conditions can be studied under which instantiations of these systems are guaranteed to be well-behaved and satisfy desirable properties. However, domain experts typically do not have the expertise to construct instantiations of formal AI systems, and especially in data-poor domains their construction therefore needs to be done mostly manually by an AI expert through a knowledge elicitation procedure in consultation with the domain expert, which is a difficult, time-consuming and error-prone task.

Accordingly, in this thesis we aim to facilitate the construction of instantiations of formal AI systems to allow domain experts to formally evaluate their problems. To this end, we study how domain knowledge captured in an initial sketch of a problem expressed using a sense-making tool can be exploited to guide the construction of formal representations within AI systems. We focus on two types of formal systems proposed in AI, namely probabilistic models, more specifically Bayesian networks (BNs) [Pearl, 1988b; Jensen and Nielsen, 2007], and computational argumentation [Prakken, 2018a]. Argumentation is particularly suited for adversarial settings, where arguments for and against claims are constructed from evidence. Arguments can then be formally evaluated on their acceptability. Probabilistic models such as BNs allow for reasoning with numeric uncertainty such as statistical and probabilistic information, thereby allowing experts to evaluate their problem in a probabilistic manner by computing probabilities of interest.

In the literature, formalisms have also been proposed that combine argumentation and probabilities, e.g. [Hunter, 2013; Prakken, 2018b; Hunter et al., 2020; Li et al., 2012; Dung and Thang, 2010]. In Chapters 4 and 5 of this thesis, we instead focus on non-probabilistic argumentation approaches and BNs, as these are widely different approaches that allow problems to be evaluated in a distinctly different manner. Probabilistic and argumentation models face different challenges in construction, and therefore these models are considered suitable for illustrating how the construction of formal representations within AI systems can be facilitated. Narrative models for reasoning about evidence (e.g. Pennington and Hastie [1993]; Wagenaar et al. [1993]) have also been proposed, in which stories, i.e. coherent sequences of events, are used to explain the evidence; however, these models are generally less formally developed than probabilistic and argumentation models (see Verheij et al. [2016]), and are therefore less suited for current purposes.

We will now first provide examples of analyses performed using sense-making tools, after which formal systems are further discussed. We then proceed with a formulation of research questions that address the difficulties regarding the construction of instantiations of these formal systems. The examples considered in the following section are taken from the legal and forensic domains. Throughout this thesis, these domains are often taken as an example, as in recent years probabilistic models and computational models of argument are increasingly being developed and used in these domains, alongside more informal sense-making tools such as mind maps (see e.g. Prakken et al. [2020]). These domains therefore serve as a suitable application domain for the approaches and formalisms proposed throughout this thesis; however, we will argue that these approaches are not limited to legal and forensic applications but are in fact applicable to any domain.

1.1 Examples of analyses performed using sense-making tools

We now present examples of analyses performed using two sense-making tools familiar to many legal experts, namely Wigmore charts [Wigmore, 1913] and mind maps [Okada et al., 2014][2]. These examples serve to illustrate the manner in which knowledge and reasoning is typically expressed using such tools and to further illustrate their merits and limitations. They are also used as running examples throughout this thesis.

1.1.1 Example of an analysis performed in a Wigmore chart

First, Wigmore charts are considered, which are diagrams familiar to many legal experts in which symbols indicating hypotheses and claims are joined by lines representing relations between these hypotheses and claims. Wigmore charts were introduced by John Henry Wigmore [1913] and were further developed and studied from an academic perspective by the so-called 'New Evidence Theorists' including Anderson, Schum and Twining [Anderson et al., 2005], who provided a modernised, more user-friendly version of Wigmore's charting method. Wigmore introduced his method as an aid in structuring a mass of evidence in a legal case in detailed way. An important aspect of his method is that it not only used for expressing supporting reasons but also for revealing possible sources of doubt. Wigmore's charts can be considered a precursor of diagrams in argument diagramming software tools [Buckingham Shum, 2003] including Araucaria [Reed and Rowe, 2007], as well as a forerunner of instantiations of formal argumentation systems [Bex et al., 2003] (see Section 1.2.1). Wigmore's method has been used to analyse complex legal cases, including the Dutch ballpoint case [Dingley, 1999] and the well-known Sacco and Vanzetti case [Kadane and Schum, 1996].

An example of one of Wigmore's original charts is depicted in Figure 1.1, which is taken from Kadane and Schum [1996, p. 69], who adapted it from Wigmore [1913, pp. 759–765]. Each circle or square in the diagram represents a unique claim, where squares denote testimonies and circles denote circumstantial claims. An arc in the chart indicates that a claim is offered as *'... evidencing, or explaining, or corroborating ...'* [Wigmore, 1913, p. 752]; hence, arcs can be regarded as indicating which claims are inferred from each other.

Example 1. *We now consider the Wigmore chart depicted in Figure 1.1 in more detail. In short, in the case defendant Oliver Hatchett (H.) was accused of murdering*

[2]See also Master's research [Timmers, 2017].

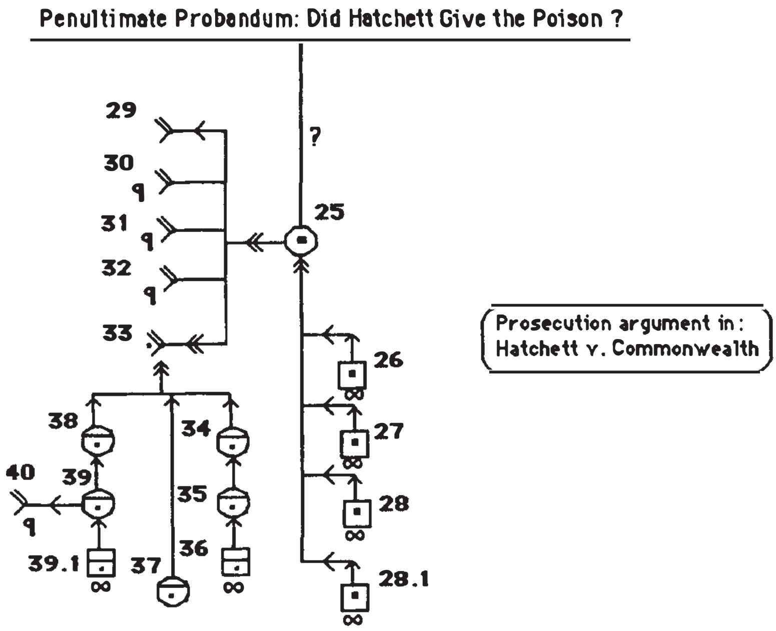


Figure 1.1: Wigmore chart for the Hatchett case, taken from Kadane and Schum [1996, p. 69], who adapted it from Wigmore [1913, pp. 759–765].

Moses Young (Y.) by giving him whiskey laced with poison. One of the penultimate claims (or probanda) in the case concerned whether or not Hatchett gave poison to the victim. The relevant elements of the key list, which indicates for every number in the chart to which claim it corresponds, are enumerated below. These are adapted from Wigmore's original key list [Wigmore, 1913, pp. 759–765]:

- *25* *Y. died apparently in good health, within three hours after drinking defendant's whiskey.*
- *26-28.1* *Different witness testimonies concerning Y.'s time of death.*
- *29* *Neither H. nor his father are shown to have possessed any strychnine to put in the drink.*
- *30* *Y. might have died by colic, from which he had often suffered.*
- *31* *Y. might have died from the former injury in his side.*
- *32* *Y. might have died of ptomaine poisoning in supper-food.*
- *33* *Y. might have died from poison put in his supper-food by third person; the only third person having access being Sallie his wife.*

The arcs between claims $26 - 28.1$ *and claim* 25 *indicate inferences from the testimonies to the claim to which is testified. Possible hypotheses are then proposed that explain* 25*, namely* $29 - 33$*, indicated by open triangles in the chart. Here, claims* $30 - 32$ *are inferred but not further backed up by further information, which is indicated by paragraph symbols underneath these claims. In the bottom-left part of the* *chart, claims are depicted that further support or oppose claim* 33. *For instance,* 38 *indicates that* 'Sallie had a plan to kill Y.'*, which is concluded from* 39*:* 'Sallie had received strychnine from H.C. three weeks before, with instructions to put it in Y.'s coffee or food.'*, which follows from testimony to this claim (*39.1*). Claim* 40 *weakens this chain of inferences, as it states that* 'Sallie's failure to use it during those three weeks' opportunity indicates abandonment of her design.'. □

1.1.2 Example of an analysis performed using a mind mapping tool

Next, we present an example of an analysis performed using a mind mapping tool [Okada et al., 2014], which is an example of a tool typically used by domain experts, for instance in crime analysis[3]. A mind map usually takes the shape of a diagram in which hypotheses and claims are represented by boxes and underlined text, and undirected edges symbolise relations between these hypotheses and claims. An example is depicted in Figure 1.2, which is based on a standard template used by the Dutch police for criminal cases involving the suspicious death of a person. In this template, four high-level hypotheses concerning the person's death are considered, namely natural death, suicide, accident and murder. For each of the four hypotheses, the crime analyst tries to construct different scenario-elements by answering the seven W questions: (1) what happened exactly?; (2) where did the event take place?; (3) when did the event take place?; (4) who were involved in the event?; (5) why did the event take place?; (6) which items were involved in the event (with)?; (7) how did the event take place (in which way)? Here, the answers to the 'Why' question are typically connected to the answers to the 'Who' question to link the different motives to the different persons. The crime analyst then uses evidence to support or oppose the different scenario-elements, indicated in the mind map by plus and minus symbols, respectively. Compared to Wigmore charts, which offer a wide range of symbols and arcs to allow users to be expressive and more precise in modelling legal reasoning, mind maps are less precise and are used to obtain an overview of different possible alternative scenarios.

Example 2. *An example of a partially filled in mind map is depicted in Figure 1.2, which also serves as our running example for Chapters 3, 4 and 5. In this example case, adapted from Bex [2011], the high-level hypotheses 'Murder' and 'Accident' are considered; for illustration purposes the details of the case have been changed. The case concerns the murder of Leo de Jager, which took place in the small Dutch town of Anjum. Leo's body was found on the property of Marjan van der E.; we are interested in her involvement in the murder, as well as Leo's cause of death. First, Marjan's involvement is considered. As a police report (*police_report*) indicates that Leo's body was found on Marjan's property, the claim* marjan_murdered_leo *is added as an answer to the 'Who' question for the high-level hypothesis 'Murder'. By*

[3]See Master's research [Timmers, 2017].

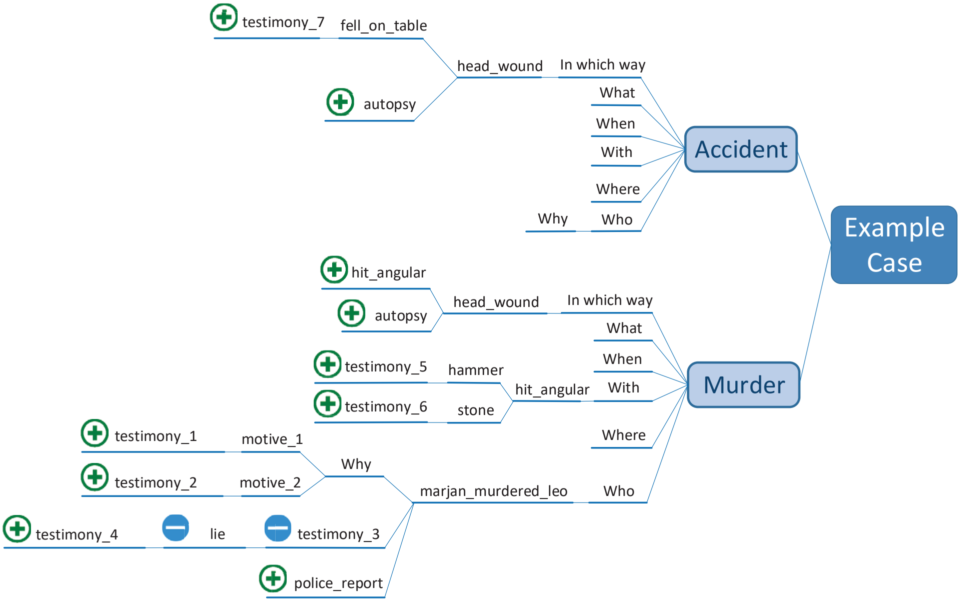


Figure 1.2: Example of a partially filled in mind map.

means of a plus symbol and an undirected edge connecting the evidence to the claim, it is indicated that the police report supports the claim that Marjan murdered Leo. Possible motives (motive_1 and motive_2) are provided as to why Marjan may have wanted to murder Leo, which are connected to the 'Why' question via undirected edges. Claims testimony_1 and testimony_2 support these two motives, indicated by the plus symbols connected to these claims. In her testimony (testimony_3), Marjan denied any involvement in the murder of Leo, which is indicated by a minus symbol. This opposes the claim that Marjan murdered Leo. Further testimony (testimony_4) indicates that Marjan had reason to lie when giving her testimony (lie). By means of a minus symbol and an undirected edge connecting lie to testimony_3, it is indicated that this claim weakens the inference step from her testimony to the claim that she did not murder Leo.

Next, Leo's cause of death is considered, where first high-level hypothesis 'Murder' is examined. According to witness testimony (testimony_5), Leo was hit with a hammer (hammer); however, according to another testimony (testimony_6), Leo was hit with a stone (stone). Claims hammer and stone are connected via undirected edges to hit_angular, which indicates that hammers and stones can generally be considered to be angular. In turn, claim hit_angular is connected to the 'With' question to indicate that it provides an answer to this question. As an answer to the 'In which way' question, it is indicated that Leo died because of a head wound (head_wound), which is again supported by the claim that Leo was hit with an angular object (hit_angular). An autopsy report (autopsy) further supports claim head_wound.

High-level hypothesis 'Accident' provides a competing alternative explanation for head_wound. *As an answer to the 'In which way' question, it is again indicated that Leo died because of a head wound and that this claim is supported by* autopsy*; however, in contrast to the answer to this question for high-level hypothesis 'Murder', it is indicated that the head wound was caused because Leo fell on a table by accident (*fell_on_table*), a claim supported by further testimony (*testimony_7*).* □

1.1.3 Limitations of sense-making tools

Sense-making tools such as mind maps and Wigmore charts support domain experts in structuring and analysing a mass of evidence and in revealing possible sources of doubt. As mentioned earlier, a limitation of these tools is that they are only intended for visualising knowledge and reasoning, and do not allow for automated reasoning or formal evaluation. Hence, evaluation of the problem under consideration takes place in the mind of the domain expert, which has its limitations such as limited working memory [Pirolli and Card, 2005] and which makes it difficult for others to understand how conclusions are precisely reached upon evaluation. Furthermore, the assumptions of domain experts underlying their analyses are typically not explicitly stated and not all information involved in performing these analyses is explicitly recorded; hence, such analyses may be misinterpreted by (other) experts involved in solving the problem. In Chapter 3 these examples are revisited in the light of our conceptual analysis of reasoning about evidence (Section 2.1) to illustrate this.

1.2 Formal systems for reasoning about evidence

We now further discuss formal systems for reasoning about evidence [Verheij et al., 2016]. In Sections 1.2.1 and 1.2.2 we provide a general description of argumentation approaches and probabilistic approaches, respectively, where we highlight their differences. Argumentation formalisms and BNs each have their individual merits and limitations, and depending on the specific application context one may be preferred over the other. Technical details of these approaches are provided in Chapter 2. In Sections 1.2.3 and 1.2.4 narrative models and connections between the different formal systems are also briefly discussed to put our work into context.

1.2.1 Argumentation models

Argumentation approaches focus on how conclusions are based on the evidence. Such approaches are particularly suited for adversarial settings such as the legal domain, where arguments for and against claims are constructed from evidence. As noted by Bex and colleagues [2003], Wigmore's charting method can be considered a forerunner of formal rule-based argumentation systems, which are the types of systems under consideration in this thesis (e.g. Modgil and Prakken [2018]; Dung et al. [2009]; Verheij [2003]; Vreeswijk [1997]). In such systems, arguments are iteratively constructed from a knowledge base by chaining so-called inference rules. Going back to the work of John Pollock [1987], these inference rules can be either strict or defeasible, where strict rules hold without exception and where for defeasible rules exceptional circumstances can be provided under which the rule may not hold. Inferences performed with strict and defeasible rules are then certain, respectively, uncertain. Arguments can be attacked by other arguments in various ways (e.g. rebuttal and undercutting attack [Pollock, 1987, 1995]).

Similar to Wigmore charts, formal argumentation approaches allow for describing how specific conclusions are supported starting from the evidence, and for revealing possible sources of doubt. However, unlike Wigmore charts, they do so in a formal, precisely defined manner. Moreover, argumentation systems allow for actually performing inference instead of only representing knowledge and reasoning. Another important difference between Wigmore's charts and aforementioned argumentation approaches is that they allow for formal evaluation. In particular, given a set of arguments and a binary attack relation over these arguments, arguments can be evaluated using Dung's [1995] argumentation semantics, where it can be calculated which arguments are accepted and which are rejected. In rule-based argumentation, efforts are made to define systems in such a way that the result of argumentation is guaranteed to be well-behaved (e.g. Caminada and Amgoud [2007]). Argumentation techniques have found applications in domains such as law, medicine and tutoring (see e.g. Modgil et al. [2013] for an overview), for instance in supporting legal reasoning about evidence [Bex et al., 2003; Reed et al., 2007], persuading human agents to change their stance on a subject [Chalaguine and Hunter, 2020], and in fraud inquiry at the Dutch National Police [Odekerken et al., 2020].

Formal argumentation approaches are typically qualitative; until recently, there was no emphasis in argumentation approaches on incorporating graded uncertainty. In probabilistic models of argument, probabilities are used to express grades of uncertainty in or about the arguments [Hunter, 2013]; in our chapter on related work (Section 8.3) several approaches to probabilistic argumentation will be reviewed. In Chapter 4 we focus on strictly qualitative argumentation.

1.2.2 Probabilistic models

Probabilistic approaches allow for reasoning with numeric uncertainty such as statistical and probabilistic information, which can increasingly be expected in many domains due to the rise of big-data analytics. For instance, in forensic science methods such as DNA analysis are increasingly being used, where the significance of a DNA match is established by calculating the probability of a match if the suspect was not the donor of the trace by consulting a DNA database. More generally, a way to probabilistically evaluate evidence is by calculating so-called likelihood ratios, which are ratios of observing the evidence under two mutually exclusive hypotheses. The prior probabilities of these hypotheses can then be 'updated' using Bayes' theorem by multiplying the ratio of these prior probabilities by the likelihood ratio to obtain a ratio of posterior probabilities of these hypotheses given the evidence. The likelihood ratio approach to evidential reasoning has found applications in domains such as medicine, forensics and law (see e.g. Fenton et al. [2016]).

The Bayesian network (BN) formalism is a particularly powerful formalism for probabilistic reasoning [Pearl, 1988b; Jensen and Nielsen, 2007] that has found applications in many fields where uncertainty and evidence plays a role [Fenton and Neil, 2012]. For instance, in recent years legal and forensic experts have increasingly developed and used BNs for the interpretation of different types of forensic trace evidence [Taroni et al., 2014], such as glass fragments, finger marks and DNA traces, traces found on adhesive tapes [Wieten et al., 2015], as well as entire legal cases [Fenton et al., 2016]. A BN consists of a graph, which captures the probabilistic independence relation among variables relevant to the domain, and locally specified (conditional) probability distributions that collectively describe a joint probability distribution. The required conditional probabilities can, for instance, be elicited from domain experts as degrees of belief or they can be estimated from data sets by calculating (frequency) statistics [Druzdzel and van der Gaag, 2000]. A BN is generally used for *probabilistic inference* [Jensen and Nielsen, 2007], that is, calculating any probability of interest from the distribution over the variables represented in the network. BNs thus allow experts to evaluate their problem in a probabilistic manner. A problem with BNs is that they are generally difficult to construct; domain experts typically do not have the expertise to construct mathematical models and misinterpret the directed arcs of a BN as non-symmetric relations of cause and effect instead of collectively encoding an independence relation [Dawid, 2010]. Especially in data-poor domains, BN construction therefore needs to be done mostly manually through a knowledge elicitation procedure in consultation with the domain expert, which is a difficult and error-prone process [van der Gaag and Helsper, 2002].

1.2.3 Narrative models

Narrative models for reasoning about evidence have also been proposed [Pennington and Hastie, 1993; Wagenaar et al., 1993], in which stories are used to explain the evidence. As mentioned earlier, these models are generally less formally developed than probabilistic and argumentation models (see Verheij et al. [2016]). In narrative models, stories are modelled as sequences of events that explain the evidence. The aim is to construct stories that are coherent in that they are complete, consistent and plausible; multiple stories are then compared on their coherence, or quality, and the extent to which they explain all the evidence in an attempt to find the 'best' story. Advantages of narrative models compared to the other models discussed in this section are that they allow for keeping a global overview of a problem by providing different explanations of observed evidence, and that the focus lies on developing alternative stories that serve as competing hypotheses about what may have happened, thereby reducing the risk of tunnel vision.

In the field of AI so-called formal-logical models of abductive reasoning (e.g. Josephson and Josephson [1994]; Console and Torasso [1991]) were developed, which basically model the process of constructing alternative stories by computing possible explanations for the available evidence. In most of these models, stories are only compared on basic criteria such as the number of assumptions, where (subset) minimal explanations are preferred; criteria by which the quality of stories can be judged such as plausibility and completeness are typically not formalised in these models. Exceptions include the model of Josephson [2002], in which plausibility and consistency are considered, and the story part of Bex' [2011] formal hybrid theory of stories and arguments, in which completeness and consistency are considered. A disadvantage of narrative models is that atomistic reasoning about a single piece of evidence and its conclusions is impossible, in contrast with argumentation models. To benefit from the advantages of both the story-based approach and the argument-based approach, Bex [2011] studied how to combine stories and arguments, in line with a more general strand of research in which formal systems are combined and the connections between formal systems are considered; this strand of research is briefly discussed next.

1.2.4 Connections between formal systems

In previous work, connections between the aforementioned formal systems for reasoning about evidence have been theoretically investigated; an overview of this research is provided by Verheij and colleagues [2016]. For instance, connections between narrative models and BNs were considered by Vlek and colleagues [2014], a formal hybrid theory of stories and arguments was proposed by Bex [2011], an argumentation-based explanation method for BNs was proposed by Timmer and colleagues [2017], constraints on BNs given information specified in arguments were derived by Bex and Renooij [2016], and a framework in which arguments, stories and probabilities are combined was proposed by Verheij [2017]. In these approaches, the central concepts of different formal systems are connected to help gain a better understanding of the relations between them. In line with this work, we consider the relations between different modes of reasoning involved in reasoning about evidence, where we put special emphasis on informal aspects. In particular, we consider both informal sense-making tools and formal systems, where we investigate how the practical construction of formal representations within AI systems can be guided by using domain knowledge specified using such tools.

1.3 Research questions

As explained above, in this thesis we aim to guide the construction of formal representations of evidential knowledge and inference by exploiting domain knowledge specified by experts in analyses performed using informal sense-making tools they are familiar with. In particular, we address the following main research question:

Research question 1 How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of formal representations within AI systems?

As discussed earlier, in this thesis we focus on the construction of formal representations within two types of AI systems, namely argumentation frameworks and BNs.

Accordingly, research question 1 can be divided in the following two subquestions:

Research question 1a How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of argumentation frameworks?

Research question 1b How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of Bayesian networks?

To answer research question 1, in Chapter 3 we study the examples from Section 1.1 in the light of the conceptual analysis of reasoning about evidence we provide in Section 2.1. Our analysis of these examples serves to illustrate that not all information involved in performing analyses using sense-making tools is explicitly recorded and that the assumptions of domain experts underlying their analyses are typically not explicitly stated. From this, we conclude that there is a gap between sense-making tools and formal systems: while the domain knowledge specified in analyses performed using such tools conveys an initial sketch of the problem, the specified knowledge is not formal enough to be directly used in guiding the construction of formal representations within AI systems. In particular, while analyses performed using such tools may be constructed according to general templates or formats, the various elements that can be incorporated in these analyses are often ambiguous and not precisely defined. Hence, in order to bridge the gap between informal sense-making tools and formal systems, we wish to formalise analyses performed using such tools as an intermediary step in a manner that allows for guiding the construction of formal representations within AI systems.

Accordingly, a first step towards answering our research questions is to propose an intermediate formalism for formally representing analyses performed using these tools in terms of the *information graph* (IG) formalism. Our IG-formalism is designed to formalise and disambiguate analyses performed using informal sense-making tools in a way that (1) allows for guiding the construction of formal representations within AI systems and that (2) is in line with the conceptual analysis of reasoning about evidence we provide in Section 2.1, while (3) allowing inference to be performed and visualised in a manner that is closely related to the way in which inference is visualised by domain experts using such tools. The IG-formalism is graph-based instead of a logic-based to remain closely related to the way analyses are visualised using aforementioned graph-based tools as well as the BN-formalism. In defining an intermediary formalism between analyses performed using informal sense-making tools and formal AI systems, we were inspired by approaches for constructing BNs from ontologies [Uschold and Gruninger, 1996], formally specified knowledge representations which capture relations between concepts in a domain, such as [Fenz, 2012] (see also Section 8.2.3), and the Argument Interchange Format (AIF) [Rahwan and Reed, 2009], an argumentation ontology that serves as an intermediary formalism between analyses performed using argument diagramming tools [Bex et al., 2003; Okada et al., 2014] and formal argumentation frameworks [Bex et al., 2013]. Compared to ontologies, our IG-formalism is tailored to precisely model the process of reasoning about evidence and hence is closely related to sense-making tools such as mind maps as well as the formal representations whose construction we wish to facilitate. Moreover, compared to ontologies our IG-formalism allows for actually performing inference instead of only representing knowledge and reasoning. In Chapter 3 we further motivate why we prefer to use the IG-formalism to other existing formalisms as an intermediary formalism between analyses performed using informal sense-making tools and formal AI systems.

An IG serves as a source of unambiguous information that can be used to guide the construction of instantiations of AI systems for which a formal reasoning mechanism is defined. In particular, we study how knowledge expressed using this formalism can be exploited to guide the construction of argumentation frameworks and BNs, thereby answering research questions 1a and 1b.

Our approach for constructing argumentation frameworks from IGs serves for constructing an initial representation that can be directly used for formal evaluation using argumentation semantics. For our BN construction approach this is not the case: as (numerical) probabilities are typically not indicated using sense-making tools, probabilities are not accounted for in our IG-formalism, and therefore our approach can only serve for deriving some qualitative constraints on the probabilities of the BN under construction. Hence, initial BNs constructed by our approach are only partially specified and cannot be directly used for probabilistic inference.

Accordingly, in this thesis we investigate how the construction of BNs can be further facilitated. Given that the two main formal AI systems under consideration in this thesis are argumentation frameworks and BNs, we study how argumentation techniques can be used to facilitate BN construction. In previous work [Bex and Renooij, 2016; Wieten et al., 2018a; Timmer et al., 2015], the problem of constructing BNs from information specified in arguments about the domain was investigated (see also Section 8.2.2). In this thesis, we consider a different approach and investigate how argumentation can be used to argue about the BN under construction instead of about the domain:

Research question 2 How can Bayesian network construction be facilitated by exploiting expert knowledge expressed as arguments about BN elements?

In BN construction, it is typically considered good practice to document the BN model itself; however, the importance of documenting reasons pro and con BN modelling decisions has received relatively little attention, while proper documentation can play a vital role in allowing experts involved in the construction and use of BNs to understand and accept them. Moreover, experts involved in the construction of BNs may disagree about modelling decisions, where existing approaches do not provide systematic means to resolve such disagreements. Since disagreements about BNs are essentially argumentative in nature, we investigate how argumentation techniques can be used to capture and help resolve conflicts about BN elements in a BN under construction. We propose an approach that is generally applicable to both

BNs constructed from IGs and BNs otherwise constructed, and that serves to facilitate both the qualitative graph-construction step and the quantitative probability elicitation step involved in BN construction.

1.4 Outline of this thesis

In Chapter 2 we provide the reader with the necessary preliminaries for the remaining chapters. In particular, we provide a conceptual analysis of reasoning about evidence in Section 2.1, where we introduce assumptions that demarcate the scope of the work presented in this thesis. This analysis serves to motivate the concepts incorporated in the IG-formalism. In Sections 2.2 and 2.3 the technical preliminaries of the two main formal AI systems under consideration in this thesis are provided, namely argumentation frameworks and BNs, respectively. In Chapter 3 we motivate and present the IG-formalism. The idea of a graph-based intermediary formalism between analyses performed using informal sense-making tools and formal AI systems was first considered by us in [Wieten et al., 2018b, 2019b] in the form of so-called 'argument graphs'. Based on this earlier work the IG-formalism was introduced in [Wieten et al., 2020, 2021a], which we extended to increase its expressivity in [Wieten et al., 2021b].

In Chapter 4 we present an argumentation formalism based on IGs and study formal properties of our approach. The work presented in that chapter is based on the work presented in [Wieten et al., 2020, 2021b]. Based on the results of Chapter 4 we formulate an answer to research question 1a.

In Chapter 5 BN construction is considered. The idea of constructing BNs by exploiting knowledge captured in argument graphs was introduced in [Wieten et al., 2018b], where in [Wieten et al., 2019b] the scope of this approach was extended and formal properties were preliminarily investigated. In [Wieten et al., 2021a] IGs were taken as a starting point, where a structured approach for automatically constructing a BN graph from an IG was proposed and its formal properties were investigated. The work presented in Chapter 5 is based on this work. Based on the results of Chapter 5 we formulate an answer to research question 1b.

In Chapter 6 we illustrate the approaches of Chapters 4 and 5 by applying them to parts of an actual legal case, namely the well-known Sacco and Vanzetti case. We then compare the results obtained by applying the BN construction approach from Chapter 5 to a BN modelling by Kadane and Schum [1996]; this part of the case study is based on a precursory case study performed in [Wieten et al., 2018b] and serves as a preliminary validation of our approach of Chapter 5.

In Chapter 7 we present our approach for capturing and resolving discussions about BNs using argumentation, based on the work published in [Wieten et al., 2019a]. Based on our results we formulate an answer to research question 2.

We discuss how the research presented in this thesis relates to other research in Chapter 8. In Chapter 9 we conclude this thesis by summarising the problem and our results, where we answer our research questions and discuss possible avenues for future research.

Chapter 2

Preliminaries

In this chapter, we provide the necessary preliminaries required in the following chapters. In Section 2.1 we provide a conceptual analysis of reasoning about evidence, where we review the terminology used to describe it and introduce assumptions that demarcate the scope of the work presented in this thesis. In Sections 2.2 and 2.3 we then provide the technical preliminaries on argumentation and BNs, respectively.

2.1 Reasoning about evidence

First, we provide a conceptual analysis of reasoning about evidence and review the terminology used to describe it. Inference is the process of drawing conclusions from premises starting from the evidence, where evidence is that what has been established with certainty in the context under consideration. For instance, in the context of a legal trial the evidence consists of that what is actually observed by a judge or jury, such as documents (e.g. police and autopsy reports) and other tangible evidence, as well as testimonial evidence [Anderson et al., 2005]. Inference is often performed using domain-specific *generalisations* [Anderson et al., 2005; Bex et al., 2003; Bex, 2011], also called defaults [Pearl, 1988a; Reiter, 1980], which capture knowledge about the world in conditional form. Generalisations can either be strict or defeasible, where defeasible generalisations are of the form *'If* $a_1, \ldots, a_n$*, then usually/normally/typically* b*'* and strict generalisations are of the form *'If* $a_1, \ldots, a_n$*, then always* b*'*. Here, claims $a_1, \ldots, a_n$ are called the *antecedents* of the generalisation and b its *consequent*, where we assume that claims are literal propositions and that generalisations have one or more antecedents and exactly one consequent. In case a generalisation has multiple antecedents, it expresses that only the antecedents together allow us to infer the consequent. We semi-formally denote generalisations as $a_1, \ldots, a_n \rightarrow b$, among other things to ease the description of examples in this section and in Section 3.1.2. For defeasible generalisations, exceptional circumstances can be provided under which the generalisation may not hold, whereas strict generalisations hold without exception. An example of a (defeasible) generalisation is *'If fire, then typically smoke'*, where *'fire'* is its antecedent and *'smoke'* its consequent.

An example of an exception to this generalisation is that sufficient oxygen is present for complete combustion to occur.

A distinction can be made between causal and evidential generalisations [Bex, 2011; Pearl, 1988a], where instead of writing these generalisations in the form *'If ..., then ...'*, causal generalisations are written as '$c_1, \ldots, c_n$ *usually/normally/typically cause* e' (e.g. *'fire typically causes smoke'*) and evidential generalisations are written as '$e_1, \ldots, e_n$ *are evidence for* c' (e.g. *'smoke is evidence for fire'*). For a causal generalisation, its antecedents express a cause for the consequent, and for an evidential generalisation, its consequent expresses the usual cause for its antecedents. In the context of commonsense reasoning about evidence, causal and evidential generalisations are often assumed to be defeasible (see e.g. Bex [2011] and Kadane and Schum [1996]); in this thesis, this assumption is also made. The examples considered throughout this thesis illustrate that causal and evidential generalisations are typically not strict[1]. Pearl [1988a, p. 264] argued that people generally consider it difficult to express knowledge using only causal generalisations, and in an empirical study, van den Braak and colleagues [2008] found that while there are situations in which subjects consistently choose either causal or evidential modelling techniques, there are also many examples in which people use both types of generalisations in their reasoning. For instance, subjects often considered testimonies to be evidential, whereas a motive for committing an act is considered a cause for committing that act. This discussion illustrates that in formal accounts of reasoning about evidence, it is important to allow for both causal and evidential generalisations [Bex, 2011].

For causal generalisations, additional circumstances, also called *enabling conditions* [Cheng and Novick, 1990; Ortiz Jr., 1999], or enablers, may be provided under which the generalisation may be used in performing inference. Causal generalisations that include enablers are of the general form $e_1, \ldots, e_m, a_1, \ldots, a_n \to b$, where $e_1, \ldots, e_m$ are its enablers and $a_1, \ldots, a_n$ its actual antecedents. For a causal generalisation, only its actual antecedents and not its enablers express a cause for the consequent. Causality is a contentious topic, and it is easy to disagree about whether an event is an actual cause or an enabler. Cheng and Novick [1990] note that an event is typically viewed as an actual cause if it describes a situation that deviates from 'normal' circumstances. For instance, lighting a match is considered a cause of fire, but the presence of oxygen is typically not consider a cause of fire as it is normal that oxygen is present. This is, however, also context-dependent, and oxygen can be considered a cause of fire in situations where oxygen is typically not present (e.g. in space). We note that generalisations capture knowledge about the world as perceived by the person stating the knowledge, and that the distinction between enablers and actual causes allows domain experts to be more expressive in stating their knowledge.

We also consider generalisations that are neither causal nor evidential. For instance, abstractions [Bex, 2011; Console and Dupré, 1994; Kautz, 1991] allow for

[1]Note that strict generalisations such as strict rules from classical logic and definitions can be expressed using strict generalisations of type 'other' and strict abstractions.

Table 2.1: Table indicating for each generalisation type whether generalisations may be defeasible or strict.

	Causal generalisations	**Evidential generalisations**	**Abstractions**	**Other generalisations**
Defeasible	V	V	V	V
Strict	X	X	V	V

reasoning at different levels of abstraction. More precisely, abstractions are of the form '$p_1, \ldots, p_n$ *can usually/normally/typically/always be considered a specialisation of* q' (e.g. guns can usually be considered deadly weapons), where antecedents $p_1, \ldots, p_n$ are considered to be more specific than the more abstract consequent q. As noted by Console and Dupré [1994], abstractions are syntactically the same as causal generalisations but they are semantically different in that the antecedents of abstractions do not express a cause for the consequent or vice versa. Abstractions may be defeasible (cf. Bex [2011]) but may also be strict (cf. Console and Dupré [1994]); an example of a strict abstraction is generalisation $lung\_cancer \rightarrow cancer$, which states that lung cancer is a type of cancer. An example of defeasible abstraction is $gun \rightarrow deadly\_weapon$, where an example of an exception to this generalisation is that the gun is a non-functional replica, or a water gun.

Another example of a different type of generalisation is a generalisation representing a mere statistical correlation, such as a correlation between homelessness and criminality. While there may be one or more confounding factors that cause both homelessness and criminality (e.g. unemployment), a domain expert may be unaware of these factors or may wish to refrain from expressing them explicitly. In this thesis, we distinguish between generalisations that are causal, evidential, abstraction, or of another type, where generalisations of type 'other' may be defeasible or strict. Specifically, as this category contains all possible types of generalisations other than causal, evidential and abstraction, we allow for the option to distinguish between strict and defeasible generalisations among these generalisations. Table 2.1 provides an overview of the different generalisation types, where for each type it is indicated whether generalisations may be defeasible or strict. The notation $\rightarrow_{\mathrm{c}}$, $\rightarrow_{\mathrm{e}}$, $\rightarrow_{\mathrm{a}}$ and $\rightarrow_{\mathrm{o}}$ is used for the different types of generalisations, respectively.

Different types of inferences can be performed with generalisations depending on whether their antecedents or consequent are *affirmed* in that they are either observed or inferred; here, a claim is *inferred* iff it is either deductively or abductively inferred, where in deductive inference the consequent is inferred from the antecedents and in abductive inference the antecedents are inferred from the consequent. These two inference types are now considered in more detail.

2.1.1 Deductive inference

Inference can be performed in a deductive fashion, where from a generalisation and by affirming the antecedents, the consequent is inferred by modus ponens on the generalisation. Note that the term 'deduction' is not consistently used in the literature, as it can either mean strict inference, in which the consequent universally holds given the antecedents (e.g. Besnard and Hunter [2009]), or defeasible inference, in which the consequent tentatively holds given the antecedents (e.g. Shanahan [1989]). To cover both meanings, in this thesis 'deduction' is used as an umbrella term for both defeasible 'forward' inference and strict 'forward' inference; hence, deduction is not necessarily a stronger or more reliable form of inference than abduction, which is a type of defeasible inference. Defeasible deduction can only be performed using defeasible generalisations (of any type) and not using strict generalisations (see Table 2.2). Strict deductive inference can only be performed using strict abstractions and strict generalisations of type 'other'. For a given instance of deduction, it will be explicitly specified whether it concerns strict or defeasible deduction.

Example 3. *Consider causal generalisation* $g\colon$ fire $\rightarrow_{\mathrm{c}}$ smoke*. By affirming* g*'s antecedent* fire*, its consequent* smoke *is defeasibly deductively inferred.* □

The following example illustrates strict deductive inference.

Example 4. *Consider strict abstraction* $g\colon$ lung_cancer $\rightarrow_{\mathrm{a}}$ cancer*. Upon observing that a person has lung cancer, we can strictly deductively infer that the person has cancer using generalisation* g*.* □

Prediction [Shanahan, 1989] is a specific type of deductive inference in which the consequent of a causal generalisation is deductively inferred by affirming its antecedents. Specifically, as the antecedents of a causal generalisation express a cause for the consequent, the consequent is said to be predicted from the antecedents in this case. Example 3 provides an example of prediction.

2.1.2 Abductive inference

Abduction [Josephson and Josephson, 1994; Console and Torasso, 1991; Console and Dupré, 1994], a type of defeasible inference, can be performed using causal generalisations and abstractions: from a causal generalisation or an abstraction and by affirming the consequent, the antecedents are inferred, since if the antecedents are true it would allow us to deductively infer the consequent modus-ponens-style. Following Josephson and Josephson [1994] and Console and Torasso [1991], in case causes $c_1, \ldots, c_n$ and $c'_1, \ldots, c'_m$ are abductively inferred from common effect e using causal generalisations $g_1\colon c_1, \ldots, c_n \rightarrow_{\mathrm{c}} e$ and $g_2\colon c'_1, \ldots, c'_m \rightarrow_{\mathrm{c}} e$, then sets $\{c_1, \ldots, c_n\}$ and $\{c'_1, \ldots, c'_m\}$ are considered to be *competing alternative explanations* for e. We assume that causes c_i (and c'_j) are not in competition among themselves.

Table 2.2: Table indicating for defeasible and strict generalisations of every generalisation type which types of inferences may be performed.

	Causal generalisations	Evidential generalisations	Defeasible abstractions	Strict abstractions	Defeasible other generalisations	Strict other generalisations
Defeasible deduction	V	V	V	X	V	X
Strict deduction	X	X	X	V	X	V
Abduction	V	X	V	V	X	X

Example 5. *Consider the following causal generalisations:*

$g_1:$ fire $\rightarrow_{\mathrm{c}}$ smoke*;*
$g_2:$ smoke_machine $\rightarrow_{\mathrm{c}}$ smoke*.*

*By affirming the common consequent (*smoke*),* fire *and* smoke_machine *are abductively inferred, which are then competing alternative explanations of* smoke. □

Abduction can also be performed using abstractions [Bex, 2011; Console and Dupré, 1994], where the used abstraction can either be defeasible (cf. Bex [2011]) or strict (cf. Console and Dupré [1994]). An example of a model including strict abstractions is that of Console and Dupré [1994], in which both explanatory axioms (comparable to causal generalisations) and abstraction axioms are used to explain observations. Multiple explanations that are inferred using abstraction axioms can then be considered competing alternative explanations. Note that an abductive inference step with a strict abstraction is still defeasible, as it concerns an inference step from the more abstract consequent to a more specific antecedent. Following Console and Dupré [1994] and Bex [2011], we allow for abductive inference using both strict and defeasible abstractions, where in performing abduction with abstractions $g_1 : p_1, \ldots, p_n \rightarrow_{\mathrm{a}} q$ and $g_2 : p'_1, \ldots, p'_m \rightarrow_{\mathrm{a}} q$ sets of antecedents $\{p_1, \ldots, p_n\}$ and $\{p'_1, \ldots, p'_m\}$ are considered to be competing alternative explanations of the common consequent q. We assume that antecedents p_i (and p'_j) are not in competition among themselves.

Example 6. *Consider the following defeasible abstractions:*

$g_1:$ gun $\rightarrow_{\mathrm{a}}$ deadly_weapon*;*
$g_2:$ knife $\rightarrow_{\mathrm{a}}$ deadly_weapon*.*

*By affirming the common consequent (*deadly_weapon*),* gun *and* knife *are abductively inferred using generalisations g_1 and g_2, which are then competing alternative explanations of* deadly_weapon*.* □

The following example illustrates abductive inference with strict abstractions.

Example 7. *Consider the following strict abstractions:*

$g'_1:$ lung_cancer $\rightarrow_{\text{a}}$ cancer*;*
$g'_2:$ colon_cancer $\rightarrow_{\text{a}}$ cancer*.*

Upon observing that a person has cancer, lung_cancer *and* colon_cancer *are abductively inferred, which are then competing alternative explanations of* cancer*.* □

2.1.3 Representing causal knowledge

Abductive inference with causal generalisations and deductive inference with evidential generalisations are related: in some cases, we will accept not only causal generalisation *'c usually/normally/typically causes e'* but also evidential generalisation *'e is evidence for c'* [Bex, 2015; Pearl, 1988a], which we will call the *evidential counterpart* of the causal generalisation. However, it can be argued that we only accept the evidential counterpart of a causal generalisation if c is the *usual* cause of e, where we assume that only one cause can be the usual cause of e.

Example 8. *Fire can be considered the usual cause of smoke, so we will accept both causal generalisation* $g:$ fire $\rightarrow_{\text{c}}$ smoke *and its evidential counterpart* $g':$ smoke $\rightarrow_{\text{e}}$ fire*. In this case, abductive inference with generalisation* g *can be encoded as deductive inference with generalisation* g'*. Because a smoke machine cannot be considered the usual cause of smoke, we will accept causal generalisation* smoke_machine $\rightarrow_{\text{c}}$ smoke *but we will not accept evidential generalisation* smoke $\rightarrow_{\text{e}}$ smoke_machine*.* □

Note that a causal generalisation g can only have an evidential counterpart g' in case g has a single antecedent, as we assume generalisations have a single consequent but multiple antecedents. Furthermore, as we assume that only one cause can be the usual cause of e, only one of the causal generalisations $c_1 \rightarrow_{\text{c}} e$ or $c_2 \rightarrow_{\text{c}} e$ can be replaced by an evidential generalisation. Hence, we do not consider c_1 and c_2 to be competing alternative explanations of e in case deductive inference is performed using evidential generalisations $e \rightarrow_{\text{e}} c_1$ and $e \rightarrow_{\text{e}} c_2$.

2.1.4 Mixed inference and inference constraints

Deduction and abduction can be iteratively performed, where *mixed* abductive-deductive inference is also possible.

Example 9. *Suppose that from the causal generalisation* $g_1:$ fire $\rightarrow_{\text{c}}$ smoke *and by affirming its consequent (*smoke*), its antecedent (*fire*) is inferred. Now, if the additional causal generalisation* $g_2:$ fire $\rightarrow_{\text{c}}$ heat *is provided, then its consequent (*heat*) can be deductively inferred (or predicted) as the antecedent (*fire*) has been previously abductively inferred.* □

2.1.4.1 Constraints on performing inference with causal and evidential generalisations

Mixed deductive inference, using both causal and evidential generalisations, can also be performed [Bex, 2015], but as noted by Pearl [1988a] care should be taken in performing mixed inference that no cause for an effect is inferred in case an alternative cause for this effect was already previously inferred.

Example 10. *Consider the example in which a causal generalisation* g_1: smoke_machine $\rightarrow_\text{c}$ smoke *and an evidential generalisation* g_2: smoke $\rightarrow_\text{e}$ fire *are provided. Deductively chaining these generalisations would make us infer that there is a fire when seeing a smoke machine, which is clearly undesirable.* □

Similarly, in performing mixed deductive-abductive inference, care should be taken that no cause for an effect is inferred in case an alternative cause for this effect was already previously inferred.

Example 11. *Consider Example 10, where instead of an evidential generalisation* g_2: smoke $\rightarrow_\text{e}$ fire *a causal generalisation* g_2: fire $\rightarrow_\text{c}$ smoke *is provided. Upon seeing a smoke machine, this would make us infer that there is a fire in case deduction and abduction are performed in sequence, which is again undesirable.* □

Accordingly, we wish to prohibit these types of inference patterns, and refer to the constraint that no cause for an effect should be inferred in case an alternative cause for this effect was already previously inferred as *Pearl's constraint* [Pearl, 1988a].

The above discussion can be extended to generalisations with multiple antecedents.

Example 12. *Suppose that the following generalisations are provided:*

g_1: high_body_temperature $\rightarrow_\text{e}$ fever;
g_2: smoke $\rightarrow_\text{c}$ coughing;
g_3: fever, coughing $\rightarrow_\text{e}$ pneumonia.

Upon observing that a person has high body temperature and that there is smoke, this would make us infer that the person has a fever and is coughing using generalisations g_1 *and* g_2*, respectively. In turn, this would make us infer that the person has pneumonia using generalisation* g_3*, which is undesirable: as a cause for* coughing *was already previously inferred (*smoke*), we should not be able to infer a different cause for* coughing *(*pneumonia*). Specifically,* fever *is in itself not a sufficient condition for inferring* pneumonia*:* coughing *is also necessary. Only in case a separate evidential generalisation* g_4: fever $\rightarrow_\text{e}$ pneumonia *is provided should we be able to infer* pneumonia. □

Similarly, problems can arise in performing inference using causal generalisations that include enabling conditions, as illustrated by the following example.

Example 13. *Consider the example in which the following causal generalisations are provided:*

g_1 : torch $\rightarrow_{\text{c}}$ fire;
g_2 : match, oxygen $\rightarrow_{\text{c}}$ fire.

In this case, the presence of oxygen is an enabler of generalisation g_2, as it cannot be considered an actual cause of fire. Upon striking a match in the presence of oxygen, we can deductively infer that there is a fire using generalisation g_2. Similar to Example 11, we should now not be able to abductively infer torch *using generalisation g_1. Similarly, performing deduction and abduction in sequence using generalisations g_1 and g_2 is undesirable.* □

To summarise this section, we wish to prohibit (1) subsequent deductive inference using a causal generalisation and an evidential generalisation, and (2) subsequent deductive and abductive inference using two causal generalisations with the same consequent. Note that, while these constraints deviate from Pearl's original constraints [Pearl, 1988a] as he only considered defaults with single antecedents, we will refer to these constraints as *Pearl's constraint* throughout this thesis.

2.1.4.2 Constraints on performing inference with abstractions

When performing inference with abstractions, care should be taken that no version of an event at a lower level of abstraction is abductively inferred if an alternative version of this event at a lower level of abstraction was already previously inferred. In particular, performing deduction and abduction in that order with two abstractions with the same consequent leads to undesirable results.

Example 14. *Consider generalisations g_1 :* gun $\rightarrow_{\text{a}}$ deadly_weapon *and g_2 :* knife $\rightarrow_{\text{a}}$ deadly_weapon *from Example 6. Upon observing that a provided object is a* gun*, this would make us deductively infer that this object is a* deadly_weapon *using generalisation g_1. Upon performing abduction with generalisation g_2, this would make us infer that the provided object is a* knife*, which is clearly undesirable.* □

Performing abduction and deduction in that order with two abstractions with the same consequent does not lead to undesirable results.

Example 15. *Consider abstractions g_2 :* knife $\rightarrow_{\text{a}}$ deadly_weapon *and g_3 :* knife $\rightarrow_{\text{a}}$ metal_object*. Upon observing* metal_object*, we can abductively infer* knife *using generalisation g_3. In turn, claim* deadly_weapon *can be deductively inferred using generalisation g_2.* □

The following example illustrates that mixed inference, using either a causal generalisation and an abstraction or an evidential generalisation and an abstraction, does not lead to undesirable results. Hence, we argue that no additional inference constraints need to be imposed.

Example 16. *Consider Example 7. Assume that in addition to strict abstractions* $g'_1\colon$ lung_cancer $\rightarrow_\mathrm{a}$ cancer *and* $g'_2\colon$ colon_cancer $\rightarrow_\mathrm{a}$ cancer*, causal generalisation* $g'_3\colon$ smoking $\rightarrow_\mathrm{c}$ cancer *is provided. Upon observing that a person smokes, we deductively infer that the person has cancer using generalisation* g'_3*. Using generalisations* g'_1 *and* g'_2*, we can then in turn abductively infer that the person has either lung cancer or colon cancer, which are then competing alternative explanations of* cancer *(see Example 7). Note that it is not undesirable to infer* lung_cancer *or* colon_cancer *from* cancer *in this case, as* smoking *and* lung_cancer *(*colon_cancer*) are not alternative explanations of* cancer*; instead,* smoking *is a cause of* cancer*, while* lung_cancer *(*colon_cancer*) is not a cause of* cancer *but instead describes claim* cancer *at a lower level of abstraction.* □

To summarise this section, we only wish to prohibit subsequent deductive and abductive inference using two abstractions with the same consequent and not other inference patterns involving abstractions. Finally, note that for generalisations of type 'other' no additional inference constraints are imposed.

2.1.5 Ambiguous inference

Situations may arise in practice in which both deduction and abduction can be performed with the same causal generalisation or abstraction; the inference type is, therefore, *ambiguous*.

Example 17. *Consider generalisation* $g_1\colon$ fire $\rightarrow_\mathrm{c}$ smoke*. Suppose* fire *and* smoke *are not observed but have been previously inferred, for instance via deduction using generalisations* $g_2\colon$ see_fire $\rightarrow_\mathrm{e}$ fire *and* $g_3\colon$ see_smoke $\rightarrow_\mathrm{e}$ smoke*, where* see_fire *and* see_smoke *are provided as evidence. Then both deduction and abduction can be performed with* g_1 *to infer* smoke *from* fire *and* fire *from* smoke*.* □

Generally, we do not wish to prohibit this type of ambiguous inference patterns as we do not consider them to be undesirable.

2.2 Argumentation

In this section, technical preliminaries on argumentation are provided. In particular, Dung's abstract approach to argumentation [Dung, 1995] (Section 2.2.1) and the ASPIC$^+$ argumentation framework [Modgil and Prakken, 2013, 2014] (Section 2.2.2) are reviewed.

2.2.1 Abstract argumentation frameworks

First, we provide Dung's [1995] definition of an *abstract argumentation framework*, which consists of a set of arguments along with a binary relation of defeat. Such a framework is fully abstract as it leaves the internal structure of arguments and the nature of the defeat relation completely unspecified. Dung's definitions for argumentation semantics can then be used to evaluate the acceptability of arguments in abstract argumentation frameworks.

Definition 1 (Abstract argumentation framework [Dung, 1995], using the terminology of [Modgil and Prakken, 2014])**.** *An* abstract argumentation framework (AF) *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*.*

Note that Dung called his relation 'attack'. Instead, we follow the terminology used in describing the ASPIC$^+$ framework (see Section 2.2.2), where the term attack is reserved for the basic notion of conflict which is then resolved into defeat using preferences. Furthermore, following ASPIC$^+$-conventions, throughout this thesis calligraphic capitals are used to denote sets of arguments, defeats, as well as other sets used in the ASPIC$^+$ framework. An AF can be visualised as a directed graph in which arguments are represented by circles and defeats are indicated by solid arcs ($\rightarrow$); an example of an AF is depicted in Figure 2.1.

The theory of abstract argumentation frameworks is built around the notion of an *extension*, which is a set of arguments that is internally coherent and defends itself against defeat.

Definition 2 (Dung extensions, after Modgil and Prakken [2014])**.** *Let* $(\mathcal{A}, \mathcal{D})$ *be an AF.*

- *A set of arguments* $\mathcal{S} \subseteq \mathcal{A}$ *is* conflict-free *if there do not exist* $A, B \in \mathcal{S}$ *such that* $(A, B) \in \mathcal{D}$*.*
- *An argument* $A \in \mathcal{A}$ *is* acceptable *with respect to some set of arguments* $\mathcal{S} \subseteq \mathcal{A}$ *iff for all arguments* B *such that* $(B, A) \in \mathcal{D}$ *there exists an argument* $C \in \mathcal{S}$ *such that* $(C, B) \in \mathcal{D}$*.*
- *A conflict-free set of arguments* $\mathcal{S} \subseteq \mathcal{A}$ *is an* admissible *extension iff every argument* $A \in \mathcal{S}$ *is acceptable with respect to* $\mathcal{S}$*.*
- *An admissible extension* $\mathcal{S}$ *is a* complete *extension iff* $A \in \mathcal{S}$ *whenever* A *is acceptable with respect to* $\mathcal{S}$*;* $\mathcal{S}$ *is the* grounded *extension iff* $\mathcal{S}$ *is the set inclusion minimal complete extension;* $\mathcal{S}$ *is a* preferred *extension iff* $\mathcal{S}$ *is a set inclusion maximal complete extension; and* $\mathcal{S}$ *is a* stable *extension iff it is preferred and* $\forall B \notin \mathcal{S}, \exists A \in \mathcal{S}$ *such that* $(A, B) \in \mathcal{D}$*.*

The acceptability of arguments in abstract argumentation frameworks can then be evaluated by establishing whether a given argument is a member of the various extensions. Arguments are then assigned a *dialectical status* that can either be *'justified'*, *'overruled'*, or *'defensible'*, where informally an argument is justified if it survived the competition, overruled if it did not survive the competition, and defensible if it is involved in a tie.

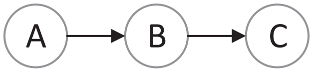


Figure 2.1: Example of an AF.

Definition 3 (Justified, overruled and defensible arguments, adapted from Prakken and Vreeswijk [2002])**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF.*

- *An argument is (i)* justified *under grounded semantics iff it is a member of the grounded extension, (ii)* overruled *under grounded semantics iff it is not justified under grounded semantics and it is defeated by an argument that is justified under grounded semantics, or (iii)* defensible *under grounded semantics iff it is neither justified nor overruled under grounded semantics.*
- *Let $T \in \{complete,\ preferred,\ stable\}$. An argument is (i)* justified *under T semantics iff it is a member of all T extensions, (ii)* overruled *under T semantics iff it is not a member of any T extension, or (iii)* defensible *under T semantics iff it is a member of some but not all T extensions.*

Example 18. *An example of an AF is depicted in Figure 2.1. In this example, set $\mathcal{S} = \{A, C\}$ is conflict-free as neither (A, C) nor (C, A) in $\mathcal{D}$. Arguments A and C are acceptable with respect to $\mathcal{S}$: for A, this holds as there is no argument in $\mathcal{A}$ defeating it, and for C, this holds because only B defeats it, where B itself is defeated by A. Hence, $\mathcal{S}$ is an admissible extension. $\mathcal{S}$ is also a complete extension, as B is not acceptable with respect to $\mathcal{S}$. Other admissible extensions such as $\emptyset$ and $\{C\}$ also exist, but these extensions are not complete as A is also acceptable with respect to $\emptyset$ and $\{C\}$. In fact, $\mathcal{S}$ is the only complete extension and, therefore, also the grounded extension and the only preferred extension. $\mathcal{S}$ is also a stable extension, as only $B \notin \mathcal{S}$, where $(A, B) \in \mathcal{D}$. It follows that under any semantics, arguments A and C are justified and argument B is overruled.* □

Compared to preferred and stable semantics, grounded semantics has stricter requirements regarding what to accept as a justified belief. As for their outcomes, these semantics mainly differ in their treatment of so-called floating arguments, as illustrated by the following example.

Example 19. *Consider the example depicted in Figure 2.2, adapted from Caminada [2006]. The complete extensions of this AF are: $\mathcal{S}_1 = \emptyset$, $\mathcal{S}_2 = \{A, D\}$ and $\mathcal{S}_3 = \{B, D\}$. Specifically, in contrast to example 18, the empty set is a complete extension as none of the arguments is acceptable with respect to the empty set. Furthermore, $\mathcal{S}_2$ is an admissible extension, as A defends itself against defeat from B by defeating B itself, and for defeat $(C, D) \in \mathcal{D}$ it holds that $A \in \mathcal{S}_2$ with $(A, C) \in \mathcal{D}$. $\mathcal{S}_2$ is then a complete extension as neither B nor C can be added to $\mathcal{S}_2$ without introducing a conflict. Similarly, $\mathcal{S}_3$ is an admissible and complete extension. Hence, the grounded extension is $\emptyset$ and under grounded semantics all arguments are defensible. $\mathcal{S}_2$ and $\mathcal{S}_3$*

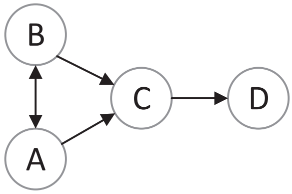


Figure 2.2: AF containing floating arguments, adapted from Caminada [2006].

are set inclusion maximal complete extensions for which it holds that the arguments in these sets defeat all arguments outside these sets; hence, $\mathcal{S}_2$ and $\mathcal{S}_3$ are preferred and stable extensions. Compared to grounded semantics, argument D is justified under preferred and stable semantics, C is overruled, and A and B are defensible. In particular, it is concluded that C is overruled, as there is no need to resolve the conflict between A and B: the status of C 'floats' on the status of A and B. Then, as C is overruled, D is justified. Thus, using grounded semantics it can be said that one is more careful in assigning the status 'justified'. □

Dung's abstract argumentation approach has been extended with new elements, for instance by adding support relations to abstract argumentation frameworks (e.g. Cayrol and Lagasquie-Schiex [2005]) or by adding preference relations (e.g. so-called preference-based argumentation frameworks, or PAFs [Amgoud and Cayrol, 2002]), probabilities (discussed in Section 8.3), or weights [Dunne et al., 2011] to AFs; a more complete overview is provided in [Beirlaen et al., 2018; Prakken, 2018a]. As discussed in the introduction of this thesis, in Chapter 4 we focus on non-probabilistic approaches to argumentation and therefore do not consider extensions of Dung's approach using probabilities. Instead of the other proposed extensions, we opt for Dung's original approach for the evaluation of arguments as it is a well-studied and widely accepted approach in the field of computational argumentation. Moreover, relations between Dung's fully abstract approach and formalisms such as ASPIC$^+$ have been previously investigated, as explained in the next subsection.

2.2.2 The ASPIC$^+$ argumentation framework

Next, the ASPIC$^+$ argumentation framework [Modgil and Prakken, 2013, 2014] is reviewed. The ASPIC$^+$ framework is an example of a framework for so-called 'structured' argumentation (see Besnard et al. [2014]) in that it is at an intermediate level of abstraction between Dung's fully abstract approach and concrete instantiating logics such as those developed in [Prakken and Sartor, 1997; Simari and Loui, 1992]. In structured argumentation, a formal language for representing knowledge is assumed, where it is specified how arguments and counterarguments can be constructed from that knowledge. An argument is then said to be structured in that the premises and conclusion of the argument are made explicit and the relation between the premises and the conclusion is formally defined. Other frameworks for structured argumentation exist, e.g. [Besnard and Hunter, 2009; Dung et al., 2009; Gordon et al., 2007; Simari and Loui, 1992; Verheij, 2003; Vreeswijk, 1997]; where applicable, we will stay as close as possible to the ASPIC$^+$ framework because it is a flexible framework that instantiates Dung's abstract approach by describing what the arguments are and how the defeat relation arises. Moreover, it has been shown [Prakken, 2010; Modgil and Prakken, 2013] that the ASPIC$^+$ framework subsumes, or at least closely approximates, other work on structured argumentation [Dung et al., 2009; Verheij, 2003]. The connection to classical-logic approaches to argumentation (e.g. Besnard and Hunter [2009]) has also been studied [Modgil and Prakken, 2013], and it was shown [van Gijzel and Prakken, 2012] that the Carneades argument model [Gordon et al., 2007] can be interpreted in terms of ASPIC$^+$. Finally, desirable properties such as key rationality postulates [Caminada and Amgoud, 2007] have been studied for the ASPIC$^+$ framework, which warrants the sound definition of specific instantiations of this framework. In Chapter 4 we define an argumentation formalism closely related to the ASPIC$^+$ framework which allows us to straightforwardly prove that instantiations of our formalism also satisfy these rationality postulates. Furthermore, in defining the approach of Chapter 7 the ASPIC$^+$ framework will be explicitly used.

2.2.2.1 Argumentation systems, knowledge bases, and arguments

We consider a simplified version of the ASPIC$^+$ framework [Modgil and Prakken, 2013, 2014]. The ASPIC$^+$ instance assumes a *logical language* $\mathcal{L}$ containing the basic elements that can be argued about, a *knowledge base* $\mathcal{K} \subseteq \mathcal{L}$ of premises and a set of *inference rules* $\mathcal{R}$ that can be chained into arguments. More specifically, inference rules are defined over $\mathcal{L}$ and are either strict or defeasible.

Formally, an ASPIC$^+$ argumentation system is defined as follows.

Definition 4 (Argumentation system, after Modgil and Prakken [2013]). *An* argumentation system *is a tuple $AS = (\mathcal{L}, \overline{\cdot}, \mathcal{R}, n)$, where:*

- *$\mathcal{L}$ is a non-empty propositional language.*
- *$\overline{\cdot}$ is a function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that:*
  - *$\phi \in \mathcal{L}$ is a* contrary *of $\psi \in \mathcal{L}$ iff $\phi \in \overline{\psi}$ and $\psi \notin \overline{\phi}$;*
  - *$\phi \in \mathcal{L}$ is a* contradictory *of $\psi \in \mathcal{L}$ (denoted by '$\phi = -\psi$') iff $\phi \in \overline{\psi}$ and $\psi \in \overline{\phi}$;*
  - *Every $\phi \in \mathcal{L}$ has at least one contradictory.*
- *$\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $s\colon \phi_1, \ldots, \phi_n \to \phi$ and $d\colon \phi_1, \ldots, \phi_n \Rightarrow \phi$ respectively, where $\phi_1, \ldots, \phi_n, \phi$ are meta-variables ranging over well-formed formulas in $\mathcal{L}$ and where $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.*
- *$n\colon \mathcal{R}_d \to \mathcal{L}$ is a naming convention for defeasible inference rules.*

Informally, $n(d)$ is a well-formed formula in $\mathcal{L}$ which states that defeasible inference rule $d \in \mathcal{R}_d$ is applicable. Furthermore, note that $-$ is not a part of the logical language $\mathcal{L}$ but a metalinguistic function symbol to obtain more concise definitions.

In contrast with Section 2.1, in which the symbol '$\rightarrow$' is semi-formally used as a connective in the notation for generalisations, in the context of ASPIC$^+$ this symbol indicates a strict inference rule. It is important to stress that inference rules in $\mathcal{R}_s$ and $\mathcal{R}_d$ are not object level formulae in the language $\mathcal{L}$, but are meta to the language. These inference rules may be domain-specific in that they reference specific formulae in $\mathcal{L}$, or they may be specified as schemes.

A contrariness function $\bar{\cdot}$ allows for defining an asymmetric notion of conflict (which can be used to model well-known constructs like negation as failure in logic programming), as well as symmetric conflict between formulae that are mutually exclusive but not necessarily exhaustive: for instance, being a bachelor and being married can be declared contradictories of each other. Classical negation, denoted by $\neg$, is a special case of the symmetric contradictory relation: $\phi \in \overline{\psi}$ iff ϕ is of the form $\neg\psi$ or ψ is of the form $\neg\phi$ (i.e. for any wff ϕ, ϕ and $\neg\phi$ are contradictories).

A knowledge base $\mathcal{K} \subseteq \mathcal{L}$ in an argumentation system AS is defined as follows.

Definition 5 (Knowledge base, after Modgil and Prakken [2013])**.** *A* knowledge base *in an argumentation system $AS = (\mathcal{L}, \bar{\cdot}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (*the axiom premises*) and $\mathcal{K}_p$ (*the ordinary premises*).*

Here, $\mathcal{K}_n$ consists of the premises which are certain and cannot be attacked. Premises in $\mathcal{K}_p$ are uncertain and can be undermined by other arguments (see Definition 8.3).

Definition 6 (Argumentation theory, after Modgil and Prakken [2013])**.** *An* argumentation theory *is a tuple $AT = (AS, \mathcal{K})$, where $AS = (\mathcal{L}, \bar{\cdot}, \mathcal{R}, n)$ is an argumentation system and $\mathcal{K}$ is a knowledge base in AS.*

Arguments are iteratively constructed from the premises in the knowledge base by chaining inference rules. In what follows, for a given argument, the operator Prem returns all formulae in $\mathcal{K}$ used to construct the argument, Conc returns its conclusion, Sub returns all its sub-arguments (including itself), and TopInf returns the last inference rule used in constructing the argument.

Definition 7 (ASPIC$^+$ argument, adapted from Modgil and Prakken [2014])**.** *An* argument *A on the basis of an argument theory AT with a knowledge base $\mathcal{K}$ and an argumentation system $AS = (\mathcal{L}, \bar{\cdot}, \mathcal{R}, n)$ is any structure obtainable by applying one or more of the following steps finitely many times:*

1. *ϕ if $\phi \in \mathcal{K}$, where:* $\textsc{Prem}(A) = \{\phi\}$; $\textsc{Conc}(A) = \phi$;
   $\textsc{Sub}(A) = \{A\}$; $\textsc{TopInf}(A) =$ *undefined.*
2. *$A_1, \ldots, A_n \Rightarrow \phi$ if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible inference rule* $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \Rightarrow \phi$ *in $\mathcal{R}_d$, where:*
   $\textsc{Prem}(A) = \textsc{Prem}(A_1) \cup \ldots \cup \textsc{Prem}(A_n)$; $\textsc{Conc}(A) = \phi$;
   $\textsc{Sub}(A) = \textsc{Sub}(A_1) \cup \ldots \cup \textsc{Sub}(A_n) \cup \{A\}$;
   $\textsc{TopInf}(A) = \textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \Rightarrow \phi$.

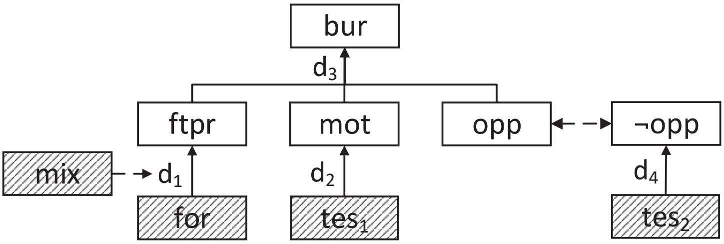


Figure 2.3: ASPIC$^+$-style argument graph for the burglary example (Example 20).

3. $A_1, \ldots, A_n \rightarrow \phi$ *if* $A_1, \ldots, A_n$ *are arguments such that there exists a strict inference rule* $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \rightarrow \phi$ *in* $\mathcal{R}_s$*, where:*
$\textsc{Prem}(A) = \textsc{Prem}(A_1) \cup \ldots \cup \textsc{Prem}(A_n)$*;* $\textsc{Conc}(A) = \phi$*;*
$\textsc{Sub}(A) = \textsc{Sub}(A_1) \cup \ldots \cup \textsc{Sub}(A_n) \cup \{A\}$*;*
$\textsc{TopInf}(A) = \textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \rightarrow \phi$*.*

Note that we overload symbols $\Rightarrow$ and $\rightarrow$ to denote an argument while it also denotes a defeasible or strict inference rule. ASPIC$^+$-style arguments in this section and in Chapter 7 are depicted informally in *ASPIC$^+$-style argument graphs*; an example of such an argument graph is depicted in Figure 2.3. Nodes in ASPIC$^+$-style argument graphs denote propositions $\phi \in \mathcal{L}$, where shaded nodes denote propositions in $\mathcal{K}_n$. Every application of an inference rule $s\colon \phi_1, \ldots, \phi_n \rightarrow \phi$ or $d\colon \phi_1, \ldots, \phi_n \Rightarrow \phi$ in $\mathcal{R}$ is indicated by a solid (hyper)arc in the graph that is directed from the nodes corresponding to $\phi_1, \ldots, \phi_n$ to the node corresponding to ϕ. Every solid (hyper)arc is annotated with the name of the applied inference rule, where d_i and s_i denote defeasible and strict rules, respectively. Attacks are indicated by dashed (hyper)arcs in the graph, as further discussed in Section 2.2.2.2.

Example 20. *Consider the following example from the legal and forensic domains, adapted from Bex and Renooij [2016]. Suppose that a burglary has taken place and that we are interested in whether a given suspect is guilty of committing the burglary (*bur*). Forensic analysis (*for*) shows a match between a pair of shoes owned by the suspect and footprints (*ftpr*) found near the crime scene. However, there is also evidence of a mix up at the forensic laboratory (*mix*): the chain-of-custody of the footprints from the crime scene to the laboratory has not been properly documented. The suspect had motive (*mot*) to commit this burglary, which is supported by at least one reliable testimony (*tes_1*). Furthermore, it is argued that the suspect had opportunity (*opp*) to commit this burglary; however, this is denied (*$\neg$opp*) in further testimony provided by the suspect (*tes_2*).*

For this example, we have the following propositional language with classical negation: $\mathcal{L} = \{\mathrm{bur}, \mathrm{for}, \mathrm{ftpr}, \mathrm{mix}, \mathrm{mot}, \mathrm{opp}, \neg\mathrm{opp}, \mathrm{tes}_1, \mathrm{tes}_2, d_1, d_2, d_3, d_4\}$*. The following domain-specific inference rules can be identified:* $\mathcal{R} = \mathcal{R}_d = \{d_1\colon \mathrm{for} \Rightarrow \mathrm{ftpr}$*;* $d_2\colon \mathrm{tes}_1 \Rightarrow \mathrm{mot}$*;* $d_3\colon \mathrm{ftpr}, \mathrm{mot}, \mathrm{opp} \Rightarrow \mathrm{bur}$*;* $d_4\colon\ \mathrm{tes}_2 \Rightarrow \neg\mathrm{opp}\}$*. The axiom and ordinary premises are* $\mathcal{K}_n = \{\mathrm{for}, \mathrm{mix}, \mathrm{tes}_1, \mathrm{tes}_2\}$ *and* $\mathcal{K}_p = \{\mathrm{opp}\}$*. Proposition* mix *can be interpreted as stating that inference rule d_1 is not applicable, as it provides an exception to the inference from* for *to* ftpr*; hence,* mix $\in \overline{d_1}$*. The constructed arguments then are:* A_1: for*;* A_2: $A_1 \Rightarrow$ ftpr*;* A_3: tes_1*;* A_4: $A_3 \Rightarrow$ mot*;* A_5: opp*;* A_6: $A_2, A_4, A_5 \Rightarrow$ bur*;* B_1: mix*;* C_1: tes_2*; and* C_2: $C_1 \Rightarrow \neg$opp*. To illustrate the operators used in Definition 7, for A_6, we have that* $\textsc{Prem}(A_6) = \{\text{for, tes}_1\text{, opp}\}$*;* $\textsc{Conc}(A_6) = \text{bur}$*;* $\textsc{Sub}(A_6) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$*;* $\textsc{TopInf}(A_6) = d_3$*.* □

2.2.2.2 Attack in ASPIC$^+$

In ASPIC$^+$, arguments can be *attacked* in three ways. We reiterate that attack in ASPIC$^+$ does not necessarily result in defeat and that preferences are applied to determine the defeat relation, as defined in Definition 9. An argument can be attacked on the application of a defeasible inference rule by providing exceptional circumstances under which it may not be applicable (*undercutting attack*), on the conclusion of a defeasible inference rule (*rebuttal*), or on an ordinary premise (*undermining attack*). The conclusion of a strict inference rule or the application of a strict inference rule cannot be attacked. Formally, attacks are defined as follows:

Definition 8 (ASPIC$^+$ attack relations, after Modgil and Prakken [2013])**.** *Let A and B be arguments constructed on the basis of an argument theory AT. Then A* attacks *B iff A undercuts, rebuts or undermines B, where:*

1. *A* undercuts *B (on B') iff* $\textsc{Conc}(A) \in \overline{n(d)}$ *for some* $B' \in \textsc{Sub}(B)$ *such that* $\textsc{TopInf}(B') = d$, $d \in \mathcal{R}_d$*.*
2. *A* rebuts *B (on B') iff* $\textsc{Conc}(A) = -\phi$ *and* $\textsc{Conc}(B') = \phi$ *for some* $B' \in \textsc{Sub}(B)$ *such that* $\textsc{TopInf}(B') = d$, $d \in \mathcal{R}_d$*. In such a case, A contrary-rebuts B iff* $\textsc{Conc}(A)$ *is a contrary of ϕ.*
3. *A* undermines *B (on B') iff* $\textsc{Conc}(A) = -\phi$ *for some* $B' \in \textsc{Sub}(B)$ *with* $\textsc{Conc}(B') = \phi \in \mathcal{K}_p$*. In such a case, A contrary-undermines B iff* $\textsc{Conc}(A)$ *is a contrary of ϕ.*

If an argument A attacks B on B, this is called a *direct* attack. If A attacks B on $B' \in \textsc{Sub}(B) \setminus \{B\}$, this is called an *indirect* attack. If A rebuts B and B rebuts A, this is called symmetric rebuttal. If A rebuts B and B does not rebut A, this is called asymmetric rebuttal. In ASPIC$^+$-style argument graphs, direct undercutting attacks are indicated by dashed hyperarcs directed from the node corresponding to $\textsc{Conc}(A)$ to a solid (hyper)arc labelled d for $\textsc{TopInf}(B) = d \in \mathcal{R}_d$. Direct rebuttals and direct undermining attacks are indicated by dashed arcs directed from the node corresponding to $\textsc{Conc}(A)$ to the node corresponding to $\textsc{Conc}(B)$.

Example 21. *Consider Example 20. Argument B_1 directly undercuts A_2 and B_1 indirectly undercuts A_6 (on A_2), as* $\textsc{Conc}(B_1) = \text{mix}$, mix $\in \overline{d_1}$*, and* $\textsc{TopInf}(A_2) = d_1 \in \mathcal{R}_d$*. Argument C_2 directly undermines A_5 and A_5 directly rebuts C_2, as* $\textsc{Conc}(A_5) = \text{opp} \in \mathcal{K}_p$ *and* $\textsc{Conc}(C_2) = \neg\text{opp}$*. C_2 then also indirectly undermines A_6 (on A_5). Direct attacks for this example are indicated in Figure 2.3.* □

2.2.2.3 Defeat and instantiating Dung's abstract approach

In ASPIC$^+$, attack is resolved into *defeat* on the basis of a binary *preference relation* $\preceq$ over the arguments. In case $A \preceq B$ and $B \npreceq A$, then B is strictly preferred to A, denoted by $A \prec B$. If an argument A undercuts, contrary-rebuts, or contrary-undermines B on B', then A is said to *preference-independent* attack B on B', otherwise A is said to *preference-dependent* attack B on B'.

Definition 9 (ASPIC$^+$ defeat relation, after Modgil and Prakken [2013])**.** *Let A attack B on B'. Then A* defeats *B iff A preference-independent attacks B on B', or A preference-dependent attacks B on B' and $A \nprec B'$.*

For further details as to why undercutting and contrary-attacks succeed as defeats independently of preferences, the reader is referred to Modgil and Prakken [2013].

Example 22. *Consider Examples 20 and 21. As undercutting attack is preference-independent, B_1 defeats A_2 and A_6. Furthermore, assuming that $A_5 \prec C_2$, it follows that C_2 defeats A_5 and A_6 and that A_5 does not defeat C_2.* □

Combining all the above, a structured argumentation framework is formally defined as follows:

Definition 10 (Structured argumentation frameworks, after Modgil and Prakken [2013])**.** *A* structured argumentation framework *defined by an argument theory AT is a triple $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$, where $\mathcal{A}$ is the smallest set of all finite arguments constructed from $\mathcal{K}$ in AS as defined by Definition 7, $\mathcal{C}$ consists of pairs of arguments (A, B), where $(A, B) \in \mathcal{C}$ iff A attacks B as defined by Definition 8, and $\preceq \subseteq \mathcal{A} \times \mathcal{A}$ is a binary preference relation over $\mathcal{A}$.*

Dung's abstract approach is then instantiated using ASPIC$^+$ arguments and the defeat relation as defined by Definition 9 as follows:

Definition 11 (Relating AFs to SAFs, after Modgil and Prakken [2014])**.** *An* abstract argumentation framework (AF) corresponding to a SAF $= (\mathcal{A}, \mathcal{C}, \preceq)$ *is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by $(\mathcal{A}, \mathcal{C}, \preceq)$.*

Note that without preferences (i.e. $\preceq = \emptyset$), it follows that $\mathcal{D} = \mathcal{C}$. Given an AF, we can use any semantics for argumentation frameworks as defined in Section 2.2.1 for determining the dialectical status of arguments.

Finally, the status of a conclusion $\phi \in \mathcal{L}$ of an argument is defined. Note that several definitions are possible; the following definition directly uses the notions of justified, defensible and overruled arguments.

Definition 12 (The status of conclusions, after Prakken and Vreeswijk [2002]). *Let* $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ *be a structured argumentation framework and let* $AF = (\mathcal{A}, \mathcal{D})$ *be the corresponding abstract argumentation framework. Let* $T \in \{$*complete, grounded, preferred, stable*$\}$. *For every* $\phi \in \mathcal{L}$*:*

1. ϕ *is* justified *under* T *semantics iff there exists an argument* $A \in \mathcal{A}$ *with* $\textsc{Conc}(A) = \phi$ *that is justified under* T *semantics;*
2. ϕ *is* defensible *under* T *semantics iff* ϕ *is not justified under* T *semantics and there exists an argument* $A \in \mathcal{A}$ *with* $\textsc{Conc}(A) = \phi$ *that is defensible under* T *semantics;*
3. ϕ *is* overruled *under* T *semantics iff* ϕ *is neither justified nor defensible under* T *semantics and all arguments with conclusion* ϕ *are overruled under* T *semantics.*

Example 23. *For Examples 20, 21 and 22, we have that* $\mathcal{A} = \{A_1, A_2, A_3, A_4, A_5, A_6, B_1, C_1, C_2\}$ *and* $\mathcal{D} = \{(B_1, A_2), (B_1, A_6), (C_2, A_5), (C_2, A_6)\}$. *Under any semantics,* A_6 *is overruled as it is defeated by justified argument* B_1. *Therefore, proposition* bur *is also overruled, as* A_6 *is the only argument in* $\mathcal{A}$ *with conclusion* bur. □

2.2.2.4 Using ASPIC$^+$ to model argument schemes

ASPIC$^+$ can be used to model argument schemes [Walton et al., 2008], which capture stereotypical defeasible patterns of argumentation in a given domain as a scheme with a set of premises and a conclusion, plus a set of critical questions that need to be answered before the scheme can be used to infer conclusions. Argument schemes and critical questions can guide the practical construction of arguments and counter-arguments as they provide guidelines for frequently occurring types of reasoning. An example is the argument scheme for arguments from expert opinion (after Prakken [2020], who adapted it from Walton and colleagues [2008]):

> E is an expert in domain D.
> E asserts that P.
> P is within D.
>
> ---
>
> Therefore, presumably, P.

This scheme includes the following critical questions:

1. How credible is E as an expert source?
2. Is E personally reliable as a source?
3. Is P consistent with what other experts assert?
4. Is E's assertion of P based on evidence?

Argument schemes can be modelled in ASPIC$^+$ by using defeasible inference rules in $\mathcal{R}_d$. Critical questions then correspond to undercutters, as they point towards exceptional circumstances under which the defeasible inference represented by the scheme may not be applicable.

2.3 Bayesian networks

Finally, Bayesian networks (BNs) [Jensen and Nielsen, 2007] are reviewed. A BN compactly represents a joint probability distribution $\Pr(\mathbf{V})$ over a finite set of discrete random variables $\mathbf{V}$. Each variable can take on a finite number of mutually exclusive and exhaustive values; we will refer to these values as the variable's *value space*. In this thesis, we often assume variables to be Boolean, where we write v to denote $\mathrm{V} = \mathit{true}$ and $\neg v$ to denote $\mathrm{V} = \mathit{false}$. There is a one-to-one correspondence between nodes and variables in BNs. Therefore, throughout this thesis the terms 'node' and 'variable' are used interchangeably. Variables and their associated nodes are indicated by capital letters, such as V_i or Bur (see e.g. Figure 2.4a). For claims and propositions (see Sections 2.1 and 2.2), as well as values of variables, lower case letters are used. Finally, boldface is used to indicate sets.

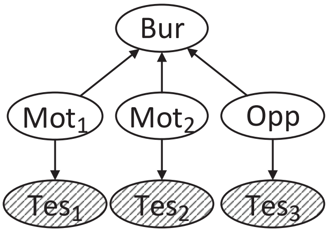
Formally, a BN is defined as follows:

Definition 13 (Bayesian network)**.** *A* Bayesian network *(BN) is a pair* $(G_{\mathcal{B}}, \Pr)$*.* $G_{\mathcal{B}}$ *is a directed acyclic graph (DAG)* $(\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ *over nodes* $\mathbf{V}$ *representing random variables.* $\mathbf{A}_{\mathcal{B}} \subseteq \mathbf{V} \times \mathbf{V}$ *is a set of ordered pairs* $(V_i,\ V_j)$ *representing arcs directed from parent* $V_i \in \mathbf{V}$ *to child* $V_j \in \mathbf{V}$*, indicated by* $V_i \to V_j$*, where* $\mathbf{Par}(V)$ *denotes the set of parents of* V *and* $\mathbf{Ch}(V)$ *denotes the set of children of* V*.* Pr *is a probability function which specifies for each variable* $V \in \mathbf{V}$ *a conditional probability table (CPT). This CPT describes the conditional probability distributions* $\Pr(V \mid x)$*, or* probability parameters*, for each possible joint value combination* x *for* $\mathbf{Par}(V)$*.*

The joint distribution $\Pr(\mathbf{V})$ now factorises into the conditional probability distributions specified for the graph. The graph in fact encodes the probabilistic independencies that hold among the represented variables (see Section 2.3.1).

The reflexive, transitive closures of V under the parent and child relations are denoted by $\mathbf{Par}^*(\mathrm{V})$ and $\mathbf{Ch}^*(\mathrm{V})$, respectively, where nodes in $\mathbf{Par}^*(\mathrm{V})$ are called *ancestors* of V and nodes in $\mathbf{Ch}^*(\mathrm{V})$ are called *descendants* of V. The set of *neighbours* of node V is defined as $\mathbf{Par}(\mathrm{V}) \cup \mathbf{Ch}(\mathrm{V})$.

A BN is generally used for *probabilistic inference* [Jensen and Nielsen, 2007], that is, calculating any probability of interest from any prior or posterior distribution over the variables represented in the network. Posterior distributions are obtained by *instantiating* one or more variables $\mathbf{E_V} \subseteq \mathbf{V}$ in that they are set to a specific value. Instantiations are also called evidence. The inference algorithms associated with the BN-formalism provide for computing probabilities of interest and for processing evidence; these algorithms constitute the basic building blocks for probabilistic reasoning with knowledge represented in the formalism. Reasoning is also possible at a more qualitative level, using the knowledge that is represented by a BN by means of its graphical structure $G_{\mathcal{B}}$ and properties of its probability function Pr. As the focus of this thesis lies on the latter, algorithms for probabilistic inference are not further discussed. Finally, we emphasise that various sets of instantiations $\mathbf{E_V}$ can be used to perform probabilistic inference with the same BN [van der Gaag and Meyer, 1996] (i.e. $\mathbf{E_V}$ is not a fixed set).



(a)

			Bur	
Mot_1	Mot_2	Opp	t	f
t	t	t	0.90	0.10
t	t	f	0.20	0.80
t	f	t	0.70	0.30
t	f	f	0.10	0.90
f	t	t	0.80	0.20
f	t	f	0.10	0.90
f	f	t	0.05	0.95
f	f	f	0.01	0.99

(b)

Figure 2.4: An example of a BN (a); CPT for node Bur (b), where Mot_1 and Mot_2 exhibit a negative product synergy wrt value Bur $=$ *true* in presence of uninstantiated parent Opp.

Example 24. *An example of a BN graph and one of its CPTs is depicted in Figure 2.4, where ovals represent nodes and nodes in a chosen set $\mathbf{E_V}$ are shaded. In this BN graph, we are interested in whether a given suspect committed a burglary (Bur). This node is connected by arcs to nodes Mot_1, Mot_2 and Opp, which describe whether the suspect had motive(s) and opportunity to commit the burglary. In turn, nodes Mot_1, Mot_2 and Opp are connected to instantiated nodes Tes_1, Tes_2 and Tes_3, which capture the testimonies provided to these claims.* $\square$

2.3.1 Bayesian network graphs

The BN graph $G_{\mathcal{B}}$ encodes the probabilistic independence relation among its variables. Independencies can be read from the graph by means of the notion of d-separation, which is defined by the notions of blocked and active chains. In the following, let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be a BN graph.

Definition 14 (Chain)**.** *A* chain $c = (V_1, Arc_1, V_2, \ldots, Arc_{n-1}, V_n)$ *is a sequence of distinct nodes* $V_1, \ldots, V_n \in \mathbf{V}$ *and arcs* $Arc_1, \ldots, Arc_{n-1} \in \mathbf{A}_{\mathcal{B}}$ *such that for every* $Arc_i, 1 \leq i < n$, *it holds that either* $Arc_i \equiv V_i \to V_{i+1}$ *or* $Arc_i \equiv V_{i+1} \to V_i$.

Definition 15 (Head-to-head node)**.** *A node $V \in \mathbf{V}$ is called a* head-to-head node *on a chain c in $G_{\mathcal{B}}$ if it has two incoming arcs on c.*

Definition 16 (Blocked chain)**.** *A chain c between nodes $V_1 \in \mathbf{V}$ and $V_2 \in \mathbf{V}$ in $G_{\mathcal{B}}$ is* blocked *by a (possibly empty) set of nodes $\mathbf{Z}$ iff it includes a node $V \notin \{V_1, V_2\}$ such that either:*

- *V is a head-to-head node on c and $\mathbf{Ch}^*(V) \cap \mathbf{Z} = \emptyset$, or;*
- *V has at most one incoming arc on c and $V \in \mathbf{Z}$.*

A chain that is not blocked by $\mathbf{Z}$ is called *active* given $\mathbf{Z}$.

Definition 17 (d-separation)**.** *Two sets of nodes $\mathbf{V_1} \subseteq \mathbf{V}$ and $\mathbf{V_2} \subseteq \mathbf{V}$ are* d-separated *by a (possibly empty) set of nodes $\mathbf{Z} \subseteq \mathbf{V}$ iff there exist no active chains between any node in $\mathbf{V_1}$ and any node in $\mathbf{V_2}$ given $\mathbf{Z}$.*

If $\mathbf{V_1}$ and $\mathbf{V_2}$ are d-separated given $\mathbf{Z} \subseteq \mathbf{V}$, then their corresponding variables are considered conditionally independent given $\mathbf{Z}$. Upon performing probabilistic inference, only nodes on active chains are involved in computations, where the set of instantiations $\mathbf{E_V}$ serves as the blocking set $\mathbf{Z}$ to determine which chains are active.

Example 25. *In Figure 2.4, given the evidence for* $\mathbf{E_V} = \{Tes_1,\ Tes_2,\ Tes_3\}$ *all chains between* Mot_1 *and* Mot_2 *are blocked, as Bur is an uninstantiated head-to-head node on chain* $(Mot_1,\ Mot_1 \rightarrow Bur,\ Bur,\ Mot_2 \rightarrow Bur,\ Mot_2)$ *without instantiated descendants (i.e.* $\mathbf{Ch}^*(Bur) \cap \mathbf{E_V} = \emptyset$*); hence, variables* Mot_1 *and* Mot_2 *are considered conditionally independent given the evidence for* $\mathbf{E_V}$*. Knowing* Mot_1*, therefore, does not affect the probabilities for* Mot_2 *and vice versa.* □

Given its Markov blanket, a node is conditionally independent from the rest of the nodes in the graph.

Definition 18 (Markov blanket)**.** *The* Markov blanket *of a node* V *is the set* $\mathbf{Ch}(V) \cup \mathbf{Par}(V) \cup \mathbf{Par}(\mathbf{Ch}(V)) \setminus \{V\}$.

Finally, we review the following concept from graph theory.

Definition 19 (Weakly connected component)**.** *Let* $G = (\mathbf{V}, \mathbf{A})$ *be a directed graph and let* $C = (\mathbf{V}^c, \mathbf{A}^c)$ *with* $\mathbf{V}^c \subseteq \mathbf{V}$ *and* $\mathbf{A}^c \subseteq (\mathbf{V}^c \times \mathbf{V}^c) \cap \mathbf{A}$ *be a sub-graph of* G*. Then* C *is a* weakly connected component *of* G *iff:*

1. *For every pair of nodes* $V_1, V_2 \in \mathbf{V}^c$ *there exists a chain between* V_1 *and* V_2 *in* C*;*
2. C *is a maximal sub-graph of* G *for which property 1 holds.*

2.3.2 Intercausal interactions and qualitative probabilistic constraints

Next, we review the concepts of intercausal interactions and qualitative probabilistic constraints. In case a head-to-head node or one of its descendants in a BN graph is instantiated, an active chain is *induced* between the parents of the head-to-head node, allowing for *intercausal interactions*. Note that, while the term 'intercausal interactions' is used, these interactions can also occur regardless of the type of relation between parents and child. If one of the parents takes on the value *true*, then the probability of another parent taking on this value as well may change, depending on the synergistic effect modelled in the CPT for the head-to-head node. In case the probability that one of the other parents takes on the value *true* decreases, this is called the *'explaining away'* effect [Druzdzel and Henrion, 1993]. For Boolean nodes, we will generally assume an ordering *true* > *false* on its values unless specified otherwise. In case this ordering is reversed, then the occurrences of these two values need to be interchanged in the equations appearing in Definitions 20 and 22. To achieve the explaining away effect between two parents V_1 and V_3 of V_2 for instantiation v_2, the CPT for V_2 needs to be constrained such that V_1 and V_3 exhibit a *negative product synergy wrt* v_2. First, we review the concept of product synergy I [Druzdzel and Henrion, 1993], which captures the special case in which all other parents of V_2 are instantiated. In the below definitions, $\Pr(v_2 \mid v_1, v_3, x)$ denotes the conditional probability of v_2 given the *conjunction* of v_1, v_3 and x.

Definition 20 (Product synergy I)**.** *Let* $\mathbf{B} = (G_{\mathcal{B}}, \Pr)$ *be a BN and let* $V_1, V_3 \in \mathbf{V}$ *be parents of* $V_2 \in \mathbf{V}$ *in* $G_{\mathcal{B}}$*. Let* $\mathbf{X} = \mathbf{Par}(V_2) \setminus \{V_1, V_3\}$ *and let* x *be the combination of observed values for* $\mathbf{X}$*. Then* V_1 *and* V_3 *exhibit a* negative product synergy *wrt* v_2*, written* $\mathbf{X}^-(\{V_1, V_3\}, v_2)$*, iff*

$$\Pr(v_2 \mid v_1, v_3, x) \cdot \Pr(v_2 \mid \neg v_1, \neg v_3, x) \leq \Pr(v_2 \mid v_1, \neg v_3, x) \cdot \Pr(v_2 \mid \neg v_1, v_3, x)$$

If $\mathbf{X} = \emptyset$, then this equation simplifies by leaving out every occurrence of x. V_1 and V_3 exhibit a *zero product synergy* wrt v_2, written $\mathbf{X}^0(\{\mathrm{V}_1,\mathrm{V}_3\}, v_2)$, if $\leq$ in the above equation is replaced by $=$. In this case, no direct intercausal interaction effect exists between parents V_1 and V_3 for value v_2 of V_2. V_1 and V_3 exhibit a *positive product synergy* wrt v_2, written $\mathbf{X}^+(\{\mathrm{V}_1,\mathrm{V}_3\}, v_2)$, if $\leq$ is replaced by $\geq$ in the above equation. In this case, the joint occurrence of the causes may be a more likely explanation of the common effect than would either of them considered individually.

Next, the case is considered in which $\mathbf{X} \neq \emptyset$ is not instantiated to a combination of values. First, we review the concept of *matrix half negative semi-definiteness*.

Definition 21 (Half negative semi-definite matrix)**.** *Let* M *be a square* $n \times n$ *matrix,* $n \geq 1$*, and let* x *be any non-negative vector* x *of* n *elements. Then* M *is called* half negative semi-definite *iff* $x^T M x \leq 0$*.*

Similarly, a square matrix M is called *half positive semi-definite* iff $x^T M x \geq 0$ for all non-negative vectors x of n elements. We now provide the definition of extended product synergy, termed *product synergy II* [Druzdzel and Henrion, 1993].

Definition 22 (Product synergy II)**.** *Let* $\mathbf{B} = (G_{\mathcal{B}}, \Pr)$ *be a BN and let* $V_1,\ V_3 \in \mathbf{V}$ *be parents of* $V_2 \in \mathbf{V}$ *in* $G_{\mathcal{B}}$*. Let* $\mathbf{X} = \mathbf{Par}(V_2) \setminus \{V_1,\ V_3\}$*. Let* n *denote the number of possible combinations of values for* $\mathbf{X}$*. Then* V_1 *and* V_3 *exhibit a* negative product synergy *wrt* v_2 *iff the* $n \times n$ *matrix* M *with elements*

$$M_{ij} = \Pr(v_2 \mid v_1, v_3, x_i) \cdot \Pr(v_2 \mid \neg v_1, \neg v_3, x_j) - \Pr(v_2 \mid v_1, \neg v_3, x_i) \cdot \Pr(v_2 \mid \neg v_1, v_3, x_j)$$

is half negative semi-definite for all combinations of values x_i *and* x_j *for* $\mathbf{X}$*.*

For a positive or zero product synergy, the matrix M has to be half positive semi-definite or zero, respectively. Note that product synergy I is a special case of product synergy II; hence, in referring to the general concept of product synergy throughout this thesis, we are referring to product synergy II.

Example 26. *Consider the BN of Figure 2.4. The entries of the CPT of Figure 2.4b are chosen such that* Mot_1 *and* Mot_2 *exhibit a negative product synergy wrt value* $Bur =$ true *in presence of uninstantiated parent Opp. Specifically, the* 2×2

matrix M consisting of the following four elements is half negative semi-definite:

$$M_{11} = 0.9 \cdot 0.05 - 0.7 \cdot 0.8 = -0.515; M_{12} = 0.9 \cdot 0.01 - 0.7 \cdot 0.1 = -0.061$$
$$M_{21} = 0.2 \cdot 0.05 - 0.1 \cdot 0.8 = -0.070; M_{22} = 0.2 \cdot 0.01 - 0.1 \cdot 0.1 = -0.008 \ \square$$

2.3.3 Bayesian network construction

BN construction is typically an iterative process [Druzdzel and van der Gaag, 2000]. After constructing an initial BN graph, it should be verified that it is acyclic and that it correctly captures the (conditional) independencies. If the graph does not yet exhibit these properties, arcs should be reversed, added or removed by the BN modeller in consultation with the domain expert. We call this the *'graph validation step'*. We note that software tools such as Matilda [Boneh et al., 2006] exist that can aid experts involved in BN construction in exploring whether (conditional) independencies are correctly modelled in the (initially) constructed BN.

The conditional probabilities of the BN are elicited in a separate quantification step. Probabilities are often taken from domain-specific literature, estimated from data sets by calculating (frequency) statistics, or elicited from domain experts [Druzdzel and van der Gaag, 2000]. Various techniques exist to obtain probabilities through expert elicitation, such as the use of numerical probability scales and reference lotteries [Renooij, 2001]. As the structure determines which conditional probabilities have to be estimated, the qualitative graph-construction step and the quantitative probability elicitation step are initially carried out sequentially. However, building a BN often requires a careful trade-off between wanting to obtain a rich and accurate model and having a model that is too computationally and representationally complex. Therefore, in practice the qualitative and quantitative steps are iteratively run through until a model is obtained that is deemed satisfactory [Druzdzel and van der Gaag, 2000].

The elicited probability estimates are inevitably inaccurate, as knowledge of the domain is typically partial and data is typically incomplete. Sensitivity analysis can be used to study the effects of uncertainties in probability assessments for a given set of nodes on a probability of interest [van der Gaag et al., 1999]. For instance, one type of sensitivity analysis involves systematically varying the probability assessments for nodes in a chosen set while keeping all other assessments fixed. Sensitivity analysis provides detailed insight into the level of accuracy that is required for the various probabilities of the network and can hence serve to facilitate the probability elicitation process. More specifically, in the initial phases of quantification only rough estimates of probabilities can be elicited, after which sensitivity analyses can be used to establish the level of accuracy that is required for the various probabilities of the network. Probability assessments can then be further refined if required.

To facilitate BN construction, construction methods have been proposed in the literature, including approaches for constructing BNs from information specified in arguments and ontologies and approaches for constructing BNs from fragments. These approaches are reviewed in Section 8.2.

Chapter 3

The information graph formalism

As illustrated in the introduction, various graph-based sense-making tools exist that allow domain experts such as legal experts to make sense of a mass of evidence in a case, including Wigmore charts [Wigmore, 1913], mind maps [Okada et al., 2014][1] and argument diagrams [Bex et al., 2003, 2013; Okada et al., 2014]. In this chapter we revisit our examples of analyses performed using informal sense-making tools from Section 1.1 in the light of our conceptual analysis of reasoning about evidence from Section 2.1 to illustrate that, when performing analyses using aforementioned tools, domain experts naturally mix deductive and abductive inference with the different types of generalisations distinguished in our conceptual analysis. In performing such analyses, the used generalisations and the inference type (deduction, abduction) are typically left implicit. Moreover, the assumptions of domain experts underlying their analyses are typically not explicitly stated and the various elements that can be incorporated in these analyses are often ambiguous and not precisely defined; hence, we conclude that, because of their informal nature, sense-making tools do not directly allow for guiding the construction of formal representations within AI systems such as argumentation frameworks [Dung, 1995] and BNs [Jensen and Nielsen, 2007].

Accordingly, in this chapter we set out to formalise and disambiguate analyses performed using informal sense-making tools in a manner that (1) allows for guiding the construction of formal representations within AI systems and that (2) is in line with our conceptual analysis of reasoning about evidence as provided in Section 2.1, while (3) allowing inference to be performed and visualised in a manner that is closely related to the way in which inference is performed and visualised by domain experts using such tools. In particular, we propose the *information graph* (IG) formalism, which formalises analyses performed using such tools in this manner.

[1] See also Master's research [Timmers, 2017].

3.1 Revisiting examples of analyses performed using sense-making tools

In this section, we revisit the examples from Section 1.1. Based on these examples, in Section 3.2 we then further motivate our IG-formalism.

3.1.1 Revisiting the Wigmore chart example

First, we revisit the Wigmore chart example from Section 1.1.1 to establish which (types of) generalisations and inferences could have been used in constructing the chart of Figure 1.1. In defining the elements of his charts, Wigmore is often ambiguous in the language he uses. For instance, an arc in the chart indicates that a claim is offered as *'... evidencing, or explaining, or corroborating ...'* [Wigmore, 1913, p. 752]. Arcs can therefore be regarded as indicating which claims are inferred from each other, where the generalisations used in performing these inferences, as well as the inference type (deduction or abduction), are not explicitly recorded in the chart. To be able to interpret whether inferences are deductive or abductive, and hence what the antecedents and consequents are of generalisations used in performing the inferences, the evidence in the chart also needs to be considered. One interpretation is that the arcs between claims 26 – 28.1 and claim 25 in Figure 1.1 represent deductive inferences from the testimonies to the claim to which is testified, where evidential generalisations of the form *'Testimony to x is evidence for x'* are used. Possible hypotheses are then proposed that explain claim 25, namely claims 29 – 33. The arcs between 25 and claims 29 – 33 can possibly be interpreted as deductive inferences using evidential generalisations; however, it seems contrived to consider the claim that Y. died in good health to be evidence for a *specific* cause of death (i.e. one of the claims 29 – 33). Instead, it makes more sense to consider that Wigmore made multiple abductive inference steps in constructing the chart. For instance, from the causal generalisation *'Y. had a former injury in his side is a cause for Y. dies'* and by affirming the consequent (25), the antecedent (31) is abductively inferred. However, establishing with certainty whether the inferences are deductive or abductive would require directly consulting John Henry Wigmore.

Lastly, we note that the manner in which claims and links conflict is also not precisely specified in Wigmore charts. For instance, claim 40 weakens the chain of inferences from 38.1 to 39 via 38; however, the precise manner in which claim 40 opposes (inferences between) claims 39.1, 39 and 38 is left unspecified.

3.1.2 Revisiting the mind map example

Next, we revisit the mind map example presented in Section 1.1.2. As the edges in a mind map are undirected, it is unclear from the graphical representation alone which types of generalisations and inferences were used in constructing the map depicted in Figure 1.2. Establishing this with certainty would require directly consulting the domain experts involved in constructing the chart. We note, however, that the reasoning performed in constructing this mind map can be interpreted in multiple ways; we illustrate this for the part of the mind map concerning Leo's cause of death. One interpretation of this part of the mind map is that the domain expert first (preliminarily) inferred that Leo died because of a head wound from the autopsy report via deductive inference using the evidential generalisation g'_1: *autopsy* $\rightarrow_{\mathrm{e}}$ *head_wound*, and then abductively inferred *hit_angular* using the causal generalisation g'_2: *hit_angular* $\rightarrow_{\mathrm{c}}$ *head_wound*. In turn, *hammer* and *stone* are abductively inferred from *hit_angular* using the abstractions g'_3: *hammer* $\rightarrow_{\mathrm{a}}$ *hit_angular* and g'_4: *stone* $\rightarrow_{\mathrm{a}}$ *hit_angular*. These two claims are then competing alternative explanations of *hit_angular* and are subsequently grounded in evidence, namely via deductive inference from the testimonies using evidential generalisations g'_5: *testimony_5* $\rightarrow_{\mathrm{e}}$ *hammer* and g'_6: *testimony_6* $\rightarrow_{\mathrm{e}}$ *stone*. An alternative interpretation is that the mind map was constructed iteratively from the evidence, where from the testimonies the claims *hammer* and *stone* are inferred via deductive inference using generalisations g'_5 and g'_6. Claim *hit_angular* is then inferred modus-ponens style: from abstractions g'_3 and g'_4 and the previously inferred antecedents, the consequent is deductively inferred. In this way, *hammer* and *stone* are not in competition for *hit_angular*.

This example illustrates that the types of generalisations and inferences involved in the analysis of a case using a mind mapping tool are typically left implicit. Similarly, the manner in which claims and links conflict is not precisely specified: a minus symbol can either indicate support for the opposing claim (e.g. *testimony_3* supports *marjan_did_not_murder_leo*) or indicate an exception to the performed inference (e.g. *lie* opposes the inference from *testimony_3* to *marjan_did_not_murder_leo*). Similarly, conflicts between competing alternative explanations such as *hit_angular* and *fell_on_table* are not explicitly indicated in the graph.

3.2 Motivating the information graph formalism

The examples from Section 3.1 make it plausible that both deductive and abductive inference are performed by domain experts when performing analyses using sense-making tools they are familiar with. In performing such analyses, the used generalisations, as well as the inference type (deduction, abduction), are left implicit. Furthermore, the manner of conflict (e.g. negatory conflict, exception-based conflict, conflict between competing alternative explanations) is typically not precisely specified and the assumptions of domain experts underlying their analyses are typically not explicitly stated, making these analyses ambiguous to interpret. For current purposes, we wish to provide a precise account of the interplay between the different types of inferences and generalisations that formalises and disambiguates these analyses in a manner that makes the used generalisations explicit. Existing formalisms that allow for both deduction and abduction with different types of information are logic-based [Bex, 2011; Poole, 1989; Shanahan, 1989]; instead, we propose the graph-based IG-formalism to remain closely related to the way analyses are visualised using aforementioned graph-based tools as well as the BN-formalism. Moreover, in contrast with existing formalisms we put special emphasis on the constraints we argue should be imposed on the types of inferences that may be performed with the different types of generalisations. Finally, compared to the ASPIC$^+$ framework (Section 2.2.2), which only allows for deductive reasoning, we allow for both deductive and abductive reasoning and introduce a new type of conflict, namely conflict between competing alternative explanations, which is currently not accounted for in that framework. Relations to existing formalisms are further discussed in Section 8.1.

An IG serves as a source of unambiguous information that can be used to guide the construction of instantiations of AI systems for which a formal reasoning mechanism is defined. In particular, in Chapter 4 we define an argumentation formalism based on IGs that allows us to assign Dung's argumentation semantics to argumentation frameworks constructed on the basis of IGs, and in Chapter 5 we demonstrate the use of the IG-formalism in guiding BN construction by serving as an intermediary formalism between analyses performed using informal sense-making tools and BNs. Viewed this way, in the context of argumentation the IG-formalism is comparable to the Argument Interchange Format (AIF) [Rahwan and Reed, 2009], an argumentation ontology that serves as an intermediary formalism between analyses performed using argument diagramming tools [Bex et al., 2003; Okada et al., 2014] and formal argumentation frameworks such as the ASPIC$^+$ framework (see Bex et al. [2013]). Compared to the AIF, our IG-formalism is tailored to model the process of reasoning about evidence in that it provides a precise account of the interplay between the different types of inferences and generalisations. Moreover, the AIF does not have an inference engine and is only intended as a notation for argumentation and arguments. More specifically, the elements of AIF graphs are typed in terms of argumentation-theoretical concepts such as inference and conflict, in contrast with the IG-formalism which allows for actually performing inference. Accordingly, we wish to refrain from using the AIF as an intermediary formalism between analyses performed using informal sense-making tools and formal AI systems. For similar reasons, we wish to refrain from using other ontologies for this purpose (see also Section 8.2).

Information graphs (IGs), which we define in Section 3.3, are knowledge representations that explicitly describe generalisations in the graph. In constructing an IG from an analysis performed using a tool, an interpretation step may be required; we provide examples of this interpretation step by discussing possible formalisations of the mind map of Section 1.1.2. In Section 3.4 we then define how deductive and abductive inferences can be read from IGs given the evidence, based on our conceptual analysis of reasoning about evidence (Section 2.1). In Section 3.5 we define sequences of propositions that are iteratively inferred from each other given the evidence, termed inference chains, for which we prove a number of properties. Finally, in Section 3.6 we discuss further extending our inference constraints.

3.3 Information graphs

First, the syntax of IGs is defined. Throughout this thesis, boldface is used to indicate sets used in the IG-formalism.

Definition 23 (Information graph)**.** *An* information graph *(IG) is a directed graph $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$, where $\mathbf{P}$ is a set of nodes representing propositions from a propositional language consisting of only literals and that is closed under classical negation, where the negation symbol is denoted by $\neg$. $\mathbf{A}_{\mathcal{I}}$ is a set of (hyper)arcs that divides into three pairwise disjoint subsets $\mathbf{G}$, $\mathbf{N}$ and $\mathbf{Exc}$ of generalisation arcs, negation arcs and exception arcs, defined in Definitions 24, 28, and 29, respectively.*

For IGs, there is a one-to-one correspondence between nodes and propositions, generalisation arcs and generalisations, exception arcs and exceptions, and negation arcs and negations. Throughout this thesis, in the context of IGs, the terms 'node' and 'proposition', 'generalisation arc' and 'generalisation', 'exception arc' and 'exception', and 'negation arc' and 'negation' are therefore used interchangeably. In figures in this thesis, rectangles represent proposition nodes. Similarly as for ASPIC$^+$, we write $p = -q$ in case $p = \neg q$ or $q = \neg p$. Finally, note that while we currently only consider classical negation, our IG-formalism may be extended in future work to allow for more general notions of conflicts such as contrariness (cf. ASPIC$^+$). Contrariness generalises the notion of negatory conflict to among other things allow conflict among more than two propositions, which can also be indicated using certain argument diagramming tools, such as Rationale (see also Bex et al. [2013]).

Definition 24 (Generalisation arc)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. A* generalisation arc *$g \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ is a directed (hyper)arc $g\colon \{p_1, \ldots, p_n\} \rightarrow p$, indicating a generalisation with antecedents $\mathbf{P_1} = \{p_1, \ldots, p_n\} \subseteq \mathbf{P}$ and consequent $p \in \mathbf{P} \setminus \mathbf{P_1}$. Here, propositions in $\mathbf{P_1}$ are called the* tails *of g, denoted by $\mathbf{Tails}(g)$, and p is called the* head *of g, denoted by $Head(g)$. $\mathbf{G}$ divides into four pairwise disjoint subsets $\mathbf{G}^{\mathrm{c}}$, $\mathbf{G}^{\mathrm{e}}$, $\mathbf{G}^{\mathrm{a}}$ and $\mathbf{G}^{\mathrm{o}}$ of causal generalisation arcs, evidential generalisation arcs, abstraction arcs, and all other types of generalisation arcs, respectively. Generalisations in $\mathbf{G}^{\mathrm{c}}$ and $\mathbf{G}^{\mathrm{e}}$ are defeasible, $\mathbf{G}^{\mathrm{a}}$ divides into disjoint subsets $\mathbf{G}^{\mathrm{a}}_{\mathrm{s}}$ and $\mathbf{G}^{\mathrm{a}}_{\mathrm{d}}$ of strict and defeasible abstraction arcs, respectively, and $\mathbf{G}^{\mathrm{o}}$ divides into disjoint subsets $\mathbf{G}^{\mathrm{o}}_{\mathrm{s}}$ and $\mathbf{G}^{\mathrm{o}}_{\mathrm{d}}$ of strict and defeasible other types of generalisation arcs, respectively. Defeasible and strict generalisations are then denoted by $\mathbf{G}_{\mathrm{d}} = \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{e}} \cup \mathbf{G}^{\mathrm{a}}_{\mathrm{d}} \cup \mathbf{G}^{\mathrm{o}}_{\mathrm{d}}$ and $\mathbf{G}_{\mathrm{s}} = \mathbf{G}^{\mathrm{a}}_{\mathrm{s}} \cup \mathbf{G}^{\mathrm{o}}_{\mathrm{s}}$. For $g \in \mathbf{G}$, $\mathbf{Tails}(g)$ divides into disjoint subsets $\mathbf{Enabler}(g)$ and $\mathbf{Ant}(g)$ of propositions representing enabling conditions and actual antecedents of the generalisation, respectively, where for $g \in \mathbf{G}^{\mathrm{c}}$ it holds that $\mathbf{Ant}(g) \neq \emptyset$ and possibly $\mathbf{Enabler}(g) = \emptyset$, and for $g \in \mathbf{G}^{\mathrm{e}} \cup \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ it holds that $\mathbf{Enabler}(g) = \emptyset$ (i.e. $\mathbf{Tails}(g) = \mathbf{Ant}(g)$).*

Curly brackets are omitted in case $|\mathbf{Tails}(g)| = 1$. In figures in this thesis, generalisation arcs are denoted by solid (hyper)arcs, which are labelled 'c' for $g \in \mathbf{G}^{\mathrm{c}}$, 'e' for $g \in \mathbf{G}^{\mathrm{e}}$, and 'a' for $g \in \mathbf{G}^{\mathrm{a}}$, where 'o' labels for $g \in \mathbf{G}^{\mathrm{o}}$ are omitted.

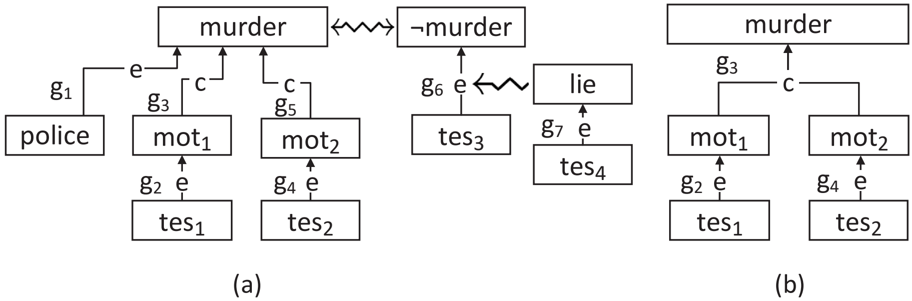


Figure 3.1: An IG corresponding to a possible interpretation of the part of the mind map of Figure 1.2 concerning Marjan's involvement in the murder (a), where $\leftrightsquigarrow$ is a negation arc and $\rightsquigarrow$ is an exception arc; adjustment to the IG of Figure 3.1a, where generalisation arc $g_3\colon \{mot_1,\ mot_2\} \rightarrow murder$ is now included instead of arcs $g_3\colon\ mot_1 \rightarrow murder$ and $g_5\colon\ mot_2 \rightarrow murder$ (b).

In accordance with our assumptions stated in Section 2.1, causal and evidential generalisations are defeasible and only causal generalisations can include enabling conditions. Abstractions and generalisations of type 'other' can either be strict or defeasible. A causal generalisation $g\colon c \rightarrow e$ may have an evidential counterpart of the form $g'\colon e \rightarrow c$ (see Section 2.1.3), but only if c is the usual cause of e. Definition 24 does not prohibit the coexistence of a causal generalisation $g\colon c \rightarrow e$ and its evidential counterpart $g'\colon e \rightarrow c$ in an IG, and inferences can be read from IGs including both generalisations without yielding anomalous results; hence, both generalisations may be included if this is considered desirable. However, it should be noted that g and g' represent the same knowledge, and that care should be taken in for instance modelling exceptions to generalisations (see Definition 29), as an exception to g can also be considered an exception to g'. Ultimately, it is the responsibility of the knowledge engineer in consultation with the domain expert to decide which knowledge to include in the IG and to ensure this knowledge is correctly and consistently represented.

In the following example, the mind map of Section 1.1.2 is modelled as an IG.

Example 27. *In Figure 3.1a, an IG is depicted for a possible interpretation of the part of the mind map of Figure 1.2 concerning Marjan's involvement in the murder. First, we consider the undirected edges connected to the testimonies and the police report in that part of the mind map. In an empirical study in the legal domain, van den Braak and colleagues [2008] found that subjects often considered testimonies to be evidential, where generalisations are of the form* 'Testimony to x is evidence for x'. *Police reports can similarly be considered evidential. The IG therefore includes generalisation arcs $g_1, g_2, g_4, g_7 \in \mathbf{G}^{\mathrm{e}}$ to denote these generalisations. As* tes_3 *concerns Marjan's testimony to denying any involvement in the murder, proposition* $\neg$ murder *is included in* $\mathbf{P}$ *and* $g_6\colon\ \mathrm{tes}_3 \rightarrow \neg\mathrm{murder}$ *in* $\mathbf{G}^{\mathrm{e}}$. *A motive for commit-*

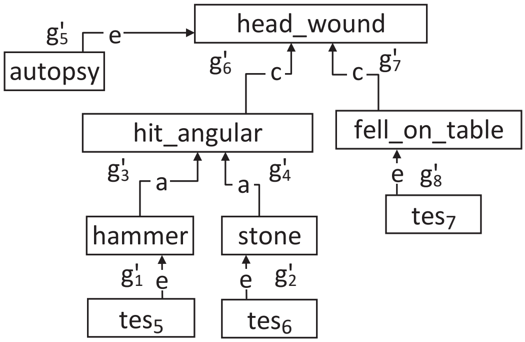


Figure 3.2: IG corresponding to a possible interpretation of the part of the mind map of Figure 1.2 concerning Leo's cause of death, where 'a' labels denote abstractions.

ting an act can be considered a cause for committing that act [van den Braak et al., 2008]. The IG therefore includes generalisation arcs $g_3\colon$ $\mathrm{mot}_1 \to \mathrm{murder}$ *and* $g_5\colon$ $\mathrm{mot}_2 \to \mathrm{murder}$ *in* $\mathbf{G}^{\mathrm{c}}$ *to denote these generalisations.*

Consider Figure 3.2, which depicts an IG for a possible interpretation of the part of the mind map of Figure 1.2 concerning Leo's cause of death. The generalisations used in the inferences from the testimonies, as well as from autopsy*, are considered to be evidential; therefore, generalisation arcs* g'_1, g'_2, g'_5 *and* g'_8 *are included in* $\mathbf{G}^{\mathrm{e}}$*. The relation between* hammer *(*stone*) and* hit_angular *is neither causal nor evidential; instead, generalisation arcs* g'_3 *and* g'_4 *are included in* $\mathbf{G}^{\mathrm{a}}_{\mathrm{d}}$ *to express that, at a higher level of abstraction, both hammers and stones can generally be considered angular objects. These generalisations are defeasible as not all hammers and stones are angular. Finally,* hit_angular *and* fell_on_table *can both be considered causes of* head_wound*; therefore, generalisation arcs* g'_6 *and* g'_7 *are included in* $\mathbf{G}^{\mathrm{c}}$*.* □

The following example illustrates generalisation arcs that include enablers.

Example 28. *Consider* $g''_7\colon \{\mathrm{fell\_on\_table}, \mathrm{no\_helmet}\} \to \mathrm{head\_wound}$ *in* $\mathbf{G}^{\mathrm{c}}$*, which is an adjustment to generalisation* g'_7 *of Example 27 which states that falling on a table causes a head wound in case you are not wearing a helmet. As in Example 27, proposition* fell_on_table *expresses a cause for* head_wound *and hence,* fell_on_table *is included in* $\mathbf{Ant}(g''_7)$*. Proposition* no_helmet *does not express a cause for* head_wound *and can thus be considered an enabler of* g''_7*; therefore,* no_helmet *is included in* $\mathbf{Enabler}(g''_7)$*. It should be noted that, while* no_helmet *does not express a cause for the consequent, it still is a necessary condition of generalisation* g''_7*.* □

Specific configurations of generalisation arcs express that two sets of propositions are *alternative explanations* of a common proposition, as captured by Definition 25. The terminology used is illustrated in Figure 3.3.

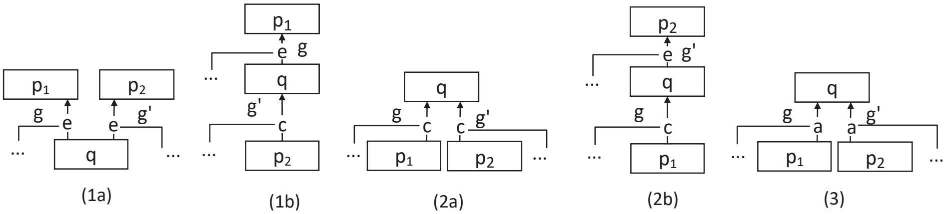


Figure 3.3: Illustration of the terminology used in Definition 25, where $p_1 \in \mathbf{P_1}$, $p_2 \in \mathbf{P_2}$.

Definition 25 (Alternative explanations)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG. Then sets* $\mathbf{P_1} \subseteq \mathbf{P}$ *and* $\mathbf{P_2} \subseteq \mathbf{P}$ *are* alternative explanations *of* $q \in \mathbf{P}$*, as indicated by generalisations* g *and* g' *in* $\mathbf{G}$*, iff one of the following holds:*

1. $\mathbf{P_1} = \{p_1\}$, $g \in \mathbf{G}^{\mathrm{e}}$, $Head(g) = p_1$, $q \in \mathbf{Tails}(g)$, *and either:*
   1a) $\mathbf{P_2} = \{p_2\}$, $g' \in \mathbf{G}^{\mathrm{e}}$, $g' \neq g$, $Head(g') = p_2$, $q \in \mathbf{Tails}(g')$, *or;*
   1b) $g' \in \mathbf{G}^{\mathrm{c}}$, $Head(g') = q$, $\mathbf{P_2} \subseteq \mathbf{Ant}(g')$.
2. $g \in \mathbf{G}^{\mathrm{c}}$, $Head(g) = q$, $\mathbf{P_1} \subseteq \mathbf{Ant}(g)$, *and either:*
   2a) $g' \in \mathbf{G}^{\mathrm{c}}$, $g' \neq g$, $Head(g') = q$, $\mathbf{P_2} \subseteq \mathbf{Ant}(g')$, *or;*
   2b) $\mathbf{P_2} = \{p_2\}$, $g' \in \mathbf{G}^{\mathrm{e}}$, $Head(g') = p_2$, $q \in \mathbf{Tails}(g')$.
3. $g \in \mathbf{G}^{\mathrm{a}}$, $Head(g) = q$, $\mathbf{P_1} \subseteq \mathbf{Tails}(g)$ *and* $g' \in \mathbf{G}^{\mathrm{a}}$, $g' \neq g$, $Head(g') = q$, $\mathbf{P_2} \subseteq \mathbf{Tails}(g')$.

Note that cases 1b and 2b are symmetrical in terms of $\mathbf{P_1}$ and $\mathbf{P_2}$ and the used generalisations; we opt to keep the distinction between these two cases as they simplify the proof of Proposition 1. For singleton sets $\mathbf{P_1} = \{p_1\}$ and $\mathbf{P_2} = \{p_2\}$, we will simply say that propositions p_1 and p_2 are alternative explanations of q instead of sets $\mathbf{P_1}$ and $\mathbf{P_2}$. In case 1a, q is an antecedent of both $g \in \mathbf{G}^{\mathrm{e}}$ and $g' \in \mathbf{G}^{\mathrm{e}}$; hence, both p_1 and p_2 are causes of q. In case 1b, propositions in $\mathbf{P_2}$ are actual antecedents and not enablers of $g' \in \mathbf{G}^{\mathrm{c}}$ and thus propositions in $\mathbf{P_2}$ express a cause for q, and q is an antecedent of $g \in \mathbf{G}^{\mathrm{e}}$ and thus p_1 is a cause of q. In case 2a, propositions in $\mathbf{P_1}$ and propositions in $\mathbf{P_2}$ are actual antecedents of $g \in \mathbf{G}^{\mathrm{c}}$ and $g' \in \mathbf{G}^{\mathrm{c}}$, respectively; hence, propositions in $\mathbf{P_1}$ and propositions in $\mathbf{P_2}$ are actual causes of q. Finally, in case 3, propositions in $\mathbf{P_1}$ and propositions in $\mathbf{P_2}$ are antecedents of abstractions $g \in \mathbf{G}^{\mathrm{a}}$ and $g' \in \mathbf{G}^{\mathrm{a}}$ with the same consequent q, and hence are considered alternative explanations of q.

Note that by Definition 25, different antecedents of the same causal generalisation or abstraction do not express alternative explanations of the consequent (enforced by assuming that $g' \neq g$), which is in accordance with our assumption that only the antecedents together allow us to infer the consequent (see Section 2.1). Also note that Definition 25 only captures that sets $\mathbf{P_1}$ and $\mathbf{P_2}$ are alternative explanations of q (where possibly $\mathbf{P_1} \cap \mathbf{P_2} \neq \emptyset$) and not whether these sets are in *competition* for q; the concept of competing alternative explanations is defined in Section 3.4.5.

Example 29. *Consider the IG of Figure 3.2. According to condition* 2*a of Definition 25,* hit_angular *and* fell_on_table *are alternative explanations of* head_wound *as indicated by generalisations g'_6 and g'_7. Similarly, according to condition* 3 *of Definition 25,* hammer *and* stone *are alternative explanations of* hit_angular *as indicated by generalisations g'_3 and g'_4.* □

Generalisation chains are sequences of generalisation arcs.

Definition 26 (Generalisation chain)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Generalisation arcs $g_1, \ldots, g_m \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ form a* generalisation chain *$[g_1, \ldots, g_m]$ in $G_{\mathcal{I}}$ iff $Head(g_{i-1}) \in \mathbf{Tails}(g_i)$ for $1 < i \leq m$.*

Note that a subchain of a generalisation chain is again a generalisation chain.

Example 30. *In the IG of Figure 3.1a, $[g_2, g_3]$ is a generalisation chain as $Head(g_2) = \text{mot}_1 \in \mathbf{Tails}(g_3)$.*

Consider the IG of Figure 3.1b, which is an adjustment to the IG of Figure 3.1a in which generalisation arc $g_3\colon \{\text{mot}_1, \text{mot}_2\} \to \text{murder}$ in $\mathbf{G}^{\text{c}}$ is included instead of two separate generalisation arcs g_3 and g_5. According to Definition 26, $[g_2, g_3]$ is a generalisation chain, but mot_2 is neither a head nor a tail of generalisation arc g_2; it suffices that $Head(g_2) = \text{mot}_1 \in \mathbf{Tails}(g_3)$. □

We introduce the following terminology regarding generalisation chains.

1. A generalisation chain $[g_1, \ldots, g_m]$ is called *non-repetitive* if $\nexists i, j \in \{1, \ldots, m\}$ such that $Head(g_i) = Head(g_j)$.
2. A generalisation chain $[g_1, \ldots, g_m]$ is called *consistent* if $\nexists i, j \in \{1, \ldots, m\}$ such that $Head(g_i) = -Head(g_j)$.

The importance of the concepts 'non-repetitive' and 'consistent' generalisation chains will become apparent in Chapter 5, namely in studying conditions on IGs under which BN graphs constructed by the approach defined in that chapter are guaranteed to be acyclic (Section 5.2.1).

We define the following notion of a *causal cycle*.

Definition 27 (Causal cycle)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Proposition $p \in \mathbf{P}$ expresses a* direct cause *for $q \in \mathbf{P}$ iff $\exists g \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ with $g \in \mathbf{G}^{\text{c}}$, $p \in \mathbf{Ant}(g)$, $q = Head(g)$ or $g \in \mathbf{G}^{\text{e}}$, $p = Head(g)$, $q \in \mathbf{Tails}(g)$. Proposition $p_1 \in \mathbf{P}$ expresses an* indirect cause *for $p_3 \in \mathbf{P}$ iff $\exists p_2 \in \mathbf{P}$, $p_2 \neq p_1, p_2 \neq p_3$, such that p_1 expresses a direct cause for p_2 and p_2 expresses a direct or indirect cause for p_3. A* causal cycle *exists in $G_{\mathcal{I}}$ iff $\exists p, q \in \mathbf{P}$ such that p expresses a direct or indirect cause for $q \in \mathbf{P}$ and q or $-q$ expresses a direct or indirect cause for p or for $-p$.*

Examples of IGs including causal cycles are provided in Figure 3.4. In this thesis, we generally assume that IGs do not include causal cycles (see also Poole [1994]).

A negation arc captures a conflict between a proposition and its negation expressed in an IG.

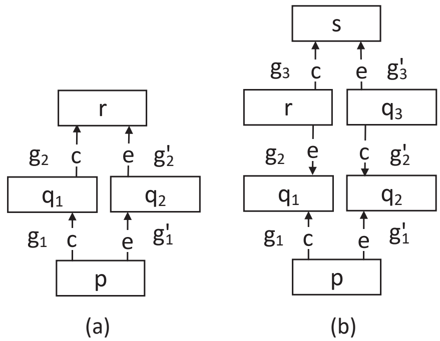


Figure 3.4: Examples of IGs including causal cycles.

Definition 28 (Negation arc)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG. A* negation arc $n \in \mathbf{N} \subseteq \mathbf{A}_{\mathcal{I}}$ *is a bidirectional arc* $n\colon p \leftrightsquigarrow q$ *in* $G_{\mathcal{I}}$ *that exists between a pair* $p, q \in \mathbf{P}$ *iff* $q = -p$*.*

Example 31. *Consider the running example. As both* murder *and* ¬murder *are included in the IG of Figure 3.1a, negation arc* $n\colon$ murder $\leftrightsquigarrow$ ¬murder *is also included in the graph.* □

As defeasible generalisations do not hold universally, exceptional circumstances can be provided under which such a generalisation may not hold; hence, we allow exceptions to generalisations in $\mathbf{G}_{\mathrm{d}}$ to be specified in IGs by means of exception arcs.

Definition 29 (Exception arc)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG. An* exception arc $exc \in \mathbf{Exc} \subseteq \mathbf{A}_{\mathcal{I}}$ *is a hyperarc* $exc\colon p \rightsquigarrow g$*, where* $p \in \mathbf{P}$ *is called an* exception *to defeasible generalisation* $g \in \mathbf{G}_{\mathrm{d}}$*.*

An exception arc directed from p to g indicates that p provides exceptional circumstances under which g may not hold.

Example 32. *In the running example, proposition* lie*, which states that Marjan had reason to lie when giving her testimony, provides an exception to evidential generalisation* $g_6\colon$ $\text{tes}_3 \rightarrow \neg\text{murder}$ *in* $\mathbf{G}^{\mathrm{e}}$*. In Figure 3.1a, this is indicated by a curved hyperarc* $exc\colon$ lie $\rightsquigarrow g_6$ *in* $\mathbf{Exc}$*.* □

3.4 Reading inferences from information graphs

We now define how deductive and abductive inferences can be performed with constructed IGs. By itself, a generalisation arc only expresses that the tails together allow us to infer the head in case this generalisation is used in deductive inference, or that the tails together can be inferred from the head in case of abductive inference. Only when considering the available evidence can directionality of inference actually be read from the graph.

Definition 30 (Evidence set)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG. An* evidence set *is a subset* $\mathbf{E_p} \subseteq \mathbf{P}$ *for which it holds that for every* $p \in \mathbf{E_p}$*,* $\neg p \notin \mathbf{E_p}$*.*

The restriction that for every $p \in \mathbf{E_p}$ it holds that $\neg p \notin \mathbf{E_p}$ ensures that not both a proposition and its negation are observed.

In figures in this thesis, nodes in $G_{\mathcal{I}}$ corresponding to elements of $\mathbf{E_p}$ are shaded and all shaded nodes correspond to elements of $\mathbf{E_p}$, comparable to the manner in which shading is used to denote evidence in BN graphs. We emphasise that various evidence sets $\mathbf{E_p}$ can be used to establish (different) inferences from the same IG.

Example 33. *In the running example, the evidence consists of the testimonies, the police report and the autopsy report. In Figure 3.5, the IG of Figure 3.1a is again depicted, with nodes in* $\mathbf{E_p} = \{\mathrm{tes}_1,\ \mathrm{tes}_2,\ \mathrm{tes}_3,\ \mathrm{tes}_4,\ \mathrm{police}\}$ *shaded. Similarly, in Figure 3.9, the IG of Figure 3.2 is again depicted, with nodes in* $\mathbf{E_p} = \{\mathrm{tes}_5,\ \mathrm{tes}_6,\ \mathrm{tes}_7,\ \mathrm{autopsy}\}$ *shaded.* □

We now define when we consider configurations of generalisation arcs and evidence to express deductive and abductive inference.

3.4.1 Deductive inference

First, we specify under which conditions we consider a configuration of generalisation arcs and evidence to express deductive inference, where strict and defeasible deduction are distinguished.

Definition 31 (Deductive inference)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. Let* $p_1, \ldots, p_n, q \in \mathbf{P}$*, with* $q \notin \mathbf{E_p}$*. Then given* $\mathbf{E_p}$*,* q *is* deductively inferred from propositions $p_1, \ldots, p_n$ using a generalisation $g\colon \{p_1, \ldots, p_n\} \to q$ in $\mathbf{G}$ *iff* $\forall p_i$*,* $i = 1, \ldots, n$*:*

1. $p_i \in \mathbf{E_p}$*, or;*
2. p_i *is deductively inferred from propositions* $r_1, \ldots, r_m \in \mathbf{P}$ *using a generalisation* $g'\colon \{r_1, \ldots, r_m\} \to p_i$*, where* $g' \notin \mathbf{G}^{\mathrm{c}}$ *if* $g \in \mathbf{G}^{\mathrm{e}}$*, or;*
3. p_i *is abductively inferred from a proposition* $r \in \mathbf{P}$ *using a generalisation* $g'\colon \{p_i, r_1, \ldots, r_m\} \to r$ *in* $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$*,* $g \neq g'$*,* $r_1, \ldots, r_m \in \mathbf{P}$ *(see Definition 32).*

Here, proposition q *is* defeasibly deductively inferred from $p_1, \ldots, p_n$*, denoted* $p_1, \ldots, p_n \twoheadrightarrow_g q$*, iff* $g \in \mathbf{G}_{\mathrm{d}}$*, and proposition* q *is* strictly deductively inferred from $p_1, \ldots, p_n$*, denoted* $p_1, \ldots, p_n \rightharpoonup_g q$*, iff* $g \in \mathbf{G}_{\mathrm{s}}$*.*

For ease of reference, symbols $\twoheadrightarrow$ and $\rightharpoonup$ are annotated with the name of the generalisation used in performing a defeasible or strict inference. In accordance with our assumptions stated in Section 2.1.1, deduction can be performed using all types of generalisations in $\mathbf{G}$, where strict deduction can only be performed using strict abstractions and strict generalisations of type 'other'. The condition $q \notin \mathbf{E_p}$ ensures that deduction cannot be performed with a generalisation to infer its consequent in case its consequent is already observed. Deduction can only be performed using a

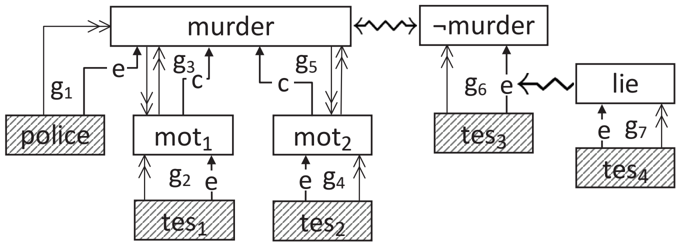


Figure 3.5: The IG of Figure 3.1a, where evidence $\mathbf{E_p}$ and resulting inference steps ($\twoheadrightarrow$) are also indicated.

generalisation $g \in \mathbf{G}$ to infer its consequent $Head(g)$ from its antecedents $\mathbf{Tails}(g)$ in case every antecedent $p_i \in \mathbf{Tails}(g)$ has been affirmed in that either p_i is observed (i.e. $p_i \in \mathbf{E_p}$), p_i itself is deductively inferred, or p_i is abductively inferred. In correspondence with Pearl's constraint (see Section 2.1.4.1), we assume in condition 2 that a proposition $q \in \mathbf{P}$ cannot be deductively inferred from $p_1, \ldots, p_n \in \mathbf{P}$ using a generalisation $g \in \mathbf{G}^{\mathrm{e}}$ if at least one of its antecedents $p_i \in \mathbf{Tails}(g)$ is deductively inferred using a generalisation $g' \in \mathbf{G}^{\mathrm{c}}$. In this case, q and $\mathbf{Ant}(g')$ are considered alternative explanations of p_i as indicated by g and g' (Definition 25, case 1b or case 2b). Condition 3 of Definition 31 is explained in Section 3.4.3, after abductive inference is defined.

Example 34. *In the IG of Figure 3.5, given* $\mathbf{E_p}$ mot_1 *and* mot_2 *are deductively inferred from* tes_1 *and* tes_2 *using generalisations* g_2 *and* g_4*, respectively, as* tes_1, $\mathrm{tes}_2 \in \mathbf{E_p}$ *(condition* 1 *of Definition 31). Similarly,* murder, ¬murder *and* lie *are deductively inferred from* police, tes_3 *and* tes_4 *using generalisations* g_1, g_6 *and* g_7*, respectively, as* police, tes_3, $\mathrm{tes}_4 \in \mathbf{E_p}$.

Proposition murder *is also deductively inferred from* mot_1 *and* mot_2 *using causal generalisations* g_3 *and* g_5*, as* mot_1 *and* mot_2 *are deductively inferred (condition* 2 *of Definition 31). This illustrates mixed deductive inference using both evidential and causal generalisations.* □

The following example illustrates strict deductive inference.

Example 35. *Consider Example 4 from Section 2.1.1. In this example, generalisation arc* g: lung_cancer → cancer *is included in* $\mathbf{G}^{\mathrm{a}}_{\mathrm{S}}$. *As* lung_cancer $\in \mathbf{E_p}$, cancer *is strictly deductively inferred from* lung_cancer *(Definition 31, condition* 1*).* □

The next example illustrates the restrictions put on performing deduction in our IG-formalism.

Example 36. *Figure 3.6a depicts an example of an IG in which* q *cannot be deductively inferred from* p *using* g_1*, as* $Head(g_1) = q \in \mathbf{E_p}$. *In Figure 3.6b,* q *cannot be deductively inferred from* p_1 *and* p_2 *using* g_1*, as* $p_2 \notin \mathbf{E_p}$ *and* p_2 *is neither deductively nor abductively inferred.*

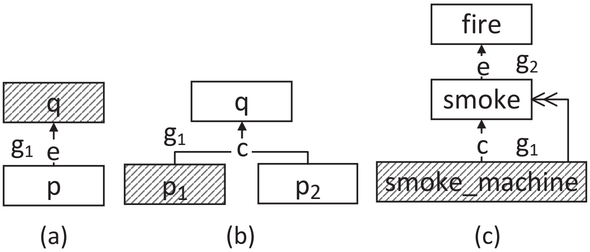


Figure 3.6: Examples of IGs illustrating the restrictions put on performing deductive inference within our IG-formalism (a-c).

In Figure 3.6c, Example 10 illustrating Pearl's constraint is modelled. As smoke_machine $\in \mathbf{E_p}$, smoke *is deductively inferred from* smoke_machine *using* g_1 *(Definition 31, condition* 1*).* fire *cannot in turn be inferred from* smoke *using* g_2 *(Definition 31, condition* 2*), as* $g_2 \in \mathbf{G}^{\mathrm{e}}$ *and* smoke *is deductively inferred using* $g_1 \in \mathbf{G}^{\mathrm{c}}$. □

3.4.2 Abductive inference

Next, we specify under which conditions we consider a configuration of generalisation arcs and evidence to express abductive inference.

Definition 32 (Abductive inference)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. Let* $p_1, \ldots, p_n, q \in \mathbf{P}$*, with* $\{p_1, \ldots, p_n\} \cap \mathbf{E_p} = \emptyset$*. Then given* $\mathbf{E_p}$, propositions $p_1, \ldots, p_n$ are abductively inferred from q using a $g\colon \{p_1, \ldots, p_n\} \rightarrow q$ in $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$*, denoted* $q \twoheadrightarrow_g p_1; \ldots ; q \twoheadrightarrow_g p_n$*, iff:*

1. *$q \in \mathbf{E_p}$, or;*
2. *q is deductively inferred from propositions $r_1, \ldots, r_m \in \mathbf{P}$ using a generalisation $g'\colon \{r_1, \ldots, r_m\} \rightarrow q$ in $\mathbf{G}$, $g \neq g'$ (see Definition 31), where $g' \notin \mathbf{G}^{\mathrm{c}}$ if $g \in \mathbf{G}^{\mathrm{c}}$ and $g' \notin \mathbf{G}^{\mathrm{a}}$ if $g \in \mathbf{G}^{\mathrm{a}}$, or;*
3. *q is abductively inferred from a proposition $r \in \mathbf{P}$ using a generalisation $g'\colon \{q, r_1, \ldots, r_m\} \rightarrow r$ in $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$, $r_1, \ldots, r_m \in \mathbf{P}$.*

In accordance with our assumptions stated in Section 2.1.2, abduction is defeasible and is modelled using only causal generalisations and abstractions. Following Console and Dupré [1994] and Bex [2011], we assume that abductive inference can be performed with both strict and defeasible abstractions, where such an inference is always defeasible as it concerns an inference from the more abstract consequent to a more specific antecedent (see Section 2.1.2). The condition $\{p_1, \ldots, p_n\} \cap \mathbf{E_p} = \emptyset$ ensures that abduction cannot be performed with a generalisation to infer its antecedents in case at least one of its antecedents is already observed. Furthermore, abductive inference can only be performed using a generalisation $g \in \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$ to infer its antecedents $\mathbf{Tails}(g)$ from its consequent $Head(g)$ in case $Head(g)$ has been affirmed in that either $Head(g)$ is observed (i.e. $Head(g) \in \mathbf{E_p}$), $Head(g)$ is deductively inferred, or $Head(g)$ is itself abductively inferred.

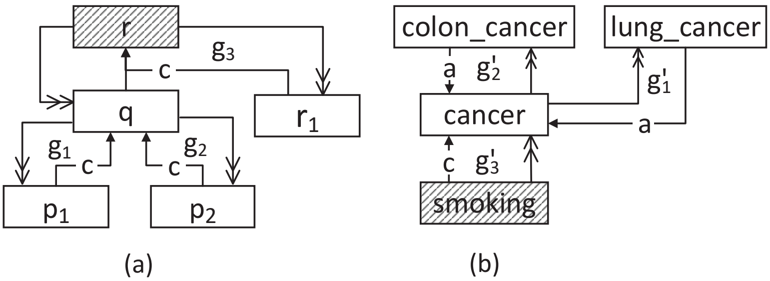


Figure 3.7: Example of an IG illustrating abduction with causal generalisations (a); example of an IG illustrating abduction with abstractions (b).

In correspondence with Pearl's constraint (see Section 2.1.4.1), we assume in condition 2 that propositions $p_1, \ldots, p_n \in \mathbf{P}$ cannot be abductively inferred from a proposition $q \in \mathbf{P}$ using a generalisation $g \in \mathbf{G}^{\mathrm{c}}$ if its consequent q is deductively inferred using a generalisation $g' \neq g$, $g' \in \mathbf{G}^{\mathrm{c}}$. In enforcing this constraint, we do not need to consider whether or not the antecedents of g or g' include enablers, as illustrated in Example 13 from Section 2.1.4.1. More specifically, in Definition 24 it is assumed that $\forall g \in \mathbf{G}^{\mathrm{c}}$, $\mathbf{Ant}(g) \neq \emptyset$; therefore, sets $\mathbf{Ant}(g)$ and $\mathbf{Ant}(g')$ are alternative explanations of q according to case 2a of Definition 25 which may not be inferred from each other by inferring q as an intermediary step. Similarly, we assume in condition 2 that $g' \notin \mathbf{G}^{\mathrm{a}}$ if $g \in \mathbf{G}^{\mathrm{a}}$ to account for our constraints on performing deduction and abduction in that order with two abstractions (see Section 2.1.4.2). In this case, $\mathbf{Tails}(g)$ and $\mathbf{Tails}(g')$ are alternative explanations of q as indicated by g and g' according to case 3 of Definition 25.

Example 37. *In the IG of Figure 3.7a, q and r_1 are abductively inferred from r using generalisation $g_3\colon \{q, r_1\} \to r$ in $\mathbf{G}^{\mathrm{c}}$ by condition 1 of Definition 32, as $r \in \mathbf{E_p}$. Then by condition 3 of Definition 32, p_1 and p_2 are abductively inferred from q using g_1 and g_2, respectively.* □

The following example further illustrates abductive inference with abstractions.

Example 38. *In Figure 3.7b, Example 16 from Section 2.1.4.2 is modelled as an IG. As* smoking $\in \mathbf{E_p}$, cancer *is deductively inferred from* smoking *using g'_3. Propositions* lung_cancer *and* colon_cancer *are then abductively inferred from* cancer *using strict abstractions g'_1 and g'_2, respectively (Definition 32, condition 2). Hence, in this example, a cause (*smoking*) for an event (*cancer*) is known, after which this event is inferred and is in turn further specified at a lower level of abstraction (*lung_cancer *or* colon_cancer*). As noted in Section 2.1.4.2, this type of mixed inference using a causal generalisation and abstractions does not lead to undesirable results.* □

The following examples illustrate that Pearl's constraint for mixed deductive-abductive inference (see Section 2.1.4.1), as well as our proposed constraints on performing inference with abstractions (see Section 2.1.4.2), are adhered to.

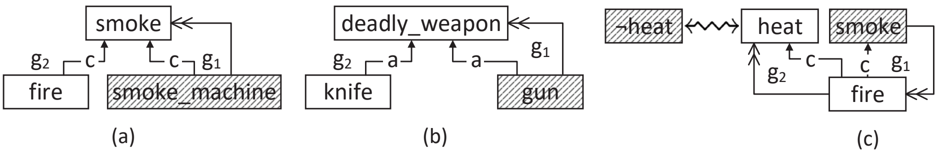


Figure 3.8: An IG illustrating Pearl's constraint for mixed deductive-abductive inference (a); an IG illustrating our inference constraints for abstractions (b); an IG illustrating mixed abductive-deductive inference (c).

Example 39. *In Figure 3.8a, Example 11 is modelled as an IG. As* smoke_machine $\in \mathbf{E_p}$, smoke *is deductively inferred from* smoke_machine *using* g_1. *Proposition* fire *cannot be inferred from* smoke, *as* $g_2 \in \mathbf{G}^{\mathrm{c}}$ *and* smoke *is deductively inferred using* $g_1 \in \mathbf{G}^{\mathrm{c}}$ *(Definition 32, condition 2).*

In Figure 3.8b, Example 14 is modelled as an IG. As gun $\in \mathbf{E_p}$, deadly_weapon *is deductively inferred from* gun *using* g_1. *Proposition* knife *cannot in turn be inferred from* deadly_weapon, *as* $g_2 \in \mathbf{G}^{\mathrm{a}}$ *and* deadly_weapon *is deductively inferred using* $g_1 \in \mathbf{G}^{\mathrm{a}}$ *(Definition 32, condition 2).* □

The following example describes the inferences that can be made based on the IG of Figure 3.2 corresponding to the part of the mind map example of Section 1.1.2 concerning Leo's cause of death.

Example 40. *Consider the IG of Figure 3.9. Given* $\mathbf{E_p} = \{\mathrm{tes}_5, \mathrm{tes}_6, \mathrm{tes}_7, \mathrm{autopsy}\}$, head_wound *is deductively inferred from* autopsy *using* g'_5. *Then,* hit_angular *and* fell_on_table *are abductively inferred from* head_wound *using* g'_6 *and* g'_7, *respectively (Definition 32, condition 2).* head_wound *is also deductively inferred from* fell_on_table *using* g'_7, *as* fell_on_table *is deductively inferred from* tes_7 *using* g'_8. *Propositions* hammer *and* stone *are abductively inferred from* hit_angular *using* g'_3 *and* g'_4, *respectively (Definition 32, condition 3).* hit_angular *is also deductively inferred from* hammer *and* stone *using* g'_3 *and* g'_4, *respectively, as* hammer *is deductively inferred from* tes_5 *using* g'_1 *and* stone *is deductively inferred from* tes_6 *using* g'_2. *Then* head_wound *is deductively inferred from* hit_angular *using* g'_6. □

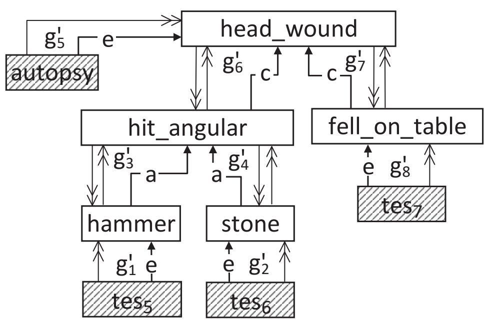


Figure 3.9: The IG of Figure 3.2, where evidence set $\mathbf{E_p}$ (shaded) and resulting inference steps ($\twoheadrightarrow$) are also indicated.

3.4.3 Mixed abductive-deductive inference

As apparent from Definitions 31 and 32, mixed abductive-deductive inference can be performed within our IG-formalism.

Example 41. *In Figure 3.8c, Example 9 from Section 2.1.4 is modelled as an IG. From* smoke, fire *is abductively inferred using* g_1, *as* smoke $\in \mathbf{E_p}$. *Then* heat *is deductively inferred (or predicted) from* fire *using* g_2 *(Definition 31, condition* 3*).*□

3.4.4 Ambiguous inference

The conditions under which we consider a configuration of generalisation arcs and evidence to express deductive and abductive inference according to Definitions 31 and 32 are not mutually exclusive. Under specific conditions, both inference types can be established from the same causal generalisation or abstraction in an IG given the provided evidence; the inference type is, therefore, ambiguous. As noted in Section 2.1.5, ambiguous inference patterns may arise in practice and, therefore, we do not wish to prohibit them from occurring.

Example 42. *Consider the IG of Figure 3.5. Given* $\mathbf{E_p}$, murder *is deductively inferred from* police *using* g_1 *and* mot_1 *and* mot_2 *are deductively inferred from* tes_1 *and* tes_2 *using* g_2 *and* g_4, *respectively. As* murder, $\text{mot}_1, \text{mot}_2 \notin \mathbf{E_p}$, murder *is deductively inferred from* mot_1 *and* mot_2 *and* mot_1 *and* mot_2 *are abductively inferred from* murder *using* g_3 *and* g_5, *respectively.* □

3.4.5 Competing alternative explanations

Next, we consider how the concept of competing alternative explanations (see Section 2.1.2) is captured within our IG-formalism.

Definition 33 (Competing alternative explanations)**.** *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. Let* $g, g' \in \mathbf{G}^{\text{c}}$ *or* $g, g' \in \mathbf{G}^{\text{a}}$. *Then given* $\mathbf{E_p}$, $\mathbf{Ant}(g)$ *is considered to be* in competition *with* $\mathbf{Ant}(g')$ *for proposition* q *in case* $\mathbf{Ant}(g)$ *and* $\mathbf{Ant}(g')$ *are alternative explanations of* q *as indicated by* g *and* g' *(Definition 25, case* 2a *or case* 3*) and in case* $\mathbf{Ant}(g)$ *and* $\mathbf{Ant}(g')$ *are abductively inferred from* q *given* $\mathbf{E_p}$ *using* g *and* g', *respectively.*

The above definition captures competition between *sets* of propositions $\mathbf{Ant}(g)$ and $\mathbf{Ant}(g')$ (where possibly $\mathbf{Ant}(g) \cap \mathbf{Ant}(g') \neq \emptyset$), as these sets are abductively inferred from q using g and g', respectively (see Section 2.1.2). More specifically, individual propositions in $\mathbf{Ant}(g)$ are not in competition with individual propositions in $\mathbf{Ant}(g')$. In case a causal generalisation or abstraction has multiple antecedents, then these antecedents are not in competition *among themselves*.

Example 43. *Consider Figure 3.10a, which depicts an adjustment to the IG of Figure 3.1a. Given* $\mathbf{E_p} = \{\text{police}\}$, *propositions* mot_1 *and* mot_2 *are abductively*

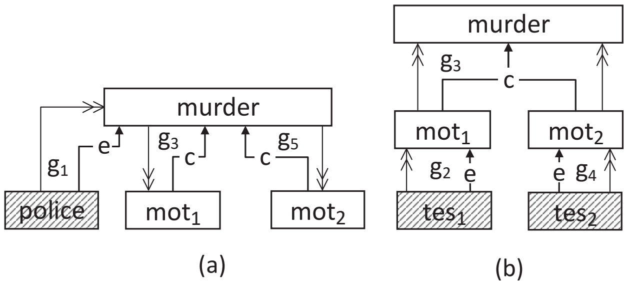


Figure 3.10: Adjustment to the IG of Figure 3.1a involving two competing alternative explanations mot_1 and mot_2 for *murder* (a); the IG of Figure 3.1b with evidence $\mathbf{E_p}$ and resulting inference steps now indicated, involving two non-competing alternative explanations mot_1 and mot_2 for *murder* (b).

inferred from murder *using g_3 and g_5, respectively, as* murder *is deductively inferred from* police *using g_1. Therefore,* mot_1 *and* mot_2 *are in competition for common effect* murder*. To allow for the possibility that both motives are true, the additional generalisation arc* $\{\text{mot}_1, \text{mot}_2\} \to$ murder *can be included in* $\mathbf{G}^{\text{c}}$*. In this case,* mot_1 *and* mot_2 *are no longer considered mutually exclusive causes for* murder*. In case only* mot_1 *and* mot_2 *together are considered a cause for* murder*, only generalisation arc* $\{\text{mot}_1, \text{mot}_2\} \to$ murder *should be included in the IG and separate generalisation arcs g_3 and g_5 should be excluded.*

In Figure 3.10b, the IG of Figure 3.1b is again depicted, where evidence $\mathbf{E_p} = \{\text{tes}_1, \text{tes}_2\}$ *and resulting inferences are also indicated. In this IG,* murder *is deductively inferred from* $\{\text{mot}_1, \text{mot}_2\}$ *given* $\mathbf{E_p}$ *using* $g_3\colon \{\text{mot}_1, \text{mot}_2\} \to$ murder *in* $\mathbf{G}^{\text{c}}$*; therefore,* mot_1 *and* mot_2 *are* not *in competition for* murder. □

Finally, note that a causal generalisation $g_1\colon c_1 \to e$ may be replaced by an evidential generalisation $g_1'\colon e \to c_1$ if c_1 is the usual cause of e, in which case abduction with g_1 can be encoded as deduction with g_1' (see Section 2.1.3). Considering the case in which only g_1 and not g_1' is included in IG $G_{\mathcal{I}}$ and additional causal generalisation $g_2\colon c_2 \to e$ is provided, then upon observing e propositions c_1 and c_2 are abductively inferred from e using generalisation arcs g_1 and g_2, respectively, which are then competing alternative explanations of e according to Definition 33. However, in case only g_1' and g_2 are included in $G_{\mathcal{I}}$ and not g_1, then upon observing e c_1 is deductively inferred from e using g_1' and c_2 is abductively inferred from e using g_2; hence, in this case propositions c_1 and c_2 are not competing alternative explanations as deduction and not abduction is performed using g_1'. Hence, if the knowledge engineer considers c_1 and c_2 to be competing alternative explanations of e, then the involved generalisations should be modelled as causal generalisations. We reiterate that it is the responsibility of the knowledge engineer in consultation with the domain expert to decide which knowledge (including conflicts) to represent in an IG and to ensure this knowledge is modelled correctly (see also Section 3.3, p. 44).

3.5 Inference chains and their properties

Finally, we introduce the notion of an *inference chain*, which describes a sequence of propositions that are iteratively inferred from each other given $\mathbf{E_p}$. We then prove a number of formal properties of inference chains. The importance of inference chains will become apparent in Chapter 5, where we compare the reasoning patterns that can be read from IGs to the reasoning patterns captured in BN graphs constructed from IGs by the approach defined in that chapter.

First, we define the concept of a *chain* for IGs.

Definition 34 (Chain in an IG). *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $\{p_1, \ldots, p_n\} \subseteq \mathbf{P}$ and let $\mathbf{G}' = \{g_1, \ldots, g_{n-1}\} \subseteq \mathbf{G}$. Then $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ is a* chain *in $G_{\mathcal{I}}$ iff for all $1 < i \leq n$ it either holds that $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$ or $Head(g_{i-1}) = p_{i-1}$, $p_i \in \mathbf{Tails}(g_{i-1})$.*

We now define when a chain in an IG is an inference chain.

Definition 35 (Inference chain). *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $\mathbf{G}' = \{g_1, \ldots, g_{n-1}\} \subseteq \mathbf{G}$, and let $\{p_1, \ldots, p_n\} \subseteq \mathbf{P}$ such that $\nexists i, j \in \{1, \ldots, n\}$ with $p_i = -p_j$, and such that $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ is a chain in $G_{\mathcal{I}}$. Let $p_1 \in \mathbf{E_p}$ or let p_1 be deductively or abductively inferred using a generalisation $g \in \mathbf{G} \setminus \mathbf{G}'$ given $\mathbf{E_p}$ (see Definitions 31 and 32). Then chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ is an* inference chain *in $G_{\mathcal{I}}$ given $\mathbf{E_p}$ iff for all $1 < i \leq n$ it holds that:*

1. *p_i is deductively inferred using generalisation $g_{i-1} \in \mathbf{G}'$ given $\mathbf{E_p}$ (see Definition 31), where $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$, or;*
2. *p_i is abductively inferred from p_{i-1} using generalisation $g_{i-1} \in \mathbf{G}'$ given $\mathbf{E_p}$ (see Definition 32), where $Head(g_{i-1}) = p_{i-1}$, $p_i \in \mathbf{Tails}(g_{i-1})$.*

We emphasise that an inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ does not only describe that p_{i-1} was used in inferring p_i for all $1 < i \leq n$; it also describes that the inference chain needs to start in a proposition p_1 that is either observed or inferred, hence the conditions regarding p_1 in Definition 35. We refer to the assumption that for inference chains $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ it holds that all p_i are distinct (enforced by assuming that $\{p_1, \ldots, p_n\} \subseteq \mathbf{P}$) as our *non-repetitiveness* assumption on inference chains. We refer to the assumption that for $\{p_1, \ldots, p_n\}$ it holds that $\nexists i, j \in \{1, \ldots, n\}$ with $p_i = -p_j$ as our *consistency* assumption on inference chains. Note that a subchain $(p_i, g_i, p_{i+1}, g_{i+1}, \ldots, p_{m-1}, g_{m-1}, p_m)$ of an inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ for $1 \leq i, m \leq n$ is again an inference chain. Finally, generalisation $g\colon \{p_1, \ldots, p_n\} \to q$ is said to be *on an inference chain given* $\mathbf{E_p}$ if there exists an inference chain (p_i, g, q) or (q, g, p_i) for $p_i \in \mathbf{Tails}(g)$ given $\mathbf{E_p}$.

Compared to generalisation chains (see Definition 26), which are solely captured by the graphical structure of IGs, inference chains can only be read from an IG by considering the evidence $\mathbf{E_p}$. In case an inference chain only describes deductive

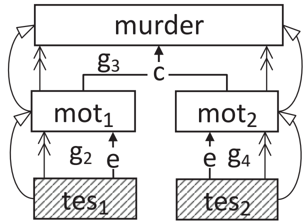


Figure 3.11: The IG of Figure 3.10b, where inference chains are also indicated by connecting arcs with open arrowheads.

inferences, then our non-repetitiveness and consistency assumptions on inference chains coincide with our non-repetitiveness and consistency assumptions on generalisation chains as described in Section 3.3 (p. 47); however, these assumptions do not coincide in case an inference chain also describes abductive inferences.

The following example illustrates the concept 'inference chain' and how it compares to the concept 'generalisation chain'.

Example 44. *In the IG of Figure 3.11,* (tes$_1$*,* g_2*,* mot$_1$*,* g_3*,* murder) *is an inference chain given* $\mathbf{E_p}$*, as* mot$_1$ *is deductively inferred from* tes$_1 \in \mathbf{E_p}$ *using* g_2*, where* $Head(g_2) =$ mot$_1$ *and* tes$_1 \in \mathbf{Tails}(g_2)$*, and* murder *is deductively inferred from* mot$_1$ *and* mot$_2$ *using* g_3*, where* $Head(g_3) =$ murder *and* mot$_1 \in \mathbf{Tails}(g_3)$*. In this IG,* $[g_2, g_3]$ *is also a generalisation chain (see Example 30). Note that the presence of this inference chain does not imply that* mot$_1$ *is by itself sufficient to infer* murder*; instead,* murder *can only be deductively inferred using* g_3 *in case both* mot$_1$ and mot$_2$ *are affirmed. The broader context in which the inference from* mot$_1$ *to* murder *is performed using* g_3 *is thus not directly apparent from this inference chain; instead, the role of* mot$_2$ *becomes apparent in considering other inference chains that can be read from this IG given* $\mathbf{E_p}$*, specifically inference chain* (tes$_2$*,* g_4*,* mot$_2$*,* g_3*,* murder)*.*

In the IG of Figure 3.5, (police*,* g_1*,* murder*,* g_3*,* mot$_1$) *is an inference chain given* $\mathbf{E_p}$*:* murder *is deductively inferred from* police $\in \mathbf{E_p}$ *using generalisation* g_1 *and* mot$_1$ *is abductively inferred from* murder *using generalisation* g_3*. However,* $[g_1, g_3]$ *is not a generalisation chain, as* $Head(g_1) =$ murder $\notin \mathbf{Tails}(g_3)$*.* □

Note that in Definition 35 it is assumed that inference chains $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ do not need to start in evidence in that it does not need to hold that $p_1 \in \mathbf{E_p}$, as long as p_1 is deductively or abductively inferred using a $g \in \mathbf{G} \setminus \mathbf{G}'$ given $\mathbf{E_p}$.

Example 45. *In Figure 3.11,* (mot$_1$*,* g_3*,* murder) *is an inference chain given* $\mathbf{E_p}$*:* murder *is deductively inferred from* mot$_1$ *and* mot$_2$ *using* g_3*. However,* mot$_1 \notin \mathbf{E_p}$*; instead,* mot$_1$ *is deductively inferred using* $g_2 \in \mathbf{G} \setminus \{g_3\}$ *given* $\mathbf{E_p}$*.* □

The next example illustrates that inference chains are generally not symmetrical.

Example 46. *In the IG of Figure 3.11,* (tes$_1$, g_2, mot$_1$, g_3, murder) *is an inference chain (see Example 44), but* (murder, g_3, mot$_1$, g_2, tes$_1$) *is not an inference chain as* mot$_1$ *cannot be inferred from* murder *using* g_3 *and* tes$_1$ *cannot be inferred from* mot$_1$ *using* g_2*.* □

We prove the following properties of inference chains that will be used in Chapter 5. Lemma 1 states that for inference chains, only the first proposition can possibly be observed.

Lemma 1. *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. Let* $p_1, \ldots, p_n \in \mathbf{P}$*,* $g_1, \ldots, g_{n-1} \in \mathbf{G}$ *and let* $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ *be an inference chain in* $G_{\mathcal{I}}$ *given* $\mathbf{E_p}$*. Then* $p_i \notin \mathbf{E_p}$ *for* $i > 1$*.*

Proof. Let $i > 1$. In case p_i is deductively inferred from p_{i-1} using g_{i-1}, then $p_i = Head(g_{i-1}) \notin \mathbf{E_p}$ per the restrictions of Definition 31. Similarly, in case p_i is abductively inferred from p_{i-1} using g_{i-1}, then $p_i \in \mathbf{Tails}(g_{i-1})$ and hence $p_i \notin \mathbf{E_p}$ per the restriction of Definition 32 that $\mathbf{Tails}(g_{i-1}) \cap \mathbf{E_p} = \emptyset$. ■

Lemma 2 states that an inference step between two consecutive propositions p_i and p_{i+1} in an inference chain can only be performed with a generalisation g_i for which it holds that $Head(g_i) = p_i$ and $p_{i+1} \in \mathbf{Tails}(g_i)$ in case g_i is a *causal* generalisation or an *abstraction*.

Lemma 2. *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. Let* $p_1, \ldots, p_n \in \mathbf{P}$*,* $g_1, \ldots, g_{n-1} \in \mathbf{G}$ *and let* $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ *be an inference chain in* $G_{\mathcal{I}}$ *given* $\mathbf{E_p}$*. Let* $i \in \{1, \ldots, n-1\}$ *and assume that* $Head(g_i) = p_i$*,* $p_{i+1} \in \mathbf{Tails}(g_i)$*. Then* $g_i \in \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$*.*

Proof. Assume that $G_{\mathcal{I}}$ includes a generalisation arc g_i with $Head(g_i) = p_i$ and $p_{i+1} \in \mathbf{Tails}(g_i)$. Then the associated generalisation's antecedent p_{i+1} cannot be inferred from its consequent p_i in case $g_i \in \mathbf{G}^{\mathrm{e}} \cup \mathbf{G}^{\mathrm{o}}$, as this would be an instance of abductive inference while per the restrictions of Definition 32 abductive inference can only be performed using generalisation arcs in $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$. ■

In performing inference care should be taken that no cause for an effect is inferred if an alternative cause for this effect was already previously inferred (i.e. Pearl's constraint, see Section 2.1.4.1). Similarly, care should be taken that no version of an event at a lower level of abstraction is inferred if an alternative version of this event at a lower level of abstraction was already previously inferred (i.e. our inference constraints for abstractions, see Section 2.1.4.2). In the context of IGs, for $g \in \mathbf{G}^{\mathrm{c}}$, propositions in $\mathbf{Ant}(g)$ express a cause for the common effect expressed by $Head(g)$, for $g \in \mathbf{G}^{\mathrm{e}}$, $Head(g)$ expresses the usual cause for propositions in $\mathbf{Tails}(g)$, and for $g \in \mathbf{G}^{\mathrm{a}}$, propositions in $\mathbf{Tails}(g)$ are at a lower level of abstraction than $Head(g)$. Hence, in defining how inferences can be read from IGs, restrictions are put in Definitions 31 and 32 such that our inference constraints (see Section 2.1.4) are adhered to. We now formally prove that the inference chains that can be read from an IG given an $\mathbf{E_p}$ indeed never violate these constraints.

First, we define the inference constraints of Section 2.1.4 in the context of IGs.

Definition 36 (Inference constraints)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $\mathbf{P_1} \subseteq \mathbf{P}$ and $\mathbf{P_2} \subseteq \mathbf{P}$ be alternative explanations of $q \in \mathbf{P}$ as indicated by generalisations g_1 and g_2 in $\mathbf{G}$ (see Definition 25). Let $p_1 \in \mathbf{P_1}$ and $p_2 \in \mathbf{P_2}$. Assume that inference chain (p_1, g_1, q) exists in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. Then chain (p_1, g_1, q, g_2, p_2) is not an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$.*

We now formally prove that these inference constraints are indeed adhered to.

Proposition 1 (Adherence to inference constraints)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Then any inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$ adheres to the inference constraints as defined in Definition 36.*

Proof. Assume that $\mathbf{P_1} \subseteq \mathbf{P}$ and $\mathbf{P_2} \subseteq \mathbf{P}$ are alternative explanations of $q \in \mathbf{P}$ as indicated by generalisations g_1 and g_2 in $\mathbf{G}$ with $p_1 \in \mathbf{P_1}$ and $p_2 \in \mathbf{P_2}$, and assume that inference chain (p_1, g_1, q) exists in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. Then we need to prove that chain (p_1, g_1, q, g_2, p_2) is not an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. In performing the inference from p_1 to q, a generalisation $g_1 \in \mathbf{G}^{\mathrm{e}}$, $q \in \mathbf{Tails}(g_1)$, $Head(g_1) = p_1$ could not have been used (Definition 25, case 1) per Lemma 2. Thus, we only need to consider cases 2 and 3 of Definition 25, which is a deductive inference with generalisation $g_1 \in \mathbf{G}^{\mathrm{c}}$, $Head(g_1) = q$, $\mathbf{P_1} \subseteq \mathbf{Ant}(g_1)$ or $g_1 \in \mathbf{G}^{\mathrm{a}}$, $Head(g_1) = q$, $\mathbf{P_1} \subseteq \mathbf{Tails}(g_1)$, respectively.

- First, consider case 2a of Definition 25 in which $g_2 \neq g_1$, $g_2 \in \mathbf{G}^{\mathrm{c}}$, $Head(g_2) = q$, $\mathbf{P_2} \subseteq \mathbf{Ant}(g_2)$. Then p_2 cannot be inferred from q using g_2, as in this case abduction would be performed with g_2 to infer p_2 from q while per the restrictions in condition 2 of Definition 32 abduction cannot be performed with g_2 as $Head(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{c}}$.
- Next, consider case 2b of Definition 25 in which $g_2 \in \mathbf{G}^{\mathrm{e}}$, $Head(g_2) = p_2$, $q \in \mathbf{Tails}(g_2)$. Then p_2 cannot be inferred from q using g_2, as in this case deductive inference would be performed with g_2 to infer p_2 from q while per the restrictions in condition 2 of Definition 31 deductive inference cannot be performed with g_2 as $q \in \mathbf{Tails}(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{c}}$.
- Finally, consider case 3 of Definition 25 in which $g_2 \neq g_1$, $g_2 \in \mathbf{G}^{\mathrm{a}}$, $Head(g_2) = q$, $\mathbf{P_2} \subseteq \mathbf{Tails}(g_2)$. Then p_2 cannot be inferred from q using g_2, as in this case abduction would be performed with g_2 to infer p_2 from q while per the restrictions in condition 2 of Definition 32 abduction cannot be performed with g_2 as $Head(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{a}}$. ■

Example 47. *In the IG of Figure 3.6c, $[g_1, g_2]$ is a generalisation chain but* (smoke_machine*, g_1,* smoke*, g_2,* fire) *is* not *an inference chain, as per Pearl's constraint* fire *cannot be deductively inferred from* smoke *using g_2.* □

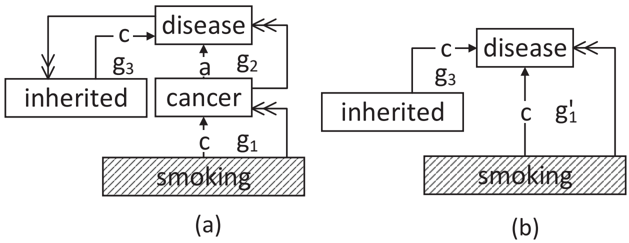


Figure 3.12: Example of an IG (a); alternative IG-modelling of the problem (b).

3.6 Discussion: additional inference constraints

In this chapter we focussed on the constraints we argue should be imposed on performing inference with pairs of generalisations (see Section 2.1.4), which cover Pearl's [1988a] original constraints (see Section 2.1.4.1) and local constraints on performing inference with abstractions (see Section 2.1.4.2). In this section, we discuss further extending our inference constraints. We provide an example of an IG for which possibly undesirable results are obtained upon performing inference with abstractions and generalisations of type 'other' using local constraints only, and discuss possible solutions that may be implemented in future work to help solve the problem.

Consider the IG depicted in Figure 3.12a. Upon observing that a person smokes (i.e. given $\mathbf{E_p} = \{smoking\}$), this would make us infer that this person has a disease that is biologically inherited (*inherited*) upon performing deduction in sequence with g_1 and g_2 and abduction with g_3. Hence, a cause for *disease* is inferred (i.e. *inherited*) while a cause for *disease* (i.e. *smoking*) is already known, which is undesirable. Similar observations can be made by replacing g_3 with g_3': *disease* $\rightarrow$ *inherited* in $\mathbf{G}^{\mathrm{e}}$, or by including $g_2 \in \mathbf{G}^{\mathrm{o}}$ instead of in $\mathbf{G}^{\mathrm{a}}$. Figure 3.12b depicts an alternative IG-modelling in which causal generalisation g_1': *smoking* $\rightarrow$ *disease* is used to express that smoking typically causes diseases. In essence, this generalisation captures the knowledge expressed by generalisations g_1 and g_2 in Figure 3.12a without making the intermediate claim *cancer* explicit. Given $\mathbf{E_p}$, *disease* is inferred from *smoking* using g_1', but *inherited* cannot in turn be abductively inferred from *disease* using g_3 per the restrictions in condition 2 of Definition 32. One may therefore argue that the undesirable results obtained for the IG of Figure 3.12a are a result of the way the available knowledge is modelled. However, a problem with considering this to be a knowledge modelling issue is that IGs including abstractions and generalisations of type 'other' would always need to be verified by a knowledge engineer, and that intermediate claims cannot always be made explicit in the manner as depicted in Figure 3.12a. Preferably, additional constraints are imposed to ensure undesirable results are not obtained upon performing inference.

Our definitions of deductive and abductive inference (Definitions 31 and 32) may be adjusted and further restricted to help solve the problem; however, this would considerably complicate our definitions, as not only single generalisations but also chains of generalisations would need to be considered. For instance, in the example of Figure 3.12a generalisation arc g_2 can arguably be replaced by a generalisation chain $[g'_2, \ldots, g'_m]$ with $g'_2, \ldots, g'_m \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $cancer \in \mathbf{Tails}(g'_2)$, $Head(g'_m) = disease$, where upon performing iterative deduction with $g_1, g'_2, \ldots, g'_m$ given $\mathbf{E_p}$, one should not in turn be allowed to perform abductive inference with g_3.

Another solution is to monitor for every proposition whether it is causally or evidentially inferred given the evidence, as in Pearl's [1988a] C-E system. In his semi-formal proposal, default rules with only single antecedents are assigned causal and evidential 'C' and 'E' labels. Each proposition is then also assigned a 'C' or 'E' label, where given the evidence a proposition is 'E-believed' if it is deductively inferred using an E-rule and 'C-believed' if it is deductively inferred using a C-rule. Deduction can always be performed with C-rules regardless of whether its antecedent is C-believed or E-believed, but deduction can only be performed with E-rules in case its antecedent is E-believed (i.e. Pearl's constraint, see Section 2.1.4.1). In his system, Pearl does not consider default rules with multiple antecedents, he does not consider abductive inference, and he does not consider (strict or defeasible) default rules that are neither causal nor evidential. In previous work, van Hooff [2004] (in Master thesis research) preliminarily investigated extending Pearl's C-E system with default rules that are neither causal nor evidential and default rules with multiple antecedents; in future work, this preliminary solution may be further elaborated and extended upon to work for current purposes. For instance, van Hooff [2004] states that a proposition deductively inferred using a default rule with a single antecedent that is neither causal nor evidential should receive the same status as the status of the antecedent of this rule. In our example, *cancer* is C-believed as it is deductively inferred using causal generalisation g_1. According to van Hooff [2004], *disease* is then also C-believed as it is deductively inferred using abstraction g_2. To ensure undesirable results are not obtained for our example, restrictions would then need to be imposed on performing abductive inference with a causal generalisation in case its head is C-believed.

3.7 Concluding remarks

In this chapter, we have set out to formalise and disambiguate analyses performed using informal sense-making tools in a manner that (1) allows for guiding the construction of formal representations within AI systems and that (2) is in line with our conceptual analysis of reasoning about evidence as provided in Section 2.1, while (3) allowing inference to be performed and visualised in a manner that is closely related to the way in which inference is performed and visualised by domain experts using such tools. In particular, we have proposed the IG-formalism, which formalises analyses performed using such tools in this manner. The IG-formalism provides a precise account of the interplay between deductive and abductive inference and causal, evidential, abstractions, and other types of information. The inference constraints we impose on our IG-formalism are based on our conceptual analysis of reasoning about evidence of Section 2.1 and are inspired by other formal systems for reasoning about evidence, e.g. [Bex, 2011, 2015; Console and Torasso, 1991; Poole, 1989], where we have accounted for constraints that are typically accounted for in these systems. Given the evidence, inference chains can be read from an IG, which are sequences of propositions that can be iteratively inferred from each other; we have formally proven that inference chains that can be read from an IG given the evidence indeed adhere to the identified inference constraints.

As illustrated through examples of analyses performed using informal sense-making tools, when performing analyses domain experts naturally mix the different types of generalisations and inferences, where the used generalisations and the inference type are left implicit. Furthermore, the manner of conflict is typically not precisely specified and the assumptions of domain experts underlying their analyses are typically not explicitly stated, making these analyses ambiguous to interpret. Our IG-formalism serves to formalise and disambiguate these analyses in a manner that makes the used generalisations and conflicts explicit, where the components and elements incorporated in the formalism are meant to represent a cross section of elements that can typically be expressed using such tools. By precisely defining the different elements incorporated in our formalism and by explicitly stating our assumptions, we expect IGs to be less ambiguous to interpret than analyses performed using tools, a claim that should be empirically evaluated in future work. Such a study may also serve to evaluate the expressivity of the IG-formalism.

In interpreting a performed analysis as an IG, an additional knowledge elicitation step may be required as the used generalisations and the manner of conflict are typically left implicit in these analyses. In this chapter, we have provided an example of this interpretation step by discussing a possible formalisation of an analysis performed using a mind mapping tool. IGs may also be directly constructed by domain experts in case work. In the following chapters, we demonstrate that our IG-formalism can be used to guide the construction of formal representations within AI systems. More specifically, in Chapter 4 we define an argumentation formalism based on IGs that allows for formal evaluation of arguments based on IGs using computational argumentation, and in Chapter 5 we demonstrate the use of the IG-formalism in guiding BN construction by serving as an intermediary formalism between analyses performed using informal sense-making tools and BNs.

Chapter 4

An argumentation formalism based on information graphs

Based on our IG-formalism from Chapter 3, we now define an argumentation formalism that allows for both deductive and abductive argumentation. Note that the IG-formalism is not an argumentation formalism; instead, in Chapter 3 we defined how inference can be performed with IGs and we defined different notions of conflicts (i.e. negatory conflict, exception-based conflict, conflict between competing alternative explanations). In the current chapter, we define an argumentation formalism based on IGs which allows us to assign a semantics to arguments constructed on the basis of IGs. More specifically, our approach generates an abstract argumentation framework (see Section 2.2.1) which thus allows arguments based on IGs to be formally evaluated according to Dung's [1995] semantics. We can then study properties of generated AFs; in particular, we prove that Caminada and Amgoud's [2007] postulates are satisfied by instantiations of our formalism, which warrants the sound definition of instantiations of our argumentation formalism and implies that anomalous results such as issues regarding inconsistency and non-closure as identified by Caminada and Amgoud [2007] are avoided.

In Section 4.1 we define arguments on the basis of a provided IG and an evidence set $\mathbf{E_p}$, which capture sequences of deductive and abductive inference applications starting with elements from $\mathbf{E_p}$. We then prove a number of formal properties of arguments, among which the property that arguments constructed on the basis of IGs conform to our inference constraints (see Section 2.1.4). In Section 4.2 we define several types of attacks between arguments based on IGs, which are based on the different types of conflicts defined for our IG-formalism. In Section 4.3 we instantiate Dung's abstract approach with arguments and attacks based on IGs. In Section 4.4 we then formally prove that key rationality postulates [Caminada and Amgoud, 2007] are satisfied by instantiations of our formalism.

4.1 Arguments

In this section, we define how arguments on the basis of an IG and an evidence set $\mathbf{E_p}$ are constructed. Here, we take inspiration from the definition of an argument as defined for the ASPIC$^+$ framework (see Section 2.2.2). By remaining close to the ASPIC$^+$ framework, this allows us to straightforwardly show that rationality postulates are satisfied for our argumentation formalism based on IGs (see Section 4.4). Furthermore, Bex [2015] previously proposed an integrated theory of causal and evidential arguments, which is a formal account of reasoning about evidence that is based on the ASPIC$^+$ framework; hence, by remaining close to the ASPIC$^+$ framework we can compare his account to ours (see Section 8.1.2).

In comparison to the operators introduced for arguments in ASPIC$^+$ (see Definition 7), the following additional operators are introduced. In what follows, for a given argument, ImmSub returns its immediate sub-arguments, Gen returns all the generalisations used in constructing the argument, TopGen returns the last generalisation used in constructing the argument, and DefInf and StInf return all the defeasible and strict inferences used in constructing the argument, respectively. Definition 37 is explained and illustrated in Examples 48 and 49.

Definition 37 (Argument)**.** *Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. An* argument A on the basis of $G_\mathcal{I}$ and $\mathbf{E_p}$ *is any structure obtainable by applying one or more of the following steps finitely many times, where steps* 2 *(i.e. step* 2a *or* 2b*) and* 3 *or vice versa are not subsequently applied using the same generalisation arc $g \in \mathbf{G}$:*

1. *p if $p \in \mathbf{E_p}$, where:* $\textsc{Prem}(A) = \{p\}$; $\textsc{Conc}(A) = p$; $\textsc{Sub}(A) = \{A\}$; $\textsc{ImmSub}(A) = \emptyset$; $\textsc{Gen}(A) = \emptyset$; $\textsc{TopGen}(A) =$ *undefined;* $\textsc{DefInf}(A) = \emptyset$; $\textsc{StInf}(A) = \emptyset$; $\textsc{TopInf}(A) =$ *undefined.*

2a. *$A_1, \ldots, A_n \twoheadrightarrow_g p$ if $A_1, \ldots, A_n$ are arguments such that p is defeasibly deductively inferred from $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)$ using a generalisation $g \colon \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)\} \to p$ according to Definition 31, where it holds that $g \in \mathbf{G}_\mathrm{d}$ and if g is of the form $g \colon c \to e$ in $\mathbf{G}^\mathrm{c}$ and its evidential counterpart $g' \colon e \to c$ is included in $\mathbf{G}^\mathrm{e}$, then $g' \notin \textsc{Gen}(A_1) \cup \ldots \cup \textsc{Gen}(A_n)$. For A, it holds that:*
$\textsc{Prem}(A) = \textsc{Prem}(A_1) \cup \ldots \cup \textsc{Prem}(A_n)$; $\textsc{Conc}(A) = p$;
$\textsc{Sub}(A) = \textsc{Sub}(A_1) \cup \ldots \cup \textsc{Sub}(A_n) \cup \{A\}$; $\textsc{ImmSub}(A) = \{A_1, \ldots, A_n\}$;
$\textsc{Gen}(A) = \textsc{Gen}(A_1) \cup \ldots \cup \textsc{Gen}(A_n) \cup \{g\}$; $\textsc{TopGen}(A) = g$;
$\textsc{DefInf}(A) = \textsc{DefInf}(A_1) \cup \ldots \textsc{DefInf}(A_n) \cup \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \twoheadrightarrow_g p\}$; $\textsc{StInf}(A) = \textsc{StInf}(A_1) \cup \ldots \textsc{StInf}(A_n)$;
$\textsc{TopInf}(A) = \textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n) \twoheadrightarrow_g p$.

2b. *$A_1, \ldots, A_n \rightharpoonup_g p$ if $A_1, \ldots, A_n$ are arguments such that p is strictly deductively inferred from $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)$ using a generalisation $g \in \mathbf{G}_\mathrm{S}$, $g \colon \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)\} \to p$ according to Definition 31, where* $\textsc{Prem}(A)$, $\textsc{Conc}(A)$, $\textsc{Sub}(A)$, $\textsc{ImmSub}(A)$, $\textsc{Gen}(A)$ *and* $\textsc{TopGen}(A)$ *are defined as in*

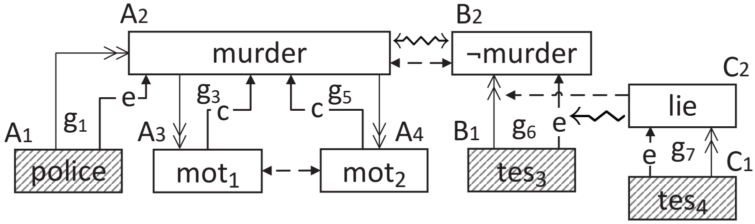


Figure 4.1: Adjustment to the IG of Figure 3.5b, where arguments and direct attacks ($\dashrightarrow$) on the basis of this IG and $\mathbf{E_p}$ are also indicated.

step 2*a, and where:*
$\text{DefInf}(A) = \text{DefInf}(A_1) \cup \ldots \text{DefInf}(A_n)$;
$\text{StInf}(A) = \text{StInf}(A_1) \cup \ldots \text{StInf}(A_n) \cup \{\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightharpoonup_g p\}$;
$\text{TopInf}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightharpoonup_g p$.

3. $A' \twoheadrightarrow_g p$ *if* A' *is an argument such that* p *is abductively inferred from* $\text{Conc}(A')$ *using a generalisation* $g \in \mathbf{G}^c \cup \mathbf{G}^a$, $g\colon \{p, p_1, \ldots, p_n\} \to \text{Conc}(A')$ *for some propositions* $p_1, \ldots, p_n \in \mathbf{P}$ *according to Definition 32, where:*
$\text{Prem}(A) = \text{Prem}(A')$; $\text{Conc}(A) = p$; $\text{Sub}(A) = \text{Sub}(A') \cup \{A\}$; $\text{ImmSub}(A) = \{A'\}$; $\text{Gen}(A) = \text{Gen}(A') \cup \{g\}$; $\text{TopGen}(A) = g$; $\text{DefInf}(A) = \text{DefInf}(A') \cup \{\text{Conc}(A') \twoheadrightarrow_g p\}$; $\text{StInf}(A) = \text{StInf}(A')$; $\text{TopInf}(A) = \text{Conc}(A') \twoheadrightarrow_g p$.

Note that we overload symbols $\twoheadrightarrow$ and $\rightharpoonup$ to denote an argument while it also denotes a defeasible or strict inference. Similar to ASPIC$^+$, the set of all arguments on the basis of $G_{\mathcal{I}}$ and $\mathbf{E_p}$ is denoted by $\mathcal{A}$.

An argument $A \in \mathcal{A}$ is called *strict* if $\text{DefInf}(A) = \emptyset$; otherwise, A is called *defeasible*. An argument $A \in \mathcal{A}$ is called a *premise argument* if only step 1 of Definition 37 is applied, *deductive* if only steps 1, 2a and 2b are applied, *abductive* if only steps 1 and 3 are applied, and *mixed* otherwise. The restriction that steps 2 (i.e. step 2a or 2*b*) and 3 or vice versa are not subsequently applied using the same generalisation arc $g \in \mathbf{G}$ ensures that cycles in which two propositions are iteratively deductively and abductively inferred from each other using the same g are avoided in argument construction. Similarly, in case causal generalisation $g\colon c \to e$ has an evidential counterpart $g'\colon e \to c$ (see Section 2.1.3, p. 20 and Section 3.3, p. 44), then the restriction in step 2a that $g' \notin \text{Gen}(A_1) \cup \ldots \cup \text{Gen}(A_n)$ ensures that cycles in which c and e are iteratively deductively inferred from each other using g' and g are avoided. Note that cycles in which c and e are iteratively deductively inferred from each other using g and g' in that order are already avoided due to the enforcement of Pearl's constraint (Definition 31, condition 2).

Example 48. *Consider the adjustment to the IG of Figure 3.5 depicted in Figure 4.1, in which arguments on the basis of this IG and* $\mathbf{E_p} = \{\text{police}, \text{tes}_3, \text{tes}_4\}$ *are also indicated. According to step* 1 *of Definition 37,* $A_1\colon$ police *is a premise argument. Based on* A_1*, deductive argument* $A_2\colon A_1 \twoheadrightarrow_{g_1}$ murder *is constructed by step* 2 *of Definition 37, as* murder *is deductively inferred from* police *using* g_1: police $\to$ murder*. Then* $A_3 \colon A_2 \twoheadrightarrow_{g_3}$ mot_1 *is a mixed argument by step* 3 *of Definition 37, as* mot_1 *is abductively inferred from* murder *using* g_3: $\text{mot}_1 \to$ murder*. To illustrate the additional operators introduced in Definition 37, for* A_3*, we have that* $\textsc{ImmSub}(A_3) = \{A_2\}$*;* $\textsc{Gen}(A_3) = \{g_1, g_3\}$*;* $\textsc{TopGen}(A_3) = g_3$*;* $\textsc{DefInf}(A_3) = \{\text{police} \twoheadrightarrow_{g_1} \text{murder}, \text{murder} \twoheadrightarrow_{g_3} \text{mot}_1\}$*;* $\textsc{StInf}(A_3) = \emptyset$*.* □

Step 3 of Definition 37 is now illustrated in more detail.

Example 49. *On the basis of the IG of Figure 3.7a and* $\mathbf{E_p} = \{r\}$*,* $A'_1 \colon r$ *is a premise argument. From* A'_1*, arguments* $A'_2 \colon A'_1 \twoheadrightarrow_{g_3} r_1$ *and* $A'_3 \colon A'_1 \twoheadrightarrow_{g_3} q$ *are constructed by step* 3 *of Definition 37, as* q *and* r_1 *are abductively inferred from* $\textsc{Conc}(A'_1)$ *using causal generalisation* $g_3 \colon \{q, r_1\} \to r$*. Then again by step* 3*,* $A'_4 \colon A'_3 \twoheadrightarrow_{g_1} p_1$ *and* $A'_5 \colon A'_3 \twoheadrightarrow_{g_2} p_2$ *are constructed using* g_1 *and* g_2*, respectively.* □

4.1.1 Properties of arguments based on IGs

We now prove a number of formal properties of arguments based on IGs. Note that the results stated below are similar to the results as proven for inference chains in Section 3.5 (i.e. Lemma 1 and Proposition 1). However, these results are not directly applicable in the context of arguments constructed on the basis of IG, as there generally does not exist a one-to-one correspondence between inference chains and arguments. The differences between arguments and inference chains are illustrated by the following example.

Example 50. *Reconsider Example 44 from Section 3.5. As discussed in this example, in the IG of Figure 3.11 chain* $c_1 = (\text{tes}_1, g_2, \text{mot}_1, g_3, \text{murder})$ *is an inference chain given* $\mathbf{E_p}$*, where the presence of this inference chain does not imply that* mot_1 *is by itself sufficient to infer* murder*. Instead,* murder *can only be deductively inferred using* g_3 *in case both* mot_1 and mot_2 *are affirmed. The broader context in which the inference from* mot_1 *to* murder *is performed using* g_3 *is thus not directly apparent from this inference chain; instead, the role of* mot_2 *becomes apparent in considering other inference chains that can be read from this IG given* $\mathbf{E_p}$*, specifically inference chain* $c_2 = (\text{tes}_2, g_4, \text{mot}_2, g_3, \text{murder})$*. In comparison, the arguments that are constructed on the basis of this IG given* $\mathbf{E_p}$ *are* $A_1 \colon \text{tes}_1$*;* $A_2 \colon A_1 \twoheadrightarrow_{g_2} \text{mot}_1$*;* $A_3 \colon \text{tes}_2$*;* $A_4 \colon A_3 \twoheadrightarrow_{g_4} \text{mot}_2$*;* $A_5 \colon A_2, A_4 \twoheadrightarrow_{g_3}$ murder*. From argument* A_5*, the broader context in which* murder *is deductively inferred using* g_3 *is directly apparent, in contrast with inference chains* c_1 *and* c_2 *upon considering them individually. This example thus illustrates that there generally does not exist a one-to-one correspondence between inference chains and arguments. In particular, argument* A_5 *cannot be interpreted as a single inference chain, and neither inference chain* c_1 *nor* c_2 *can be interpreted as one of the arguments* $A_1, \ldots, A_5$*.* □

Lemma 3 states that the conclusions of deductive, abductive, and mixed arguments constructed in our argumentation formalism based on IGs are not observed.

Lemma 3. *Let $\mathcal{A}$ be a set of arguments on the basis of IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ and evidence set $\mathbf{E_p}$. Let $A \in \mathcal{A}$ be a deductive, abductive, or mixed argument. Then* $\textsc{Conc}(A) \notin \mathbf{E_p}$.

Proof. As A is not a premise argument, step 2a, step 2b or step 3 of Definition 37 is applied last in constructing A. In case step 2a or 2b of Definition 37 is applied last, then $\exists g \in \mathbf{G}$ such that $Head(g) = \textsc{Conc}(A)$ is deductively inferred using $\textsc{TopGen}(A) = g$ according to Definition 31. Hence, per the restrictions of Definition 31, $Head(g) = \textsc{Conc}(A) \notin \mathbf{E_p}$. In case step 3 of Definition 37 is applied last, then $\exists g \in \mathbf{G}$ such that $\textsc{Conc}(A) \in \mathbf{Tails}(g)$ is abductively inferred using $\textsc{TopGen}(A) = g$ according to Definition 32. Hence, $\textsc{Conc}(A) \notin \mathbf{E_p}$ per the restriction of Definition 32 that $\mathbf{Tails}(g) \cap \mathbf{E_p} = \emptyset$. ■

In performing inference care should be taken that no cause for an effect is inferred in case an alternative cause for this effect was already previously inferred (i.e. Pearl's constraint, see Section 2.1.4.1). Similarly, care should be taken that no version of an event at a lower level of abstraction is inferred if an alternative version of this event at a lower level of abstraction was already previously inferred (i.e. our inference constraints for abstractions, see Section 2.1.4.2). In constructing non-premise arguments based on IGs, step 2a, 2b, or 3 of Definition 37 is applied last, where in applying one of these steps a deductive or abductive inference is performed to infer the conclusion of the argument from the conclusions of its immediate sub-arguments. Hence, we need to prove that the inference constraints of Section 2.1.4 are never violated in constructing sequences of arguments on the basis of IGs using these steps. First, we formally define the inference constraints of Section 2.1.4 in the context of arguments constructed on the basis of IGs.

Definition 38 (Inference constraint)**.** *Let $\mathcal{A}$ be a set of arguments on the basis of IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ and evidence set $\mathbf{E_p}$. Let $\mathbf{P_1} \subseteq \mathbf{P}$ and $\mathbf{P_2} \subseteq \mathbf{P}$ be alternative explanations of $q \in \mathbf{P}$ as indicated by generalisations g_1 and g_2 in $\mathbf{G}$ (see Definition 25). Let $p_1 \in \mathbf{P_1}$ and $p_2 \in \mathbf{P_2}$. If arguments A and B exist in $\mathcal{A}$ with $\textsc{Conc}(B) = q$, $A \in \textsc{ImmSub}(B)$, and $\textsc{Conc}(A) = p_1$, then there does not exist an argument $C \in \mathcal{A}$ with $B \in \textsc{ImmSub}(C)$ and $\textsc{Conc}(C) = p_2$.*

We now formally prove that this inference constraint is indeed adhered to.

Proposition 2 (Adherence to inference constraint)**.** *Let $\mathcal{A}$ be a set of arguments on the basis of IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ and evidence set $\mathbf{E_p}$. Then $\mathcal{A}$ adheres to the inference constraint as defined in Definition 38.*

Proof. Assume that $\mathbf{P_1} \subseteq \mathbf{P}$ and $\mathbf{P_2} \subseteq \mathbf{P}$ are alternative explanations of $q \in \mathbf{P}$ as indicated by generalisations $g_1, g_2 \in \mathbf{G}$ with $p_1 \in \mathbf{P_1}$ and $p_2 \in \mathbf{P_2}$, and assume that arguments $A, B \in \mathcal{A}$ exist with $\textsc{Conc}(B) = q$, $A \in \textsc{ImmSub}(B)$, $\textsc{Conc}(A) = p_1$. Then we need to prove that no argument C exists in $\mathcal{A}$ with $B \in \textsc{ImmSub}(C)$ and $\textsc{Conc}(C) = p_2$. In constructing argument B, either step 2a, step 2b or step 3 of

Definition 37 is applied last, where generalisation g_1 is used to infer $\textsc{Conc}(B) = q$. Here, g_1 cannot be of the form $g_1 \in \mathbf{G}^{\mathrm{e}}$, $q \in \mathbf{Tails}(g_1)$, $Head(g_1) = p_1$ (Definition 25, case 1) as in this case antecedent q of g_1 is inferred from consequent p_1 of g_1, which would be an instance of abductive inference while per the restrictions of Definition 32 abductive inference can only be performed using generalisations in $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$. More specifically, argument B cannot be constructed by applying step 2a, 2b and 3 of Definition 37 last if g_1 is of that form. Thus, we only need to consider cases 2 and 3 of Definition 25, where a generalisation $g_1 \in \mathbf{G}^{\mathrm{c}}$, $Head(g_1) = q$, $\mathbf{P_1} \subseteq \mathbf{Ant}(g_1)$ respectively a generalisation $g_1 \in \mathbf{G}^{\mathrm{a}}$, $Head(g_1) = q$, $\mathbf{P_1} \subseteq \mathbf{Tails}(g_1)$ is used to construct B, namely by applying step 2a or 2b of Definition 37 last to deductively infer $\textsc{Conc}(B) = q$. We now show that for the given options for g_1, no argument C with $B \in \textsc{ImmSub}(C)$, $\textsc{Conc}(C) = p_2$ can be constructed using g_2.

- First, consider case 2a of Definition 25 in which $g_2 \neq g_1$, $g_2 \in \mathbf{G}^{\mathrm{c}}$, $Head(g_2) = q$, $\mathbf{P_2} \subseteq \mathbf{Ant}(g_2)$. Then no argument C with $B \in \textsc{ImmSub}(C)$, $\textsc{Conc}(C) = p_2$ can be constructed using g_2, as in this case abduction would be performed with g_2 to infer p_2 from q while per the restrictions in condition 2 of Definition 32 abduction cannot be performed with g_2 as $Head(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{c}}$. In particular, step 3 of Definition 37 cannot be applied in constructing C using g_2. Furthermore, neither step 2a nor step 2b of Definition 37 can be applied in constructing C using g_2, as these steps specify deductive and not abductive inferences.

- Next, consider case 2b of Definition 25 in which $g_2 \in \mathbf{G}^{\mathrm{e}}$, $Head(g_2) = p_2$, $q \in \mathbf{Tails}(g_2)$. Then no argument C with $B \in \textsc{ImmSub}(C)$, $\textsc{Conc}(C) = p_2$ can be constructed using g_2, as in this case deduction would be performed with g_2 to infer p_2 while per the restrictions in condition 2 of Definition 31 deduction cannot be performed with g_2 as $q \in \mathbf{Tails}(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{c}}$. In particular, step 2a of Definition 37 cannot be applied in constructing C using g_2. Furthermore, step 2b cannot be applied in constructing C using g_2, as this step can only be applied using strict generalisations and $g_2 \notin \mathbf{G}_{\mathrm{S}}$, and step 3 cannot be applied in constructing C using g_2, as this step specifies an abductive and not a deductive inference.

- Finally, consider case 3 of Definition 25 in which $g_2 \neq g_1$, $g_2 \in \mathbf{G}^{\mathrm{a}}$, $Head(g_2) = q$, $\mathbf{P_2} \subseteq \mathbf{Tails}(g_2)$. Then no argument C with $B \in \textsc{ImmSub}(C)$, $\textsc{Conc}(C) = p_2$ can be constructed using g_2, as in this case abduction would be performed with g_2 to infer p_2 from q while per the restrictions in condition 2 of Definition 32 abduction cannot be performed with g_2 as $Head(g_2)$ was previously deductively inferred using $g_1 \in \mathbf{G}^{\mathrm{a}}$. In particular, step 3 of Definition 37 cannot be applied in constructing C using g_2. Furthermore, neither step 2a nor step 2b of Definition 37 can be applied in constructing C using g_2, as these steps specify deductive and not abductive inferences. ∎

4.2 Attack

In this section, several types of attacks between arguments on the basis of IGs are defined. Among the types of attacks that are typically distinguished in structured argumentation (for instance in ASPIC$^+$) are rebuttal, undermining, and undercutting attack. Of these types of attacks, we only consider rebuttal and undercutting attack and not undermining attacks, as in IGs we assume that all premises are certain and cannot be attacked (cf. ASPIC$^+$'s axiom premises). We also distinguish a fourth type of attack, namely alternative attack, a concept inspired by [Bench-Capon and Prakken, 2006; Bex, 2015] based on the notion of alternative explanations that captures conflicts between abductively inferred conclusions. In our argumentation formalism, attacks directly follow from the constructed arguments and the specified exception and negation arcs in an IG.

First, we define the general notion of attack, after which the different types of attacks are defined.

Definition 39 (Attack)**.** *Let $\mathcal{A}$ be a set of arguments on the basis of IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $A, B \in \mathcal{A}$. Then A* attacks *B iff A rebuts B, A undercuts B, or A alternative attacks B, as defined in Definitions 40, 41 and 42, respectively.*

4.2.1 Rebuttal attack

First, rebuttal attack is defined in a manner comparable to the way as it is defined for ASPIC$^+$ (see Definition 8.2).

Definition 40 (Rebuttal attack)**.** *Let $\mathcal{A}$ be a set of arguments on the basis of IG $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ and evidence set $\mathbf{E_p}$. Let $A, B, B' \in \mathcal{A}$ with $B' \in \textsc{Sub}(B)$. Then A rebuts B (on B') iff there exists a negation arc $n\colon \textsc{Conc}(A) \leftrightsquigarrow \textsc{Conc}(B')$ in $\mathbf{N}$ and B' is of the form $B''_1, \ldots, B''_n \twoheadrightarrow_g p$ for some $B''_1, \ldots, B''_n \in \mathcal{A}$, $p \in \mathbf{P}$.*

Note that, as it is assumed that B' is of the form $B''_1, \ldots, B''_n \twoheadrightarrow_g p$ (i.e. $\textsc{TopInf}(B')$ is defeasible), it holds that B' is a deductive, abductive, or mixed argument; hence, by Lemma 3, $\textsc{Conc}(B') \notin \mathbf{E_p}$. Furthermore, while a negation arc expresses a symmetric conflict, our definition of rebuttal attack allows for both symmetric or asymmetric rebuttal, as illustrated by the following example.

Example 51. *Consider the IG of Figure 4.1. Let A_1, A_2 be the arguments introduced in Example 48. Let $B_1\colon$ tes_3 and let $B_2\colon B_1 \twoheadrightarrow_{g_6} \neg\text{murder}$. Then A_2 rebuts B_2 (on B_2) and B_2 rebuts A_2 (on A_2), as $\textsc{Conc}(A_2) = \text{murder}$, $\textsc{Conc}(B_2) = \neg\text{murder}$, $n\colon$ $\text{murder} \leftrightsquigarrow \neg\text{murder}$ in $\mathbf{N}$, $\textsc{TopInf}(A_2)$ is defeasible and $\textsc{TopInf}(B_2)$ is defeasible. This symmetric rebuttal is indicated in Figure 4.1 by means of a bidirectional dashed arc between these propositions.*

Consider again Example 41, in which heat *is predicted from* fire*. Assume that contrary to this prediction we observe that there is no heat ($\neg\text{heat} \in \mathbf{E_p}$). Let $A'_1\colon$*

smoke*;* A_2': $A_1' \twoheadrightarrow_{g_1}$ fire*;* A_3': $A_2' \twoheadrightarrow_{g_2}$ heat*;* B_1': $\neg$heat*. Then* B_1' *rebuts* A_3' *(on* A_3'*), but* A_3' *does not rebut* B_1' *as* B_1' *is not of the form* $B_1'', \ldots, B_n'' \twoheadrightarrow_g p$ *for some* $B_1'', \ldots, B_n'' \in \mathcal{A}$, $p \in \mathbf{P}$ *(i.e.* B_1' *is a premise argument).* □

4.2.2 Undercutting attack

Next, undercutting attack is considered. In our argumentation formalism based on IGs, undercutting attacks follow from the specified exception arcs in $G_{\mathcal{I}}$. Specifically, as an exception arc directed from $p \in \mathbf{P}$ to $g \in \mathbf{G}_{\mathrm{d}}$ specifies an exception to defeasible generalisation g, an argument $A \in \mathcal{A}$ with $\textsc{Conc}(A) = p$ undercuts an argument $B \in \mathcal{A}$ with $g \in \textsc{Gen}(B)$.

Definition 41 (Undercutting attack)**.** *Let* $\mathcal{A}$ *be a set of arguments on the basis of IG* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *and evidence set* $\mathbf{E_p}$*. Let* $A, B, B' \in \mathcal{A}$ *with* $B' \in \textsc{Sub}(B)$*. Then* A undercuts B *(on* B'*) iff there exists an exception arc* $exc \in \mathbf{Exc}$ *such that* $exc\colon \textsc{Conc}(A) \rightsquigarrow g$ *and* $\textsc{TopGen}(B') = g \in \mathbf{G}_{\mathrm{d}}$.

Undercutting attack is illustrated by the following example.

Example 52. *Consider the IG of Figure 4.1. Let* B_1, B_2 *be the arguments introduced in Example 51. Let* C_1: tes$_4$*;* C_2: $C_1 \twoheadrightarrow_{g_7}$ lie*. Then* C_2 *undercuts* B_2 *(on* B_2*), as* exc: lie $\rightsquigarrow g_6$ *in* **Exc** *and* $\textsc{TopGen}(B_2) = g_6$*. This attack is indicated in Figure 4.1 by a dashed arc directed from* lie *to inference* tes$_3 \twoheadrightarrow_{g_6} \neg$murder*.* □

4.2.3 Alternative attack

Lastly, alternative attack is defined, a concept based on the notion of alternative explanations that captures conflicts between abductively inferred conclusions.

Definition 42 (Alternative attack)**.** *Let* $\mathcal{A}$ *be a set of arguments on the basis of IG* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *and evidence set* $\mathbf{E_p}$*. Let* $\mathbf{P_1} \subseteq \mathbf{P}$ *and* $\mathbf{P_2} \subseteq \mathbf{P}$ *be alternative explanations of* $q \in \mathbf{P}$ *as indicated by generalisations* g *and* g' *in* $\mathbf{G}$*, where either* $g, g' \in \mathbf{G}^{\mathrm{c}}$ *(Definition 25, case 2a) or* $g, g' \in \mathbf{G}^{\mathrm{a}}$ *(Definition 25, case 3). Let* $p_1 \in \mathbf{P_1}$ *and* $p_2 \in \mathbf{P_2}$*. Let* $A, B, B' \in \mathcal{A}$ *with* $B' \in \textsc{Sub}(B)$*. Then* A *alternative attacks* B *(on* B'*) iff there exists an argument* $C \in \textsc{ImmSub}(A) \cap \textsc{ImmSub}(B')$ *such that* $\textsc{Conc}(A) = p_1$ *and* $\textsc{Conc}(B') = p_2$ *are abductively inferred from* $\textsc{Conc}(C) = q$ *using generalisations* g *and* g'*, respectively.*

Note that A only alternative attacks B on B' iff $\textsc{TopInf}(B')$ is an abductive inference and hence iff the last used inference in constructing B' is defeasible. Furthermore, unlike direct rebuttal attack, which can either be symmetric or asymmetric, direct alternative attack is always symmetric in that A alternative attacks B on B iff B alternative attacks A on A.

Under the conditions set out in Definition 42, arguments $A_i\colon C \twoheadrightarrow_g p_i$ for $p_i \in \mathbf{Ant}(g)$ constructed from C via abductive inference using g are involved in alternative attack with $A'_j\colon C \twoheadrightarrow_{g'} p'_j$ for $p'_j \in \mathbf{Ant}(g')$ constructed from C via abductive inference using g'. As alternative attack is based on Definition 25, arguments are only involved in alternative attack iff their conclusions are elements of two sets that are alternative explanations according to that definition, as illustrated by the following example.

Example 53. *Consider the IG of Figure 4.1. Let A_1, A_2, A_3 be the arguments introduced in Example 48, and let $A_4\colon A_2 \twoheadrightarrow_{g_5} \mathrm{mot}_2$, where mot_2 is abductively inferred from* murder*. Then A_3 and A_4 are involved in alternative attack, as* mot_1 *and* mot_2 *are alternative explanations of* murder *as indicated by generalisations g_3 and g_5 in $\mathbf{G}^{\mathrm{c}}$ (Definition 25, case 2a) and as $A_2 \in \textsc{ImmSub}(A_3) \cap \textsc{ImmSub}(A_4)$, where $\textsc{Conc}(A_3) = \mathrm{mot}_1$ and $\textsc{Conc}(A_4) = \mathrm{mot}_2$ are abductively inferred from $\textsc{Conc}(A_2) = \mathrm{murder}$ using g_3 and g_5, respectively. This attack is indicated in Figure 4.1 by means of a bidirectional dashed arc between the conclusions of A_3 and A_4.*

Consider the IG of Figure 3.9. Given $\mathbf{E_p}$, arguments $D_1\colon$ autopsy; $D_2\colon D_1 \twoheadrightarrow_{g'_5}$ head_wound; $D_3\colon D_2 \twoheadrightarrow_{g'_6}$ hit_angular; $D_4\colon D_2 \twoheadrightarrow_{g'_7}$ fell_on_table; $D_5\colon D_3 \twoheadrightarrow_{g'_3}$ hammer; and $D_6\colon D_3 \twoheadrightarrow_{g'_4}$ stone are constructed. Here, hit_angular *and* fell_on_table *are abductively inferred from* head_wound *using g'_6 and g'_7, respectively, and* hammer *and* stone *are abductively inferred from* hit_angular *using g'_3 and g'_4, respectively. Then D_3 alternative attacks D_4 (on D_4) and D_4 alternative attacks D_3 (on D_3), as $\textsc{Conc}(D_3) = \text{hit\_angular}$ and $\textsc{Conc}(D_4) = \text{fell\_on\_table}$ are alternative explanations of $\textsc{Conc}(D_2) = \text{head\_wound}$ as indicated by g'_6 and g'_7 in $\mathbf{G}^{\mathrm{c}}$ (Definition 25, case 2a). As $D_3 \in \textsc{Sub}(D_5)$ and $D_3 \in \textsc{Sub}(D_6)$, D_4 also alternative attacks D_5 and D_6 (on D_3). Finally, D_5 alternative attacks D_6 (on D_6) and D_6 alternative attacks D_5 (on D_5), as $\textsc{Conc}(D_5) = \text{hammer}$ and $\textsc{Conc}(D_6) = \text{stone}$ are alternative explanations of $\textsc{Conc}(D_3) = \text{hit\_angular}$ as indicated by g'_3 and g'_4 in $\mathbf{G}^{\mathrm{a}}$ (Definition 25, case 3).*

Consider Example 13 from Section 2.1.4.1. Assume that in addition to generalisations g_1 and g_2, evidential generalisation $g_3\colon$ see_fire $\rightarrow$ fire is provided. Given $\mathbf{E_p} = \{\text{see\_fire}\}$, arguments $E_1\colon$ see_fire; $E_2\colon E_1 \twoheadrightarrow_{g_3}$ fire; $E_3\colon E_2 \twoheadrightarrow_{g_1}$ torch; $E_4\colon E_2 \twoheadrightarrow_{g_2}$ match; and $E_5\colon E_2 \twoheadrightarrow_{g_2}$ oxygen are constructed. Then E_3 and E_4 are involved in alternative attack, as $\textsc{Conc}(E_3) = \text{torch}$, $\textsc{Conc}(E_4) = \text{match}$, and $\mathbf{Ant}(g_1) = \{\text{torch}\}$ and $\mathbf{Ant}(g_2) = \{\text{match}\}$ are alternative explanations of $\textsc{Conc}(E_2) = \text{fire}$ as indicated by g_1 and g_2 in $\mathbf{G}^{\mathrm{c}}$ (Definition 25, case 2a), where torch *and* match *are abductively inferred from* fire *using g_1 and g_2, respectively. E_3 is not involved in alternative attack with E_5, as $\textsc{Conc}(E_5) = \text{oxygen}$ is not an element of $\mathbf{Ant}(g_2)$ but instead oxygen $\in \mathbf{Enabler}(g_2)$.*

Consider Figure 3.7a. Let A'_1, A'_2, A'_3 be as defined in Example 49. Then A'_2 and A'_3 are not *involved in alternative attack, as $r_1 = \textsc{Conc}(A'_2)$ and $q = \textsc{Conc}(A'_3)$ are abductively inferred from $r = \textsc{Conc}(A'_1)$ using the* same *generalisation g_3; specifically, in case 2a of Definition 25 it is assumed that $g \neq g'$, and hence r_1 and q are not alternative explanations of r by that definition.* □

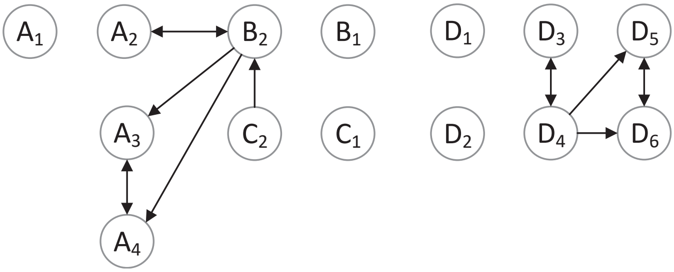


Figure 4.2: AF corresponding to the IGs of Figures 3.9 and 4.1.

4.3 Argument evaluation

We now instantiate Dung's abstract approach (see Section 2.2.1) with arguments and attacks based on IGs.

Definition 43 (Argumentation framework)**.** *Let* $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ *be an IG, and let* $\mathbf{E_p} \subseteq \mathbf{P}$ *be an evidence set. An* argumentation framework (AF) defined by $G_\mathcal{I}$ and $\mathbf{E_p}$ *is a pair* $(\mathcal{A}, \mathcal{D})$*, where* $\mathcal{A}$ *is the set of all arguments on the basis of* $G_\mathcal{I}$ *and* $\mathbf{E_p}$ *as defined by Definition 37 and where* $(A, B) \in \mathcal{D}$ *iff* $A, B \in \mathcal{A}$ *and* A *attacks* B *(see Definition 39).*

Given an AF, we can use any semantics for AFs for determining the dialectical status of arguments (see Section 2.2.1). In our IG-formalism, we opted not to account for preferences, as these are typically not indicated using sense-making tools. As the components of our argumentation formalism based on IGs are directly defined based on the elements that are accounted for in our IG-formalism, preferences are currently not accounted for in our argumentation formalism. As shown in work on structured argumentation with preferences [Prakken, 2012; Modgil and Prakken, 2013], the structure of arguments is crucial in determining how preferences must be applied to attacks and one should be cautious in extending AFs with additional elements without taking the structure of arguments into account. More specifically, it is shown by [Prakken, 2012; Modgil and Prakken, 2013] that the use of PAFs leads to violation of the sub-argument closure and consistency postulates [Caminada and Amgoud, 2007] (see also Section 4.4), among other things because PAFs cannot express how and at which points arguments attack each other. In future work, our argumentation formalism based on IGs may be extended to account for preferences at the structured level (cf. Modgil and Prakken [2013]).

We now illustrate the evaluation of arguments based on IGs through our running example.

Example 54. *Consider Examples 48, 51, 52 and 53, in which arguments* $\mathcal{A} = \{A_1, A_2, A_3, A_4, B_1, B_2, C_1, C_2, D_1, D_2, D_3, D_4, D_5, D_6\}$ *were introduced. The binary defeat relation over* $\mathcal{A}$ *follows directly from the attacks introduced in these examples:* $\mathcal{D} = \{(A_3, A_4), (A_4, A_3), (A_2, B_2), (B_2, A_2), (B_2, A_3), (B_2, A_4), (C_2, B_2), (D_3, D_4),$ $(D_4, D_3), (D_4, D_5), (D_4, D_6), (D_5, D_6), (D_6, D_5)\}$*. These arguments and defeats are* *depicted in Figure 4.2. The complete extensions of* $(\mathcal{A}, \mathcal{D})$ *are:*

$\mathcal{S}_1 = \{A_1, A_2, B_1, C_1, C_2, D_1, D_2\}$;
$\mathcal{S}_2 = \{A_1, A_2, A_3, B_1, C_1, C_2, D_1, D_2\}$;
$\mathcal{S}_3 = \{A_1, A_2, A_4, B_1, C_1, C_2, D_1, D_2\}$;
$\mathcal{S}_4 = \{A_1, A_2, B_1, C_1, C_2, D_1, D_2, D_3, D_6\}$;
$\mathcal{S}_5 = \{A_1, A_2, B_1, C_1, C_2, D_1, D_2, D_4\}$;
$\mathcal{S}_6 = \{A_1, A_2, A_3, B_1, C_1, C_2, D_1, D_2, D_3, D_5\}$;
$\mathcal{S}_7 = \{A_1, A_2, A_4, B_1, C_1, C_2, D_1, D_2, D_3, D_5\}$;
$\mathcal{S}_8 = \{A_1, A_2, A_3, B_1, C_1, C_2, D_1, D_2, D_3, D_6\}$;
$\mathcal{S}_9 = \{A_1, A_2, A_4, B_1, C_1, C_2, D_1, D_2, D_3, D_6\}$;
$\mathcal{S}_{10} = \{A_1, A_2, A_3, B_1, C_1, C_2, D_1, D_2, D_4\}$;
$\mathcal{S}_{11} = \{A_1, A_2, A_4, B_1, C_1, C_2, D_1, D_2, D_4\}$.

Under complete semantics, arguments $A_1, A_2, B_1, C_1, C_2, D_1, D_2$ *are justified as they are members of all complete extensions,* B_2 *is overruled as it is defeated by a justified argument (i.e.* C_2*), and* $A_3, A_4, D_3, D_4, D_5, D_6$ *are defensible. For the other semantics, the same statuses are assigned; for grounded semantics, this is the case as* $\mathcal{S}_1$ *is the set inclusion minimal complete extension. For preferred and stable semantics, note that* $\mathcal{S}_6$*,* $\mathcal{S}_7$*,* $\mathcal{S}_8$*,* $\mathcal{S}_9$*,* $\mathcal{S}_{10}$ *and* $\mathcal{S}_{11}$ *are set inclusion maximal complete extensions for which it holds that* $\forall B \notin \mathcal{S}_i, \exists A \in \mathcal{S}_i$ *such that* $(A, B) \in \mathcal{D}$ *for* $6 \leq i \leq 11$*; hence,* $\mathcal{S}_6$*,* $\mathcal{S}_7$*,* $\mathcal{S}_8$*,* $\mathcal{S}_9$*,* $\mathcal{S}_{10}$ *and* $\mathcal{S}_{11}$ *are preferred and stable extensions.* □

4.4 Satisfying rationality postulates

Caminada and Amgoud [2007] studied rule-based argumentation systems and identified conditions under which unintuitive and undesirable results are obtained upon performing inference. They then defined principles, called rationality postulates, that can be used to judge the quality of a given rule-based argumentation system. More specifically, so-called consistency and closure postulates were formulated for systems allowing for strict and defeasible inferences. Since these postulates are widely accepted as important desiderata for structured argumentation formalisms, we prove in this section that these postulates are satisfied by instantiations of our argumentation formalism based on IGs.

4.4.1 Comparison of our argumentation formalism based on IGs to the ASPIC$^+$ argumentation framework

In proving satisfaction of Caminada and Amgoud's [2007] rationality postulates, we follow Modgil and Prakken [2013], who proved satisfaction of these postulates for the ASPIC$^+$ framework. As noted earlier in this chapter, in defining our argumentation formalism based on IGs we were inspired by the definitions of argument and attack as given in [Modgil and Prakken, 2013]. In Definition 37, we defined how arguments on the basis of an IG and an evidence set $\mathbf{E_p}$ are constructed. In step 2a of Definition 37, it is specified that an argument A with $\textsc{Conc}(A) = p$

can be constructed from arguments $A_1, \ldots, A_n$ if p is defeasibly deductively inferred from $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)$ according to Definition 31 using a generalisation $g\colon \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)\} \rightarrow p$ in $\mathbf{G}_{\mathrm{d}}$. Hence, in terms of the terminology used in the ASPIC$^+$ framework, generalisations in $\mathbf{G}_{\mathrm{d}}$ can be interpreted as *domain-specific defeasible inference rules*[1] in ASPIC$^+$'s $\mathcal{R}_d$ that are applied when constructing arguments. Similarly, in step 2b of Definition 37 it is specified that an argument A with $\textsc{Conc}(A) = p$ can be constructed from $A_1, \ldots, A_n$ if p is strictly deductively inferred from $\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)$ according to Definition 31 using a generalisation $g\colon \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n)\} \rightarrow p$ in $\mathbf{G}_{\mathrm{s}}$. Hence, generalisations in $\mathbf{G}_{\mathrm{s}}$ can be interpreted as domain-specific *strict* inference rules in ASPIC$^+$'s $\mathcal{R}_s$. Finally, in step 3 it is specified that an argument A with $\textsc{Conc}(A) = p$ can be constructed from an argument A' if p is abductively inferred from $\textsc{Conc}(A')$ according to Definition 32 using a generalisation $g \in \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$, $g\colon \{p, p_1, \ldots, p_n\} \rightarrow \textsc{Conc}(A')$ for some propositions $p_1, \ldots, p_n \in \mathbf{P}$. Therefore, besides specifying aforementioned domain-specific defeasible and strict deduction rules, generalisations $g\colon \{q_1, \ldots, q_n\} \rightarrow q$ in $\mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}}$ also specify domain-specific *abduction* rules in ASPIC$^+$'s $\mathcal{R}_d$, namely for every $i \in \{1, \ldots, n\}$ a rule can be specified in $\mathcal{R}_d$ that states that q_i can be defeasibly inferred from q.

Considering the different types of attacks that are defined in Section 4.2, rebuttal as defined in Section 4.2.1 is identical to rebuttal as defined for a special case of ASPIC$^+$, namely one in which conflict is based on the standard classical notion of negation. Undercutting as defined in Section 4.2.2 is a special case of undercutting as defined for ASPIC$^+$, as we only consider undercutters of defeasible inferences in case an exception is provided to a defeasible generalisation used in a defeasible inference step. As preferences are not accounted for in our argumentation formalism based on IGs, attack is then resolved into defeat without considering preferences, as defined by Definition 43 in Section 4.3.

Thus, of the types of attacks that are considered in our argumentation formalism, only alternative attack is not accounted for in ASPIC$^+$. Furthermore, in comparison to our argumentation formalism, Modgil and Prakken do not impose any additional restrictions on argument construction. Hence, to prove that instantiations of our argumentation formalism based on IGs satisfy rationality postulates, in Section 4.4.3 we focus on showing how alternative attack and the additional restrictions that are imposed on argument construction in our argumentation formalism can be taken account in the results and proofs provided in [Modgil and Prakken, 2013].

4.4.2 Additional definitions and assumptions

Following Modgil and Prakken [2013], we introduce the following definitions. We define what it means for a set of propositions to be closed under strict generalisations.

[1]For further details on using ASPIC$^+$ to model domain-specific defeasible and strict inference rules, the reader is referred to [Modgil and Prakken, 2014].

Definition 44 (Closure under strict generalisations)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG and let $\mathbf{P}' \subseteq \mathbf{P}$. Then the* closure of $\mathbf{P}'$ under strict generalisations*, denoted* $\text{CL}(\mathbf{P}')$*, is the smallest set containing $\mathbf{P}'$ and the consequent $Head(g)$ of any $g \in \mathbf{G}_{\text{S}}$ whose antecedents $\mathbf{Tails}(g)$ are in $\text{CL}(\mathbf{P}')$.*

Next, the terms directly consistent and indirectly consistent set are defined.

Definition 45 (Directly consistent set)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG and let $\mathbf{P}' \subseteq \mathbf{P}$. Then $\mathbf{P}'$ is* directly consistent *iff $\nexists p, q \in \mathbf{P}'$ such that $p = -q$.*

A set $\mathbf{P}'$ is indirectly consistent if its closure under strict generalisations is directly consistent.

Definition 46 (Indirectly consistent set)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG and let $\mathbf{P}' \subseteq \mathbf{P}$. Then $\mathbf{P}'$ is* indirectly consistent *iff $\text{CL}(\mathbf{P}')$ is directly consistent.*

As noted by Caminada and Amgoud [2007], one should search for ways to alter or constrain one's argumentation formalism in such a way that rationality postulates are satisfied. Accordingly, following Modgil and Prakken [2013] we assume that IGs and evidence sets satisfy a number of properties. Similar to ASPIC^{+}, we leave the user free to make choices as to the strict and defeasible generalisations to include in $\mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ and the observations to include in $\mathbf{E_p}$; however, some care needs to be taken in making these choices to ensure that the result of argumentation is guaranteed to be well-behaved. Specifically, to ensure rationality postulates are satisfied, we assume that evidence sets $\mathbf{E_p}$ are indirectly consistent (referred to as the *axiom consistency* assumption), and we assume that $\mathbf{G}$ is closed under transposition. Note that per definition every evidence set $\mathbf{E_p} \subseteq \mathbf{P}$ is a directly consistent set, as it is assumed in Definition 30 that for every $p \in \mathbf{E_p}$, $\neg p \notin \mathbf{E_p}$. Furthermore, all examples of IGs provided in this thesis are axiom consistent, as they do not include generalisations $g \in \mathbf{G}_{\text{S}}$ for which $\mathbf{Tails}(g) \subseteq \mathbf{E_p}$. Closure under transposition is one of the solutions proposed by Caminada and Amgoud to 'repair' an argumentation system to ensure rationality postulates are satisfied [Caminada and Amgoud, 2007, p. 16], as it can help generate rules needed to obtain an intuitive outcome.

Definition 47 (Closure under transposition)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. A strict generalisation $g' \in \mathbf{G}_{\text{S}}$ is a* transposition *of $g\colon \{p_1, \ldots, p_n\} \rightarrow p$ in $\mathbf{G}_{\text{S}}$ iff g' is of the form $\{p_1, \ldots, p_{i-1}, -p, p_{i+1}, \ldots, p_n\} \rightarrow -p_i$ for some $1 \leq i \leq n$. We say that $\mathbf{G}$ is* closed under transposition *iff for all strict generalisations $g \in \mathbf{G}_{\text{S}}$, the transpositions of g are also in $\mathbf{G}_{\text{S}}$.*

An AF $(\mathcal{A}, \mathcal{D})$ defined by an IG $G_{\mathcal{I}}$ that is axiom consistent and for which $\mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ is closed under transposition is said to be *well defined*. In the remainder of this section, we assume that any given AF $(\mathcal{A}, \mathcal{D})$ is well defined. Note that most examples of IGs provided in this thesis only include defeasible generalisations and not strict generalisations, and thus that AFs defined by these IGs are well defined. Furthermore, note that our assumption that IGs do not include causal cycles (see

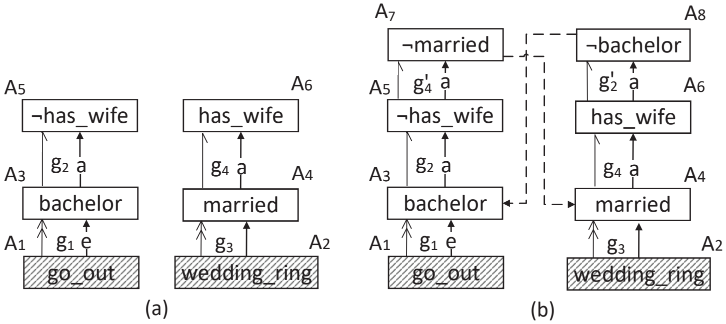


Figure 4.3: Example of an IG for which **G** is not closed under transposition (a); adjustment to this IG, in which additional generalisations are included such that **G** is closed under transposition (b).

Section 3.3, p. 47) can be concurrently assumed with our assumption that IGs are closed under transposition, as the former assumption only concerns the existence of causal and evidential generalisations, which are defeasible by assumption. The following example, adapted from Caminada and Amgoud [2007], illustrates closure under transposition and how ensuring it can help repair an argumentation system.

Example 55. *In the IG depicted in Figure 4.3a, strict abstractions* g_2: bachelor $\rightarrow$ ¬has_wife *and* g_4: married $\rightarrow$ has_wife *are included.* **G** *is not closed under transposition, as generalisations* has_wife $\rightarrow$ ¬bachelor *and* ¬has_wife $\rightarrow$ ¬married *are not included. Arguments* A_5 *and* A_6 *constructed on the basis of this IG have strict top inferences, as only step 2b of Definition 37 can be applied in constructing* A_5 *from* A_3 *and* A_6 *from* A_4 *using* g_2 *and* g_4 *in* $\mathbf{G}_\mathrm{S}$*, respectively. Note that, as* $\textsc{TopInf}(A_5)$ *and* $\textsc{TopInf}(A_6)$ *are strict,* A_5 *and* A_6 *are not involved in rebuttal. In fact,* $\mathcal{D} = \emptyset$ *for the AF corresponding to this IG, and hence under any semantics both* A_5 *and* A_6 *are justified. Thus, contradictory propositions* has_wife *and* ¬has_wife *are both justified at the same time, which is clearly undesirable and among other things violates the direct consistency postulate (see Theorem 1). In the IG depicted in Figure 4.3b,* **G** *is closed under transposition as additional generalisations* has_wife $\rightarrow$ ¬bachelor *and* ¬has_wife $\rightarrow$ ¬married *are now included. In the corresponding AF,* A_7 *directly rebuts* A_4 *and* A_8 *directly rebuts* A_3 *as* $\textsc{TopInf}(A_3)$ *and* $\textsc{TopInf}(A_4)$ *are defeasible. Then* A_7 *indirectly rebuts* A_6 *(on* A_4*) and* A_8 *indirectly rebuts* A_5 *(on* A_3*). Therefore, for this AF the more intuitive outcome is obtained that* A_5 *and* A_6 *cannot both be in the same extension at the same time.* □

Lastly, the following definitions introduce some terminology used in the below results. Following Modgil and Prakken [2018], we define strict continuations in a slightly different way than in [Modgil and Prakken, 2013], but as noted by Modgil and Prakken [2018] this does not affect the proofs stated in [Modgil and Prakken, 2013].

Definition 48 (Strict continuations)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. The set of* strict continuations *of a set of arguments from $\mathcal{A}$ is the smallest set satisfying the following conditions:*

1. *Any argument A is a strict continuation of $\{A\}$.*
2. *If $A_1, \ldots, A_n$ are arguments and $S_1, \ldots, S_n$ are sets of arguments such that for every $i \in \{1, \ldots, n\}$, A_i is a strict continuation of S_i and $\{B_{n+1}, \ldots, B_m\}$ is a (possibly empty) set of strict arguments, and $g\colon \{\textsc{Conc}(A_1), \ldots, \textsc{Conc}(A_n), \textsc{Conc}(B_{n+1}), \ldots, \textsc{Conc}(B_m)\} \to p$ is a strict generalisation in $\mathbf{G}_\mathrm{S}$, then argument $A_1, \ldots, A_n, B_{n+1}, \ldots, B_m \rightharpoonup_g p$ constructed from $A_1, \ldots, A_n, B_{n+1}, \ldots, B_m$ using g by applying step 2b of Definition 37 is a strict continuation of $S_1 \cup \ldots \cup S_n$.*

The maximal fallible sub-arguments of an argument B are those with the 'last' defeasible inferences in B. That is, they are the maximal sub-arguments of B on which B can be attacked.

Definition 49 (Maximal fallible sub-arguments)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. The set $M(B)$ of the maximal fallible sub-arguments of B is defined such that for any $B' \in \textsc{Sub}(B)$, it holds that $B' \in M(B)$ iff:*

1. *$\textsc{TopInf}(B')$ is defeasible, and;*
2. *There is no $B'' \in \textsc{Sub}(B)$ such that $B'' \neq B$, $B' \in \textsc{Sub}(B'')$ and B'' satisfies condition 1.*

4.4.3 Proofs

We prove satisfaction of Caminada and Amgoud's consistency and closure postulates for complete semantics, which implies satisfaction of these postulates for grounded, preferred, and stable semantics. Caminada and Amgoud [2007] also proposed postulates for the intersection of extensions and their conclusion sets, but since their satisfaction directly follows from satisfaction of the postulates for individual extensions, these postulates will not be reconsidered.

First, a number of intermediate properties are proven. The intermediate result stated in Lemma 4 is identical to Lemma 37 of Modgil and Prakken [2013], namely that any strict continuation B of a set of arguments $\{A_1, \ldots, A_n\}$ is acceptable with respect to a set $\mathcal{S}$ if all A_i are acceptable with respect to $\mathcal{S}$. The proof follows similar to Lemma 37 of Modgil and Prakken [2013], where alternative attack is now also considered.

Lemma 4. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $B \in \mathcal{A}$ be a strict continuation of $\{A_1, \ldots, A_n\}$, and for $i = 1, \ldots, n$, let A_i be acceptable with respect to $\mathcal{S} \subseteq \mathcal{A}$. Then B is acceptable with respect to $\mathcal{S}$.*

Proof. Let A be any argument such that $(A, B) \in \mathcal{D}$. By Definition 39, A attacks B iff A rebuts B (on B'), A undercuts B (on B'), or A alternative attacks B (on B') for some $B' \in \textsc{Sub}(B)$ (see Definitions 40, 41, and 42). Here, it holds that $\textsc{TopInf}(B')$ is defeasible; more specifically:

1. By Definition 40, A rebuts B (on B') iff B' is of the form $B_1'', \ldots, B_n'' \twoheadrightarrow_g p$ for some $B_1'', \ldots, B_n'' \in \mathcal{A}$, $p \in \mathbf{P}$ and hence iff $\textsc{TopInf}(B')$ is defeasible, and;
2. By Definition 41, A undercuts B (on B') iff there exists an exception arc $exc \in \mathbf{Exc}$ such that $exc\colon \textsc{Conc}(A) \rightsquigarrow g$ and $\textsc{TopGen}(B') = g \in \mathbf{G}_\mathrm{d}$. Hence, in constructing B' step 2b cannot be applied last, as this step can only be applied with strict generalisations $g \in \mathbf{G}_\mathrm{s}$. Therefore, step 2a of step 3 of Definition 37 is applied last in constructing B'. Thus, the last used inference in constructing B' is a defeasible deductive inference using $\textsc{TopGen}(B') = g$ (step 2a of Definition 37) or an abductive inference using $\textsc{TopGen}(B') = g$ (step 3 of Definition 37), and hence $\textsc{TopInf}(B')$ is defeasible, and;
3. By Definition 42, A alternative attacks B (on B') iff $\textsc{TopInf}(B')$ is an abductive inference and hence iff $\textsc{TopInf}(B')$ is defeasible.

Hence, by definition of strict continuations (Definition 48), it must be that $(A, B) \in \mathcal{D}$ iff $(A, A_i) \in \mathcal{D}$ for some (possibly more than one) $A_i \in \{A_1, \ldots, A_n\}$. Specifically, if A does not undercut, rebut or alternative attack some A_i, then this contradicts that $(A, B) \in \mathcal{D}$. Thus, we have shown that if $(A, B) \in \mathcal{D}$, then $(A, A_i) \in \mathcal{D}$ for some $A_i \in \{A_1, \ldots, A_n\}$. By assumption, A_i is acceptable with respect to $\mathcal{S}$, thus $\exists C \in \mathcal{S}$ such that $(C, A) \in \mathcal{D}$. Thus, B is acceptable with respect to $\mathcal{S}$. ■

The intermediate result stated in Lemma 5 is similar to Proposition 8 of Modgil and Prakken [2013]. Compared to Proposition 8 of Modgil and Prakken, in which no assumptions are made regarding A, we now assume that A is defeasible with a strict top inference or that A is strict, as these are the only cases needed in our proof of Theorem 1. As Modgil and Prakken do not impose any restrictions on argument construction in their formalism, a result proven by Caminada and Amgoud [2007] (i.e. their Lemma 6) can be directly used to complete their proof. Below, we show that the restrictions that are imposed on argument construction in our argumentation formalism based on IGs do not restrict the construction of strict continuations, and hence that the proof can similarly be completed.

Lemma 5. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $A, B \in \mathcal{A}$ such that B is defeasible,* $\textsc{Conc}(A) = -\textsc{Conc}(B)$. *Let A be strict or let A be defeasible with* $\textsc{TopInf}(A)$ *strict. Then for all $B' \in M(B)$, there exists a strict continuation A^+ of $(M(B) \setminus \{B'\}) \cup \{A\}$ such that A^+ rebuts B on B'.*

Proof. Let A be strict or let A be defeasible with $\textsc{TopInf}(A)$ strict. Let B be defeasible, and let $\textsc{Conc}(A) = -\textsc{Conc}(B)$. First, note that according to Definition 48, any strict continuation of a given set of arguments from $\mathcal{A}$ is either (1) A if the set of arguments under consideration is $\{A\}$ (Definition 48, condition 1), or (2) is constructed by applying step 2b of Definition 37 one or more (but finitely many) times (Definition 48, condition 2). As restrictions are imposed on argument construction in our argumentation formalism based on IGs, we first show that in constructing any strict continuation A^+ of $(M(B) \setminus \{B'\}) \cup \{A\}$, step 2b of Definition 37 can be applied without restrictions.

Generally, in applying step 2b of Definition 37, an argument C with $\textsc{Conc}(C) = p$ is constructed from arguments $C_1, \ldots, C_n$ by strictly deductively inferring p from $\textsc{Conc}(C_1), \ldots, \textsc{Conc}(C_n)$ according to Definition 31 using a generalisation g: $\textsc{Conc}(C_1), \ldots, \textsc{Conc}(C_n) \rightarrow p$ in $\mathbf{G}_\mathrm{S}$. In Definition 31, no constraints are imposed on performing deduction with strict generalisations $g \in \mathbf{G}_\mathrm{S}$; in particular, the only constraint that is imposed is in condition 2 of this definition, where constraints are imposed on performing deduction with *defeasible* generalisations in $\mathbf{G}^\mathrm{e}$ (i.e. Pearl's constraint). The only other case in which step 2b of Definition 37 cannot be applied in constructing an argument C using a $g \in \mathbf{G}_\mathrm{S}$ is in case the same g was already used in the previous construction step to construct an argument $C' \in \textsc{ImmSub}(C)$, namely by applying step 3 of Definition 37. Now again consider argument A. By assumption, A is strict or $\textsc{TopInf}(A)$ strict, and therefore step 3 of Definition 37, which specifies a defeasible inference, could not have been applied last in constructing A; therefore, no restrictions are imposed on constructing strict continuations A^+ of $(M(B) \setminus \{B'\}) \cup \{A\}$ in our argumentation formalism. By assumption, $(\mathcal{A}, \mathcal{D})$ is well defined and, therefore, closed under transposition; hence, by straightforward generalisation of Lemma 6 in [Caminada and Amgoud, 2007] one can construct a strict continuation A^+ that continues $(M(B) \setminus \{B'\}) \cup \{A\}$ with strict inferences and that concludes $-\textsc{Conc}(B')$. By construction of $M(B)$, B' has a defeasible top inference and therefore A^+ rebuts B'. But then A^+ also rebuts B. ■

The intermediate result stated in Lemma 6 is identical to Lemma 38 of Modgil and Prakken [2013].

Lemma 6. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $A \in \mathcal{A}$ be acceptable with respect to admissible extension $\mathcal{S} \subseteq \mathcal{A}$. Let $\mathcal{S}' = \mathcal{S} \cup \{A\}$. Then $\forall B \in \mathcal{S}'$, neither $(A, B) \in \mathcal{D}$ nor $(B, A) \in \mathcal{D}$.*

Proof. Suppose for contradiction that: (1) $\exists B \in \mathcal{S}'$ such that $(A, B) \in \mathcal{D}$. As $B \in \mathcal{S}'$, it follows that B is acceptable with respect to $\mathcal{S}$, as either $B = A$, which is acceptable with respect to $\mathcal{S}$ by assumption, or B is an element of admissible extension $\mathcal{S}$. Hence, $\exists C \in \mathcal{S}$ such that $(C, A) \in \mathcal{D}$. Then, as A is acceptable with respect to $\mathcal{S}$, $\exists D \in \mathcal{S}$ such that $(D, C) \in \mathcal{D}$, contradicting $\mathcal{S}$ is conflict-free; (2) $\exists B \in \mathcal{S}'$ such that $(B, A) \in \mathcal{D}$. As A is acceptable with respect to $\mathcal{S}$, $\exists C \in \mathcal{S}$ such that $(C, B) \in \mathcal{D}$, contradicting $\mathcal{S}$ is conflict-free. ■

The result stated in Lemma 7 is identical to Lemma 35-2 of Modgil and Prakken [2013], namely that an argument A defeats an argument B iff A defeats some sub-argument B' of B. Compared to Lemma 35-2 of Modgil and Prakken [2013], alternative attack is now also considered in the proof.

Lemma 7. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $A, B \in \mathcal{A}$. Then $(A, B) \in \mathcal{D}$ iff $(A, B') \in \mathcal{D}$ for some $B' \in \textsc{Sub}(B)$.*

Proof. By Definition 39, $(A, B) \in \mathcal{D}$ iff A rebuts B (on B'), A undercuts B (on B'), or A alternative attacks B (on B') for some $B' \in \textsc{Sub}(B)$ (see Definitions 40, 41, and 42); hence, also $(A, B') \in \mathcal{D}$. ■

The intermediate result stated in Lemma 8 is identical to Proposition 10 of Modgil and Prakken [2013].

Lemma 8. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_{\mathcal{I}}$ and evidence set $\mathbf{E_p}$. Let $A \in \mathcal{A}$ be acceptable wrt admissible extension $\mathcal{S} \subseteq \mathcal{A}$. Then $\mathcal{S}' = \mathcal{S} \cup \{A\}$ is conflict-free.*

Proof. We need to show that there do not exist $B, C \in \mathcal{S}'$ such that $(B, C) \in \mathcal{D}$. As $\mathcal{S}$ is an admissible extension, $\mathcal{S}$ is conflict free: hence, there do not exist $B, C \in \mathcal{S}$ such that $(B, C) \in \mathcal{D}$. Thus, we need to show that $(A, A) \notin \mathcal{D}$, and neither $(A, B) \in \mathcal{D}$ nor $(B, A) \in \mathcal{D}$ for all $B \in \mathcal{S}$. As by assumption A is acceptable with respect to $\mathcal{S}$, this follows directly from Lemma 6. ■

Theorem 1, corresponding to the direct consistency postulate, states that the conclusions of arguments in an admissible extension (and so by implication in a complete extension) are directly consistent. The conclusions of arguments in an extension should not be contradictory, as this leads to what Caminada and Amgoud call 'absurdities' [Caminada and Amgoud, 2007, p. 15] in that two contradictory statements can then be justified at the same time.

Theorem 1 (Direct consistency)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_{\mathcal{I}}$ and evidence set $\mathbf{E_p}$. Then for all admissible extensions $\mathcal{S}$ of AF it holds that the set $\{\textsc{Conc}(A) \mid A \in \mathcal{S}\}$ is directly consistent.*

Proof. Let $\mathcal{S}$ be an admissible extension of AF and let $A, B \in \mathcal{S}$. We show that if $\textsc{Conc}(A) = q$, $\textsc{Conc}(B) = r$ with $q = -r$ (i.e. $\{\textsc{Conc}(A) \mid A \in \mathcal{S}\}$ is not directly consistent), then this leads to a contradiction:

1. If A is a strict argument, and:
   1.1 if B is also a strict argument, then this contradicts our *axiom consistency* assumption on evidence sets $\mathbf{E_p}$;
   1.2 if B is a defeasible argument, and:
      1.2.1 if B has a defeasible top inference, then A rebuts B (on B) by Definition 40, as a negation arc n: $\textsc{Conc}(A) \leftrightsquigarrow \textsc{Conc}(B)$ exists in $\mathbf{N}$ (as $q = -r$). Hence, this contradicts $\mathcal{S}$ is conflict-free.
      1.2.2 if B has a strict top inference, then by Lemma 5 there exists a strict continuation A^+ of $(M(B) \setminus \{B'\}) \cup \{A\}$ for every $B' \in M(B)$ such that A^+ rebuts B on B'; hence, $(A^+, B) \in \mathcal{D}$. By our Lemma 4, A^+ is acceptable with respect to $\mathcal{S}$, and by Lemma 8, $\mathcal{S} \cup \{A^+\}$ is conflict-free, contradicting that $(A^+, B) \in \mathcal{D}$.
2. If A is a defeasible argument and B is a strict argument, then the result follows similar to case 1.2 with the roles of arguments A and B reversed.
3. If A and B are defeasible arguments, and:
   3.1 if $\textsc{TopInf}(A)$ or $\textsc{TopInf}(B)$ is defeasible, then the result follows similar to case 1.2.1 (either with the roles of arguments A and B as they currently are or with their roles reversed).
   3.2 if $\textsc{TopInf}(A)$ and $\textsc{TopInf}(B)$ are strict, then the result follows similar to case 1.2.2. ■

The result stated in Lemma 9 is identical to Lemma 35-3 of Modgil and Prakken [2013].

Lemma 9. *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $\mathcal{S} \subseteq \mathcal{A}$ and let $A \in \mathcal{S}$ with $A' \in \textsc{Sub}(A)$. Then A' is acceptable with respect to $\mathcal{S}$ if A is acceptable with respect to $\mathcal{S}$.*

Proof. Assume that A is acceptable with respect to $\mathcal{S}$. We need to prove that for every argument B such that $(B, A') \in \mathcal{D}$, $\exists C \in \mathcal{S}$ such that $(C, B) \in \mathcal{D}$. Let $B \in \mathcal{A}$ and assume that $(B, A') \in \mathcal{D}$. By Lemma 7, $(B, A) \in \mathcal{D}$. Then, as A is acceptable with respect to $\mathcal{S}$, $\exists C \in \mathcal{S}$ such that $(C, B) \in \mathcal{D}$. Hence, A' is acceptable with respect to $\mathcal{S}$. ■

Below, Caminada and Amgoud's [2007] closure and indirect consistency postulates are stated. Informally, the closure postulates state that the conclusions returned by an argumentation system should be 'complete' [Caminada and Amgoud, 2007, p. 16]. The sub-argument closure postulate states that for any argument A in a complete extension $\mathcal{S}$, all sub-arguments of A are also in $\mathcal{S}$.

Theorem 2 (Sub-argument closure)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Then for all complete extensions $\mathcal{S}$ of AF it holds that if an argument A is in $\mathcal{S}$ then all sub-arguments $A' \in \textsc{Sub}(A)$ of A are in $\mathcal{S}$.*

Proof. Let $\mathcal{S}$ be a complete extension of AF, let $A \in \mathcal{S}$ and let $A' \in \textsc{Sub}(A)$. Then A' is acceptable with respect to $\mathcal{S}$ by Lemma 9. Then $\mathcal{S} \cup \{A'\}$ is conflict-free by Lemma 8. Hence, since $\mathcal{S}$ is complete, it holds that $A' \in \mathcal{S}$. ■

Theorem 3, corresponding to the strict closure postulate, states that the conclusions of arguments in a complete extension are closed under strict inference.

Theorem 3 (Closure under strict inferences)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $\mathcal{S}$ be a complete extension of AF. Then $\{\textsc{Conc}(A) \mid A \in \mathcal{S}\} = \mathrm{CL}(\{\textsc{Conc}(A) \mid A \in \mathcal{S}\})$.*

Proof. It suffices to show that any strict continuation X of $\{A \mid A \in \mathcal{S}\}$ is in $\mathcal{S}$. By Lemma 4, any such X is acceptable wrt $\mathcal{S}$. By Lemma 8, $\mathcal{S} \cup \{X\}$ is conflict-free. Hence, since $\mathcal{S}$ is complete, $X \in \mathcal{S}$. ■

Finally, Theorem 4, corresponding to the indirect consistency postulate, states the mutual consistency of the strict closure of conclusions of arguments in a complete extension.

Theorem 4 (Indirect consistency)**.** *Let $(\mathcal{A}, \mathcal{D})$ be an AF defined by IG $G_\mathcal{I}$ and evidence set $\mathbf{E_p}$. Let $\mathcal{S}$ be a complete extension of AF. Then $\{\textsc{Conc}(A) \mid A \in \mathcal{S}\}$ is indirectly consistent.*

Proof. The result follows from Theorems 1 and 3. ■

To conclude this section, we have shown that instantiations of our argumentation formalism based on IGs satisfy Caminada and Amgoud's [2007] consistency and closure postulates. Satisfaction of these postulates warrants the sound definition of instantiations of our argumentation formalism and implies that anomalous results as identified by [Caminada and Amgoud, 2007] are avoided.

4.5 Discussion and concluding remarks

In this chapter, we have proposed an argumentation formalism that allows for both deductive and abductive argumentation, the latter of which has received relatively little attention in argumentation. Our argumentation formalism is based on our IG-formalism and generates an argumentation framework [Dung, 1995]. By formalising analyses performed by domain experts using the informal sense-making tools they are familiar with (e.g. mind maps) as IGs as an intermediary step, this therefore allows us to assign Dung's argumentation semantics to argumentation frameworks constructed on the basis of IGs. Besides allowing for rebuttal and undercutting attack, which are among the types of attacks that are typically distinguished in structured argumentation, we have also defined the notion of alternative attack among arguments based on IGs, a concept based on the notion of alternative explanations that is inspired by [Bench-Capon and Prakken, 2006; Bex, 2015]. Alternative attack captures a crucial aspect of abductive reasoning, namely that of conflict between abductively inferred conclusions [Console and Dupré, 1994; Josephson and Josephson, 1994]. We have contributed to the literature on computational argumentation by allowing for the formal evaluation of arguments involved in this type of conflict.

We have proven a number of formal properties of our approach. We have proven that arguments constructed in our argumentation formalism based on IGs adhere to the identified constraints on performing inferences with causal, evidential, abstractions, and other types of information (see Section 2.1.4). Moreover, we have shown that instantiations of our argumentation formalism satisfy key rationality postulates [Caminada and Amgoud, 2007], which warrants the sound definition of instantiations of our argumentation formalism and implies that anomalous results such as issues regarding inconsistency and non-closure as identified by [Caminada and Amgoud, 2007] are avoided.

Of the concepts defined in Chapter 3, only competing alternative explanations (Section 3.4.5), inference chains (Section 3.5) and some of the concepts regarding generalisation chains (Section 3.3, p. 47) have not been used in defining our argumentation formalism based on IGs. The importance of these concepts will become apparent in the following chapter, in which we consider BN construction from IGs.

Chapter 5

Constructing Bayesian networks from information graphs

In the preceding chapter, we demonstrated the use of our IG-formalism in guiding the construction of AFs. By formalising analyses performed using sense-making tools as IGs as an intermediary step, our approach allows for the formal evaluation of AFs on the basis of IGs using Dung's [1995] argumentation semantics. In the current chapter we provide another application of our IG-formalism, where we demonstrate the use of the IG-formalism in guiding BN construction. We propose a structured approach for automatically constructing a directed BN graph from an IG. In our approach, we focus on exploiting the knowledge expressed in an IG to constrain the graphical structure of the BN and the conditional independence relation it encodes by means of the d-separation criterion. Moreover, we demonstrate that the inferences that can be read from an IG given the evidence provide for qualitative constraints on the probability distribution represented by the BN.

We expect direct IG construction to be more straightforward than direct BN construction for domain experts unfamiliar with the BN-formalism, a claim that should be empirically evaluated in future research. We believe this to be a plausible assumption, however, among other things due to the fact that the arcs of a BN are easily misinterpreted by domain experts unfamiliar with BNs as non-symmetric relations of cause and effect instead of collectively encoding an independence relation [Dawid, 2010], making manual BN construction a difficult, time-consuming and error-prone process (see also van der Gaag and Helsper [2002]). Moreover, it is justified to assume that information regarding causality is present in the domain expert's original analysis (see Bex [2011] and van den Braak et al. [2008]), and in manual BN graph construction, conditional independencies are typically not directly elicited, but instead the *notion of causality* is commonly used as a guiding principle [Fenton and Neil, 2012; Jensen and Nielsen, 2007].

In IGs, causality information is made explicit by means of causal and evidential generalisations and can thus be directly used in BN graph construction. We first focus on IGs that only include such generalisations without enablers in Section 5.1, where we prove a number of properties of our approach in Section 5.2. We formally prove that BN graphs constructed by our approach capture reasoning patterns similar to those represented by the original IG. Moreover, we identify conditions under which the fully automatically constructed initial graph is guaranteed to be a DAG, and identify bounds on the complexity of probabilistic inference in BNs constructed by our approach. In Sections 5.3 and 5.4 we then discuss extending our approach to IGs including abstractions and generalisations of type 'other', as well as generalisations that include enablers.

5.1 Constructing BNs from IGs conform the notion of causality

In this section, we motivate and present our approach for constructing BN graphs from IGs, where we first consider the special case in which IGs only include causal and evidential generalisations without enablers. In Sections 5.1.1 and 5.1.2 we motivate the steps of our approach for automatically constructing an initial BN graph from an IG; the approach itself is presented in Section 5.1.3. In Section 5.1.4 we then explain and illustrate the steps of our approach with several examples.

5.1.1 Extracting a BN graph from an IG

First, we consider the graphical structure $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ of the BN. For constructing a BN graph from an IG, the IG's structure $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ is used, specifically the propositions $\mathbf{P}$, generalisations $\mathbf{G}$, exceptions $\mathbf{Exc}$ and negations $\mathbf{N}$ expressed in the graph.

Information in proposition nodes. For every proposition $p \in \mathbf{P}$ in an IG, we propose to form a single BN node in $\mathbf{V}$ describing both values p and $\neg p$, as captured by step 1 of our approach (see Section 5.1.3). By this step, two propositions $p, -p \in \mathbf{P}$ involved in negation are captured as two mutually exclusive values of the same node. Negation arcs present in an IG can thus be disregarded in construction of the BN graph, as such arcs are drawn between a pair of propositions $p, q \in \mathbf{P}$ iff $q = -p$ and are therefore captured in the definition of the BN nodes.

Information in causal and evidential generalisations. As noted above, in the manual construction of BN graphs arcs are typically directed using the notion of causality as a guiding principle. Specifically, if the domain expert indicates that p or $\neg p$ typically causes q or $\neg q$, then the arc is set from node P to node Q. By following this heuristic, causes form a head-to-head connection in the node corresponding to their common effect. As such, possible interactions between causes, for example due to the fact that they could be in competition, can be directly captured in the CPT for this node. Hence, we propose to use the same heuristic in automatically directing arcs, where we exploit causality information explicitly expressed in an IG by means of causal and evidential generalisations. Specifically, arcs in the BN graph are set in the same direction as generalisation arcs in $\mathbf{G}^{\mathrm{c}}$ and in the opposite direction for generalisation arcs in $\mathbf{G}^{\mathrm{e}}$. This is captured by step 2 of our approach.

Information in exceptions. Arcs in **Exc** denote exceptions to generalisations. For instance, if a generalisation is in the evidential direction, then an exception suggests an alternative explanation for the same effect (see also Bex [2015, p. 15]). Multiple exceptions to an evidential generalisation then express different alternative explanations for the same effect. Exceptions to causal generalisations do not suggest alternative explanations for the same effect, but do possibly interact with them (examples are provided in Section 5.1.4.2). Accordingly, we propose to enable capturing possible interactions between an exception and a generalisation arc, if any, in the CPTs for head-to-head nodes formed in the BN graph. This is captured by step 3 of our approach.

5.1.2 Extracting qualitative probabilistic constraints from IGs

By itself, a generalisation arc only captures knowledge about the world in conditional form; only when considering the available evidence $\mathbf{E_p}$ in the IG can directionality of inference be read from the graph. In comparison, from a BN graph we can read the chains between nodes that are active given the evidence and will be exploited to propagate the evidence upon probabilistic inference. In our approach, we want to ensure that the sequences of propositions that can be iteratively inferred from each other given $\mathbf{E_p}$ in an IG (i.e. inference chains, see Section 3.5) are captured in the BN graph by means of active chains given the available evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p} \subseteq \mathbf{P}$. In Section 5.2.2 we formally prove that BN graphs constructed by our approach indeed allow reasoning patterns similar to the inference chains that can be read from the original IG given the evidence.

Exploiting competing alternative explanations. Probabilistic constraints on the BN under construction are derived by considering the inferences that can be read from an IG given $\mathbf{E_p}$. In case the tails of two causal generalisations are abductively inferred from the common head given $\mathbf{E_p}$, these sets of tails are competing alternative explanations for the common effect expressed by the head (see Definition 33 from Section 3.4.5, illustrated in Figure 3.10a and explained in Example 43). In this case, we propose to constrain the CPT for the variable corresponding to the head such that the explaining away effect can occur between the variables corresponding to the tails of the generalisations, as captured by step 5a. In case the tails of a *single* causal generalisation are abductively inferred from the head given $\mathbf{E_p}$, then the tails are not in competition *among themselves* and the explaining away effect should not occur, as captured by step 5b. Similarly, the tails of a causal or evidential generalisation are not in competition among themselves in case the head is deductively inferred from the tails given $\mathbf{E_p}$, which is captured by the same step (illustrated and explained in Section 3.4.5, Figure 3.10b and Example 43). In step 5b, probabilistic constraints for causal generalisations are defined on CPTs for nodes in the BN under construction, but for evidential generalisations constraints are defined that cannot be directly imposed on one of the CPTs as divergent connections instead of convergent connections are formed. For evidential generalisations, we impose the constraint that the probability that the head is true given one of its tails should not decrease in the presence of one of its other tails.

We note that various evidence sets $\mathbf{E_p}$ can be used to establish inferences from the same IG, and thus that, depending on $\mathbf{E_p}$, (non-conflicting) constraints may be derived on different CPTs or different (conditional) probabilities of the BN under construction. The structure of the BN does not depend on $\mathbf{E_p}$, as the IG's structure is used in BN graph construction and not the IG's inferences.

Exploiting interactions between exceptions and generalisations. The presence of an exception to a generalisation g weakens a deductive or abductive inference performed with g. Depending on whether the tails of g are abductively inferred from the head given $\mathbf{E_p}$ or the head is deductively inferred from the tails given $\mathbf{E_p}$, different probabilistic constraints are derived, as captured by step 6 of our approach.

5.1.3 Steps of the approach

In this subsection, we present the steps of our approach. Let $\mathrm{Var}\colon \mathbf{P} \to \mathbf{V}$ be an operator mapping every proposition p or $\neg p \in \mathbf{P}$ in an IG to a BN node $\mathrm{Var}(p) = \mathrm{Var}(\neg p) \in \mathbf{V}$ describing values p and $\neg p$. For an IG $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ with $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^{\mathrm{c}}$, a BN graph $G_\mathcal{B} = (\mathbf{V}, \mathbf{A}_\mathcal{B})$ is constructed as follows:

1) $\forall p, \neg p \in \mathbf{P}$, include $\mathrm{Var}(p)$ in $\mathbf{V}$; if p or $\neg p \in \mathbf{E_p}$, also include $\mathrm{Var}(p)$ in $\mathbf{E_V}$.
2) For every generalisation arc $g\colon \{p_1, \ldots, p_n\} \to p$:
   2a) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(p_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_\mathcal{B}$.
   2b) If $g \in \mathbf{G}^{\mathrm{c}}$, include $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ in $\mathbf{A}_\mathcal{B}$.
3) For every exception arc $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}$ with $g\colon \{q_1, \ldots, q_n\} \to q$:
   3a) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_\mathcal{B}$.
   3b) If $g \in \mathbf{G}^{\mathrm{c}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q)$ in $\mathbf{A}_\mathcal{B}$.

While our approach exploits the domain knowledge captured in the IG in constructing an initial BN graph, the IG may lack information needed to prevent cycles and unwarranted (in)dependencies in the obtained BN graph; hence, we include the following validation step, which is standard in BN construction:

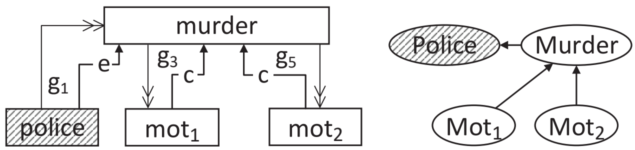
4) Verify the properties of the constructed graph $G_\mathcal{B}$ by applying the standard graph validation step (see Section 2.3.3).

Finally, we define several probabilistic constraints on the BN under construction:

5) For every generalisation arc $g\colon \mathbf{P_1} \to q$ in $\mathbf{G}$, $\mathbf{P_1} = \{p_1, \ldots, p_n\} \subseteq \mathbf{P}$:
   5a) $\forall g'\colon \mathbf{Q} \to q$ in $\mathbf{G}$, $\mathbf{Q} = \{q_1, \ldots, q_m\} \subseteq \mathbf{P}$, $g \neq g'$ such that both $g, g' \in \mathbf{G}^{\mathrm{c}}$ and for which, given $\mathbf{E_p}$, $\mathbf{P_1}$ and $\mathbf{Q}$ are competing alternative explanations for the common effect expressed by q (see Definition 33), constrain the CPT for $\mathrm{Var}(q)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p_i), \mathrm{Var}(q_j)\}, q)$ for $p_i \in \mathbf{P_1} \setminus \mathbf{Q}$, $q_j \in \mathbf{Q} \setminus \mathbf{P_1}$.
   5b) If $g \in \mathbf{G}^{\mathrm{c}}$ is on an inference chain given $\mathbf{E_p}$ (see Section 3.5, p. 56), constrain the CPT for $\mathrm{Var}(q)$ such that $\mathbf{X}^{\delta}(\{\mathrm{Var}(p_i), \mathrm{Var}(p_j)\}, q)$ with $\delta \neq -$, $p_i, p_j \in \mathbf{P_1}$, $p_i \neq p_j$. If $g \in \mathbf{G}^{\mathrm{e}}$ is on an inference chain given $\mathbf{E_p}$, constrain the probabilities of the BN such that $\Pr(q \mid p_i, p_j) \not< \Pr(q \mid p_i, \neg p_j)$ for $p_i, p_j \in \mathbf{P_1}$, $p_i \neq p_j$.
6) For every $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}$ with $p \in \mathbf{P}$ and $g\colon \{q_1, \ldots, q_n\} \to q$ in $\mathbf{G}$:
   6a) If $g \in \mathbf{G}^{\mathrm{e}}$ and q is deductively inferred from $q_1, \ldots, q_n$ given $\mathbf{E_p}$ using g, constrain the CPT for $\mathrm{Var}(q_i)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p), \mathrm{Var}(q)\}, q_i)$, $i = 1, \ldots, n$. If in addition $\exists exc'\colon p' \rightsquigarrow g$ in $\mathbf{Exc}$, further constrain the CPT for $\mathrm{Var}(q_i)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p), \mathrm{Var}(p')\}, q_i)$, $i = 1, \ldots, n$.
   6b) If $g \in \mathbf{G}^{\mathrm{c}}$ and q is deductively inferred from $q_1, \ldots, q_n$ given $\mathbf{E_p}$ using g, constrain the CPT for $\mathrm{Var}(q)$ such that $\Pr(q \mid p, q_1, \ldots, q_n) < \Pr(q \mid \neg p, q_1, \ldots, q_n)$.
   6c) If $g \in \mathbf{G}^{\mathrm{c}}$ and $q_1, \ldots, q_n$ are abductively inferred from q given $\mathbf{E_p}$ using g, constrain the probabilities of the BN such that $\Pr(q_i \mid p, q) < \Pr(q_i \mid \neg p, q)$, $i = 1, \ldots, n$.

We reiterate that the initially constructed BN by our approach should always be verified by the BN modeller in consultation with the domain expert, which includes verifying the derived probabilistic constraints. After this verification step, the derived constraints can be used in subsequent probability assessment, thereby partially simplifying it. In particular, since we are considering BN construction in data-poor domains the required conditional probabilities will often need to be elicited from domain experts, where it can be monitored whether the assessed conditional probabilities satisfy the derived probabilistic constraints.

We note that the above probabilistic constraints concern (intercausal) interactions between individual nodes and not sets, as to the best of our knowledge no approaches have been proposed in the literature that allow for capturing interactions between sets of parents of a node. The type of competition between sets of proposition nodes in an IG as captured by Definition 33 can, therefore, not be straightforwardly captured between variables in a corresponding BN; instead, in step 5a we propose to constrain the CPT for $\mathrm{Var}(q)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p_i), \mathrm{Var}(q_j)\}, q)$ for *pairs* of propositions $p_i \in \mathbf{P_1} \setminus \mathbf{Q}$, $q_j \in \mathbf{Q} \setminus \mathbf{P_1}$, where the intersection of $\mathbf{P_1}$ and $\mathbf{Q}$ is not considered. Similarly, in step 6a interactions between pairs of nodes and not sets are considered. In future work, it can be investigated whether the concept of product synergy can be extended to sets of nodes.



		Murder	
Mot_1	Mot_2	t	f
t	t	0.4	0.6
t	f	0.6	0.4
f	t	0.5	0.5
f	f	0.1	0.9

(a) (b) (c)

Figure 5.1: The IG of Figure 3.10a (a); the BN graph constructed from this IG by our approach (b); a possible CPT for node Murder (c).

Upon using our approach, arcs in the BN under construction are directed from cause to effect; therefore, for nodes in the BN under construction for which probabilistic constraints are directly imposed on the CPTs the necessary conditional probabilities can be elicited in the form of likelihood ratios, which as noted in the introduction of this thesis are commonly used in probabilistic evaluation in domains such as forensics, law and medicine (see Section 1.2.2). In some cases probabilistic constraints in the above steps are defined that cannot be directly imposed on one of the CPTs for nodes in the BN under construction. We note that approaches have been proposed that allow one to also use these probabilistic constraints in an elicitation procedure for obtaining the required local probability distributions [Druzdzel and van der Gaag, 1995].

5.1.4 Explanation and illustration of the approach

In this section, we explain and illustrate the steps of our approach through our running example. In Section 5.1.4.1 we illustrate that steps $1-2$ of our approach suffice for constructing BN graphs from restricted IGs not including exception arcs, where the (conditional) probabilities of the BN under construction should be constrained according to step 5. In Section 5.1.4.2 we then illustrate that the BN under construction needs to be further constrained in case exception arcs are present in the IG; this is accounted for in steps 3 and 6 of our approach.

5.1.4.1 Explanation and illustration of steps $1-2$ and 5

First, we explain and illustrate the main idea behind our approach by applying it to the IG depicted in Figure 5.1a.

Steps $1-2$. The first step is to capture every proposition in $G_{\mathcal{I}}$ and its negation as two mutually exclusive values of the same BN node in $G_{\mathcal{B}}$. In steps 2a and 2b, arcs in the BN graph are directed using the notion of causality in that for every $g \in \mathbf{G}^{\mathrm{c}}$, arcs in the BN graph are directed from nodes corresponding to $\mathbf{Tails}(g)$ to $\mathrm{Var}(Head(g))$, and vice versa for $g \in \mathbf{G}^{\mathrm{e}}$. This formalises the approach typically taken in the manual construction of BN graphs, namely that of setting arcs in the causal direction as a guiding principle [Fenton and Neil, 2012; Jensen and Nielsen, 2007]. The resulting BN graph is depicted in Figure 5.1b.

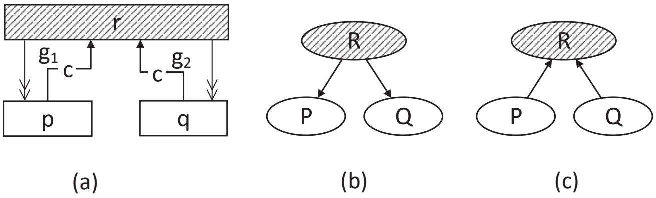
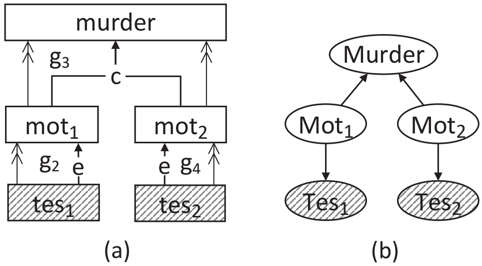


Figure 5.2: Example of an IG (a); the BN graph constructed by directing arcs according to the *inferences* that can be read from this IG given $\mathbf{E_p}$ (b); the BN graph constructed by directing arcs according to the *generalisations* in the IG (c).

Step 5a. The inferences that can be read from an IG given the evidence allow us to derive constraints on the (conditional) probabilities of the BN under construction. In the IG of Figure 5.1a, given $\mathbf{E_p} = \{police\}$ propositions mot_1 and mot_2 are abductively inferred from *murder* using g_3 and g_5, respectively, as given $\mathbf{E_p}$ *murder* is deductively inferred from *police* using g_1. Therefore, mot_1 and mot_2 are competing alternative explanations for common effect *murder* in that accepting one explanation will diminish our belief in the other (see Definition 33). We propose to link this type of intercausal interaction in IGs to the explaining away effect in BNs. Specifically, as proposed in step 5a of our approach, the CPT for Murder should be constrained such that $\mathbf{X}^-(\{\mathrm{Mot}_1, \mathrm{Mot}_2\}, murder)$. Note that the IG only informs us that there should be a negative product synergy between Mot_1 and Mot_2 wrt value Murder $=$ *true*; it does not inform us whether a product synergy should also be exhibited between these variables wrt value Murder $=$ *false*, as proposition $\neg murder$ does not appear in the IG. Figure 5.1c depicts a possible CPT for Murder, where $\mathbf{X}^-(\{\mathrm{Mot}_1, \mathrm{Mot}_2\}, murder)$ as $0.4 \cdot 0.1 \leq 0.6 \cdot 0.5$. However, as $0.6 \cdot 0.9 \geq 0.4 \cdot 0.5$, it also holds that $\mathbf{X}^+(\{\mathrm{Mot}_1, \mathrm{Mot}_2\}, \neg murder)$. Care should be taken, therefore, in eliciting the involved probabilities, as it may be undesirable that a positive product synergy for value $\neg murder$ is exhibited.

By following steps 2a and 2b of our approach, causes automatically form a head-to-head connection in the node corresponding to their common effect for any given IG; interactions between causes in an IG, for instance because they are competing alternative explanations for the common effect, can, therefore, always be directly captured in the CPT for the node corresponding to the common effect. We note that directing arcs in the BN graph in the same direction as the *inferences* that can be read from an IG given the evidence would lead to undesirable results. Consider the IG depicted in Figure 5.2a. By directing arcs according to the inferences that can be read from this IG given $\mathbf{E_p}$, the BN graph of Figure 5.2b is constructed. In the IG of Figure 5.2a, p and q are competing alternative explanations for common effect r given $\mathbf{E_p}$; however, this competition cannot be directly captured in the CPT for node R in the BN graph of Figure 5.2b as a divergent connection is formed. Moreover, all chains between P and Q are blocked given $\mathbf{E_V} = \{\mathrm{R}\}$; hence, interactions between causes expressed in an IG cannot always be captured by directing arcs in a corresponding BN graph according to the induced inferences in an IG.



		Murder
Mot_1	Mot_2	t
t	t	0.8
t	f	0.2
f	t	0.2
f	f	0.1

(c)

Figure 5.3: The IG of Figure 3.10b (a); the BN graph constructed from this IG (b); a possible CPT for node Murder (c).

Step 5b. Next, consider the IG of Figure 5.3a. Given $\mathbf{E_p}$, *murder* is deductively inferred from mot_1 and mot_2 using g_3 (i.e. g_3 is on an inference chain given $\mathbf{E_p}$). Therefore, mot_1 and mot_2 are not competing alternative explanations for *murder* in this IG. By following steps $1-2$ of our approach, the BN graph of Figure 5.3b is constructed. As mot_1 and mot_2 are not competing alternative explanations for *murder* in this example, we need to assure that the explaining away effect cannot occur between Mot_1 and Mot_2 for value Murder $=$ *true*. This can be achieved by constraining the CPT for Murder such that $\mathbf{X}^{\delta}(\{\text{Mot}_1, \text{Mot}_2\}, murder)$ for $\delta \neq -$, as captured by step 5b of our approach. Specifically, a negative product synergy should not be exhibited, as mot_1 and mot_2 are not competing alternative explanations for the common effect; hence, either a zero or a positive product synergy should be exhibited. Figure 5.3c depicts a possible CPT for Murder, where $\mathbf{X}^{+}(\{\text{Mot}_1, \text{Mot}_2\}, murder)$ as $0.8 \cdot 0.1 \geq 0.2 \cdot 0.2$.

In the IG of Figure 5.4a, an IG including a single evidential generalisation arc g_2 is depicted. Given $\mathbf{E_p}$, mot_1 is deductively inferred from tes_1 and tes_2 using g_2. Therefore, tes_1 and tes_2 are not competing alternative explanations for mot_1 in this IG. The BN graph constructed from this IG is depicted in Figure 5.4b. Note that, in contrast to the BN graph of Figure 5.3b, a head-to-head node is not formed and probabilistic constraints are derived by step 5b that are not defined on the CPTs for nodes in the BN under construction. Instead, probabilistic constraints on the probabilities of the BN are derived, namely $\Pr(mot_1 \mid tes_1, tes_2) \not< \Pr(mot_1 \mid tes_1, \neg tes_2)$ and $\Pr(mot_1 \mid tes_1, tes_2) \not< \Pr(mot_1 \mid \neg tes_1, tes_2)$.

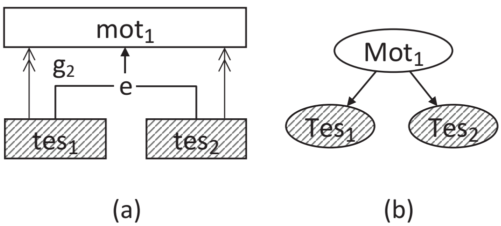
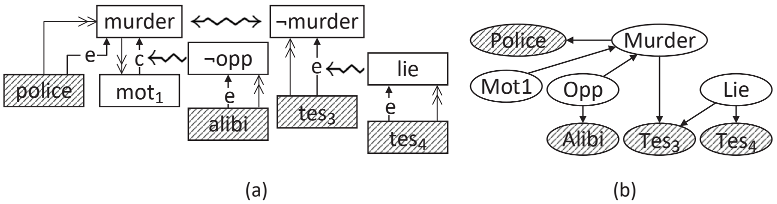


Figure 5.4: Adjustment to the IG of Figure 5.3a including an evidential generalisation arc with multiple tails (a); the BN graph constructed from this IG (b).



		Tes$_3$
Murder	Lie	t
f	t	0.2
f	f	0.8
t	t	0.3
t	f	0.01

(c)

Figure 5.5: IG including exceptions to generalisations in $\mathbf{G}^{\mathrm{e}}$ and $\mathbf{G}^{\mathrm{c}}$ (a); the BN graph constructed from this IG (b); a possible CPT for node Tes$_3$ (c).

5.1.4.2 Explanation and illustration of steps 3 and 6

Next, IGs including exception arcs are considered.

Step 3a. In Figure 5.5a, an example of an IG is depicted in which exceptions to both an evidential and a causal generalisation are provided. Proposition *lie*, which states that Marjan had reason to lie when giving her testimony, provides an exception to the evidential generalisation $tes_3 \to \neg murder$. Since tes_3 is either the result of Marjan truly not committing the murder or due to a lie, $\neg murder$ and *lie* can be seen as competing alternative explanations for Marjan's testimony; we illustrate and explain this by discussing an alternative modelling of the same IG, depicted in Figure 5.6b. In the IG of Figure 5.6b, evidential generalisation arc $tes_3 \to \neg murder$ of Figure 5.6a is replaced by causal generalisation arc $\neg murder \to tes_3$ (see also Section 2.1.3). Furthermore, as exception *lie* can be considered a cause for tes_3, causal generalisation arc $lie \to tes_3$ is included in the IG of Figure 5.6b. Given $\mathbf{E_p}$, propositions $\neg murder$ and *lie* are abductively inferred from tes_3, and hence these propositions are competing alternative explanations for tes_3. Generally, exceptions to an evidential generalisation can be considered competing alternative explanations for the common effects expressed by the antecedents of the generalisation (see also Bex [2015, p. 15]). We therefore propose to enable capturing such interactions between an exception and an evidential generalisation by forming head-to-head nodes in the nodes corresponding to the tails of the generalisation arc. By step 2a of our approach, the BN graph under construction includes arc Murder $\to$ Tes$_3$. A head-to-head node can, therefore, be formed in node Tes$_3$ by adding additional arc Lie $\to$ Tes$_3$ to the BN graph; this is captured by step 3a of our approach.

Step 6a. In the IG of Figure 5.5a, given $\mathbf{E_p} = \{police, alibi, tes_3, tes_4\}$, $\neg murder$ is inferred from tes_3. As proposition *lie* provides an exception to the generalisation used in performing this inference and thereby weakens this inference, we propose to constrain the CPT for Tes$_3$ such that the explaining away effect can occur between Lie and Murder for value Tes$_3$ = *true*. This is achieved by constraining the CPT for Tes$_3$ such that $\mathbf{X}^-(\{\text{Lie}, \text{Murder}\}, tes_3)$, as captured by step 6a of our approach. In this particular example, $\neg murder$ is one of the possible causes of tes_3; therefore,

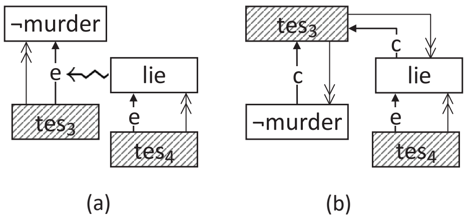


Figure 5.6: Part of the IG of Figure 5.5a (a); alternative modelling of this IG fragment (b).

for variable Murder the ordering *false* > *true* is assumed. For example, the CPT for Tes_3 can be chosen as in Figure 5.5c, as in this case it holds that $\Pr(tes_3 \mid \neg murder, lie) \cdot \Pr(tes_3 \mid murder, \neg lie) = 0.2 \cdot 0.01 \leq \Pr(tes_3 \mid \neg murder, \neg lie) \cdot \Pr(tes_3 \mid murder, lie) = 0.8 \cdot 0.3$.

We note that multiple exceptions to an evidential generalisation arc g express different alternative explanations for the common effects expressed by $\mathbf{Tails}(g)$. In particular, multiple exceptions to an evidential generalisation arc can be alternatively modelled using causal generalisations in a similar manner as illustrated for a single exception in Figure 5.6, where each exception expresses a different alternative explanation for the common effects expressed by $\mathbf{Tails}(g)$. We therefore propose to constrain the CPTs for the nodes corresponding to the tails such that a negative product synergy is exhibited between the nodes corresponding to each pair of exceptions, as captured by step 6a of our approach.

Step 3b. In the IG of Figure 5.5a, proposition $\neg\, opp$, which states that Marjan did not have opportunity to commit the murder as she has an alibi (*alibi*), provides an exception to the causal generalisation arc $mot_1 \rightarrow murder$. In contrast with the exception to the evidential generalisation arc, this exception cannot be considered a competing alternative explanation for the tail of the generalisation arc; the absence of opportunity cannot be considered a cause for motive. Instead, it allows us to infer that Marjan did not murder Leo ($\neg murder$). For exceptions to generalisations $g \in \mathbf{G}^{\mathrm{c}}$, we therefore propose to form a head-to-head node in $\mathrm{Var}(Head(g))$ as opposed to in $\mathrm{Var}(p_i)$ for $p_i \in \mathbf{Tails}(g)$. By step 2b of our approach, the BN graph under construction includes arc $\mathrm{Mot}_1 \rightarrow \mathrm{Murder}$. A head-to-head node can, therefore, be formed in Murder by adding additional arc $\mathrm{Opp} \rightarrow \mathrm{Murder}$ to the BN graph; this is captured by step 3b of our approach. The corresponding BN graph is depicted in Figure 5.5b.

Steps 6b-c. Bex and Renooij [2016] previously noted that, for deduction, the presence of a proposition opposing an inference step from $q_1, \ldots, q_n$ to q should decrease the probability that q is true. We propose to take a similar approach for exceptions to causal generalisations. For generalisations $q_1, \ldots, q_n \rightarrow q$ in $\mathbf{G}^{\mathrm{c}}$ for which q is deductively inferred from $q_1, \ldots, q_n$ given $\mathbf{E_p}$ using g in presence of an exception p, we propose to constrain the CPT for $\mathrm{Var}(q)$ such that $\Pr(q \mid p, q_1, \ldots, q_n) < \Pr(q \mid \neg p, q_1, \ldots, q_n)$, as captured by step 6b of our approach. In case $q_1, \ldots, q_n$ are abductively inferred from q given $\mathbf{E_p}$ using g, the probability that q_i is true given q should decrease in the presence of an exception p for $i = 1, \ldots, n$. Accordingly, we propose to constrain the probabilities of the BN such that $\Pr(q_i \mid p, q) < \Pr(q_i \mid \neg p, q)$, $i = 1, \ldots, n$, as captured by step 6c of our approach. Note that the latter constraints cannot be directly imposed on the CPTs for nodes $\mathrm{Var}(p)$, $\mathrm{Var}(q)$, or $\mathrm{Var}(q_i)$, as nodes $\mathrm{Var}(q_i)$ and $\mathrm{Var}(p)$ are parents of node $\mathrm{Var}(q)$ by steps 2b and 3b of our approach.

5.2 Properties of the approach

In this section, we prove a number of formal properties of our approach. In Section 5.2.1 we study conditions on IGs under which the fully automatically constructed initial BN graph is guaranteed to be acyclic. In Section 5.2.2 we prove that, as intended, BN graphs constructed by our approach capture all inference chains that can be read from an IG given the evidence in the form of induced active chains. In Section 5.2.3 we look into the size of the CPTs and complexity of probabilistic inference in BN graphs constructed by our approach. Finally, in Section 5.2.4 we look into mapping properties of our approach; specifically, we investigate conditions under which the same BN graph is constructed from different IGs by our approach, and discuss ways by which a distinction can be made in the (conditional) probabilities of the BN under construction.

5.2.1 Constructing acyclic graphs

In this section, we study conditions under which the initial graph constructed by steps $1-3$ of our approach is guaranteed to be a DAG. Hence, under these conditions the (manual) verification step of whether the obtained graph contains cycles (part of step 4 of our approach) can be skipped.

Conditions a) and b) of Proposition 3 concern the existence of exception arcs in IGs. Specifically, cycles are possibly introduced within weakly connected components of the BN graph under construction in step 3 of our approach in case exception arcs exist within weakly connected components of IGs (condition a). Furthermore, cycles are also possibly introduced in the BN graph from a node V_1 in one weakly connected component via a node V_2 in another weakly connected component in this step in case exception arcs exist between propositions in separate weakly connected components of IGs (condition b). Examples of IGs violating these conditions are provided after the formal result. The terms 'non-repetitive' and 'consistent' generalisation chains used in condition c) were introduced in Section 3.3 (p. 47); informally, for IGs adhering to condition c) the possibility of using a proposition p to deductively infer itself or $-p$ is excluded.

Proposition 3. *Let* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *be an IG with* $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ *and for which* $\mathbf{Enabler}(g) = \emptyset$ *for every* $g \in \mathbf{G}^{\mathrm{c}}$. *Let* $G_{\mathcal{I}}^{*} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}}^{*})$ *be the possibly disconnected sub-graph of* $G_{\mathcal{I}}$ *with* $\mathbf{A}_{\mathcal{I}}^{*} = \mathbf{A}_{\mathcal{I}} \setminus \mathbf{Exc}$. *Let* $\mathbf{C} = \{C = (\mathbf{P}^{\mathrm{C}}, \mathbf{A}_{\mathcal{I}}^{\mathrm{C}}) \mid \mathbf{P}^{\mathrm{C}} \subseteq \mathbf{P}, \mathbf{A}_{\mathcal{I}}^{\mathrm{C}} \subseteq \mathbf{A}_{\mathcal{I}}^{*}, C$ *is a weakly connected component of* $G_{\mathcal{I}}^{*}\}$ *be the set of* IG components. *Assume that the following conditions are satisfied:*

a) For any IG component $C \in \mathbf{C}$, *there does not exist an exc*: $p \rightsquigarrow g$ *in* $\mathbf{Exc}$ *with* $p \in \mathbf{P}^{\mathrm{C}}$, $g \in \mathbf{A}_{\mathcal{I}}^{\mathrm{C}}$.

b) For every pair of IG components $C_1, C_2 \in \mathbf{C}$, *there does not exist both an* $exc_1 \colon p_1 \rightsquigarrow g_1$ *in* $\mathbf{Exc}$ *with* $p_1 \in \mathbf{P}^{\mathrm{C}_1}$, $g_1 \in \mathbf{A}_{\mathcal{I}}^{\mathrm{C}_2}$ *and an* $exc_2 \colon p_2 \rightsquigarrow g_2$ *in* $\mathbf{Exc}$ *with* $p_2 \in \mathbf{P}^{\mathrm{C}_2}$, $g_2 \in \mathbf{A}_{\mathcal{I}}^{\mathrm{C}_1}$.

c) Generalisation chains in $G_{\mathcal{I}}$ *are non-repetitive and consistent (Section 3.3, p. 47).*

Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ *be the graph constructed from* $G_{\mathcal{I}}$ *according to steps* $1-3$ *of our approach. Then* $G_{\mathcal{B}}$ *is a DAG.*

Proof. By setting arcs in $\mathbf{A}_{\mathcal{B}}$ per step 2 of our approach, no cycles are introduced. Specifically, our non-repetitiveness and consistency assumptions on generalisation chains (condition c) jointly assume that for every $p \in \mathbf{P}$ there does not exist a generalisation chain $[g_1, \ldots, g_m]$ with $p \in \mathbf{Tails}(g_1)$ such that either $Head(g_m) = p$ or $Head(g_m) = -p$. Therefore, no chain of arcs exists in $\mathbf{A}_{\mathcal{B}}$ from a node P to itself. The only other case in which cycles are possibly introduced in $G_{\mathcal{B}}$ is when a causal cycle exists in $G_{\mathcal{I}}$, which is also prohibited by assumption (see Section 3.3, p. 47).

We now prove that if $C \in \mathbf{C}$ is an IG component of $G_{\mathcal{I}}$, then the BN segment C' obtained from C after step 2 is a weakly connected component of the thus far constructed BN graph $G_{\mathcal{B}}$. Let $C \in \mathbf{C}$ be an IG component of $G_{\mathcal{I}}$. Then propositions within C are interconnected by arcs in $\mathbf{G} \cup \mathbf{N} \subseteq \mathbf{A}_{\mathcal{I}}$ but are not connected to other propositions in the supergraph $G_{\mathcal{I}}$; therefore, corresponding nodes in BN segment C' are interconnected but are not connected to other nodes in supergraph $G_{\mathcal{B}}$. This is the case as per step 2, $\mathbf{A}_{\mathcal{B}}$ only includes arcs between the variables corresponding to $\mathbf{Tails}(g)$ and $Head(g)$ for every $g \in \mathbf{G}$ in the IG; no arcs are introduced corresponding to negation arcs $n \in \mathbf{N}$. We then call C' the weakly connected component *corresponding to* IG component C.

In step 3 of our approach, additional arcs are included in $\mathbf{A}_{\mathcal{B}}$ for every $exc \in \mathbf{Exc}$ in the IG. We now prove that no cycles are introduced within the weakly connected components of BN graph $G_{\mathcal{B}}$ or from a node V_1 in one weakly connected component to itself via a node V_2 in another weakly connected component of $G_{\mathcal{B}}$ in step 3. Under condition a), no cycles are introduced within a weakly connected component C' of $G_{\mathcal{B}}$ in this step. Specifically, C' contains no cycles after step 2 and no cycles are introduced in C' in step 3 as no exception arc is directed from a $p \in \mathbf{P}^{\mathrm{C}}$ to a $g \in \mathbf{A}_{\mathcal{I}}^{\mathrm{C}}$ in corresponding IG-component C. Furthermore, for every pair of IG components C_1 and C_2 of $G_{\mathcal{I}}$ with corresponding weakly connected components C_1' and C_2' of $G_{\mathcal{B}}$, no cycles are introduced from a node $\mathrm{V}_1 \in C_1'$ to itself via a node $\mathrm{V}_2 \in C_2'$ under condition b). The resulting BN graph is therefore acyclic. ∎

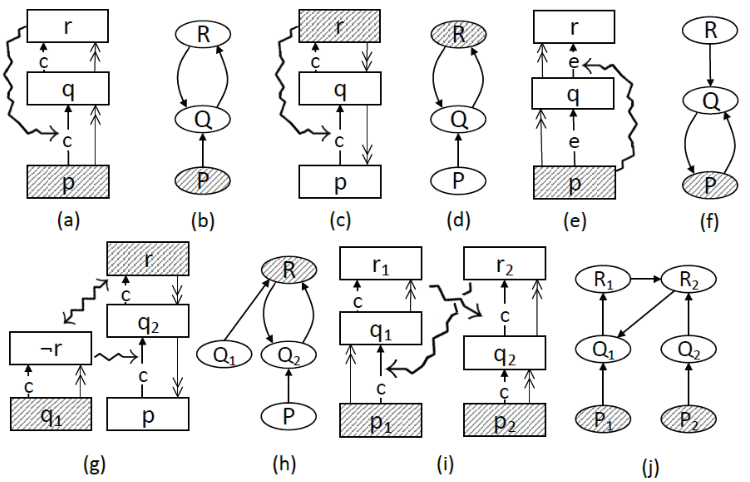


Figure 5.7: Examples of IGs (a, c, e, g, i) for which a cyclic graph is constructed by steps $1-3$ of our approach (b, d, f, h, j).

Figures 5.7a, 5.7c, 5.7e and 5.7g depict examples of IGs that do not satisfy condition a) of Proposition 3 and hence result in cyclic graphs. An IG violating only condition a) may contain:

1a) A generalisation chain $[g_1, \ldots, g_m]$, $g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{c}}$ and an exception arc $exc\colon Head(g_j) \rightsquigarrow g_i$ for $1 \leq i < j \leq m$ (see Figures 5.7a and 5.7c), or;

1b) A generalisation chain $[g_1, \ldots, g_m]$, $g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{e}}$ and an exception arc $exc\colon Head(g_i) \rightsquigarrow g_j$ for $1 \leq i < j \leq m$ (see Figure 5.7e), or;

2) Propositions r, $\neg r$ with $n\colon r \leftrightsquigarrow \neg r$ in $\mathbf{N}$, where $\neg r$ provides an exception to a generalisation g_i in a generalisation chain $[g_1, \ldots, g_m]$ with either:
   2a) $Head(g_m) = r$ and $g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{c}}$ (see Figure 5.7g), or;
   2b) $r \in \mathbf{Tails}(g_1)$ and $g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{e}}$.

For 1a), $Head(g_j)$ poses an exception to a generalisation that is used in iteratively inferring $Head(g_j)$ in case solely deductive inferences are performed with the generalisations in the chain, as illustrated in Figure 5.7a. In case solely abductive inferences are performed with the generalisations in the chain, $Head(g_j)$ instead poses an exception to a generalisation that is used to iteratively abductively infer another proposition from $Head(g_j)$, as illustrated in Figure 5.7c. Similarly, for 1b) $Head(g_i)$ poses an exception to a generalisation that is used to iteratively deductively infer another proposition from $Head(g_i)$, as only deduction can be performed with evidential generalisations; this is illustrated in Figure 5.7e. For 2), an example is provided in Figure 5.7g. Finally, an example of an IG violating condition b) of Proposition 3 is provided in Figure 5.7i.

In the validation step that follows the initial construction of BN graphs corresponding to IGs violating conditions a), b) and c) of Proposition 3, arcs can be reversed or removed to make these graphs acyclic. The choice of arc to reverse or remove will depend on its effect on active chains, including those between nodes not directly incident on the arc. We note that this type of (manual) verification is standard in BN construction, especially in data-poor domains. While the domain knowledge expressed in the original IG has been exploited to construct an initial BN graph, additional knowledge may need to be elicited to obtain a valid graph.

5.2.2 Capturing induced inference chains as active chains

In this section, we study whether BN graphs constructed by our approach capture reasoning patterns similar to those that can be read from the original IG given the evidence. Recall that an IG, by means of its inference chains (see Section 3.5), describes sequences of propositions that can be iteratively inferred from each other given the available evidence. In comparison, from a BN graph we can read the chains between nodes that are active given the evidence and will be exploited to propagate the evidence upon probabilistic inference. Note that, similar to active chains for BNs, inference chains do not need to start in evidence (see Example 45), but that in contrast to active chains inference chains are generally not symmetrical (see Example 46). We now formally prove that all inference chains that can be read from an IG given the evidence $\mathbf{E_p} \subseteq \mathbf{P}$ are captured in the BN graph by means of active chains given the available evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p}$. This result implies that, for every inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ given $\mathbf{E_p}$, $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ are not d-separated given the evidence for $\mathbf{E_V}$.

Proposition 4. *Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG with $\mathbf{G}^\mathrm{a} = \mathbf{G}^\mathrm{o} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^\mathrm{c}$. Let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set, and let $G_\mathcal{B} = (\mathbf{V}, \mathbf{A}_\mathcal{B})$ be the BN graph constructed from $G_\mathcal{I}$ according to steps $1-3$ of our approach. Let $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ be any inference chain that can be read from $G_\mathcal{I}$ given $\mathbf{E_p}$. Then there exists an active chain between Var(p_1) and Var(p_n) in $G_\mathcal{B}$ given the evidence for $\mathbf{E_V}$.*

Proof. Following steps $1-2$ of our approach, a sequence of nodes and arcs is formed between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ in $G_\mathcal{B}$, as for every g_i, $1 \leq i < n$ arcs between $\mathbf{Tails}(g_i)$ and $Head(g_i)$ are added to $\mathbf{A}_\mathcal{B}$. By our non-repetitiveness and consistency assumptions on inference chains, this is a sequence of *distinct* nodes and arcs and thus a chain in $G_\mathcal{B}$. We now prove that this chain in the BN graph is active given $\mathbf{E_V}$, as all options to block a chain do not occur. First, note that per Lemma 1 it holds that $p_i \notin \mathbf{E_p}$ for $i > 1$; therefore, corresponding nodes $\mathrm{Var}(p_i)$ in the BN graph are not instantiated and hence do not block chains. Possibly only $p_1 \in \mathbf{E_p}$. However, in this case, the corresponding node $\mathrm{Var}(p_1)$ is an end-point of the chain which, therefore, does not block it. Hence, chains between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ are never blocked by $\mathbf{E_V}$.

The only other option to block a chain occurs in case it includes an uninstantiated head-to-head node without instantiated descendants. Consider subchain $(p_{i-1}, g_{i-1}, p_i, g_i, p_{i+1})$ of our inference chain for an arbitrary $1 < i < n$. We show that the corresponding chain in the BN graph does not include a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$. Note that by steps 2a and 2b of our approach, a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ is only formed by setting arcs corresponding to generalisation arc g_{i-1} and g_i in case:

1. $g_{i-1} \in \mathbf{G}^{\mathrm{e}}$, $Head(g_{i-1}) = p_{i-1}$, $p_i \in \mathbf{Tails}(g_{i-1})$, and either:
   1a) $g_i \in \mathbf{G}^{\mathrm{e}}$, $Head(g_i) = p_{i+1}$, $p_i \in \mathbf{Tails}(g_i)$, or;
   1b) $g_i \in \mathbf{G}^{\mathrm{c}}$, $Head(g_i) = p_i$, $p_{i+1} \in \mathbf{Tails}(g_i)$.
2. $g_{i-1} \in \mathbf{G}^{\mathrm{c}}$, $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$, and either:
   2a) $g_i \in \mathbf{G}^{\mathrm{c}}$, $Head(g_i) = p_i$, $p_{i+1} \in \mathbf{Tails}(g_i)$, or;
   2b) $g_i \in \mathbf{G}^{\mathrm{e}}$, $Head(g_i) = p_{i+1}$, $p_i \in \mathbf{Tails}(g_i)$.

However, in performing the inference steps from p_{i-1} to p_i and from p_i to p_{i+1} none of these combinations of generalisations could have been used, as proven in Proposition 1. Thus the chain between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ corresponding to inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ does not include a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ for $1 < i < n$ and is therefore never blocked.

Finally, in step 3 $\mathbf{A}_{\mathcal{B}}$ is extended for exception arcs. This step does not change the chains formed between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ in step 2, which therefore remain active given $\mathbf{E}_{\mathbf{V}}$. ■

The implication in the other direction of Proposition 4 does not generally hold. Specifically, it does not generally hold that for every induced active chain in a BN graph constructed from an IG $G_{\mathcal{I}}$, there exists a corresponding induced inference chain in $G_{\mathcal{I}}$. For instance, since the notion of an active chain is a symmetrical concept, a BN graph will also capture reasoning patterns in the direction opposite of the inference chains that can be read from an IG. As inference chains are generally not symmetrical (see Example 46), reasoning patterns may appear in the BN graph that do not appear in the original IG.

5.2.3 Size and complexity of constructed BNs

The following properties concern the size and complexity of the resulting BN model. Proposition 5 gives an upper-bound on the total number of nodes and arcs introduced in a BN graph constructed from an IG by our approach.

Proposition 5. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG with $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^{\mathrm{c}}$. Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the BN graph constructed from $G_{\mathcal{I}}$ according to steps $1-3$ of our approach. Let $\mathbf{Exc}^{\mathrm{e}}$ and $\mathbf{Exc}^{\mathrm{c}}$ be disjoint subsets of $\mathbf{Exc}$ consisting of exceptions to generalisation arcs in $\mathbf{G}^{\mathrm{e}}$ and $\mathbf{G}^{\mathrm{c}}$, respectively. Then $|\mathbf{V}| = |\mathbf{P}| - |\{p \mid p \in \mathbf{P} \text{ and } \neg p \in \mathbf{P}\}|$ and $|\mathbf{A}_{\mathcal{B}}| \leq \sum\limits_{g \in \mathbf{G}} |\mathbf{Tails}(g)| + |\mathbf{Exc}^{\mathrm{c}}| + \sum\limits_{p \rightsquigarrow g \text{ in } \mathbf{Exc}^{\mathrm{e}}} |\mathbf{Tails}(g)|$.*

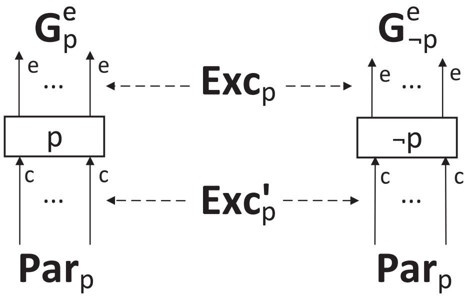


Figure 5.8: Illustration of the terminology used in Proposition 6.

Proof. By step 1 of our approach, both p and its negation are mapped to the same node $\mathrm{Var}(p) = \mathrm{Var}(\neg p) \in \mathbf{V}$. Therefore, the exact number of nodes introduced in this step is $|\mathbf{P} \setminus \{p \mid p \in \mathbf{P} \text{ and } -p \in \mathbf{P}\}|$. In step 2, at most $|\mathbf{Tails}(g)|$ arcs are added to $\mathbf{A}_{\mathcal{B}}$ for every $g \in \mathbf{G}$. For every $exc \in \mathbf{Exc}^{\mathrm{c}}$, one additional arc is added to $\mathbf{A}_{\mathcal{B}}$ in step 3b. For every $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}^{\mathrm{e}}$, at most $|\mathbf{Tails}(g)|$ arcs are added to $\mathbf{A}_{\mathcal{B}}$ in step 3a. ■

Proposition 6 gives an upper-bound on the number of parents introduced by our approach for each node $\mathrm{Var}(p)$ in $\mathbf{V}$, which bounds both the size of the CPTs and the complexity of probabilistic inference in the BN [Darwiche, 2009, pp. 141–142]. Informally, this bound captures the number of generalisation arcs and exception arcs that involve either proposition p or $\neg p$. The terminology used in Proposition 6 is illustrated in Figure 5.8.

Proposition 6. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG with $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^{\mathrm{c}}$. Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the BN graph constructed from $G_{\mathcal{I}}$ according to steps $1-3$ of our approach. For every $p \in \mathbf{P}$, let $\mathbf{Par}_p = \{p_i \mid p_i \in \mathbf{Tails}(g), g \in \mathbf{G}^{\mathrm{c}}, Head(g) \in \{p, \neg p\}\}$. Let $\mathbf{G}^{\mathrm{e}}_p$ be a subset of $\mathbf{G}^{\mathrm{e}}$, where $g \in \mathbf{G}^{\mathrm{e}}_p$ iff $p \in \mathbf{Tails}(g)$. Let $\mathbf{Exc}_p \subseteq \mathbf{Exc}$ be the subset of exception arcs directed to a $g \in \mathbf{G}^{\mathrm{e}}_p$ or a $g \in \mathbf{G}^{\mathrm{e}}_{\neg p}$. Similarly, let $\mathbf{Exc}'_p \subseteq \mathbf{Exc}$ be the subset of exception arcs directed to a $g \in \mathbf{G}^{\mathrm{c}}$ for which $Head(g) \in \{p, \neg p\}$. Then an upper-bound for the number of parents of Var(p) is:*

$$|\mathbf{Par}_p| + |\mathbf{Exc}_p| + |\mathbf{Exc}'_p| + |\mathbf{G}^{\mathrm{e}}_p| + |\mathbf{G}^{\mathrm{e}}_{\neg p}|$$

Proof. For every $g \in \mathbf{G}^{\mathrm{c}}$ with $Head(g) \in \{p, \neg p\}$, $\mathrm{Var}(p)$ has at most $|\mathbf{Tails}(g)|$ parents by step 2b of our approach; hence the term $|\mathbf{Par}_p|$. By steps 3a and 3b, $\mathbf{A}_{\mathcal{B}}$ includes a single arc directed towards $\mathrm{Var}(p)$ for every exception exc in $\mathbf{Exc}_p$ or in $\mathbf{Exc}'_p$, respectively; hence the terms $|\mathbf{Exc}_p|$ and $|\mathbf{Exc}'_p|$. For every $g \in \mathbf{G}^{\mathrm{e}}$ with p or $\neg p$ in $\mathbf{Tails}(g)$, a single arc directed towards $\mathrm{Var}(p)$ is included in $\mathbf{A}_{\mathcal{B}}$ by step 2a of our approach; hence the terms $|\mathbf{G}^{\mathrm{e}}_p|$ and $|\mathbf{G}^{\mathrm{e}}_{\neg p}|$. ■

Note that, in case $\mathbf{G}^{\mathrm{e}}_p = \mathbf{G}^{\mathrm{e}}_{\neg p} = \emptyset$, it follows that $\mathbf{Exc}_p = \emptyset$; hence, terms $|\mathbf{G}^{\mathrm{e}}_p|$, $|\mathbf{G}^{\mathrm{e}}_{\neg p}|$ and $|\mathbf{Exc}_p|$ are equal to zero in this case. Similarly, $\mathbf{Par}_p$ may be empty, in which case $\mathbf{Exc}'_p = \emptyset$ and terms $|\mathbf{Par}_p|$ and $|\mathbf{Exc}'_p|$ are equal to zero.

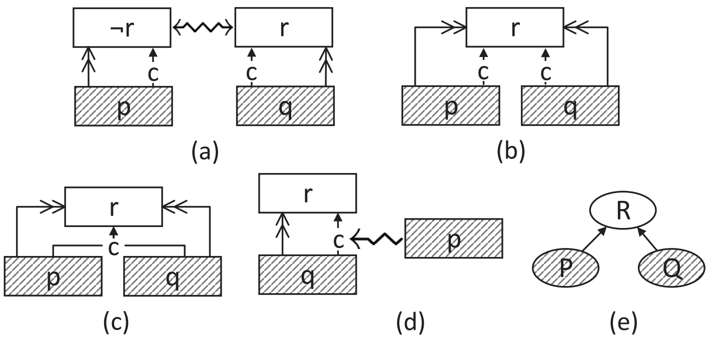


Figure 5.9: Examples of IGs (a-d) for which the same BN graph (e) is constructed by our approach.

5.2.4 Mapping properties and probabilistic constraints

Finally, we investigate conditions under which the same BN graph is constructed from different IGs by our approach, and discuss ways by which a distinction can be made between these different cases in the (conditional) probabilities of the BN under construction. First, we prove in Proposition 7 that for every finite BN graph $G_{\mathcal{B}}$, there exists a finite IG such that this IG is mapped to $G_{\mathcal{B}}$ by our approach.

Proposition 7. *Let* $\mathbf{IG}$ *be the space of finite IGs* $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ *with* $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ *and for which* $\mathbf{Enabler}(g) = \emptyset$ *for every* $g \in \mathbf{G}^{\mathrm{c}}$*. Let* $\mathbf{BN}$ *be the space of finite BN graphs. Let* $\mathcal{F} \colon \mathbf{IG} \to \mathbf{BN}$ *be the function defined by steps* $1-3$ *of our approach. Then* $\mathcal{F}$ *is a surjection.*

Proof. Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be a BN graph in $\mathbf{BN}$. Then we need to find at least one IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}}) \in \mathbf{IG}$ such that $\mathcal{F}(G_{\mathcal{I}}) = G_{\mathcal{B}}$. Define $G_{\mathcal{I}}$ as follows. For every node $\mathrm{P} \in \mathbf{V}$, include proposition $p \in \mathbf{P}$. For every arc $\mathrm{P}_1 \to \mathrm{P}_2 \in \mathbf{A}_{\mathcal{B}}$, include generalisation arc $g \colon p_1 \to p_2$ in $\mathbf{G}^{\mathrm{c}}$. Then $\mathcal{F}(G_{\mathcal{I}}) = G_{\mathcal{B}}$ by steps 1 and 2b. ■

However, $\mathcal{F}$ is not an injection. Figures 5.9a-d depict examples of IGs for which the same BN graph, namely the graph depicted in Figure 5.9e, is constructed by $\mathcal{F}$. Possible differences between these IGs can be captured in the (conditional) probabilities of the BN under construction. In Figure 5.9a, a negation arc is drawn between r and $\neg r$. A possible probabilistic interpretation is that this IG informs us on probabilities $\Pr(r \mid p, q)$ and $\Pr(\neg r \mid p, q)$ (see also Prakken [2018b]), where a preference for r over $\neg r$ defines an ordering on these two probabilities. For reasons mentioned in Section 4.3, we have currently not accounted for preferences in our IG-formalism; hence, possible probabilistic constraints resulting from such preferences are not further discussed. In Figure 5.9b, p and q can each be considered sufficient for deductively inferring r, while in Figure 5.9c both p and q are needed. A possible probabilistic interpretation of the IG in Figure 5.9b is that it only informs us on probabilities $\Pr(r \mid p)$ and $\Pr(r \mid q)$ and not on $\Pr(r \mid p, q)$, while the reverse holds for the IG of Figure 5.9c. Figure 5.9c is distinguished from Figure 5.9a, as Figure 5.9c only informs us on $\Pr(r \mid p, q)$ while Figure 5.9a also informs us on $\Pr(r \mid q)$ and $\Pr(\neg r \mid p)$. For exception arcs, specific probabilistic constraints are derived, as captured by step 6 of our approach. Specifically, in the example of Figure 5.9d, constraint $\Pr(r \mid p, q) < \Pr(r \mid \neg p, q)$ is derived.

5.3 Extending the approach to any IG

We now discuss extending our approach to general IGs as defined in Chapter 3, i.e. including abstractions and generalisations of type 'other', as well as causal generalisations that include enablers. For our extended approach we wish to preserve the desirable properties as proven for our original approach, namely acyclicity of the initially constructed graph (Proposition 3) and the properties regarding the captured reasoning patterns in BNs constructed from IGs (Proposition 4) as much as possible. We only focus on deriving the graphical structure of the BN, where we show that for the solution proposed, inference chains in general IGs which only describe deductive inferences given the evidence are captured as induced active chains in the constructed BN graph. Furthermore, we identify conditions on general IGs under which initial BN graphs constructed by the extended approach are guaranteed to be acyclic.

In Section 5.3.1 we motivate the steps of our extended approach. The extended approach itself is presented in Section 5.3.2. In Section 5.3.3 we explain and illustrate the steps of the extended approach with examples. In Section 5.4 we then prove a number of formal properties of our extended approach.

5.3.1 Extracting a BN graph from a general IG

In this section, we motivate the steps of our extended approach for constructing BN graphs from general IGs.

Information in enabling conditions. For causal generalisations that include enabling conditions, we propose to direct arcs conform the notion of causality, in the same way as specified in step 2b of our original approach (see Section 5.1.3). Specifically, for every generalisation arc $g\colon \{p_1, \ldots, p_n\} \rightarrow p$ in $\mathbf{G}^{\mathrm{c}}$, arcs $\mathrm{Var}(p_i) \rightarrow \mathrm{Var}(p)$, $i = 1, \ldots, n$ are included in $\mathbf{A}_{\mathcal{B}}$ regardless of whether p_i is an enabler or not. This is captured by step 2b$'$ of our extended approach.

Information in abstractions and generalisations of type 'other'. In the manual construction of BN graphs, arcs are typically directed using the notion of causality as a guiding principle; however, non-causal relations are also considered in the literature. For instance, in the BN construction approach of Neil and colleagues [2000] not only causal but also definitional relations are considered, in which arcs in the BN graph are oriented in the direction in which a sub-attribute (or combination of sub-attributes) defines an attribute. Directing arcs in this manner increases the interpretability of the graph; hence, for abstractions and generalisations of type 'other' we propose to direct arcs in a similar manner. Specifically, for generalisation arcs $g\colon \{p_1, \ldots, p_n\} \to p$ in $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, we propose to include arcs $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$, in a similar manner as specified for causal generalisations in step 2b′; this is captured by step 2c′ of the extended approach. By directing arcs in this manner for multiple abstractions with the same head, head-to-head connections are formed in the node corresponding to the head which allows for directly capturing possible interactions between the tails in the CPT for this node, for example due to the fact that they could be competing alternative explanations.

Steps 2d′ and 4a′ are included in our extended approach to guarantee that the initially constructed BN graph is guaranteed to be acyclic and to ensure that inference chains which only describe deductive inferences given the evidence are captured as active chains in the BN graph given the evidence; these steps are motivated and explained in Section 5.3.3.

Information in exceptions. Similar to exceptions to causal generalisations, exceptions to abstractions and generalisations of type 'other' do not suggest competing alternative explanations for the same effect, but do possibly interact with them. Accordingly, we propose to enable capturing possible interactions between an exception and a generalisation arc, if any, in the CPTs for head-to-head nodes formed in the BN graph. This is captured by step 3′ of our extended approach.

5.3.2 Steps of the extended approach

In this subsection, we present the steps of our extended approach. Note that steps 1′, 2a′, 2b′ and 3a′ of the extended approach are identical to steps 1, 2a, 2b and 3a of the original approach (see Section 5.1.3). For an IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$, a BN graph $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ is constructed as follows:

1′) $\forall p, \neg p \in \mathbf{P}$, include $\mathrm{Var}(p)$ in $\mathbf{V}$; if p or $\neg p \in \mathbf{E_p}$, also include $\mathrm{Var}(p)$ in $\mathbf{E_V}$.

2′) For every generalisation arc $g\colon \{p_1, \ldots, p_n\} \to p$:
  - 2a′) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(p_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
  - 2b′) If $g \in \mathbf{G}^{\mathrm{c}}$, include $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
  - 2c′) If $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ and $\nexists g_1 \in \mathbf{G}^{\mathrm{e}}$ such that $[g, g_1]$ is a generalisation chain, include $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
  - 2d′) If $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ and $\exists g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{e}}$ such that $[g_1, \ldots, g_m]$ is a maximal generalisation chain of evidential generalisation arcs in $G_{\mathcal{I}}$ following g, include $\mathrm{Var}(p_i) \to \mathrm{Var}(\mathrm{Head}(g_m))$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.

3′) For every exception arc $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}$ with $g\colon \{q_1, \ldots, q_n\} \to q$:
  - 3a′) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
  - 3b′) If $g \notin \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q)$ in $\mathbf{A}_{\mathcal{B}}$.

4′) Verify the properties of the constructed graph $G_{\mathcal{B}}$:
  - 4a′) Break cycles in $G_{\mathcal{B}}$ introduced in step 2′ by so-called *interceptors* in $G_{\mathcal{I}}$ (see Section 5.3.3, Definition 51 for further details).
  - 4b′) Apply the standard graph validation step (see Section 2.3.3).

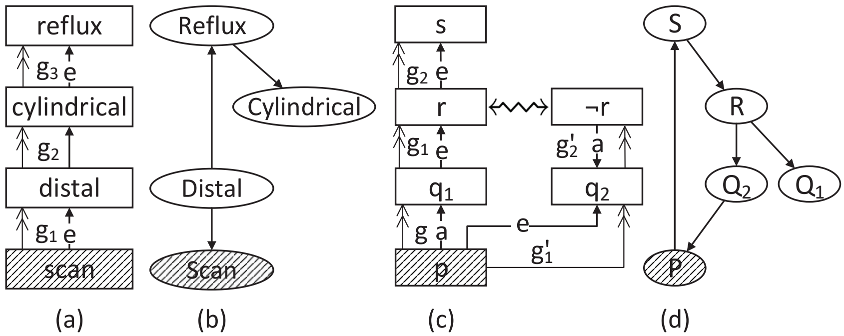


Figure 5.10: Example of an IG (a); BN graph constructed by steps $1-2\text{a}'$ and $2\text{d}'$ of the extended approach (b); example of an IG (c) for which a cycle is introduced upon directing arcs according to step $2'$ of the extended approach (d).

5.3.3 Explanation and illustration of the extended approach

In this section, we explain and illustrate the steps of our extended approach.

Step 2b′. The steps of our original approach do not need to be reconsidered for causal generalisations that include enablers, as in Definitions 31 and 32 the same constraints are imposed on performing deductive or abductive inference with a causal generalisation g regardless of whether or not $\textbf{Enabler}(g) = \emptyset$. Therefore, for causal generalisations that include enabling conditions arcs we propose to direct arcs in the BN graph in the same way as specified in step 2b of our original approach; this is captured by step 2b′ of our extended approach.

Steps 2d′ and 4a′. In Figures 5.10a-b, step 2d′ of the extended approach is illustrated for a medical example (taken from van der Gaag and Helsper [2002]). After performing a CT scan (*scan*) on a patient who has severe difficulty swallowing, it is established that a tumour is present in the lower (*distal*) part of his oesophagus. Clinical studies indicate a strong correlation between the location of an oesophageal tumour and its cell type; however, neither can be considered a cause of the other. Distal tumours generally consist of cylindrical cells (*cylindrical*), often formed as a result of frequent gastric reflux (*reflux*). In the IG of Figure 5.10a, generalisation arcs $g_1, g_3 \in \mathbf{G}^{\mathrm{e}}$ and $g_2 \in \mathbf{G}^{\mathrm{o}}$ are included. Given $\mathbf{E_p} = \{scan\}$, (*scan*, g_1, *distal*, g_2, *cylindrical*, g_3, *reflux*) is an inference chain, as no constraints are imposed in our IG-formalism on deductively chaining a generalisation of type 'other' and an evidential generalisation (i.e. g_2 and g_3). The BN graph constructed from this IG by steps $1-2\text{a}'$ and $2\text{d}'$ of the extended approach is depicted in Figure 5.10b. As arcs Distal $\rightarrow$ Reflux and Reflux $\rightarrow$ Cylindrical are included in $\mathbf{A}_{\mathcal{B}}$ and the involved nodes are not instantiated, active chains exist between Distal and Reflux and between Distal and Cylindrical. Hence, all inference chains that can be read from the IG given $\mathbf{E_p}$

are captured as active chains in the constructed BN graph given the evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p}$. Note that we do not wish to set arcs as per step 2c′ of the extended approach, as in this case head-to-head node Distal → Cylindrical ← Reflux would be formed which would block the chain between Distal and Reflux.

The technical construction described by step 2d′ is designed to ensure that inference chains which only describe deductive inferences given the evidence are captured as active chains for IGs including abstractions and generalisations of type 'other'. However, under specific conditions cycles are introduced in the BN graph by directing arcs in this manner, as illustrated in Figures 5.10c-d. In the IG of Figure 5.10c, generalisation arc $g \in \mathbf{G}^{\mathrm{a}}$ is followed by a maximal generalisation chain of evidential generalisation arcs $[g_1, g_2]$. By step 2d′ of the extended approach, arc $P \to S$ is included in the corresponding BN graph depicted in Figure 5.10d to ensure an active chain exists between nodes P and S given $\mathbf{E_V}$. By steps 2a′ and 2c′, arcs $S \to R$, $R \to Q_2$ and $Q_2 \to P$ are also included in $\mathbf{A}_{\mathcal{B}}$, introducing a cycle in $G_{\mathcal{B}}$.

Generally, cycles are introduced in the BN graph by directing arcs according to step 2′ of the extended approach in case a so-called *interceptor* exists in the IG. To define interceptors, we first define the terms 'direct precursor' and 'indirect precursor', inspired by the terms 'direct cause' and 'indirect cause' introduced in the definition of a causal cycle (Definition 27, Section 3.3). Specifically, instead of only considering $g \in \mathbf{G}^{\mathrm{c}}$, $p \in \mathbf{Ant}(g)$ as in the definition of a 'direct cause', in the definition of a 'direct precursor' we also consider $g \in \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $p \in \mathbf{Tails}(g)$, thereby including abstractions, generalisations of type 'other and causal generalisations that include enabling conditions in our definitions.

Definition 50 (Direct and indirect precursors)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Proposition $p \in \mathbf{P}$ is a* direct precursor *of $q \in \mathbf{P}$ iff $\exists g \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ with $g \in \mathbf{G}^{\mathrm{c}} \cup \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $p \in \mathbf{Tails}(g)$, $q = Head(g)$ or $g \in \mathbf{G}^{\mathrm{e}}$, $p = Head(g)$, $q \in \mathbf{Tails}(g)$. Proposition $p_1 \in \mathbf{P}$ is an* indirect precursor *of $p_3 \in \mathbf{P}$ iff $\exists p_2 \in \mathbf{P}$, $p_2 \neq p_1, p_2 \neq p_3$, such that p_1 is a direct precursor of p_2 and p_2 is a direct or indirect precursor of p_3.*

Example 56. *In the IG of Figure 5.10c, proposition $\neg r$ is a direct precursor of q_2, as $g_2' \in \mathbf{G}^{\mathrm{a}}$, $\neg r \in \mathbf{Tails}(g_2')$, $q_2 = Head(g_2')$. Proposition q_2 is a direct precursor of p, as $g_1' \in \mathbf{G}^{\mathrm{e}}$, $p \in \mathbf{Tails}(g_1')$, $q_2 = Head(g_1')$. Then $\neg r$ is an indirect precursor of p, as $\neg r$ is a direct precursor of q_2 and q_2 is a direct precursor of p.* □

An interceptor is then defined as follows.

Definition 51 (Interceptor)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Let $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ and assume that $\exists g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{e}}$ such that $[g_1, \ldots, g_m]$ is a maximal generalisation chain of evidential generalisation arcs in $G_{\mathcal{I}}$ following g. Then proposition p is an* interceptor *of generalisation chain $[g, g_1, \ldots, g_m]$ iff p is a direct or indirect precursor of a $q \in \mathbf{Tails}(g)$ and $p = Head(g_j)$ or $p = -Head(g_j)$ for a $j \in \{1, \ldots, m\}$.*

Example 57. *In the IG of Figure 5.10c, proposition $\neg r$ is an interceptor of generalisation chain $[g, g_1, g_2]$, as $\neg r$ is an indirect precursor of $p \in \mathbf{Tails}(g)$ and $\neg r = -Head(g_1)$.* □

Cycles in the BN graph introduced by interceptors in the IG can be broken by removing specific arcs introduced in step 2d′ of the extended approach. For instance, in the example of Figure 5.10d, arc $P \rightarrow S$ can be safely removed, as an active chain already exists between P and S via R and Q_2. In general, cycles are broken in step 4a′ by removing arcs $Var(q) \rightarrow Var(Head(g_m))$ from $\mathbf{A}_{\mathcal{B}}$ introduced in step 2d′ $\forall q \in \mathbf{Tails}(g)$ for which a direct or indirect precursor p of q exists such that p is an interceptor of generalisation chain $[g, g_1, \ldots, g_m]$.

5.4 Properties of the extended approach

In this section, we prove a number of formal properties of our extended approach. We focus on formally proving that our proposed technical solution allows for capturing inference chains in general IGs which only describe deductive inferences given the evidence as active chains in the constructed BN graph given the evidence (Section 5.4.1). Furthermore, we study conditions on general IGs under which the initially constructed BN graph is guaranteed to be acyclic (Section 5.4.2).

5.4.1 Capturing induced inference chains as active chains

In this section, we study whether BN graphs constructed by our extended approach capture reasoning patterns similar to those that can be read from the original IG given the evidence. In Proposition 8 we formally prove that inference chains in general IGs which only describe deductive inferences given the evidence are captured as active chains in the constructed BN graph given the evidence. Hence, for general IGs that only include generalisations with which deductive inference is performed given $\mathbf{E_p} \subseteq \mathbf{P}$, all inference chains that can be read from the IG given $\mathbf{E_p}$ are captured as active chains given the evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p}$.

Proposition 8. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the BN graph constructed from $G_{\mathcal{I}}$ according to steps $1' - 4a'$ of the extended approach. Let $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ be an inference chain that can be read from $G_{\mathcal{I}}$ given $\mathbf{E_p}$, and assume that for every $1 < i \leq n$, p_i is deductively inferred given $\mathbf{E_p}$ using generalisation g_{i-1} with $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$. Then there exists an active chain between Var(p_1) and Var(p_n) in $G_{\mathcal{B}}$ given the evidence for $\mathbf{E_V}$.*

Proof. For the case $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^{\mathrm{c}}$, the proof reduces to the proof of Proposition 4. Thus, assume that $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}} \neq \emptyset$ and that causal generalisations may include enabling conditions. First, note that for generalisations $g \in \mathbf{G}^{\mathrm{c}}$ the same constraints on performing deduction and abduction are imposed in Definitions 31 and 32 regardless of whether or not $\mathbf{Enabler}(g) = \emptyset$, and that for causal generalisations that include enabling conditions arcs in the BN graph are directed by step 2b′ in the same way as specified in step 2b of our original approach. Therefore, for IGs that include causal generalisations with enabling conditions the proof again reduces to the proof of Proposition 4. Now, consider the case that for every $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $\nexists g_1 \in \mathbf{G}^{\mathrm{e}}$ such that $[g, g_1]$ is a generalisation chain. Following steps $1' - 2'$ of the extended approach, a sequence of nodes and arcs is formed between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ in $G_{\mathcal{B}}$, as for every g_i, $1 \leq i < n$, arcs between $\mathbf{Tails}(g_i)$ and $Head(g_i)$ are added to $\mathbf{A}_{\mathcal{B}}$. By our non-repetitiveness and consistency assumptions on inference chains, this is a sequence of *distinct* nodes and arcs and thus a chain in $G_{\mathcal{B}}$. We now prove that this chain in the BN graph is active given $\mathbf{E_V}$, as all options to block a chain do not occur.

First, note that, identical to Proposition 4, nodes $\mathrm{Var}(p_i)$ in the BN graph are not instantiated and hence do not block chains between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$. Possibly only $p_1 \in \mathbf{E_p}$. However, in this case, the corresponding node $\mathrm{Var}(p_1)$ is an end-point of the chain which, therefore, does not block it. The only other option to block a chain occurs in case it includes an uninstantiated head-to-head node without instantiated descendants. Consider subchain $(p_{i-1}, g_{i-1}, p_i, g_i, p_{i+1})$ of our inference chain for an arbitrary $1 < i < n$. We show that the corresponding chain in the BN graph does not include a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$. By assumption, given $\mathbf{E_p}$, p_i is deductively inferred from p_{i-1} using g_{i-1} and p_{i+1} is deductively inferred from p_i using g_i, where for g_{i-1} it holds that $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$ and for g_i it holds that $Head(g_i) = p_{i+1}$, $p_i \in \mathbf{Tails}(g_i)$. Note that by steps 2a$'$, 2b$'$ and 2c$'$ of the extended approach, a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ is then only formed in case $g_{i-1} \in \mathbf{G}^{\mathrm{c}}$ and $g_i \in \mathbf{G}^{\mathrm{e}}$. However, in performing the deductive inference steps from p_{i-1} to p_i and from p_i to p_{i+1}, generalisations of this form could not have been used, as proven in Proposition 1. Specifically, this would violate Pearl's constraint for deductive inference as enforced in condition 2 of Definition 31. Thus the chain between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ corresponding to inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ does not include a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ for $1 < i < n$ and is therefore never blocked.

Now, assume that a $g_j \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $1 \leq j < n$, is followed by a chain of generalisation arcs in $\mathbf{G}^{\mathrm{e}}$, and let $[g_{j+1}, \ldots, g_{j+l}]$ be a maximal such chain. We consider the cases $j + l > n - 1$ and $j + l \leq n - 1$ separately; the difference between these cases is illustrated in Figure 5.11. First, the case $j + l > n - 1$ is considered. By step 2d$'$, arcs $Var(p) \to Var(Head(g_{j+l}))$ $\forall p \in \mathbf{Tails}(g_j)$ are introduced in $\mathbf{A}_{\mathcal{B}}$. By step 2a$'$, $\mathbf{A}_{\mathcal{B}}$ in addition includes a directed path from $Var(Head(g_{j+l}))$ to $Var(Head(g_{n-1}))$, as $g_{j+1}, \ldots, g_{j+l} \in \mathbf{G}^{\mathrm{e}}$. Therefore, chains between nodes in $\{Var(p) \mid p \in \mathbf{Tails}(g_j)\}$ and $Var(Head(g_{n-1}))$ via $Var(Head(g_{j+l}))$ are active given $\mathbf{E_V}$, as $Var(Head(g_{j+l}))$ is not a head-to-head node on these chains. Next, case $j + l \leq n - 1$ is considered. By step 2d$'$, arcs $Var(p) \to Var(Head(g_{j+l}))$ $\forall p \in \mathbf{Tails}(g_j)$ are introduced in $\mathbf{A}_{\mathcal{B}}$; therefore, active chains exist between all nodes in $\{Var(p) \mid p \in \mathbf{Tails}(g_j)\}$ and $Var(Head(g_{j+l}))$ given $\mathbf{E_V}$ after this step. As $[g_{j+1}, \ldots, g_{j+l}]$ is a maximal generalisation chain of evidential generalisation arcs following g_j, g_{j+l} is not followed by an evidential generalisation arc. By repeatedly

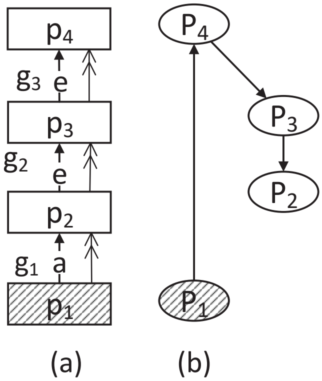


Figure 5.11: An example of an IG, where $g_1 \in \mathbf{G}^{\mathrm{a}}$ (a); the corresponding BN graph constructed by the extended approach (b). In case inference chain $(p_1, g_1, p_2, g_2, p_3)$ is considered ($n-1=2$, $j=1$, $l=2$), an active chain exists between $Var(p_1)$ (the tail of g_j) and $Var(Head(g_{n-1}))$, as arc $P_1 \to P_4$ is included by step 2d′ and arc $P_4 \to P_3$ is included in $\mathbf{A}_{\mathcal{B}}$ by step 2a′; this illustrates the case in which $j+l > n-1$ in the proof of Proposition 8. In case inference chain $(p_1, g_1, p_2, g_2, p_3, g_3, p_4)$ is considered ($n-1=3$, $j=1$, $l=2$), an active chain exists between $Var(p_1)$ and $Var(Head(g_{n-1}))$, as arc $P_1 \to P_4$ is included in $\mathbf{A}_{\mathcal{B}}$ by step 2d′; this illustrates the case in which $j+l \leq n-1$ in the proof of Proposition 8.

applying the same argument for all generalisation arcs in $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ followed by chains of generalisation arcs in $\mathbf{G}^{\mathrm{e}}$, an active chain therefore also exists between $Var(p_1)$ and $Var(p_n)$ given $\mathbf{E_V}$.

In step 4a′, a specific subset of the arcs introduced in step 2d′ is removed; we prove that for every $g_j \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $1 \leq j < n$, followed by a maximal generalisation chain of evidential generalisation arcs $[g_{j+1}, \ldots, g_{j+l}]$ active chains continue to exist between nodes in $\{Var(p) \mid p \in \mathbf{Tails}(g_j)\}$ and $Var(Head(g_{j+l}))$, and therefore between $Var(p_1)$ and $Var(p_n)$, despite of this arc removal sub-step. Arcs introduced in step 2d′ are removed in case there exists an interceptor r of generalisation chain $[g_j, g_{j+1}, \ldots, g_{j+l}]$ (see Definition 51). Assume that $r = Head(g_k)$ or $r = -Head(g_k)$ for a $k \in \{j+1, \ldots, j+l\}$ and that r is a direct or indirect precursor of a proposition $q \in \mathbf{Tails}(g_j)$. Then arc $Var(q) \to Var(Head(g_{j+l}))$ is removed in step 4a′ (see Section 5.3.3, p. 104); however, as a directed path from $Var(Head(g_{j+l}))$ to $Var(q)$ via $Var(Head(g_k))$ exists in $\mathbf{A}_{\mathcal{B}}$ by step 2′ as r is a direct or indirect precursor of q and $g_{k+1}, \ldots, g_{j+l} \in \mathbf{G}^{\mathrm{e}}$, an active chain still exists between nodes $Var(q)$ and $Var(Head(g_{j+l}))$ given $\mathbf{E_V}$.

Finally, in step 3′ $\mathbf{A}_{\mathcal{B}}$ is extended for exception arcs. Similar to Proposition 4, this step does not change the chains formed between Var(p_1) and Var(p_n) in step 2′, which therefore remain active given $\mathbf{E_V}$. ■

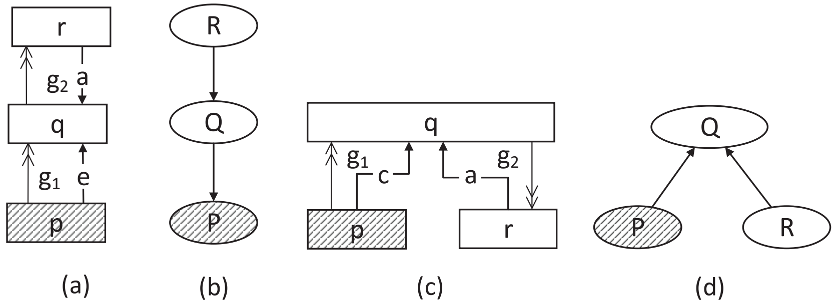


Figure 5.12: Example of an IG, where given $\mathbf{E_p}$, r is abductively inferred from q using abstraction g_2 (a); BN graph constructed by our extended approach (b); another example of an IG, where given $\mathbf{E_p}$, r is abductively inferred from q using abstraction g_2 (c); BN graph constructed by our extended approach in which head-to-head node $P \rightarrow Q \leftarrow R$ is formed (d).

We now illustrate that the result stated in Proposition 8 is not guaranteed to hold for inference chains which also describe abductive inferences given the evidence. Figures 5.12a-b depict an example for which the desirable properties as proven for our original approach are still preserved. In particular, as no uninstantiated head-to-head nodes without instantiated descendants are formed, all inference chains that can be read from the IG are captured as active chains in the constructed BN graph given the evidence. Specifically, inference chain (p, g_1, q, g_2, r) is captured as an active chain in the constructed BN graph, while the inference step from q to r given $\mathbf{E_p}$ is an abductive inference with abstraction g_2.

In Figure 5.12c an example of a general IG is depicted for which the desirable properties as proven for our original approach are not preserved. Given $\mathbf{E_p}$, q is deductively inferred from p using g_1. In turn, r is abductively inferred from q using g_2; specifically, no constraints are imposed in our IG-formalism on performing deduction and abduction in that order with a causal generalisation and an abstraction. By step 2b$'$ of the extended approach, arc $P \rightarrow Q$ is included in the BN graph of Figure 5.12d. By introducing arc $R \rightarrow Q$, in accordance with step 2c$'$ of the extended approach, head-to-head node $P \rightarrow Q \leftarrow R$ is formed. Hence, while (p, g_1, q, g_2, r) is an inference chain given $\mathbf{E_p}$, all chains between P and R are blocked given $\mathbf{E_V}$. Similar observations can be made by including g_1 in $\mathbf{G}^{\mathrm{O}}$ instead of in $\mathbf{G}^{\mathrm{C}}$, as no constraints are imposed in our IG-formalism on performing deduction and abduction in that order with a generalisation of type 'other' and an abstraction. Similarly, deduction and abduction can be performed in that order in case $g_1 \in \mathbf{G}^{\mathrm{O}}$, $g_2 \in \mathbf{G}^{\mathrm{C}}$. In future work, it may be investigated how our approach can be extended or adjusted such that inference chains in general IGs which also describe abductive inferences given the evidence are captured as active chains in the constructed BN graph.

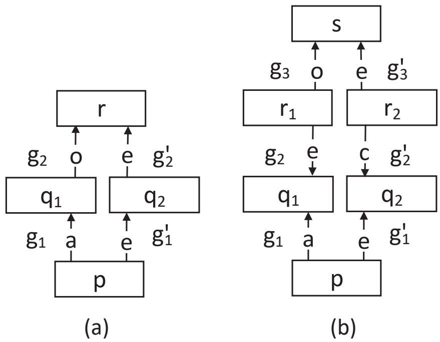


Figure 5.13: Examples of IGs including IG-cycles.

5.4.2 Constructing acyclic graphs

In this section, we study conditions on general IGs under which the initial graph constructed by steps $1' - 4a'$ of the extended approach is guaranteed to be a DAG. Conditions a), b) and c) of Proposition 9 are identical to conditions a), b) and c) of Proposition 3. As noted in the proof of Proposition 3, cycles are possibly introduced in a BN graph $G_{\mathcal{B}}$ constructed by our original approach when a causal cycle exists in $G_{\mathcal{I}}$ (see Section 3.3, Definition 27). To ensure cycles are not introduced in BN graphs $G_{\mathcal{B}}$ constructed by our extended approach, in condition d) of Proposition 9 it is assumed that general IGs do not include so-called *IG-cycles* (see Definition 52). Our definition of an IG-cycle extends the definition of a causal cycle (Definition 27) by also taking into account generalisation arcs in $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ and causal generalisations that include enabling conditions.

Definition 52 (IG-cycle)**.** *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. An* IG-cycle *exists in $G_{\mathcal{I}}$ iff $\exists p, q \in \mathbf{P}$ such that p is a direct or indirect precursor of q (see Definition 50) and q or $-q$ is a direct or indirect precursor of p or of $-p$.*

Example 58. *An IG-cycle exists in the IG of Figure 5.13a, as p is an indirect precursor of r and r is an indirect precursor of p. Similarly, an IG-cycle exists in the IG of Figure 5.13b, as p is an indirect precursor of s and s is an indirect precursor of p.* □

Note that an IG-cycle is not a cycle in the IG itself; instead, it specifies a special case under which cycles are introduced in a BN graph constructed from a general IG by our extended approach. Furthermore, note that, unlike causal cycles, for which it conceptually makes sense to assume IGs do not include them, this is not the case for IG-cycles; instead, condition d) of Proposition 9 poses a technical constraint to ensure acyclic graphs are constructed by our approach.

We now prove that under conditions a) to d) of Proposition 9, the initial graph constructed by steps $1' - 4a'$ of the extended approach is guaranteed to be a DAG.

Proposition 9. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Let $G^*_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}^*_{\mathcal{I}})$ be the possibly disconnected sub-graph of $G_{\mathcal{I}}$ with $\mathbf{A}^*_{\mathcal{I}} = \mathbf{A}_{\mathcal{I}} \setminus \mathbf{Exc}$. Let $\mathbf{C} = \{C = (\mathbf{P}^{\mathrm{c}}, \mathbf{A}^{\mathrm{c}}_{\mathcal{I}}) \mid \mathbf{P}^{\mathrm{c}} \subseteq \mathbf{P}, \mathbf{A}^{\mathrm{c}}_{\mathcal{I}} \subseteq \mathbf{A}^*_{\mathcal{I}}, C$ is a weakly connected component of $G^*_{\mathcal{I}}\}$ be the set of* IG components. *Assume that the following conditions are satisfied:*

a) For any IG component $C \in \mathbf{C}$, there does not exist an exc: $p \rightsquigarrow g$ in $\mathbf{Exc}$ with $p \in \mathbf{P}^{\mathrm{c}}$, $g \in \mathbf{A}^{\mathrm{c}}_{\mathcal{I}}$.

b) For every pair of IG components $C_1, C_2 \in \mathbf{C}$, there does not exist both an $exc_1\colon p_1 \rightsquigarrow g_1$ in $\mathbf{Exc}$ with $p_1 \in \mathbf{P}^{\mathrm{c}_1}$, $g_1 \in \mathbf{A}^{\mathrm{c}_2}_{\mathcal{I}}$ and an $exc_2\colon p_2 \rightsquigarrow g_2$ in $\mathbf{Exc}$ with $p_2 \in \mathbf{P}^{\mathrm{c}_2}$, $g_2 \in \mathbf{A}^{\mathrm{c}_1}_{\mathcal{I}}$.

c) Generalisation chains in $G_{\mathcal{I}}$ are non-repetitive and consistent (Section 3.3, p. 47).

d) IG $G_{\mathcal{I}}$ does not include an IG-cycle (see Definition 52).

Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the graph constructed from $G_{\mathcal{I}}$ according to steps $1' - 4a'$ of the extended approach. Then $G_{\mathcal{B}}$ is a DAG.

Proof. For the case $\mathbf{G}^{\mathrm{a}} = \mathbf{G}^{\mathrm{o}} = \emptyset$ and for which $\mathbf{Enabler}(g) = \emptyset$ for every $g \in \mathbf{G}^{\mathrm{c}}$, the proof reduces to the proof of Proposition 3. Thus assume that $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}} \neq \emptyset$ and that causal generalisations may include enabling conditions. Then for generalisations $g\colon \{p_1, \ldots, p_n\} \to p$ in $\mathbf{G}^{\mathrm{e}}$, arcs $\mathrm{Var}(p) \to \mathrm{Var}(p_i)$, $i = 1, \ldots, n$ are included in $\mathbf{A}_{\mathcal{B}}$ by step 2a$'$ and for generalisations $g\colon \{p_1, \ldots, p_n\} \to p$ in $\mathbf{G}^{\mathrm{c}}$, arcs $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ are included in $\mathbf{A}_{\mathcal{B}}$ by step 2b$'$ regardless of whether p_i is an enabler or not. Consider the case that for every $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, $\nexists g_1 \in \mathbf{G}^{\mathrm{e}}$ such that $[g, g_1]$ is a generalisation chain. Then furthermore, by step 2c$'$ arcs in $\mathbf{A}_{\mathcal{B}}$ are set similarly for $g \in \mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$ as for $g \in \mathbf{G}^{\mathrm{c}}$. Hence, similar to our assumption that no causal cycles exist in $G_{\mathcal{I}}$, we now have to assume that no IG-cycle exists in $G_{\mathcal{I}}$ (condition d) to avoid introducing cycles in $G_{\mathcal{B}}$.

Now, assume that there exists a $g\colon \{q_1, \ldots, q_n\} \to r$ in $\mathbf{G}^{\mathrm{a}} \cup \mathbf{G}^{\mathrm{o}}$, and that $\exists g_1, \ldots, g_m \in \mathbf{G}^{\mathrm{e}}$ such that $[g_1, \ldots, g_m]$ is a maximal generalisation chain of evidential generalisation arcs in $G_{\mathcal{I}}$ following g. By step 2d$'$ of the extended approach, $\mathbf{A}_{\mathcal{B}}$ includes arcs $Var(q_i) \to Var(Head(g_m))$ for $i = 1, \ldots, n$. Under specific conditions, cycles are introduced in $G_{\mathcal{B}}$ by this step, namely when an interceptor p of $[g, g_1, \ldots, g_m]$ exists (see Definition 51). Assume that $p = Head(g_j)$ or $p = -Head(g_j)$ for a $j \in \{1, \ldots, m\}$ and that p is a direct or indirect precursor of a proposition $q \in \mathbf{Tails}(g)$. Then a sequence of arcs directed from $Var(Head(g_m))$ to $Var(Head(g_j))$ is introduced in $\mathbf{A}_{\mathcal{B}}$ by step 2a$'$ corresponding to evidential generalisations $g_{j+1}, \ldots, g_m$. Furthermore, a sequence of arcs directed from $Var(Head(g_j))$ to $Var(q)$ is introduced in $\mathbf{A}_{\mathcal{B}}$ by step 2$'$, as p is a direct or indirect precursor of q; therefore, a cycle from node $Var(q)$ to itself is introduced. However, these cycles are broken in step 4a$'$ by removing arc $Var(q) \to Var(Head(g_m))$ from $\mathbf{A}_{\mathcal{B}}$ introduced in step 2d$'$ (see Section 5.3.3, p. 104).

In step 3$'$ of the extended approach, additional arcs are included in $\mathbf{A}_{\mathcal{B}}$ for every $exc \in \mathbf{Exc}$ in the IG. Under conditions a) and b), the proof is completed in a similar manner as the proof of Proposition 3 by considering the weakly connected components of BN graph $G_{\mathcal{B}}$. ■

5.5 Discussion and concluding remarks

In this chapter, we have introduced a BN graph construction approach that exploits a notion of causality as expressed in the generalisations and conflicts of an IG. Moreover, by considering the inferences that can be read from such an IG given the evidence, some qualitative constraints on the (conditional) probabilities of the BN under construction are derived. We have formally proven that, as intended, BN graphs constructed by our approach capture all inference chains that can be read from the IG given the evidence in the form of induced active chains (Proposition 4). We have identified conditions on the IG under which the fully automatically constructed initial graph is guaranteed to be a DAG (Proposition 3), simplifying the (manual) verification step. We have identified bounds on the size of the CPTs and the complexity of probabilistic inference in BNs constructed by our approach (Propositions 5 and 6). Lastly, we have investigated mapping properties of our approach (Proposition 7).

Our IG-formalism is intended to facilitate the construction of BNs by serving as an intermediary formalism between analyses performed using informal sense-making tools and BNs. As mentioned earlier, we expect direct IG construction to be more straightforward than direct BN construction for domain experts unfamiliar with the BN-formalism, a claim that should be empirically evaluated in future work. Our approach may be evaluated in future work by assessing the quality of BNs constructed from IGs. Since we are considering BN construction in data-poor domains, we assume that there is insufficient data to learn a reliable BN from and that such a BN is therefore not available for comparison. A quality assessment should therefore mainly be based on compliance with best practice guidelines for BN construction [Neil et al., 2000][1].

In future work, more probabilistic constraints may be derived on BNs, for instance by interpreting defeasible inferences that can be read from an IG given the evidence as qualitative influences [Wellman, 1990]. Specifically, variable P is said to have a positive qualitative influence on variable Q if $\Pr(q \mid p) \geq \Pr(q \mid \neg p)$ and a negative qualitative influence if $\Pr(q \mid p) \leq \Pr(q \mid \neg p)$. Interpreting all defeasible inferences between propositions $p_1, \ldots, p_n$ and q that can be read from an IG as positive qualitative influences and all defeasible inferences between propositions $p_1, \ldots, p_n$ and $\neg q$ as negative qualitative influences, a fully specified qualitative probabilistic network (QPN) [Wellman, 1990] may be constructed by our approach which can be used for qualitative probabilistic inference given multiple observations [Renooij, 2001]. Quantification of QPNs can then be performed incrementally by specifying probability intervals for CPTs for nodes in the graph as an intermediary step, resulting in so-called semi-qualitative probabilistic networks [Renooij and van der Gaag, 2002] that can also be used for probabilistic inference. Alternatively, a credal network [Cozman, 2000] can be constructed [de Campos and Cozman, 2005].

In Sections 5.3 and 5.4 we discussed extending our approach to IGs including

[1] See also Master's research [de Leeuw, 2020].

abstractions and generalisations of type ‘other’, as well as generalisations that include enablers. We have formally proven for such general IGs that initial BN graphs constructed by our extended approach are guaranteed to be acyclic (Proposition 9) and that inference chains in such IGs which only describe deductive inferences given the evidence are captured as active chains in the BN graph given the evidence (Proposition 8). As discussed, the latter property is not guaranteed to hold for inference chains which also describe abductive inferences. Therefore, our extended approach does not fully preserve the desirable properties as proven for our original approach regarding the captured reasoning patterns in BNs constructed from IGs. In its current form, therefore, BNs constructed using our extended approach do not fully benefit from the knowledge available in an IG. However, our description of the problem for the general case and our solution for a special case can support others in finding a more general solution in future work. Regardless of how much knowledge is being exploited, our IG-formalism, together with our BN construction approach, allow us to construct at least an initial BN graph from a domain expert’s initial analysis; it thereby simplifies the BN elicitation process. We note that BN construction is an iterative process in which both the domain expert and BN modeller should stay involved; this also holds when applying our approach, as even the provided IG may be incomplete or may be subject to change over time. In Chapter 7 we propose an approach that can aid experts in this iterative process. The approach allows experts to reason and argue about a BN under construction, where possible conflicts are resolved as much as possible using computational argumentation. In particular, this approach can be applied to both BNs constructed from IGs via the approach presented in the current chapter and BNs otherwise constructed, and serves to facilitate both the qualitative graph-construction step and the quantitative probability elicitation step involved in BN construction.

Chapter 6

Case study: the Sacco and Vanzetti case

In this chapter, we apply the approaches from Chapters 4 and 5 to parts of an actual legal case, namely the well-known Sacco and Vanzetti case. The case concerns Sacco and Vanzetti, who were convicted for shooting and killing payroll guard Berardelli during a robbery in South Braintree, Massachusetts on 15 April 1920; a detailed description of the case is provided by Kadane and Schum [1996]. Kadane and Schum performed a probabilistic analysis of this case by first constructing Wigmore charts of aspects of the case and then manually constructed corresponding BNs by assessing the modelled independence relation and assessing the necessary (conditional) probabilities. In this chapter, we illustrate and perform a preliminary validation of our BN graph construction approach from Chapter 5 by formalising one of Kadane and Schum's Wigmore charts (chart 25, Kadane and Schum [1996, pp. 330–331]) as an IG, where we compare the obtained BN graph to their BN graph. In addition, we illustrate our approach of Chapter 4 by constructing and evaluating arguments based on the same IG.

The chapter is structured as follows. In Section 6.1 Kadane and Schum's Wigmore chart concerning Sacco's consciousness of guilt is presented, where a possible formalisation of this Wigmore chart as an IG is provided in Section 6.2. In Section 6.3 we then apply the approach from Chapter 5 to this IG and compare the obtained BN graph to that of Kadane and Schum. In Section 6.4 we illustrate our approach of Chapter 4 by constructing and evaluating arguments based on the same IG. In Section 6.5 we then conclude the case study.

6.1 Wigmore chart concerning Sacco's consciousness of guilt

According to Kadane and Schum, the ultimate claim under consideration in the Sacco and Vanzetti case is Π_3, which states that *'It was Sacco who, with the assistance of Vanzetti, intentionally fired shots that took the life of Berardelli during the robbery and shooting that took place in South Brain tree.'* In the prosecution's case against Sacco and Vanzetti, their alleged consciousness of guilt in the South Braintree crime played an important role. However, as noted by Kadane and Schum the inferences made based on the available evidence for this part of the case are not particularly strong; a significant part of Kadane and Schum's analysis is, therefore, devoted to this part of the case. During their arrest, Sacco and Vanzetti were armed. According to the two arresting officers, Connolly and Spear, Sacco and Vanzetti made suspicious hand movements, from which the prosecution concluded that they intended to draw their concealed weapons in order to escape their arrest. This suggests that they were conscious of having committed a criminal act. In the remainder of this chapter, we only consider this part of the case.

In Figure 6.1, a modernised Wigmore chart concerning Sacco's consciousness of guilt is depicted, adapted from Kadane and Schum [1996, pp. 330–331]. Compared to Kadane and Schum's original chart, we consider a subset of the mapped claims; in particular, additional claims regarding Sacco's political beliefs (claims $471 - 480$ in the original chart) and claims that were provided post-trial by historians are not considered. On the right-hand side the corresponding key list is depicted. As noted by Kadane and Schum [1996, p. 88], vertical arcs between nodes in their version of Wigmore's charts indicate inferences between corresponding claims, where the generalisations used in performing these inferences are not explicitly recorded in the chart. Instead, in their analysis of the case some of the used generalisations are indicated in the text (see e.g. Kadane and Schum [1996, pp. 97–98]). For instance, generalisations used in the inferences from the provided testimonies are of the general form *'If a person testifying under oath tells us that event E occurred, then this event (probably, usually, often, etc,) did occur.'* [Kadane and Schum, 1996, p. 88]. As noted by Kadane and Schum [1996, pp. 74–76], in constructing their charts abduction is in some instances performed to generate interim hypotheses between the evidence and the ultimate claim Π_3. However, Kadane and Schum do not explicitly indicate which inferences in their charts are abductive and which are deductive.

In their version of Wigmore charts, Kadane and Schum make a distinction between directly relevant and ancillary claims[1], where the role of an ancillary claim is to show why a generalisation holds or fails in a particular situation [Kadane and Schum, 1996, p. 53]. Directly relevant and ancillary claims provided by the defence

[1] Kadane and Schum [1996] use the terms 'directly relevant evidence' and 'ancillary evidence'. To avoid confusion with the manner in which the term 'evidence' is used in this thesis (i.e. that what has been established with certainty), we instead use the term 'claim'.

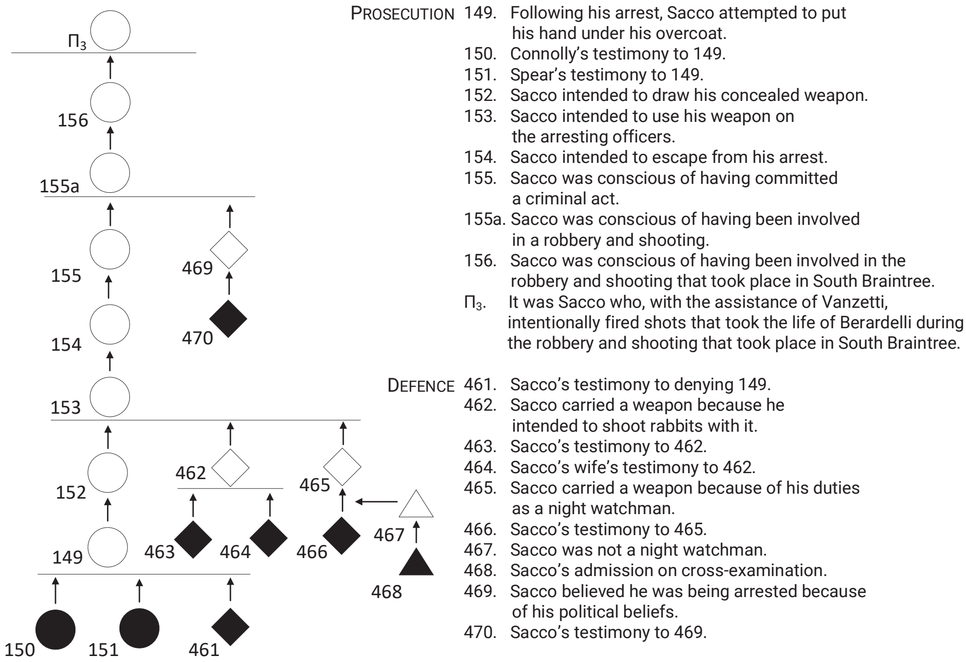


Figure 6.1: Wigmore chart concerning Sacco's consciousness of guilt, along with the corresponding key list, adapted from Kadane and Schum [1996, pp. 330–331].

are represented as diamonds and triangles, respectively; for the prosecution, these are represented as circles and squares, respectively. Note that in the Wigmore chart of Figure 6.1, all claims provided by the prosecution are directly relevant. All nodes in Kadane and Schum's charts indicate either directly relevant or ancillary claims and nodes corresponding to the evidence are shaded. An arc directed from a node corresponding to an ancillary claim to an arc between two or more claims indicates that this ancillary claim either supports or weakens the applicability of the generalisation in the inference at hand [Kadane and Schum, 1996, p. 87]. Finally, horizontal lines in the Wigmore chart indicate that information needs to be combined to draw a conclusion.

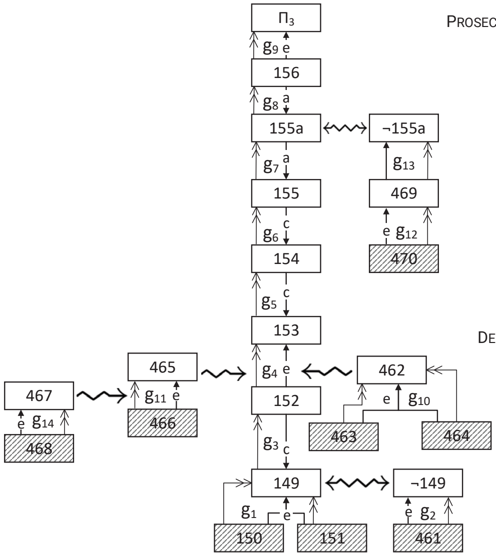
6.2 Formalising the Wigmore chart as an IG

We now provide a possible formalisation of Kadane and Schum's Wigmore chart of Figure 6.1 as an IG. In Figure 6.2, an IG is depicted for a possible interpretation of this Wigmore chart. For every claim p in the Wigmore chart, a proposition node p is included in $\mathbf{P}$. In establishing which generalisations could have been used in performing the inferences indicated in the chart, we take the following general approach. In case generalisations are explicitly indicated by Kadane and Schum in the text, then these generalisations are used; otherwise, we first establish whether or not there is a causal relation between the nodes in the chart, and if so, what the direction of causality is. To aid in this process, we determine whether sequences of described events can be interpreted as instances of so-called story schemes [Bex and Verheij, 2012], which capture stereotypical patterns of causal reasoning. In case p usually/normally/typically causes q, then we establish whether p can be considered the usual cause for q. If this is the case, then evidential generalisation $q \rightarrow p$ is included in $\mathbf{G}^{\mathrm{e}}$ to explicitly capture in the IG that p is considered the usual cause of q; otherwise, causal generalisation $p \rightarrow q$ is included in $\mathbf{G}^{\mathrm{c}}$ (see also Section 2.1.3). If a relation cannot be interpreted in a causal manner, then we establish whether it can be considered an abstraction and whether this generalisation is strict or defeasible. If a relation can neither be classified as causal nor evidential nor an abstraction, then a generalisation of type 'other' is included in the IG, where we again establish whether the generalisation under consideration is strict or defeasible.

As noted by Kadane and Schum [1996, p. 88], the generalisations used in the inferences from the provided testimonies are evidential (see Section 6.1). As propositions $150, 151, 463, 464$, $466, 468$ and 470 denote testimonies, the IG includes generalisation arcs $g_1\colon \{150, 151\} \rightarrow 149$, $g_{10}\colon \{463, 464\} \rightarrow 462$, $g_{11}\colon 466 \rightarrow 465$, $g_{12}\colon 470 \rightarrow 469$ and $g_{14}\colon 468 \rightarrow 467$ in $\mathbf{G}^{\mathrm{e}}$. Here, testimonies 150, 151 and 463, 464 are combined in the antecedents of generalisations g_1 and g_{10}, respectively, as these sets of propositions concern testimonies to the same claim.

The manner in which claims and links conflict is not precisely specified in Kadane and Schum's Wigmore charts, as also observed by Bex and colleagues [2003] in formalising such charts as Pollock-style arguments [Pollock, 1995]. As we wish to formalise the Wigmore chart of Figure 6.1 as an IG, we consider how possible conflicts between claims proposed by the prosecution and defence can be interpreted in terms of the conflict relations defined in Section 3.3. As 461 concerns Sacco's testimony to denying 149, proposition $\neg 149$ is included in $\mathbf{P}$, generalisation arc $g_2\colon 461 \rightarrow \neg 149$ is included in $\mathbf{G}^{\mathrm{e}}$, and negation arc $n_1\colon 149 \leftrightsquigarrow \neg 149$ is included in $\mathbf{N}$.

Kadane and Schum do not indicate which (types of) generalisations were used in performing the inferences between propositions 149 and Π_3. We note that the inferences between 149 and 155 fit a so-called episode scheme for intentional actions [Bex, 2011, p. 64], a story scheme in which someone's psychological state causes them to form certain goals, which in turn lead to actions that have consequences. In this case, Sacco intended to escape from his arrest (154; goal) as he was conscious of having committed a criminal act (155; psychological state); therefore, we consider 155 to typically cause 154. Sacco's intention to use his weapon (153) can then be considered a sub-goal of 154 and his intention to draw his concealed weapon (152) a further sub-goal of 153. Sacco's intention to draw his weapon (152) caused Sacco to attempt to put his hand under his overcoat (149; action); more specifically, we consider 152 to typically cause 149. Finally, we consider 153 to be the usual cause for 152, as the usual cause for wanting to draw a weapon is wanting to use this weapon;



PROSECUTION
149. Following his arrest, Sacco attempted to put his hand under his overcoat.
150. Connolly's testimony to 149.
151. Spear's testimony to 149.
152. Sacco intended to draw his concealed weapon.
153. Sacco intended to use his weapon on the arresting officers.
154. Sacco intended to escape from his arrest.
155. Sacco was conscious of having committed a criminal act.
155a. Sacco was conscious of having been involved in a robbery and shooting.
156. Sacco was conscious of having been involved in the robbery and shooting that took place in South Braintree.
Π_3. It was Sacco who, with the assistance of Vanzetti, intentionally fired shots that took the life of Berardelli during the robbery and shooting that took place in South Braintree.

DEFENCE
461. Sacco's testimony to denying 149.
462. Sacco carried a weapon because he intended to shoot rabbits with it.
463. Sacco's testimony to 462.
464. Sacco's wife's testimony to 462.
465. Sacco carried a weapon because of his duties as a night watchman.
466. Sacco's testimony to 465.
467. Sacco was not a night watchman.
468. Sacco's admission on cross-examination.
469. Sacco believed he was being arrested because of his political beliefs.
470. Sacco's testimony to 469.

Figure 6.2: An IG corresponding to a possible interpretation of the Wigmore chart of Figure 6.1, along with the corresponding key list.

we therefore include $g_4\colon 152 \to 153$ in $\mathbf{G}^{\mathrm{e}}$. Generalisation arcs $g_3\colon 152 \to 149$, $g_5\colon 154 \to 153$ and $g_6\colon 155 \to 154$ are then included in $\mathbf{G}^{\mathrm{c}}$, as we do not consider their antecedents to express the usual cause for their consequents. Alternatively, it may be argued that some (or all) of these relations are evidential. Below, we show that similar inferences can be performed with the constructed IG and that the same BN graph and AF are constructed from the IG regardless of whether these relations are interpreted as causal or evidential.

The relations between propositions 155, 155a and 156 cannot be interpreted in a causal manner. Instead, proposition 155 can be considered an abstraction of 155a: being involved in a robbery and shooting can generally be considered committing a criminal act. The involved generalisation is defeasible: involvement in a robbery and shooting does not imply that this involvement is of a criminal nature, as it may also imply that the person under consideration is the victim. Proposition 155a can be considered a strict abstraction of 156, as at a higher level of abstraction being conscious of having been involved in the specific robbery and shooting that took place in South Braintree can be considered being conscious of having been involved in a robbery and shooting. Π_3 can be considered a cause of 156; more specifically, committing a robbery and shooting typically causes a person (in this case Sacco) to be conscious of having been involved in this act. Moreover, Π_3 can be considered the *usual* cause for 156; therefore, generalisation $g_7\colon 155a \to 155$ is included in $\mathbf{G}^{\mathrm{a}}_{\mathrm{d}}$, $g_8\colon 156 \to 155a$ in $\mathbf{G}^{\mathrm{a}}_{\mathrm{s}}$, and $g_9\colon 156 \to \Pi_3$ in $\mathbf{G}^{\mathrm{e}}$.

From 469 (Sacco believed he was being arrested because of his political beliefs), we conclude that Sacco was not conscious of having been involved in a robbery and shooting ($\neg 155a$). We consider the relation between 469 and $\neg 155a$ to be defeasible and neither causal nor evidential nor an abstraction, and therefore include $g_{13}\colon 469 \rightarrow \neg 155a$ in $\mathbf{G}^{\mathrm{o}}_{\mathrm{d}}$. Arc $n_2\colon 155a \leftrightsquigarrow \neg 155a$ is then included in $\mathbf{N}$.

In Kadane and Schum's Wigmore chart, it is indicated that 467 is an ancillary claim that weakens (or supports) the applicability of generalisation $g_{11}\colon 466 \rightarrow 465$ in the inference from 466 to 465. In this particular instance, 467 can be interpreted as an exception to generalisation g_{11}, as the claim that Sacco was not a night watchman indicates that Sacco's veracity in providing his testimony about the reason for carrying a weapon is questionable. Therefore, we include $exc_1\colon 467 \rightsquigarrow g_{11}$ in $\mathbf{Exc}$.

Finally, the conflicts between the defence's claims 462 and 465 and the prosecution's claims 152 and 153 are considered. Multiple interpretations are possible. One possible interpretation is that 462 and 465 indicate exceptions to generalisation $g_4\colon 152 \rightarrow 153$ in $\mathbf{G}^{\mathrm{e}}$. Specifically, 462 and 465 can be considered competing alternative explanations for 152: as Sacco carried his weapon for an innocent reason (462 or 465), this caused him to draw his weapon (152) with the intention of surrendering it. In Figure 6.2, these exceptions are indicated by curved hyperarcs $exc_2\colon 462 \rightsquigarrow g_4$ and $exc_3\colon 465 \rightsquigarrow g_4$ in $\mathbf{Exc}$. An alternative interpretation is that 462 and 465 indicate support for the negation of proposition 153: as Sacco carried his weapon for an innocent reason (either 462 or 465), he intended to surrender his weapon and, therefore, did not intend to use it ($\neg 153$). Accordingly, generalisations $g_{15}\colon 462 \rightarrow \neg 153$ and $g_{16}\colon 465 \rightarrow \neg 153$ can be included, as depicted in the adjusted IG of Figure 6.3a. As these generalisations are defeasible and neither causal nor evidential nor an abstraction, g_{15} and g_{16} are included in $\mathbf{G}^{\mathrm{o}}_{\mathrm{d}}$. Negation arc $n_3\colon 153 \leftrightsquigarrow \neg 153$ is then included in $\mathbf{N}$.

In the Wigmore chart of Figure 6.1, the evidence consists of the testimonies; hence, $\mathbf{E_p} = \{150, 151, 461, 463, 464, 466, 468, 470\}$. Given $\mathbf{E_p}$, the inferences that can be read from the IG of Figure 6.2 coincide with the inferences indicated in the Wigmore chart. Specifically, given $\mathbf{E_p}$, propositions 149, $\neg 149$, 462, 465, 467 and 469 are defeasibly deductively inferred from 150 and 151, 461, 463 and 464, 466, 468 and 470 using generalisations g_1, g_2, g_{10}, g_{11}, g_{14} and g_{12}, respectively. Proposition 152 is then abductively inferred from 149 using g_3, as 149 is deductively inferred. Propositions 153, 154, 155, 155a, 156 and Π_3 are then iteratively defeasibly inferred using generalisations g_4, g_5, g_6, g_7, g_8 and g_9, respectively. Finally, from 469, $\neg$155a is defeasibly deductively inferred using g_{13}, as 469 is deductively inferred.

As mentioned earlier, instead of including causal generalisations $g_3\colon 152 \rightarrow 149$, $g_5\colon 154 \rightarrow 153$ and $g_6\colon 155 \rightarrow 154$, an alternative interpretation is that the antecedents of these generalisations express the usual cause for their consequents; accordingly, evidential generalisations $g'_3\colon 149 \rightarrow 152$, $g'_5\colon 153 \rightarrow 154$ and $g'_6\colon 154 \rightarrow 155$ may instead be included. Similar inferences can then be performed with the constructed IG given $\mathbf{E_p}$; specifically, propositions 152, 153, 154 and 155 are then iteratively defeasibly deductively inferred given $\mathbf{E_p}$ using g'_3, g_4, g'_5 and g'_6 instead of that some of these inferences are abductive.

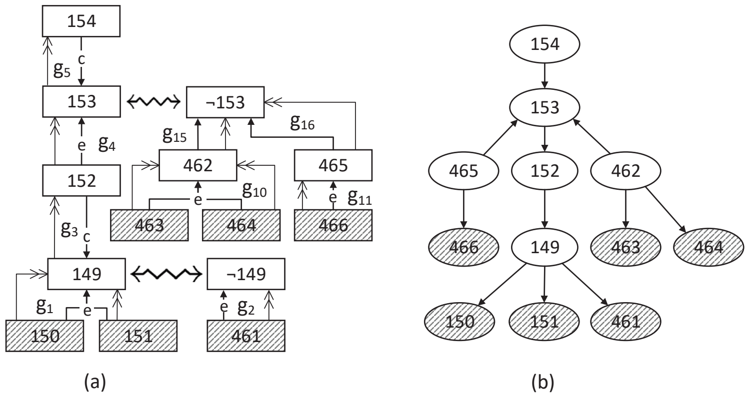


Figure 6.3: Adjustment to part of the IG of Figure 6.2, where 462 and 465 indicate support for ¬153 (a); corresponding BN fragment constructed by our approach (b).

6.3 Constructing a BN graph from the IG

We now apply our BN graph construction approach from Chapter 5 to the IG of Figure 6.2 and compare the obtained graph to that of Kadane and Schum.

6.3.1 Applying our BN graph construction approach

By applying our BN graph construction approach from Chapter 5 to the IG of Figure 6.2, the BN graph depicted in Figure 6.4b is obtained. By step 1 of our approach, every proposition and its negation are captured as two mutually exclusive values of the same node. Arcs in the BN graph corresponding to generalisation arcs in $\mathbf{G}^{\mathrm{e}} \cup \mathbf{G}^{\mathrm{c}}$ are then directed according to step 2. Furthermore, for abstractions $g_7\colon 155a \rightarrow 155$ and $g_8\colon 156 \rightarrow 155a$ in $\mathbf{G}^{\mathrm{a}}$, arcs $155a \rightarrow 155$ and $156 \rightarrow 155a$ are included in $\mathbf{A}_{\mathcal{B}}$, and for $g_{13}\colon 469 \rightarrow \neg 155a$ in $\mathbf{G}^{\mathrm{o}}_{\mathrm{d}}$, arc $469 \rightarrow 155a$ is included in $\mathbf{A}_{\mathcal{B}}$, as discussed in Section 5.3.

Additional arcs are then added to $\mathbf{A}_{\mathcal{B}}$ for every exception arc in **Exc** by step 3 of our approach. Specifically, $exc_1\colon 467 \rightsquigarrow g_{11}$, $exc_2\colon 462 \rightsquigarrow g_4$ and $exc_3\colon 465 \rightsquigarrow g_4$ are specified in the IG, where $g_{11}, g_4 \in \mathbf{G}^{\mathrm{e}}$; therefore, additional arcs $467 \rightarrow 466$, $465 \rightarrow 152$ and $462 \rightarrow 152$ are included in $\mathbf{A}_{\mathcal{B}}$ by step 3a.

Note that in case causal generalisations g_3, g_5 and/or g_6 are replaced by evidential generalisations g'_3, g'_5 and/or g'_6, the same BN graph is obtained by our approach. More specifically, by step 2b of our approach, arc $\mathrm{Var}(p) \rightarrow \mathrm{Var}(q)$ is included for every causal generalisation $g\colon p \rightarrow q$, where the same arc is included in $\mathbf{A}_{\mathcal{B}}$ by step 2a of our approach for every evidential generalisation $g\colon q \rightarrow p$.

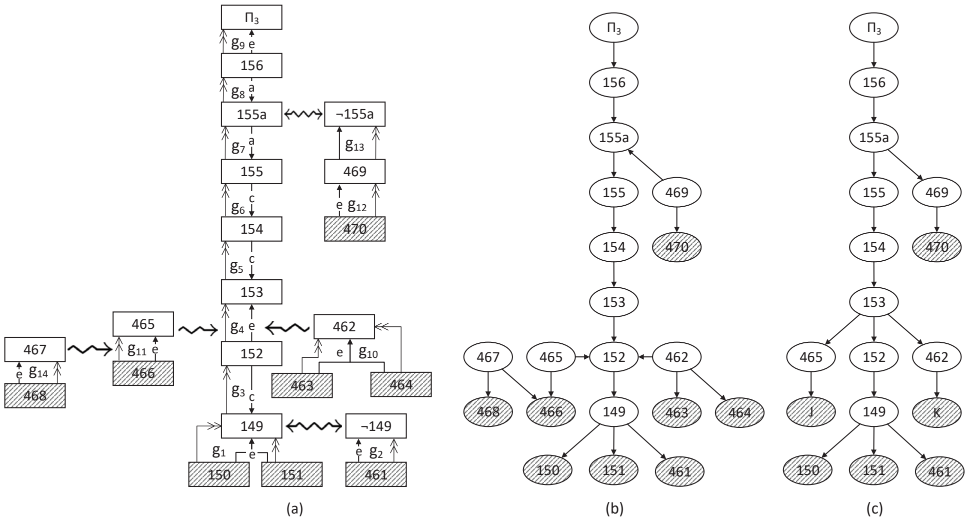


Figure 6.4: The IG of Figure 6.2 (a); the corresponding BN graph constructed according to our approach (b); adaptation of the BN graph constructed by Kadane and Schum [1996, p. 232] (c).

6.3.2 Comparison to Kadane and Schum's BN graph

The structure of the obtained graph is largely identical to that of the BN graph that Kadane and Schum manually constructed for this part of the case, depicted in Figure 6.4c; the differences and similarities between the two BN graphs are now discussed. First, note that Kadane and Schum aggregate nodes 463 and 464 into a single Boolean node K. Similarly, nodes 466, 467 and 468 are aggregated into Boolean node J; possible intercausal effects between 467 and 465 can, therefore, not be explicitly captured in their BN. While aggregation as performed by Kadane and Schum reduces the number of conditional probabilities to be assessed, we prefer to explicitly capture all elements of the IG in the corresponding BN graph to prevent loss of information. The only case in which IG elements are aggregated by our approach is when two propositions p and $\neg p$ appear in the graph, which are then captured as two values of the same node. We note that, by step 6a of our approach, constraints on the CPTs of the BN under construction are automatically obtained, which partially simplifies subsequent probability assessment. Specifically, a head-to-head node is formed in 466, which allows for directly capturing possible interactions between 465 and 467. By step 6a, constraint $\mathbf{X}^{-}(\{465, 467\}, 466 = true)$ is derived on the CPT for node 466. For instance, entries for this CPT can be chosen as follows: $\Pr(466 \mid 465, 467) = 0$, $\Pr(466 \mid \neg 465, \neg 467) = 0.4, \Pr(466 \mid 465, \neg 467) = 0.9, \Pr(466 \mid \neg 465, 467) = 0.2$, as in this case $0 \cdot 0.4 \leq 0.9 \cdot 0.2$. Note that the conditioned event of conditional probability $\Pr(466 \mid 465, 467)$ cannot actually occur in practice, as Sacco cannot both be and not be a night watchman at the same time. Hence, the exact number to which this conditional probability is set is irrelevant:

we choose to set $\Pr(466 \mid 465, 467) = 0$. In case Sacco was indeed a night watchman (467 is not true) but Sacco did not carry a weapon because of this reason (465 is not true), then we find it plausible that Sacco was lying under oath in providing his testimony ($\Pr(466 \mid \neg 465, \neg 467) = 0.4$); more specifically, as he was indeed a night watchman, he can use this as an excuse to claim that he carried his weapon because of this reason. In case Sacco was a night watchman (467 is not true) and Sacco actually carried a weapon because of his duties as a night watchman (465 is true), then we consider the event that Sacco testifies to this claim (466) to be very likely ($\Pr(466 \mid 465, \neg 467) = 0.9$). Finally, in case Sacco was not a night watchman (467 is true) and Sacco did not carry his weapon because of his duties as a night watchman (465 is not true), then we set $\Pr(466 \mid \neg 465, 467) = 0.2$ to again take into account the probability that Sacco may be lying under oath. We believe this probability to be lower than $\Pr(466 \mid \neg 465, \neg 467)$, as we consider it less likely for Sacco to come up with the explanation that he carried his weapon because of his duties as a night watchman if he was in fact not a night watchman.

In the BN graph of Figure 6.4b, a head-to-head node is also formed in node 152, which allows for directly capturing possible interactions between 462, 465 and 153. These interactions cannot be captured in the BN graph of Figure 6.4c, as in this graph arcs $153 \to 465$ and $153 \to 462$ are included instead of arcs $465 \to 152$ and $462 \to 152$. By step 6a, constraints $\mathbf{X}^-(\{462, 153\}, 152 = \mathit{true})$, $\mathbf{X}^-(\{465, 153\}, 152 = \mathit{true})$ and $\mathbf{X}^-(\{465, 462\}, 152 = \mathit{true})$ are derived on the CPT for node 152 in our BN graph. Note that in the BN graph of Kadane and Schum, variables 462 and 465 are conditionally independent from 152 given 153; therefore, in contrast with our BN under construction, for Kadane and Schum's BN it needs to hold that $\Pr(152 \mid 462, 465, 153) = \Pr(152 \mid 153)$. As the entries for the CPT for node 152 in our BN cannot be compared to that of Kadane and Schum, the assessment of the involved conditional probabilities is not further discussed.

In the IG of Figure 6.4a, given $\mathbf{E_p}$, proposition 149 is deductively inferred from 150 and 151 using g_1 and proposition 462 is deductively inferred from 463 and 464 using g_{10}. By step 5b of our approach, the following probabilistic constraints on the probabilities of the BN are derived: $\Pr(149 \mid 150, 151) \not< \Pr(149 \mid 150, \neg 151)$; $\Pr(149 \mid 150, 151) \not< \Pr(149 \mid \neg 150, 151)$; $\Pr(462 \mid 463, 464) \not< \Pr(462 \mid 463, \neg 464)$ and $\Pr(462 \mid 463, 464) \not< \Pr(462 \mid \neg 463, 464)$. As mentioned earlier, these probabilistic constraints can be used in an elicitation procedure for further quantifying the BN under construction [Druzdzel and van der Gaag, 1995].

We note that for every active chain that exists between two nodes in the BN graph of Figure 6.4b given the evidence, there exists an active chain between these nodes in the BN graph of Figure 6.4c given the evidence and vice versa; therefore, given $\mathbf{E_V}$, similar probabilistic inferences can be performed in both BN graphs, besides the aforementioned differences. More specifically, as 152 has an instantiated descendant in the BN graph of Figure 6.4b, chains between 465 and 462 are active. Furthermore, as no head-to-head node is formed among nodes 155, 155a and 156 in the BN graph of Figure 6.4b, chains between 155 and 156 are active. We note that the IG under consideration is of the special class of IGs including abstractions with which abductive inference is performed given $\mathbf{E_p}$ (i.e. abstractions $g_7\colon 155a \rightarrow 155$ and $g_8\colon 156 \rightarrow 155a$) and for which a BN graph is constructed that captures all inference chains that can be read from the IG given the evidence as induced active chains (as discussed in Section 5.4.1). The only other difference with the BN graph of Figure 6.4c is that the arc between nodes 155a and 469 is reversed; a head-to-head node is therefore formed in node 155a in the BN graph of Figure 6.4b, but as $155a$ has an instantiated descendant the chain between nodes 469 and 156 is active.

From the alternative IG fragment of Figure 6.3a, the BN fragment of Figure 6.3b is constructed. Compared to the BN graph of Figure 6.4b, a head-to-head node is formed in 153 instead of in 152, which allows for capturing possible interactions between 462, 465 and 154 instead of between 462, 465, and 153. This illustrates that depending on the modelling choices made in constructing an IG, different BN graphs may be constructed.

6.4 Constructing an AF from the IG

To further illustrate how AFs can be constructed based on IGs, in Section 6.4.1 we apply our approach from Chapter 4 to the IG of Figure 6.4a. In Section 6.4.2 we then evaluate the constructed arguments.

6.4.1 Applying our AF construction approach

Consider Figure 6.5, in which arguments constructed on the basis of the IG of Figure 6.4a using our approach from Chapter 4 are indicated. According to step 1 of Definition 37, $A_1\colon 150$ and $A_2\colon 151$ are premise arguments. Based on A_1 and A_2, defeasible deductive argument $A_3\colon A_1, A_2 \twoheadrightarrow_{g_1} 149$ is constructed by step 2a of Definition 37, as 149 is defeasibly deductively inferred from 150 and 151 using $g_1 \in \mathbf{G}^{\mathrm{e}}$. Argument $A_4\colon A_3 \twoheadrightarrow_{g_3} 152$ then is a defeasible mixed argument by step 3 of Definition 37, as 152 is abductively inferred from 149 using g_3. Arguments $A_5\colon A_4 \twoheadrightarrow_{g_4} 153$; $A_6\colon A_5 \twoheadrightarrow_{g_5} 154$; $A_7\colon A_6 \twoheadrightarrow_{g_6} 155$; $A_8\colon A_7 \twoheadrightarrow_{g_7} 155\mathrm{a}$; $A_9\colon A_8 \twoheadrightarrow_{g_8} 156$ and $A_{10}\colon A_9 \twoheadrightarrow_{g_9} \Pi_3$ similarly are defeasible mixed arguments.

Let $B_1\colon 461$ and let $B_2\colon B_1 \twoheadrightarrow_{g_2} \neg 149$. Then B_2 rebuts A_3 (on A_3) and A_3 rebuts B_2 (on B_2), as $\textsc{Conc}(A_3) = 149$, $\textsc{Conc}(B_2) = \neg 149$ (and hence $n\colon 149 \leftrightsquigarrow \neg 149$ in $\mathbf{N}$), where $\textsc{TopInf}(A_3) = 150, 151 \twoheadrightarrow_{g_1} 149$ and $\textsc{TopInf}(B_2) = 461 \twoheadrightarrow_{g_2} \neg 149$ are defeasible. As rebuttal is defined on sub-arguments, B_2 also asymmetrically rebuts A_i for $i \geq 4$. Similarly, let $B_3\colon 470$; $B_4\colon B_3 \twoheadrightarrow_{g_{12}} 469$; $B_5\colon B_4 \twoheadrightarrow_{g_{13}} \neg 155\mathrm{a}$. Then B_5 rebuts A_8 (on A_8) and A_8 rebuts B_5 (on B_5). Again, as rebuttal is defined on sub-arguments, B_5 also asymmetrically rebuts A_9 and A_{10}.

Let $C_1\colon 466$; $C_2\colon C_1 \twoheadrightarrow_{g_{11}} 465$. Then C_2 undercuts A_5 (on A_5), as $exc_3\colon 465 \rightsquigarrow g_4$ in $\mathbf{Exc}$ and $\textsc{TopGen}(A_5) = g_4 \in \mathbf{G}^{\mathrm{e}}$. As undercutting attack is defined on sub-arguments, C_2 also attacks A_i for $i \geq 6$. Similarly, let $C_3\colon 463$; $C_4\colon 464$; $C_5\colon C_3, C_4 \twoheadrightarrow_{g_{10}} 462$. Then C_5 undercuts A_5 (on A_5), as $exc_2\colon 462 \rightsquigarrow g_4$ in $\mathbf{Exc}$ and $\textsc{TopGen}(A_5) = g_4$. C_5 then also attacks A_i for $i \geq 6$. Lastly, let $D_1\colon 468$ and

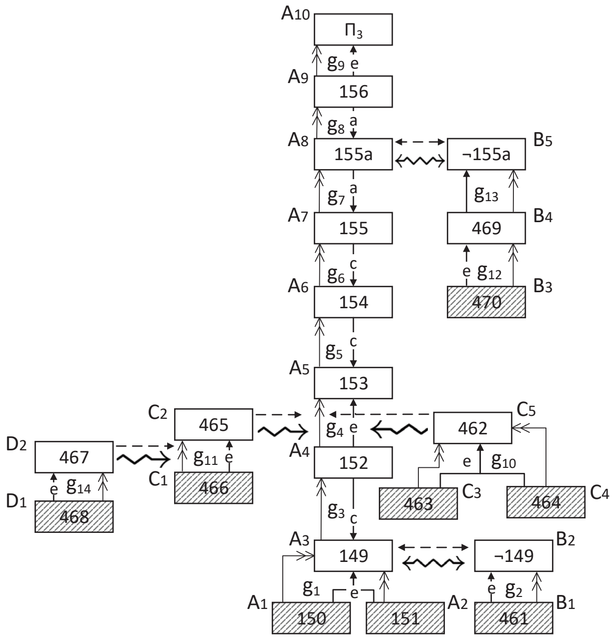


Figure 6.5: Annotation of the IG of Figure 6.4a, where arguments and direct attacks (- - →) on the basis of the IG and $\mathbf{E_p}$ are also indicated.

$D_2\colon D_1 \twoheadrightarrow_{g_{14}} 467$, then D_2 undercuts C_2 (on C_2) as $exc_1\colon 467 \rightsquigarrow g_{11}$ in **Exc** and $\textsc{TopGen}(C_2) = g_{11}$.

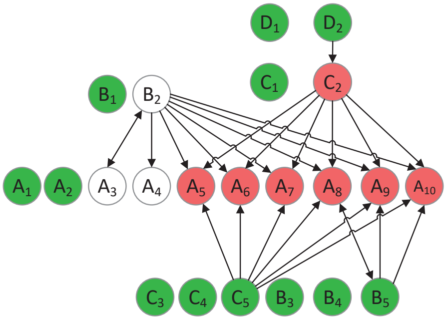
Finally, note that in case causal generalisations g_3, g_5 and/or g_6 are replaced by evidential generalisations g'_3, g'_5 and/or g'_6, similar inferences can be performed with the IG given $\mathbf{E_p}$ (see Section 6.2); therefore, arguments A_4–A_7 are again constructed according to this alternative interpretation and the same AF is obtained.

6.4.2 Argument evaluation

Next, the constructed arguments are evaluated. The constructed AF $(\mathcal{A}, \mathcal{D})$ is visualised in Figure 6.6a. The complete extensions of $(\mathcal{A}, \mathcal{D})$ are:

$\mathcal{S}_1 = \{A_1, A_2, B_1, B_3, B_4, B_5, C_1, C_3, C_4, C_5, D_1, D_2\}$;
$\mathcal{S}_2 = \{A_1, A_2, A_3, A_4, B_1, B_3, B_4, B_5, C_1, C_3, C_4, C_5, D_1, D_2\}$;
$\mathcal{S}_3 = \{A_1, A_2, B_1, B_2, B_3, B_4, B_5, C_1, C_3, C_4, C_5, D_1, D_2\}$.

Under complete semantics, $A_1, A_2, B_1, B_3, B_4, B_5, C_1, C_3, C_4, C_5, D_1, D_2$ are justified as they are members of all complete extensions, $A_5, A_6, A_7, A_8, A_9, A_{10}$ are overruled as they are defeated by justified argument C_5, C_2 is overruled as it is defeated by justified argument D_2, and A_3, A_4, and B_2 are defensible. For the other semantics, the same statuses are assigned; for grounded semantics, this is the case as $\mathcal{S}_1$ is the set inclusion minimal complete extension. For preferred and stable semantics, note that $\mathcal{S}_2$ and $\mathcal{S}_3$ are set inclusion maximal complete extensions for



(a)

Dialectical status	Claims
Justified	150 (A_1), 151 (A_2), 461 (B_1), 470 (B_3), 469 (B_4), ¬155a (B_5), 466 (C_1), 463 (C_3), 464 (C_4), 462 (C_5), 468 (D_1), 467 (D_2)
Defensible	149 (A_3), 152 (A_4), ¬149 (B_2)
Overruled	465 (C_2) 153 (A_5), 154 (A_6), 155 (A_7), 155a (A_8), 156 (A_9), Π_3 (A_{10})

(b)

Figure 6.6: AF corresponding to the IG of Figure 6.5, where justified, overruled and defensible arguments are indicated by green, red, and white circles, respectively (a); the dialectical status of the conclusions of the arguments in the AF, where the arguments corresponding to the conclusions are indicated in parentheses (b).

which it holds that $\forall B \notin \mathcal{S}_i, \exists A \in \mathcal{S}_i$ such that $(A, B) \in \mathcal{D}$ for $i = 2, 3$; hence, $\mathcal{S}_2$ and $\mathcal{S}_3$ are preferred and stable extensions.

The dialectical status of the conclusions of the arguments in the AF of Figure 6.6a are depicted in Figure 6.6b. The status of the ultimate claim Π_3 in the case is overruled, as argument A_{10} is the only argument with conclusion Π_3 and A_{10} is overruled. In particular, A_{10} is overruled as it is undercut on A_5 by justified argument C_5 with justified conclusion 462. The attack of C_5 on A_5 captures that, because Sacco carried his weapon as he intended to shoot rabbits with it (462), this caused him to draw his weapon (152) with the intention of surrendering it. A_{10} is also undercut by argument C_2 with conclusion 465 (Sacco carried a weapon because of his duties as a night watchman), but C_2 is overruled (and hence 465 is also overruled) as it is itself undercut by justified argument D_2 with justified conclusion 467 (Sacco was not a night watchman). Hence, the crucial argument responsible for the status of A_{10} is argument C_5. Similarly, intermediate conclusions $153 - 156$ that were iteratively used in inferring Π_3 are overruled, as arguments $A_5 - A_9$ are overruled. The claims that Sacco attempted to put his hand under his overcoat (149) and the negation of this claim (¬149) are defensible as B_2 and A_3 are defensible, and the claim that Sacco intended to draw his concealed weapon (152) is defensible as A_4 is defensible. All other conclusions in $\mathbf{P}$ are justified.

Finally, note that, while arguments constructed on the basis of the IG of Figure 6.4a can be directly evaluated using Dung's semantics, the BN constructed from this IG cannot be directly used for probabilistic inference. More specifically, the BN is partially specified as only qualitative probabilistic constraints and no exact probabilities are derived on the BN under construction. Moreover, the derived qualitative probabilistic constraints are only a subset of those required for the specification of a QPN [Wellman, 1990] (see also Section 5.5). Therefore, the way in which the constructed AF and BN are evaluated cannot be compared for the current case.

6.5 Concluding remarks

In this chapter, we have illustrated our approaches from Chapters 4 and 5 and performed a preliminary validation of our approach from Chapter 5 by means of a case study. We have provided a possible interpretation of Kadane and Schum's Wigmore chart as an IG, which illustrates that the IG-formalism is sufficiently expressive to model a complex case in a precise way. We have then applied our approaches from Chapters 4 and 5 to the constructed IG. Upon comparing the BN graph constructed by our approach from Chapter 5 to Kadane and Schum's manually constructed BN graph, we have concluded that the graphs are largely identical and that similar probabilistic inferences can be performed for the case at hand. As Kadane and Schum provided a thorough and extensive probabilistic analysis of the case, these similarities are a positive result of our validation and offer a preliminary indication that BNs constructed from IGs by our approach are of good quality. Moreover, the differences obtained illustrate that our approach may provide a more principled way of constructing BN graphs than the manner in which Kadane and Schum constructed their BNs. In particular, Kadane and Schum in some cases aggregated multiple claims in the Wigmore chart into single nodes in the BN graph, while by applying our approach all elements of the IG are explicitly captured in the corresponding BN graph to prevent loss of information. Furthermore, in comparison to the BN graph of Kadane and Schum head-to-head nodes are formed in our BN graph, which allows for directly capturing possible interactions between nodes in the graph.

Chapter 7

Supporting discussions about Bayesian networks using argumentation

As discussed in the introduction, BNs have found applications in many fields where uncertainty and evidence plays a role, including medicine, forensics and law [Fenton and Neil, 2012]. Although in BN construction it is good practice to document the model itself, the importance of documenting design decisions has received little attention. Such decisions, including the (possibly conflicting) reasons behind them, are important for experts involved in the construction and use of probabilistic models to understand and accept them. Moreover, when disagreements arise between experts involved in BN construction, there are no systematic means to resolve them.

An example of a tool that does support experts in documenting their BN modelling decisions is that of Yet and colleagues [2017] in the medical domain. Their tool allows BN developers to document the (clinical) knowledge, including conflicts, underlying the constructed BN in a queryable OWL ontology, which users can examine through an automatically generated free text web page. The tool, however, does not provide experts the ability to resolve disagreements and is not based on an argumentation model. Since disagreements about probabilistic models are essentially argumentative in nature, we prefer to use an approach for capturing and resolving conflicts based on argumentation. Other approaches that support experts in documenting their BN modelling decisions include other ontology-based approaches that allow experts to document the assumptions and background knowledge behind a BN under construction [Helsper and van der Gaag, 2007; van der Gaag and Tabachneck-Schijf, 2010] and a textual annotation tool [Antal et al., 2001] that allows for annotating different elements of BN graphs with background knowledge. However, these approaches similarly are not argumentation-based and do not allow experts to resolve disagreements.

Keppens [2014] recently proposed an argumentation-based approach to criticise and resolve discussions about a probability distribution. In his approach, convex sets of conditional probability values are calculated from the probabilistic constraints admitted by justified arguments posed by experts; hence, for a fixed BN graph-structure these convex sets can be used in probabilistic inference using credal networks [Cozman, 2000]. Keppens' approach can be used when discussions about a BN only concern its parameterisation. However, disagreements may also concern a BN's graph-structure including its variables, arcs, and variables' value spaces, which in this chapter we do not assume to be Boolean.

Accordingly, in this chapter we propose an approach that helps to capture and resolve disagreements among experts concerning any BN element. To this end, we allow experts to explicitly express their reasons pro and con modelling decisions regarding the structure and parameterisation of a (fully or partially specified) BN using argumentation. Disagreements are resolved as much as possible by utilising preferences that are specified over the arguments by the experts. The version of the ASPIC$^+$ framework as presented in Section 2.2.2 is used to formally specify our approach, which is more general purpose than our specific argumentation formalism from Chapter 4. Our approach is based on an argument-based analysis of an actual disagreement about a forensic BN for the interpretation of finger marks. In Section 7.1 our analysis of this disagreement is presented, on the basis of which we propose our argumentation-based approach for capturing and resolving conflicts about BN elements in Section 7.2.

7.1 Disagreements about a forensic finger mark Bayesian network

In this section, we analyse an actual disagreement about a BN to identify where disagreements about BNs typically arise and how such disagreements are typically expressed and resolved manually. Doekhie [2012] (in Master thesis research) and Haraksim and colleagues [2012] constructed a BN for the forensic interpretation of two finger marks[1], described in Section 7.1.1. Doshi [2013] (in Master thesis research) criticised this BN and proposed adjustments that address the identified shortcomings. In Section 7.1.2 we analyse the argumentation structure of Doekhie and colleagues' modelling decisions and Doshi's criticism on these decisions. While the main objective of this analysis is to identify where disagreements about BNs typically arise and how such disagreements are resolved manually, we also analyse the extent to which the constructed arguments can be classified as instances of existing or newly proposed argument schemes (Section 2.2.2.4) or as applications of critical questions of these schemes. These schemes and questions can be used alongside those proposed by Keppens [2014] and Prakken [2020] to guide the practical construction of arguments and counterarguments regarding BN elements.

[1] Compared to a fingerprint, which is a print taken from a suspect at the police station, a finger mark is a mark recovered from a crime scene.

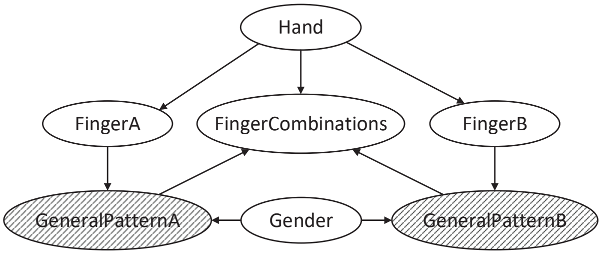


Figure 7.1: BN for the interpretation of two finger marks at finger level, constructed by Doekhie [2012] (in Master thesis research) and Haraksim and colleagues [2012].

7.1.1 BN for the interpretation of two finger marks

The BN constructed by Doekhie [2012] (in Master thesis research) and Haraksim and colleagues [2012] is used to evaluate from which fingers two finger marks recovered from a crime scene originated, where the assumption is made that these two finger marks are left behind by two consecutive fingers of the same hand in the act of a single touch. Specifically, if the fingers on the hands of a person are labelled 1 through 10, then the BN is used to calculate the (posterior) probability that the two marks originated from a specific configuration of consecutive fingers. In a police investigation, a fingerprint expert enters the marks into a software system that automatically compares them to a fingerprint database. Knowing beforehand from which finger a mark most probably originated can considerably narrow down the search and, in turn, speed up the matching process. It should be noted that the constructed BN cannot be used to evaluate from which person the finger marks originated; BNs at person level instead of at finger level are used for this purpose (see e.g. Taroni et al. [2014]).

Doekhie and colleagues' BN is depicted in Figure 7.1. FingerCombinations is the variable of interest for which we wish to obtain a posterior distribution. This variable describes eight values, corresponding to the eight possible combinations of fingers from which the two marks originated. Specifically, these values are 1 & 2, ..., 4 & 5, 6 & 7, ..., 9 & 10, where the first number denotes the finger number from which finger mark A originated and the second number indicates the finger number from which finger mark B originated. The FingerA and FingerB variables themselves each describe ten values, corresponding to the ten possible fingers from which a mark can originate. The GeneralPatternA and GeneralPatternB variables each describe a number of different values corresponding to the general patterns which are typically observed in finger marks and fingerprints, such as loops, whorls and arches. Upon using the BN in practice in a given case, these variables are instantiated to the general patterns observed in marks A and B to obtain a posterior distribution over the FingerCombinations variable, i.e. $\mathbf{E_V} = \{$GeneralPatternA, GeneralPatternB$\}$. The Hand variable accounts for the hand from which the two finger marks originated. This variable describes two values, namely *left_hand* and *right_hand*. Finally, the Gender variable accounts for the gender of the donor from which the two marks originated. According to Doekhie and colleagues, this variable describes three values: *male*, *female* and *unlabelled*.

The CPT for the Gender variable is filled using frequency statistics obtained from a *fingerprint database* (D_1). This database contains data on each subject's gender and the finger from which each print originated, as well as the general pattern of each fingerprint as labelled by a fingerprint examiner. In some cases, the subject's gender was not documented in the database and the prints were classified as 'unlabelled'. Frequency statistics from D_1 are also used to fill the CPTs for the GeneralPatternA, GeneralPatternB and FingerCombinations variables. The manner in which these frequencies are chosen by Doekhie and colleagues is not further discussed here, as the parameterisation of these variables is not criticised by Doshi.

The CPTs for the Hand, FingerA and FingerB variables are filled using frequency statistics from a *finger mark database* (D_2). This database contains similar data as D_1, except that the data is obtained from a large number of finger marks recovered from crime scenes instead of from fingerprints. As noted by Haraksim and colleagues [2012], frequency statistics provide for a more informed prior than a uniform prior, which assigns equal prior probabilities to each finger or hand. From D_2, it can, for instance, be seen that marks originating from the thumb and index finger are recovered more often from crime scenes than marks originating from other fingers.

7.1.2 Doshi's criticism on Doekhie and colleagues' BN

In this section, we analyse the argumentation structure of Doekhie and colleagues' modelling decisions (Section 7.1.1) and Doshi's criticism on these decisions. In the ASPIC$^+$-style argument graphs depicted throughout this section, propositions corresponding to Doekhie and colleagues' claims are indicated by plain boxes and propositions corresponding to Doshi's claims are indicated by thick boxes.

7.1.2.1 Relevance of the Hand variable

Haraksim and colleagues [2012, p. 102] state that frequency statistics from finger mark database D_2 indicate that the probability distribution over the Hand variable is non-uniform, from which they conclude that this variable should be included in the BN and that this non-uniform distribution should be used for the parameterisation of this variable. This can be interpreted as an *argument from data set* (defined below): as D_2 implies a property of the Hand variable, a probability distribution (or BN) over a set of variables including the Hand variable should be constrained by this property. Doekhie and colleagues' arguments are depicted in the centre of Figure 7.2 (inference d_1).

The argument scheme for arguments from data sets we propose is a generalisation of the scheme originally proposed by Keppens [2014, p. 260]. Keppens' scheme can only be used to reason about the source of a specific parameterisation. We generalise this scheme such that it can also be used to reason about general properties

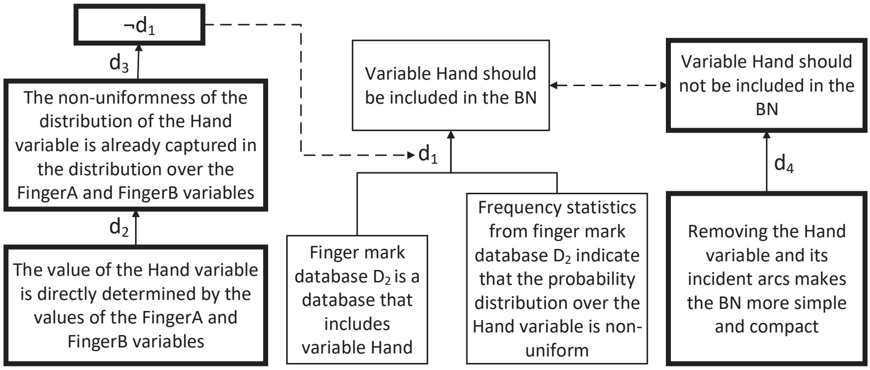


Figure 7.2: Argument-based analysis of Doekhie and colleagues' modelling decision for including the Hand variable and Doshi's criticism.

of a BN, such as whether or not a variable should be included in the BN or what its value space should be:

S is a data set that includes variable(s) $V_1, \ldots, V_n$.
S implies a property *Prop* of a subset $\mathbf{V}$ of $V_1, \ldots, V_n$.

Therefore, a BN with variables $\mathbf{V}$ may be constrained by property *Prop*.

This scheme includes the following critical questions (adapted from Keppens [2014, p. 261]):

1. Does data set S cover all variables and values of variables *necessary* to identify the relevant circumstances covered by property *Prop*?

1′. Does data set S cover all variables and values of variables *sufficient* to identify the relevant circumstances covered by property *Prop*?

2. Is the population considered in data set S representative for the population under investigation in the present case?
3. Is the volume and precision of data set S consistent with the precision of property *Prop*?
4. Is the observation of property *Prop* in data set S consistent with other data sets?

Doshi criticises Doekhie and colleagues' modelling decision by stating that the value of the Hand variable is directly determined by the values of the FingerA and FingerB variables. Specifically, knowing from which finger a mark originated implies that we know from which hand the mark originated. From this claim, it follows that the non-uniformness of the distribution over the Hand variable is already captured in the distribution over the FingerA and FingerB variables. This can be interpreted as an undercutter of the argument from data set as posed by Doekhie and colleagues; specifically, it is an instance of critical question 1′ of this argument scheme. This undercutter is depicted on the left-hand side of Figure 7.2.

Doshi provides additional reasons for not including the Hand variable by claiming that this makes the BN more simple and compact. To capture this, we propose the following *argument scheme for reduced complexity*:

V is a variable in BN **B**.
Removing variable V and its incident arcs from **B** makes **B** less computationally and representationally complex.

Therefore, variable V should be removed from **B**.

We propose the following corresponding critical questions for this scheme:

1. Is variable V relevant for computing the posterior distribution over variables of interest in **B**?
2. Does the complexity gain from removing V compensate for the loss of accuracy and completeness of **B**?

Doshi's argument for reduced complexity is depicted on the right-hand side of Figure 7.2. This argument and Doekhie and colleagues' argument then rebut. Weighing the reasons pro and con, Doshi concludes that the Hand variable along with its incident arcs should be removed.

7.1.2.2 Possible dependency between the GeneralPatternA and GeneralPatternB variables

Based on discussions with fingerprint examiners, Doshi found that there exists no significant dependency (or correlation) between the general patterns that exist on different fingers. Doshi criticises Doekhie and colleagues' BN by stating that the GeneralPatternA and GeneralPatternB variables are possibly dependent in their BN. The fingerprint examiners' claim that the GeneralPatternA and GeneralPatternB variables are independent can be interpreted as an argument from expert opinion (see Section 2.2.2.4). The argument for the claim that the GeneralPatternA and GeneralPatternB variables are independent and the argument for the claim that these variables can be dependent then rebut. Note that the latter claim is not explicitly made by Doekhie and colleagues; it is, however, implied by the structure of the BN they constructed. Doekhie and colleagues' implicit argument, as well as Doshi's counter-argument, are depicted on the right-hand side of Figure 7.3.

Doshi notes that the GeneralPatternA and GeneralPatternB variables can be made independent in Doekhie and colleagues' network by reversing the directions of arcs[2] between GeneralPatternX and FingerX, for both X = A and X = B. However, this would (1) not be in agreement with the perceived direction of causality between

[2] Note that by reversing the arcs between GeneralPatternX and FingerX for both X = A and X = B, an active chain still exists between GeneralPatternA and GeneralPatternB, namely via Gender; these variables, therefore, remain possibly dependent. In this section, we only concern ourselves with modelling Doekhie and colleagues' and Doshi's arguments, and we do not pose further (possible) (counter)arguments.

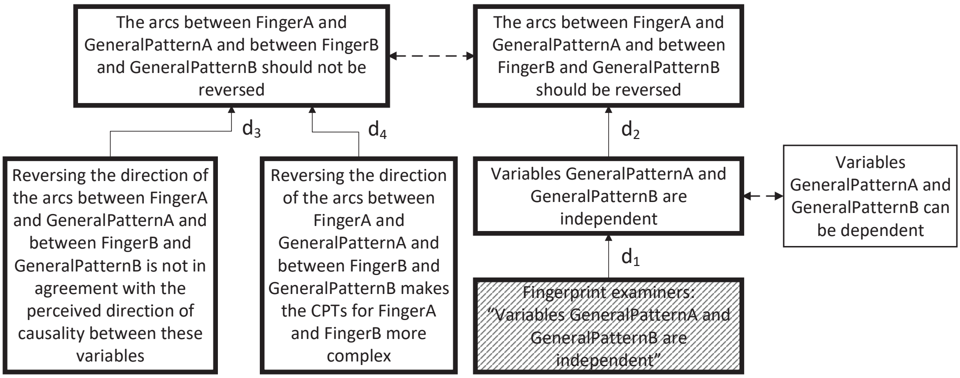


Figure 7.3: Argument-based analysis of Doshi's criticism on the possible dependency between the GeneralPatternA and GeneralPatternB variables.

these variables (inference d_3); and (2) would increase the number of probabilities that need to be estimated in the CPTs for the FingerA and FingerB variables and, therefore, would make these CPTs more complex (inference d_4). The latter argument can be interpreted as an instance of a variation on the argument scheme for reduced complexity presented in Section 7.1.2.1. Instead of concerning the removal of a variable, this variation concerns the reversal of an arc between two variables. Specifically, if reversing an arc between two variables makes the BN less computationally and representationally complex, this arc should be reversed. We replace critical question 1 by 'Does reversing the arc change the independence relation represented by the BN graph?'. Critical question 2 can be directly applied to this scheme by replacing the words 'removing V' by the words 'reversing the arc'.

Doshi's argument for reversing arc directions and his arguments for keeping the original arc directions then rebut, as depicted in Figure 7.3.

7.1.2.3 Conditional independence of the FingerCombinations variable and the FingerA and FingerB variables

Upon considering the BN graph of Figure 7.1, Doshi notes that the FingerCombinations variable is conditionally independent from the FingerA and FingerB variables given $\mathbf{Z} = \{\text{Hand, GeneralPatternA, GeneralPatternB}\}$ (its Markov blanket, see Definition 18). Doshi criticises this modelling decision by stating that, knowing from which two fingers the marks originated, we should be able to infer the combination of fingers that was used. Therefore, the FingerCombinations variable and the FingerA and FingerB variables should not be conditionally independent given $\mathbf{Z}$. An argument-based analysis of Doshi's criticism is depicted in Figure 7.4. Note that the argument on the left-hand side was not explicitly made by Doekhie and colleagues. Instead, it follows implicitly from the structure of Doekhie and colleagues' BN. Doekhie and colleagues' argument and Doshi's argument then rebut.

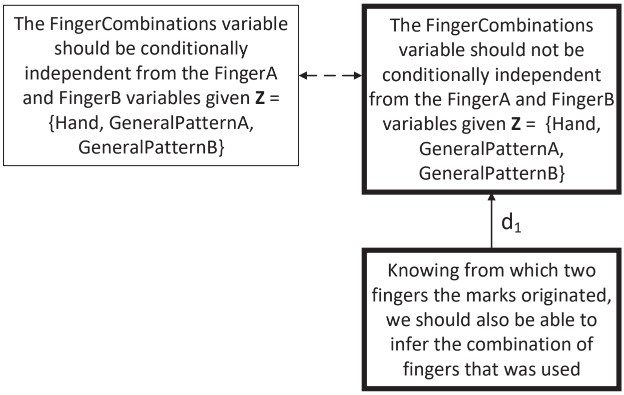


Figure 7.4: Argument-based analysis of Doshi's criticism on the independence of the FingerCombinations variable and the FingerA and FingerB variables given **Z** = {Hand, GeneralPatternA, GeneralPatternB}.

7.1.2.4 The value space of the Gender variable

Doekhie and colleagues base their design choice for using values *male*, *female* and *unlabelled* as the value space of the Gender variable on the observation that the Gender variable can take on these three values in D_1. This can be interpreted as an argument from data set; it is depicted in Figures 7.5 and 7.6 (inference d_1), where an undercutter of this argument is depicted in Figure 7.5 and a rebuttal to this argument is depicted in Figure 7.6. First, the undercutting attack is considered. Doshi criticises Doekhie and colleagues' decision by arguing that, in reality, people are either male or female, and that 'unlabelled' merely refers to the fact that data is missing with respect to this variable in D_1. Therefore (inference d_2), Doshi concludes that values *male*, *female* and *unlabelled* are not mutually exclusive. By furthermore stating that the values of BN variables should be mutually exclusive, Doshi poses (inference d_3) an exception to the argument from data set as posed by Doekhie and

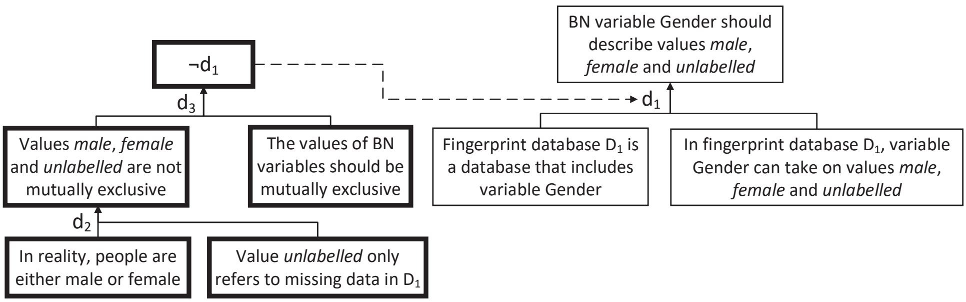


Figure 7.5: Doshi's undercutter of Doekhie and colleagues' argument for using values *male*, *female* and *unlabelled* as the value space of the Gender variable.

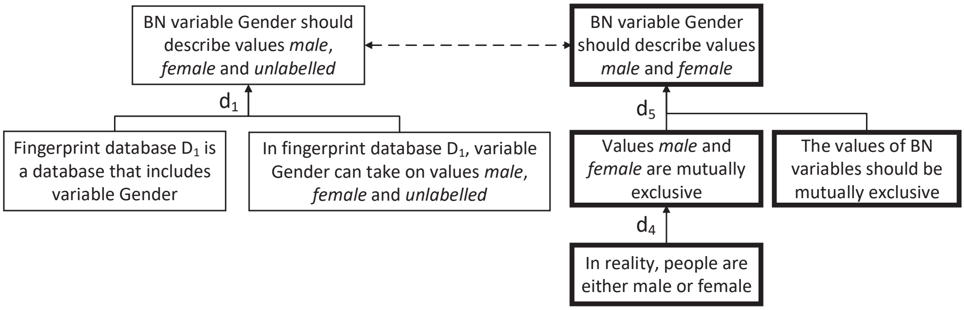


Figure 7.6: Doshi's rebuttal to Doekhie and colleagues' argument for using values *male*, *female* and *unlabelled* as the value space of the Gender variable.

colleagues, which can be interpreted as an instance of critical question $1'$ of this argument scheme. These arguments are depicted on the left-hand side of Figure 7.5.

In addition, Doshi claims that the Gender variable should instead describe values *male* and *female*. First, he claims that values *male* and *female* are mutually exclusive, which is based on his claim that, in reality, people are either male or female. This argument is depicted on the right-hand side of Figure 7.6 (inference d_4). From this claim and by again stating that the values of BN variables should be mutually exclusive, Doshi concludes that Gender should describe values *male* and *female* (d_5). This argument and Doekhie and colleagues' argument then rebut.

Based on his criticism, Doshi proposed to adjust the BN of Doekhie and colleagues by removing value *unlabelled* from the value space of the Gender variable.

7.1.2.5 Parameterisation of the FingerA and FingerB variables

Doekhie and colleagues' motivation for using frequency statistics from D_2 for the parameterisation of the FingerA and FingerB variables can be considered an argument from data set. This argument is depicted in the centre of Figure 7.7 (inference d_1). Doshi criticises this modelling decision by stating that D_2 cannot be used for establishing the relevant frequencies for these variables, as this database does not contain data regarding consecutiveness of finger marks. This can be interpreted as an undercutter of Doekhie and colleagues' argument from data set; specifically, it is an instance of critical question 1 of this argument scheme. Doshi's argument is depicted on the left-hand side of Figure 7.7 (inference d_2).

Doshi instead proposed to use uniform distributions for both values of the Hand variable for the CPTs for these variables. His reason for using uniform distributions is that no databases currently exist that contain data regarding consecutiveness of finger marks and that the CPTs for the FingerA and FingerB variables should, therefore, be uninformative (for now). These arguments are depicted on the right-hand side of Figure 7.7; Doekhie and colleagues' argument and Doshi's argument based on inferences d_3 and d_4 then rebut.

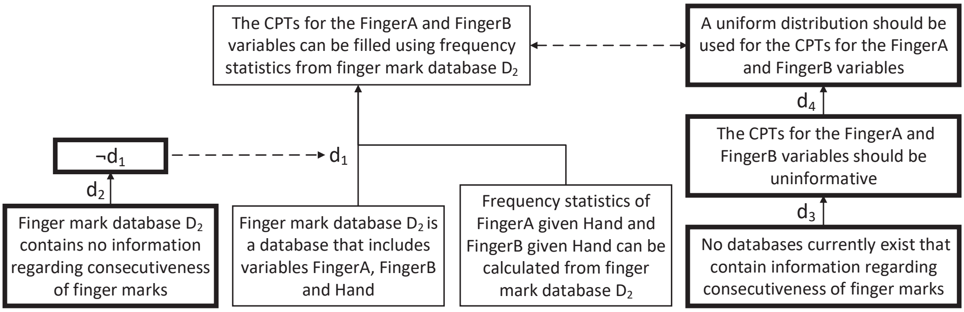


Figure 7.7: Argument-based analysis of Doekhie and colleagues' modelling decision regarding the parameterisation of the FingerA and FingerB variables and Doshi's criticism on this decision.

7.2 An argumentation-based approach to supporting discussions about Bayesian networks

In this section, we propose an argumentation-based approach that can be used to capture and help resolve conflicts about BN elements in a BN under construction. Our approach is based on our analysis of the disagreement of Section 7.1. The approach consists of two phases that are iteratively run through. In the first phase, experts are allowed to construct arguments pro and con the outcomes of modelling decisions underlying a given fully or partially specified BN, as well as specify preferences over these arguments. Arguments can be constructed regarding the inclusion or exclusion of different types of BN elements (described below) which may or not be in the existing BN. In the second phase, conflicts are then resolved by using the dialectical status of the constructed arguments to derive probabilistic and structural constraints on the BN (Sections 7.2.1 and 7.2.2).

Our sub-division of BN elements is inspired by our analysis of the disagreement of Section 7.1, by the work of Pitchforth and Mengersen on the validation of expert-elicited BNs [Pitchforth and Mengersen, 2013] and by the work of Yet and colleagues [2017]. Below, the types of arguments that can be constructed regarding each type of BN element are listed:

- Arguments regarding the existence of variables (Section 7.1.2.1). Conclusions of such arguments are of the form ϕ_V = *'Include V in* $\mathbf{V}$*'* and $\neg\phi_V$ = *'Exclude V from* $\mathbf{V}$*'*.
- Arguments regarding whether a variable is observable or not. Typically, a fixed set of variables $\mathbf{E_V}$ is observed and instantiated upon using the BN in practice in a given case. Conclusions of such arguments are of the form $\phi^V_{\mathbf{E_V}}$ = *'Include V in* $\mathbf{E_V}$*'* and $\neg\phi^V_{\mathbf{E_V}}$ = *'Exclude V from* $\mathbf{E_V}$*'*.

- Arguments regarding (conditional) (in)dependencies (Sections 7.1.2.2 and 7.1.2.3). Arguments of this type concern the inclusion or exclusion of a (conditional) independence constraint $I = I(V_1, V_2, \mathbf{Z})$ in the set of (conditional) independence constraints $\mathbf{I}$ to which the BN should adhere to, where $I(V_1, V_2, \mathbf{Z})$ states that V_1 and V_2 are conditionally independent given a (possibly empty) subset of variables $\mathbf{Z} \subseteq \mathbf{V}$. In particular, conclusions of such arguments are of the form $\phi_I^{V_1,V_2,\mathbf{Z}} =$ *'Include* $I = I(V_1, V_2, \mathbf{Z})$ *in* $\mathbf{I}$*'* and $\neg\phi_I^{V_1,V_2,\mathbf{Z}} =$ *'Exclude* $I = I(V_1, V_2, \mathbf{Z})$ *from* $\mathbf{I}$*'*.
- Arguments regarding the existence of arcs (Section 7.1.2.2). Conclusions of such arguments are of the form $\phi^{V_1 \to V_2} =$ *'Include arc* $V_1 \to V_2$ *in* $\mathbf{A}_{\mathcal{B}}$*'* and $\neg\phi^{V_1 \to V_2} =$ *'Exclude arc* $V_1 \to V_2$ *from* $\mathbf{A}_{\mathcal{B}}$*'*. Arguments for arc reversal can then be indirectly constructed by constructing arguments with conclusions $\neg\phi^{V_1 \to V_2}$ and $\phi^{V_2 \to V_1}$.
- Arguments regarding the value spaces of variables (Section 7.1.2.4). Conclusions of such arguments are of the form $\phi_{S_1}^V =$ *'Use value space 1 for variable V'*. Such arguments may either be attacked by constructing an argument for its negation $\neg\phi_{S_1}^V =$ *'Do not use value space 1 for variable V'* or by constructing an argument for an alternative value space $\phi_{S_2}^V =$ *'Use value space 2 for variable V'*, where $\phi_{S_1}^V$ and $\phi_{S_2}^V$ are declared contradictories of each other in $\mathcal{L}$.
- Arguments regarding the parameterisation of variables (Section 7.1.2.5). Conclusions of such arguments are denoted by ϕ_{Par}^V and are similar in form to arguments regarding value spaces of variables, and can be attacked similarly, by replacing the words 'value space' by the word 'parameterisation'. Here, we assume that a conflict concerns *part* of the CPT for V. Specifically, let $\mathrm{V}_1, \ldots, \mathrm{V}_n$ be the parents of V. Then conflicts regarding the parameterisation of V concern a specific distribution $\Pr(\mathrm{V} \mid \mathrm{V}_1 = v_1, \ldots, \mathrm{V}_n = v_n)$ for given values v_1 of $\mathrm{V}_1, \ldots, v_n$ of V_n. We note that probabilities $\Pr(\mathrm{V} = v \mid \mathrm{V}_1 = v_1, \ldots, \mathrm{V}_n = v_n)$ should sum to 1 when summing over all possible values v of V; conflicts regarding the parameterisation of a variable, therefore, never concern a single value of its CPT.

It should be noted that we do not propose a dialogue protocol. Instead, we first allow experts to specify arguments regarding all the different types of BN elements described above along with a preference relation over those arguments, after which a priority ordering over the arguments corresponding to the different types of BN elements is applied to establish which conflicts, if any, should be resolved first. This process is repeated until a satisfactory BN is obtained.

7.2.1 Priority ordering for conflict resolution

The idea of a priority ordering is based on the observation that BN elements from different types are dependent on one another. For instance, it only makes sense to resolve conflicts regarding the value space of variable V iff V's existence is justified.

We propose the following priority ordering over the types of BN elements:

P_1. Existence of variables in $\mathbf{V}$.
P_2. Inclusion of variables in $\mathbf{E_V}$.
P_3. Conditional independencies between variables given $\mathbf{Z} = \mathbf{E_V}$.
P_4. Conditional independencies between variables given subsets $\mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z} \neq \mathbf{E_V}$.
P_5. Existence of arcs between variables.
P_6. Value spaces of variables.
P_7. Parameterisation of variables.

A BN graph, by means of its arcs and their directions, represents an independence relation, so the (conditional) independencies implied by the graph should be verified. We note that, upon constructing BNs, arcs are typically added first, after which the implied independence relation is verified. In our approach, we instead prioritise the resolution of conflicts regarding independencies (P_3 and P_4) over those concerning arc directions (P_5), as arguments of the former type should generate new arguments for the existence of arcs and their directions at P_5 that together realise the conditional independence relation implied by the arguments at P_3 and P_4. Since an unjustified independence assumption can affect the behaviour of the BN, independencies exploited in deriving conclusions in an actual case are most important to verify. As such, we prioritise arguments regarding conditional independencies given $\mathbf{E_V}$ (P_3) over conditional independencies given subsets $\mathbf{Z} \subseteq \mathbf{V}$, $\mathbf{Z} \neq \mathbf{E_V}$ (P_4).

After obtaining a fully specified BN graph, conflicts regarding the value spaces of variables are considered (P_6). Lastly, at P_7 conflicts regarding the parameterisation of variables are resolved. Specifically, the parameterisation of a variable V can be considered iff V's existence, V's value space, the existence of V's incoming arcs, and the value spaces of V's parents are justified. We note that, alternatively, conflicts regarding value spaces can be considered directly after resolving conflicts regarding the existence of variables at P_1, as the BN elements at $P_2 - P_5$ are independent of the value spaces of variables and the same BN would be obtained via our approach using this alternative ordering. We opt to consider conflicts regarding value spaces at P_6 to allow for a more straightforward specification and explanation of the second phase of our approach, as conflicts regarding value spaces of variables and the parameterisation of variables (at P_7) are resolved similarly.

We introduce the following notation. With $\mathcal{A}_{P_i}$ we denote the set of arguments whose conclusions concern BN elements at priority level P_i, $i \in \{1, \ldots, 7\}$. We assume that $\mathcal{A}_{P_i}$ divides into disjoint subsets of arguments $\mathcal{A}^e_{P_i}$ concerning BN element e at priority level P_i. For instance, $\mathcal{A}^V_{P_1}$ only contains those arguments with conclusions ϕ_V and $\neg\phi_V$ for the inclusion/exclusion of variable V.

7.2.2 Constraint table and default choices

In Figure 7.8, the second phase of our approach, concerning conflict resolution, is summarised in pseudocode; throughout this section, we will explain and refer to

```
1  function ConflictResolution(BN, Args):
2      Input: Partially or fully specified BN
3      Input: Set of arguments Args with preferences
4      Output: Partially or fully specified BN
5      for i ∈ {1, . . . , 7} do
6          foreach BN element e at priority level P_i do
7              Calculate dialectical status of arguments in A^e_{P_i}
8              Consult constraint table:
9              if constraint table states "Include e" then
10                 Include e
11             end
12             else if constraint table states "Exclude e" then
13                 Exclude e
14                 Retain documentation regarding e
15                 Disregard arguments in A^e_{P_j} for j > i
16             end
17             else if constraint table states "No constraint" then
18                 Ask expert to further specify A^e_{P_i}:
19                 if A^e_{P_i} is further specified then
20                     goto line 7
21                 end
22                 else
23                     Resort to default choice
24                 end
25             end
26         end
27     end
28 end
```

Figure 7.8: Pseudocode of our argumentation-based approach to resolving conflicts about BN elements.

the lines in this pseudocode. As input for this phase, either a fully or partially specified BN is used, along with a set of arguments and preferences as specified in the first phase of our approach (lines 2 and 3 of the pseudocode). The process of conflict resolution at priority level P_i starts by calculating the dialectical status of conclusions in $\mathcal{A}^e_{P_i}$ for each BN element e (line 7 of the pseudocode). Conflicts are then resolved by consulting Table 7.1, which throughout this chapter will be referred to as a *constraint table* (line 8 of the pseudocode). Depending on the priority level, different rows of the constraint table are used. At priority levels $P_1 - P_5$, the set of arguments $\mathcal{A}^e_{P_i}$ can either contain arguments with ultimate conclusions ϕ_e only, $\neg\phi_e$ only, or both. For these priority levels, rows $1 - 3$ of the constraint table are used. At priority levels $P_6 - P_7$, arguments can also be constructed for claims regarding alternative value spaces or parameterisations, indicated by $\phi_{e'}$; row 4 is also used at these priority levels. The second column of the constraint table indicates the possible configurations of dialectical status of the claims in the first column. The third column then indicates the corresponding constraints on the BN for each of these configurations. Entries with an asterisk indicate BN constraints at priority

Table 7.1: Constraint table for including BN element e. Entries with an asterisk indicate BN constraints at priority levels $P_6 - P_7$, whereas for the same dialectical status configurations at priority levels $P_1 - P_5$ entries without an asterisk are used.

Proposed claims	Dialectical status	BN constraint	Default choice
ϕ_e and $\neg\phi_e$	ϕ_e justified, $\neg\phi_e$ overruled ϕ_e overruled, $\neg\phi_e$ justified ϕ_e defensible, $\neg\phi_e$ defensible ϕ_e overruled, $\neg\phi_e$ overruled	Include e Exclude e / No constraint* No constraint Further specification needed	No default choice No default choice / Boolean variable or uniform distribution* Original modelling decision No default choice
ϕ_e only	ϕ_e justified ϕ_e overruled ϕ_e defensible	Include e Exclude e / No constraint* No constraint	No default choice No default choice / Boolean variable or uniform distribution* Original modelling decision
$\neg\phi_e$ only	$\neg\phi_e$ justified $\neg\phi_e$ overruled $\neg\phi_e$ defensible	Exclude e / No constraint* Include e / No constraint* No constraint	No default choice / Boolean variable or uniform distribution* No default choice / Boolean variable or uniform distribution* Original modelling decision
ϕ_e and $\phi_{e'}$	ϕ_e justified, $\phi_{e'}$ overruled ϕ_e overruled, $\phi_{e'}$ justified ϕ_e defensible, $\phi_{e'}$ defensible ϕ_e overruled, $\phi_{e'}$ overruled	Include e Include e' No constraint No constraint	No default choice No default choice Original value space or average of the two distributions Boolean variable or uniform distribution

levels $P_6 - P_7$, whereas for the same dialectical status configurations at priority levels $P_1 - P_5$ entries without an asterisk are used.

Depending on the entry in the third column of the table, different actions should be taken. If the entry in the constraint table reads 'Include e' or 'Exclude e', then the BN element should be included in the BN, respectively excluded from the BN (lines 9-11 respectively 12-16 of the pseudocode). In the latter case, the arguments regarding this variable should still be retained for documentation purposes in order to keep a 'chain-of-custody' of the changes made to the BN. If an expert wishes to provide further reasons as to why this element should still be included, they can then review the existing arguments regarding this BN element and supplement it with additional arguments or preferences. Furthermore, in case BN element e is excluded, then arguments of inferior priority concerning e should be disregarded in that these arguments should not be evaluated or used in conflict resolution. For instance, if a variable V is removed from the BN at priority level P_1, then arguments in $\mathcal{A}_{P_j}$ for $j > 1$ that concern this variable should be disregarded. Arguments for a conditional independency at P_3 and P_4 should then be disregarded in case V is an element of the set of variables $\mathbf{Z} \subseteq \mathbf{V}$ under consideration or if V takes on the role of V_1 or V_2 in the argument.

For some configurations of dialectical status, a (univocal) constraint on the BN cannot be derived (explained in more detail per priority level below). Considering priority levels $P_1 - P_5$, two BNs are then possible: one in which e is included and one from which e is excluded. One possible approach would be to retain two BNs at this point, one including e and one excluding e, and to continue resolving conflicts for both BNs by considering all BN elements at all priority levels. Following this approach, a list (or tree) of candidate BNs is obtained after attempting to resolve all conflicts, from which the experts can choose one BN to continue with. However, for larger BNs the space of candidate BNs corresponding to a set of arguments would quickly become large. Moreover, if a list of candidate BNs is presented to the experts, the differences between BNs in the list also need to be explained to them in order for them to make a decision for one of these BNs, which would require additional machinery.

We opt for an alternative approach in which the experts are asked to specify further information every time a univocal constraint cannot be derived (line 18 of the pseudocode). Specifically, if no constraint is derived regarding BN element e, then the experts are asked to specify additional arguments to supplement $\mathcal{A}^e_{P_i}$, or to (further) specify preferences over $\mathcal{A}^e_{P_i}$. In case the experts do not further specify their arguments or preferences, a default choice is made by the approach (lines 22-24 of the pseudocode). What this default choice entails differs per priority level and per configuration of dialectical status, as indicated in the fourth column of the constraint table. We note that default choices are only made in case a more informed choice cannot be made based on the specified information, and that, informally, the more complete the specification of the arguments and preferences, the less the approach will resort to making a default choice.

In case the experts further specify their arguments or preferences, then they can recalculate the dialectical status of ϕ_e and/or $\neg\phi_e$ and establish the corresponding constraints. This process reiterates until BN element e is either included or excluded, or a default choice is made.

In the next subsections, the different configurations of dialectical status per priority level and their corresponding constraints are discussed in more detail.

Conflicts about variables. At priority level P_1, conflicts regarding the existence of variables are considered. For resolving conflicts at this priority level, rows $1-3$ of the constraint table are used. For row 1, in case claim ϕ_V for including variable V is justified and $\neg\phi_V$ is overruled, then this variable is included in $\mathbf{V}$. In case $\neg\phi_V$ is justified and ϕ_V is overruled, variable V and all its incidents arcs are excluded from $\mathbf{V}$. In case ϕ_V and $\neg\phi_V$ are defensible, no constraint is derived; in case no further information is specified by the experts, then the default choice is to resort to using the original modelling decision, i.e. if variable V was included in/excluded from the original BN, then it should be included in/excluded from the resulting BN. Lastly, if both ϕ_V and $\neg\phi_V$ are overruled, then neither the choice for including nor excluding the variable as a default is warranted. In this case, the experts should always further specify their arguments and preferences to be able to derive a constraint (entry 'Further specification needed' in the constraint table). Rows 2 and 3 of Table 7.1 are interpreted similarly.

Example 59. *As an example of an application of our approach to resolving conflicts about variables, we consider the argument-based analysis of Section 7.1.2.1. In Figure 7.2, Doekhie and colleagues' argument for including the Hand variable in the BN* *is undercut and, therefore, overruled under any semantics, as undercutting attack is preference-independent. It follows that Doshi's argument for excluding this variable is justified, as its only defeater is overruled. From the constraint table, it therefore follows that the Hand variable should not be included in the BN graph. In case we consider the set of arguments excluding Doshi's undercutter, then both claims for including and excluding the variable are defensible and no univocal constraint is derived. Doshi could in this case, for instance, specify that his argument for exclusion of the Hand variable is strictly preferred to Doekhie and colleagues' argument for inclusion of this variable. The dialectical status of the corresponding arguments is then recalculated; Doshi's argument for excluding the Hand variable is now justified, from which a constraint on the BN to exclude this variable is derived. In case both Doshi and Doekhie do not further specify their arguments or preferences, then an informed choice cannot be made by the approach and the original modelling decision is used as a default (i.e. variable Hand is included in* $\mathbf{V}$*).* □

After resolving all conflicts at priority level P_1, conflicts at P_2, regarding the set of observable variables $\mathbf{E_V}$, are considered. Here, the terms 'Include V' and 'Exclude V' now refer to including V in and excluding V from the set $\mathbf{E_V}$ rather than set $\mathbf{V}$.

Conflicts about independencies and arcs. At priority levels P_3 and P_4, conflicts regarding (conditional) independencies between variables are resolved using rows $1-3$ of the constraint table. For row 1, in case $\phi_I^{V_1,V_2,\mathbf{Z}}$ is justified and $\neg\phi_I^{V_1,V_2,\mathbf{Z}}$ is overruled, then (conditional) independence constraint $I = I(V_1, V_2, \mathbf{Z})$ should be included in the set of (conditional) independence constraints $\mathbf{I}$ to which the BN should adhere to. In case $\neg\phi_I^{V_1,V_2,\mathbf{Z}}$ is justified and $\phi_I^{V_1,V_2,\mathbf{Z}}$ is overruled, then this independence constraint should not be included in $\mathbf{I}$. In case $\phi_I^{V_1,V_2,\mathbf{Z}}$ and $\neg\phi_I^{V_1,V_2,\mathbf{Z}}$ are defensible, then the default choice is to resort to the modelling decision implied by the original BN, that is, if V_1 and V_2 are conditionally independent given $\mathbf{Z}$ then I should be included in $\mathbf{I}$ and otherwise I can be excluded. We note that, if a variable V is removed from or added to $\mathbf{E_V}$ at priority level P_2, then arguments of inferior priority concerning $\mathbf{E_V}$ (i.e. at P_3) should be disregarded in conflict resolution.

Example 60. *Consider the argument-based analysis of Section 7.1.2.3. The arguments depicted in Figure 7.4 concern priority level* P_4*, as Doshi considers the conditional independencies represented by the BN graph for a set of variables* $\mathbf{Z} = \{$*Hand, GeneralPatternA, GeneralPatternB*$\}$*. As Doshi does not express an explicit preference, the conclusions of the rebutting arguments in Figure 7.4 are defensible under any semantics and no univocal constraint is derived. In case both Doshi and Doekhie do not further specify their arguments or preferences, then the default choice is to include* $I_1 = I(\mathit{FingerCombinations}, \mathit{FingerA}, \mathbf{Z})$ *and* $I_2 = I(\mathit{FingerCombinations}, \mathit{FingerB}, \mathbf{Z})$ *in* $\mathbf{I}$*, as these conditional independence constraints are implied by the structure of the original BN graph. In case Doshi expresses a strict preference for either of these arguments, then the chosen argument will be justified and the other* *overruled, in which case a constraint is derived. We note that upon using the BN of Doekhie and colleagues in an actual case, typically only the GeneralPatternA and GeneralPatternB variables are instantiated, while the Hand variable is typically not instantiated. Doekhie and colleagues could, therefore, possibly counter Doshi's arguments at* P_4 *by constructing arguments for* $\mathbf{E_V} = \{GeneralPatternA, GeneralPatternB\}$ *at priority level* P_2 *and arguments at* P_3 *for a conditional dependency between the FingerA and FingerB variables and the FingerCombinations variable given* $\mathbf{E_V}$*. These arguments would then be prioritised over Doshi's arguments at* P_4 *in conflict resolution.* □

After resolving all conflicts at P_3 and P_4, our approach should generate new arguments for the existence of arcs and their directions at P_5 that together realise the conditional independence relation implied by **I**. One way to achieve this is by manually eliciting arc directions, as illustrated by the following example.

Example 61. *Consider the argument-based analysis presented in Section 7.1.2.2. In Figure 7.3, arguments are depicted for the claims that GeneralPatternA and GeneralPatternB should be independent and can be dependent. Assuming that Doshi strictly prefers his argument over Doekhie and colleagues' implicit argument, his argument is justified under any semantics. The corresponding independence constraint is, therefore, added to the set of independence constraints the BN should adhere to. Doshi then proceeds by attempting to find a configuration of arc directions under which GeneralPatternA and GeneralPatternB are independent, which generates arguments at* P_5*, specifically arguments for Doshi's claims that the arcs between FingerA and GeneralPatternA and between FingerB and GeneralPatternB should be reversed. The generated arguments at* P_5 *and the arguments that were already proposed by Doshi are then collectively considered, where conflicts at* P_5 *are resolved using rows* $1-3$ *of the constraint table.* □

Alternatively, so-called *structure learning algorithms* can be used to learn the structure of a BN graph [Jensen and Nielsen, 2007, Chapter 7]. These algorithms can aid experts unfamiliar with the BN-formalism in finding a BN graph with d-separation properties corresponding to their set of independence constraints **I**.

Finally, we note that (conditional) independencies can also be enforced by constraining the specified probabilities, and that arguments at P_3 and P_4 could, therefore, also possibly generate new arguments at P_7 that together with the arguments for the existence of arcs and their directions at P_5 help realise the conditional independence relation implied by the arguments at P_3 and P_4.

Conflicts about value spaces. To resolve conflicts at priority level P_6, all rows of the constraint table are used. Note that this table can only be used if arguments for at most two different value spaces for a variable V are constructed; similar constraint tables can be constructed in case more than two value spaces are proposed. In row 1 of Table 7.1, if ϕ_S^V is justified and $\neg\phi_S^V$ is overruled, then value space S should be used for V. In case ϕ_S^V and $\neg\phi_S^V$ are defensible, then no univocal constraint is derived. In this case, the default choice is to resort to using the value space of this variable as specified in the original BN. In case ϕ_S^V is overruled and $\neg\phi_S^V$ is justified, then also no constraint is derived. Specifically, the only conclusion that can be drawn from the information specified by the experts is that value space S should not be used, as arguments for alternative value spaces are not provided. In this case, the default choice is to use a Boolean variable. However, if ϕ_S^V already concerns a Boolean value space, then the experts should always further specify their arguments and preferences. Rows 2 and 3 of Table 7.1 are interpreted similarly.

In the fourth row of Table 7.1, two alternative value spaces S_1 and S_2 are under consideration and arguments are constructed for both claims $\phi_{S_1}^V$ and $\phi_{S_2}^V$. In case one of these claims is justified, the other is overruled; this results in the choice for using the value space posed in the conclusion of the justified argument. In case $\phi_{S_1}^V$ and $\phi_{S_2}^V$ are both defensible, then the default choice is to resort to using the value space of this variable as specified in the original BN. In case both $\phi_{S_1}^V$ and $\phi_{S_2}^V$ are overruled, then neither value space should be used. If possible, the default choice is to again resort to using a Boolean value space for V.

Example 62. *As an example of our approach to resolving conflicts about value spaces, we consider the argument-based analysis presented in Section 7.1.2.4. In Figures 7.5 and 7.6, arguments for two different value spaces are depicted, where the argument for Doekhie and colleagues' value space is overruled under any semantics as it is undercut. It follows that the argument for Doshi's value space is justified under any semantics, as its only defeater is overruled. From the fourth row of the constraint table, it therefore follows that Doshi's value space should be used.* □

We reiterate that the two phases or our approach may need to be iteratively run through to obtain a satisfactory BN. For instance, consider the example in which an expert wishes to include value *other* in variable Gender's value space. The expert may then specify arguments for the inclusion of this value, after which the value may be included after conflict resolution. Suppose that upon knowing that this value is included, the expert wishes to include additional variables and arcs that connect to the Gender variable to be able to specify the conditional probabilities involving this value in more detail. In case arguments for these BN elements are not specified in the current iteration, then these may be specified in the next iteration of the approach, after which these elements may be included after conflict resolution.

Conflicts about parameterisations. Conflicts at priority level P_7 are resolved similarly as conflicts at P_6, where the default choice is to resort to using a uniform distribution in case the claims for the provided parameterisations are overruled and no further information is specified. Although a uniform distribution can easily be criticised, we opt for using this distribution as a more informed choice cannot be made based on the specified information. The default choice is also to resort to using a uniform distribution even if the original distribution was uniform, instead of always requiring the experts to further specify their arguments and preferences in this case. Arguments in $\mathcal{A}^{V}_{P_7}$ regarding the parameterisation of a variable V should be disregarded in conflict resolution if one of V's incoming arcs, V's value space or one of V's parents' value spaces is overruled at P_5 or P_6. In this case, the default choice is to resort to using a uniform distribution for V. Finally, in case two alternative parameterisations Par_1 and Par_2 of variable V are under consideration and both $\phi^{V}_{Par_1}$ and $\phi^{V}_{Par_2}$ are defensible, then the default choice is to use the average of the two proposed distributions as this is more informative than using a uniform distribution or the distribution of the variable as specified in the original BN.

Example 63. *Consider the argument-based analysis presented in Section 7.1.2.5. In Figure 7.7, arguments for two different parameterisations are provided. Similar to the example discussed in the previous subsection, Doshi's argument is justified and Doekhie and colleagues' argument is overruled under any semantics; therefore, Doshi's parameterisation should be used.* □

7.3 Discussion and concluding remarks

In this chapter, we have proposed an approach for capturing and resolving conflicts about BN elements using computational argumentation. Our approach allows experts to document their reasons pro and con modelling decisions and their preferences in a structured manner using argumentation. The dialectical status of the constructed arguments is then used to derive probabilistic and structural constraints on the BN. Our approach always returns a fully or partially specified BN. Starting with a fully specified BN, a fully specified BN is returned by resorting to default choices in case a univocal constraint cannot be derived and the arguments and preferences are not further specified. Our approach can possibly be extended in future research by making explicit to the experts in which ways their arguments and preferences can be further specified to obtain a univocal constraint. This would be an application of research on so-called resolution semantics; see e.g. [Modgil and Prakken, 2012]. As discussed earlier, instead of aggregating possibly incompatible expert advice into one BN model, an alternative approach would be to retain a list (or tree) of candidate BNs, where each BN corresponds to a Dung extension. From a theoretical perspective, it may be worthwhile to study such an approach in future work. For instance, it can then be established for which (types of) extensions the same BN is obtained or a fragment of the obtained BN is identical.

In related research, Nielsen and Parsons [2007] presented a framework which allows multiple agents equipped with a BN to agree on a possible consensus BN graph using computational argumentation. Their approach differs from ours in that our goal is not to propose a method for fusing different BNs. The approach of Neil and colleagues [2019] allows multiple independently built BNs presented by different parties in a legal case to be compared and 'averaged' by a trier of fact, with no attempt to make them consistent in terms of structure or parameterisation. In their approach, the trier of fact assigns prior probabilities to the different BN models, which are then updated based on claims presented by different parties during the trial process. Models that are more heavily disconfirmed by these claims are assigned lower weights as model plausibility measures, which are then subsequently used in the adopted Bayesian model comparison and averaging approach to calculate posterior probabilities of guilt and innocence of the defendant. Their approach differs from ours in that we consider the context in which disagreements about a single BN model are to be resolved instead of the context in which different BN models are to be weighed and averaged.

In other related work, Nicholson and colleagues [2020] proposed BARD (Bayesian ARgumentation via Delphi), which is both a methodology and a software tool that supports groups of users unfamiliar with the BN-formalism to collaboratively build a consensus BN in a user-friendly manner. More specifically, users are guided through a structured (but informal) Delphi-like elicitation protocol (i.e. a systematic approach for combining multiple perspectives in a democratic, reasoned, and iterative manner), where in an online platform they can incrementally and iteratively build and evaluate their individual BNs and where feedback can be sought through communication with other group members and a facilitator. In their approach, the facilitator manually synthesises and combines the group's work at every step to develop a coherent solution that reflects the group's thoughts. In contrast to this informal approach, our approach is formally specified, where possibly incompatible expert advice is (partly automatically) aggregated into one BN model by using an algorithm based on formal (computational) argumentation.

Our approach is based on an argument-based analysis of an actual disagreement about a forensic finger mark BN. In related work, Prakken [2020] performed a similar argument-based analysis of actual court discussions regarding Bayesian analyses of criminal cases. However, he mainly concerned himself with establishing the usefulness of argumentation in structuring this kind of discussion, while our analysis instead serves to identify how disagreements about BN elements are typically expressed and resolved manually by experts, and how this process can be precisely specified and (partly) automated using formal (computational) argumentation. Such a specification allows formal properties of our approach to be studied in future work. Moreover, in future work our formal model can be the basis for developing and implementing software tools for supporting discussions about BNs between experts and for communicating their BNs and discussions to others. For instance, in the legal and forensic domains such software tools would allow forensic experts and crime analysts to communicate their BNs and discussions to each other and to judges and prosecutors.

Chapter 8

Related research

In this chapter, various formalisms and approaches related to those proposed and considered throughout this thesis are discussed. Note that relations between the IG-formalism and ASPIC$^+$ and our argumentation formalism based on IGs and ASPIC$^+$ were already discussed in Section 3.2 and Section 4.4.1, respectively. Moreover, in Section 3.2 we provided reasons for preferring the IG-formalism to other formalisms for reasoning about evidence as an intermediary formalism between analyses performed using sense-making tools and formal AI systems. In that section, a comparison of our IG-formalism to the AIF [Rahwan and Reed, 2009] was also made.

To put our IG-formalism (Chapter 3) and our argumentation formalism based on IGs (Chapter 4) further into context, in Section 8.1 related formalisms for inference and argumentation with causality and other types of information are discussed. More specifically, we discuss work by which we were inspired and compare our argumentation formalism based on IGs to argumentation formalisms other than ASPIC$^+$. In Section 8.2 related work on BN construction is considered, in particular approaches for constructing BNs from information specified in arguments and ontologies that are closely related to our BN graph construction approach of Chapter 5. In Section 8.3 approaches to probabilistic argumentation are considered, which combine argumentation and probabilities. Finally, in Section 8.4 argumentation-based explanation methods for BNs are discussed. Note that approaches related to our method for capturing and resolving disagreements about BNs using argumentation proposed in Chapter 7 were already discussed in motivating and discussing our method in the introduction and conclusion of that chapter.

8.1 Inference and argumentation with causality and other information

In Chapter 3 we presented the graph-based IG-formalism for deduction and abduction with causal, evidential, abstraction, and other types of generalisations, as well as causal generalisations that include enabling conditions. In this section, work related to the IG-formalism and the argumentation formalism based on it are discussed.

8.1.1 Inference with causality and other information

Our IG-formalism is inspired by previous work on abduction, including work on formal-logical models of causal-abductive reasoning in which causal rules are used to explain observations; examples of such models are the models of Josephson and Josephson [1994] and Console and Torasso [1991], as well as the work of Kakas and colleagues [1993] on abductive logic programming, in which causal knowledge is expressed using abductive logic programming rules.

In formal-logical models of causal-abductive reasoning, a domain theory $\mathbf{T}$ consisting of causal rules of the form $c_1, \ldots, c_n \rightarrow_c e$ applicable to the domain is specified, along with a set of observations $\mathbf{O}$ and a set of abducibles $\mathbf{C}$, both consisting of literal propositions. Given a specified subset $\mathbf{F} \subseteq \{o \mid o \in \mathbf{O}, o \text{ is a positive literal}\}$ of positive observations, possible explanations for $\mathbf{F}$ are then computed using the specified rules in $\mathbf{T}$. More precisely, an *abduction problem* is described by $< \mathbf{T}, \mathbf{C}, \mathbf{O}, \mathbf{F} >$, where the objective is to find the 'best' explanation for $\mathbf{F}$. Here, an *explanation* for $\mathbf{F}$ is a set $\mathbf{E} \subseteq \mathbf{C}$ which taken together with the rules in $\mathbf{T}$ allows for inferring the propositions in $\mathbf{F}$ modus-ponens-style. Subset-minimal explanations are called *preferred* explanations. In case multiple preferred explanations for $\mathbf{F}$ are computed, then these are in competition with each other. Inspired by these models, in Section 3.4.5 we defined how the notion of competing alternative explanations can be captured in our IG-formalism.

Example 64. *Consider the following example, in which* $\mathbf{T}$ *contains the following causal rules:*

r_1: fire $\rightarrow_\text{c}$ smoke
r_2: smoke_machine $\rightarrow_\text{c}$ smoke;

Suppose that $\mathbf{O} = \mathbf{F} = \{\text{smoke}\}$ *and* $\mathbf{C} = \{\text{fire, smoke\_machine, smoke}\}$*, then* $\mathbf{E_1} = \{\text{fire}\}$*;* $\mathbf{E_2} = \{\text{smoke\_machine}\}$*;* $\mathbf{E_3} = \{\text{fire, smoke}\}$*;* $\mathbf{E_4} = \{\text{smoke\_machine, smoke}\}$ *are explanations for* $\mathbf{F}$*, where* $\mathbf{E_1}$ *and* $\mathbf{E_2}$ *are preferred explanations that are in competition with each other.* □

Another logic-based approach, proposed by Console and Dupré [1994], similarly only allows for abductive reasoning, yet also allows for performing abduction with strict abstractions instead of only with causal rules (see also Section 2.1.2). A graph-based instead of a logic-based model for automated diagnosis similar to the aforementioned approaches was proposed by Lucas [1998]. In comparison, our IG-formalism allows for performing both abductive and deductive inference, is graph-based and allows for reasoning using more than just causal rules.

In the approach of Shanahan [1989] both deductive and abductive inference can be performed (called prediction and explanation by Shanahan, respectively); however, the aim of this approach is different than ours in that it serves to reason about the time at which events occurred using an adaptation of Kowalski and Sergot's [1986] Event Calculus. In comparison to our IG-formalism, Shanahan's system is more restricted as we provide a more general account of reasoning with modelled causal (and other types of) knowledge. Poole's [1989] Theorist framework is a formal-logical model of causal-abductive reasoning that is extended with an additional component that allows for predictive reasoning. More specifically, after finding an explanation **E** for **F** for a specified abduction problem, a query can be performed to establish whether a given proposition can be predicted given **T** and **E**. This is illustrated through the following example.

Example 65. *Consider Example 64. Assume that additional causal rule* r_3: fire $\rightarrow_{\mathrm{C}}$ heat *is provided and suppose that we wish to query whether proposition* heat *can be predicted given* **T** *and the different computed explanations. For explanations* $\mathbf{E_1}$ = {fire} *and* $\mathbf{E_3}$ = {fire, smoke} *for* **F** = {smoke}, *the system returns* yes *as an answer, along with rule* r_3 *used in predicting* heat. *For explanations* $\mathbf{E_2}$ = {smoke_machine} *and* $\mathbf{E_4}$ = {smoke_machine, smoke} *for* **F** = {smoke}, *answer* no *is returned.* □

Compared to our IG-formalism, the formalisms of Shanahan [1989] and Poole [1989] are logic-based instead of graph-based. Moreover, these formalisms only allow for reasoning with causal rules and do not allow for most types of mixed deductive-abductive inference; complications with reasoning using both causal and evidential defaults as identified by Pearl [1988a] and complications with mixed deductive-abductive discussed in Section 2.1.4.1 are thus avoided.

In our IG-formalism, we distinguished between actual antecedents and enablers of causal generalisations, where for a causal generalisation only its actual antecedents and not its enablers express a cause for the consequent. In making this distinction, we were inspired by Ortiz Jr. [1999], who also makes the distinction between enablers and actual antecedents of causal rules in his system. Compared to our IG-formalism, the aim of this approach is different in that Ortiz Jr. aims to identify the (causal) role that an event played in some broad nexus of events instead of providing a way to reason with modelled causal (and other types of) knowledge.

Besides Lucas' [1998] graph-based approach, other graph-based formalisms for reasoning with causality information have been proposed, notably Pearl's [2009] causal diagrams. Pearl provides a framework for causal inference in which diagrams are queried to determine if the assumptions available are sufficient for identifying causal effects. Compared to our IG-formalism, the aim of this framework is different in that it serves to identify causality (similar to the aforementioned approach of Ortiz Jr. [1999]) instead of providing a way to reason with modelled causal (and other types of) knowledge. Furthermore, causal diagrams require probabilistic quantification to be queried, while IGs are qualitative.

8.1.2 Argumentation with causality and other information

In Chapter 4 we proposed an argumentation formalism based on IGs that allows for both deductive and abductive argumentation and which instantiates Dung's [1995] abstract approach. Earlier work by Bex [2011, 2015] is related, although only his integrated theory [Bex, 2015] is purely argumentation-based. The hybrid theory proposed by Bex [2011] is a formal account of reasoning about evidence which combines a formal-logical model of abductive reasoning (see Section 8.1.1) with a framework for structured argumentation based on the ASPIC$^+$ framework. In Bex' hybrid theory, abductive and deductive inference are used in constructing causal stories (see Section 1.2.3) and evidential arguments, where defeasible abstractions can be used to connect one or more events in a story to more abstract versions of these events. In Bex' hybrid theory, arguments and stories are completely separate entities with their own definitions related to conflict and evaluation. The theory does not allow for most types of mixed inference with causal and evidential generalisations and abstractions, and thus largely avoids the problems associated with mixed inference as identified by Pearl [1988a] and in this thesis (see Section 2.1.4). Moreover, our argumentation formalism based on IGs allows for the construction and evaluation of both deductive and abductive arguments.

Building on his hybrid theory, Bex proposed his integrated theory of causal and evidential arguments [Bex, 2015], which is based on the ASPIC$^+$ framework extended with a notion of alternative attack. In Bex' integrated theory, causal and evidential inference rules (comparable to causal and evidential generalisations) are defined and arguments are constructed by forward chaining such inference rules. Hence, Bex' integrated theory does not allow for performing abductive inference. The notion of alternative attack as defined in Chapter 4 is inspired by Bex' account. Bex also defines alternative attack in the context in which evidential inferences are performed to infer causes c_1 and c_2 from effect e. By contrast, we do not consider c_1 and c_2 to be competing alternative explanations of e in case deduction is performed using evidential generalisations $e \rightarrow_{\mathrm{e}} c_1$ and $e \rightarrow_{\mathrm{e}} c_2$; more specifically, in case two causal generalisations $c_1 \rightarrow_{\mathrm{c}} e$ and $c_2 \rightarrow_{\mathrm{c}} e$ are provided, only one cause (either c_1 or c_2) can be the usual cause of e and therefore only one of the causal generalisations can be replaced by an evidential generalisation (see also Section 2.1.3). Bex allows for the construction of arguments that violate Pearl's constraint, which in his framework are self-defeating arguments. In our formalisms we constrain deductive and abductive inference such that Pearl's constraint is never violated by constructing arguments, as we generally consider these inference patterns to be undesirable. Finally, satisfaction of rationality postulates was not proven by Bex.

Bench-Capon and Prakken [2006] introduced a logic for defeasible argumentation that is essentially a preliminary version of ASPIC$^+$ and offered a formalisation of Aristotle's practical syllogism. Their approach allows for reasoning about alternative goals and values to justify actions. While actual abduction (i.e. 'backward' inference with causal rules) is not performed using their approach, the process of generating alternative sub-goals and actions from goals can be considered akin to performing abduction. In formalising this syllogism, Bench-Capon and Prakken only consider the abductive nature of reasoning about desires on the basis of beliefs and goals, whereas we offer a general account of abductive (and deductive) argumentation. In defining alternative attack in Chapter 4, we were inspired by Bench-Capon and Prakken's definition, as well as Bex' [2015] definition, of the same concept.

The adaptive logic framework [Batens, 2007] offers a general framework for defeasible reasoning. Several forms of defeasible reasoning have been explicated in different types of adaptive logics in the framework, including defeasible deduction [Straßer, 2014] and abduction [Meheus and Batens, 2006]. In recent studies, it was shown that adaptive logics can be mapped to accounts of structured argumentation [Borg, 2020; Heyninck and Straßer, 2016]. In comparison to our proposed formalisms, adaptive logics that combine deductive and abductive reasoning have not been proposed and different types of information (e.g. causal and evidential information) are not distinguished in the proposed logics.

Booth and colleagues [2014] and Sakama [2018] studied abduction in the context of abstract argumentation [Dung, 1995], where they see abduction as the problem of finding changes to an AF $(\mathcal{A}, \mathcal{D})$ with the goal of explaining observations that are substantiated by making arguments accepted. More specifically, an observation translates into a set of arguments $X \subseteq \mathcal{A}$, where additions and removals of arguments and defeats from the AF serve to make the arguments in X accepted. Each set of changes that makes the arguments in X accepted then acts as an explanation for the observation. Booth and colleagues [2014] in addition show how their model of abduction in abstract argumentation can be instantiated with abductive logic programs [Kakas et al., 1993]. In comparison to our argumentation formalism, these approaches consider argumentation at the abstract and not at the structured level and only consider abduction and not deduction, where the instantiation of Booth and colleagues [2014] is not an approach for structured argumentation but a logical model of causal-abductive reasoning (see also Section 8.1.1).

8.1.3 Concluding remarks

In this section, we have compared our IG-formalism and our argumentation formalism based on IGs to formalisms from the literature. In comparison to our IG-formalism, formal-logical models of abductive reasoning discussed in Section 8.1.1 only allow for performing abduction and not deduction. Related formalisms that allow for both deduction and abduction with different types of information are logic-based instead of graph-based like our IG-formalism, and do not consider the constraints on performing inference we argue should be imposed (see Section 2.1.4). These formalisms are also in some respects less expressive than our IG-formalism as most only allow for performing inference using causal rules.

In comparison to our argumentation formalism based on IGs, the formalisms discussed in Section 8.1.1 do not directly allow for formal evaluation using com-

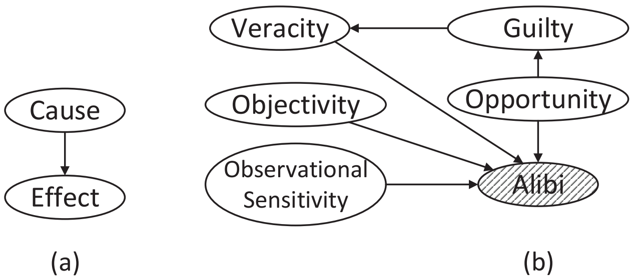


Figure 8.1: The cause-consequence idiom, adapted from Neil and colleagues [2000, p. 273] (a); the alibi idiom, adapted from Lagnado and colleagues [2013, p. 57] (b).

putational argumentation. Furthermore, most argumentation formalisms discussed in Section 8.1.2 do not allow for the construction and evaluation of both deductive and abductive arguments. In comparison, we have proposed an argumentation formalism that allows for the evaluation of deductive and abductive arguments in one unifying framework and that takes into account the constraints on performing inference we argue should be imposed. The closest to our argumentation formalism is Bex' [2015] integrated theory of causal and evidential arguments; the differences between his theory and our formalism were discussed in Section 8.1.2.

8.2 Bayesian network construction

To facilitate BN construction, construction methods have been proposed in the literature.

8.2.1 Constructing Bayesian networks from fragments

Throughout the literature, many (often domain-specific) BN fragments and modules, also called idioms, have been proposed, which capture generic patterns of frequently occurring types of knowledge and reasoning. In manual BN graph construction, instantiations of proposed idioms can be gradually incorporated in the BN under construction and can hence serve as building blocks to facilitate BN construction. Neil and colleagues [2000] developed idioms that are generally applicable to any domain; for instance, the cause-consequence idiom depicted in Figure 8.1a allows for straightforwardly modelling a causal process from cause(s) to effect, where the arc direction reflects this process. In the legal domain, Fenton and colleagues [2013], Lagnado and colleagues [2013] and Hepler and colleagues [2007] proposed BN fragments to model recurring patterns of legal reasoning. For instance, the alibi idiom by Lagnado and colleagues [2013] depicted in Figure 8.1b can be used in legal cases in which an alibi is provided by the defendant. In case an alibi is true, it rules out the opportunity for committing the crime, which in turn absolves the defendant of guilt; hence, nodes Alibi, Opportunity and Guilty and the arcs between them are

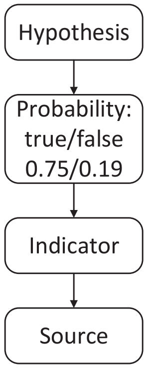


Figure 8.2: Module in an HMF, adapted from van Gosliga and van de Voorde [2008].

included in the idiom. Nodes Objectivity, Veracity and Observational Sensitivity are included to take into account the reliability of the provided alibi. The arc between nodes Guilty and Veracity is included to take into account that the defendant is more likely to lie in providing their alibi if they are guilty rather than if they are innocent. Laskey and Mahoney [1997] proposed several BN fragments in the domain of military situation assessment, and studied how these fragments can be combined to construct more complex networks.

Vlek and colleagues [2014] extended this work with BN idioms based on story schemes. More specifically, inspired by the idiom-based approaches of Fenton and colleagues [2013] and Lagnado and colleagues [2013] they proposed so-called narrative idioms for representing stories and their quality (see Section 1.2.3) in a BN, as well as a procedure for using stories to guide the construction of a BN. The proposed idioms can be used as building blocks in BN construction to represent alternative stories that serve as competing hypotheses about what may have happened, which are not necessarily jointly exhaustive. In the construction process, the network is annotated such that stories can later be extracted from the BN to form a report about the content of the BN.

To facilitate incremental BN construction, Object-Oriented BNs (OOBNs) were introduced by Koller and Pfeffer [1997] and later applied by Hepler and colleagues [2007] in the legal domain. With OOBNs, it becomes possible to incrementally construct a BN top-down, using fragments and modules such as proposed throughout the literature to gradually construct a network. Unlike our BN construction approach of Chapter 5, OOBNs do not provide an automated way of constructing BN graphs; instead, OOBNs allow experts to more quickly construct a BN manually by allowing recurrent patterns to be incorporated.

The concept of reusable network fragments was also the basis of Hypothesis Management Frameworks (HMFs) proposed by van Gosliga and van de Voorde [2008]. HMFs are network structures that allow for straightforwardly capturing knowledge regarding various hypotheses. HMFs are constructed in a modular way, where modules are of the form depicted in Figure 8.2. Hypothesis variables in an HMF can be supported or opposed by indicator variables, where source variables are used to express the reliability of sources related to an indicator. Instantiations of such modules can then be extended by adding alternative hypotheses and additional indicators. Probability nodes in the module do not represent variables, but are instead only used to textually display the conditional probabilities that the indicator is true given that the hypothesis is true and given that the hypothesis is false. Probability nodes are used as they supposedly enhance the users' understanding of an HMF and as they allow multiple experts to work on the same HMF simultaneously, delegating the different tasks involved in construction to different experts. Note that an HMF is not a BN, as probability nodes do not represent variables; instead, HMFs can be converted to BNs.

In contrast with the manual fragment-based approaches for BN graph construction discussed in this section, our approach from Chapter 5 allows for automatically constructing an initial BN graph from an IG that satisfies a number of desirable properties, for instance regarding the represented independence relation, where generalisations and conflicts can be incorporated and combined in an IG without having to conform to any predefined pattern or configuration.

8.2.2 Constructing Bayesian networks from arguments

In this subsection, approaches for constructing BNs from information specified in arguments are discussed. Bex and Renooij [2016] identified constraints on BNs given arguments constructed in ASPIC$^+$, based on the inferences on which arguments are built and the existing conflicts between arguments. These constraints suffice for constructing an undirected skeleton of a BN graph. However, for setting arc directions Bex and Renooij resort to using the commonly used notion of causality as guiding principle. The resulting BN graph then has to be verified and refined manually in terms of the independence relation it represents. In later work we refined this approach by specifying the directions in which arcs should be directed in BN graphs corresponding to structured arguments [Wieten et al., 2018a]. This was achieved by comparing the reasoning patterns captured by the BN graph to the reasoning captured by the original arguments. In comparison to the approach of Chapter 5, in which BN graphs are constructed from information specified in IGs, in these approaches ASPIC$^+$ is taken as a starting point for BN graph construction; for reasons mentioned in Section 3.2, we wish to refrain from using ASPIC$^+$ as an intermediary formalism in BN construction in this thesis.

Timmer and colleagues [2015] proposed a BN idiom (see Section 8.2.1) based on a specific type of evidential argument scheme, namely the argument scheme from position to know (see e.g. Walton et al. [2008]). Timmer [2017, Chapter 5] notes that this approach can possibly be extended to causal and other types of evidential argument schemes. By contrast, our approach of Chapter 5 can be generally applied to IGs including a mixture of both causal, evidential, abstractions and other types of generalisations, as well as generalisations that include enabling conditions.

8.2.3 Constructing Bayesian networks from ontologies

Work on the construction of BNs from information represented in ontologies [Uschold and Gruninger, 1996] is related to our BN construction approach based on IGs (Chapter 5). The approaches discussed in the current subsection differ from the approaches proposed by Yet and colleagues [2017], Helsper and van der Gaag [2007] and van der Gaag and Tabachneck-Schijf [2010] discussed in the introduction of Chapter 7 in that in the approaches currently under consideration the concepts and relations represented in an ontology are used for constructing a BN instead of for documenting the background knowledge behind a BN under construction.

The approach of Helsper and van der Gaag [2002] for constructing BNs from information represented in ontologies is fully manual. In this paper, they provide guidelines for manually incorporating the knowledge expressed in an ontology into a BN. Approaches for semi-automatically constructing BNs from information represented in ontologies have also been proposed [Fenz, 2012; Ramírez-Noriega et al., 2019]. To apply these approaches in practice, the problem under consideration first needs to be specified in the formal ontology language required as input. Informal sense-making tools such as mind maps as considered in this thesis similarly do not directly allow for guiding BN construction due to their informal nature, but are instead first formalised as IGs. In contrast with ontologies, these tools are used to capture inferences made with causal, evidential, abstractions and other types of information instead of with generic relations between concepts. In the approach of Fenz [2012], an initial BN graph is automatically constructed after a manual selection of relevant concepts and relations from an OWL ontology. Specifically, concepts are mapped to nodes in the BN graph and the direction of the relation between two concepts is used to direct arcs between corresponding nodes in the graph as a first heuristic. However, properties regarding the represented independence relation are not investigated; instead, Fenz notes that the obtained BN graph needs to be verified and refined manually by the BN modeller. Ramírez-Noriega and colleagues [2019] proposed a similar approach in the domain of intelligent tutoring systems, where the focus lies on obtaining the quantitative part of the BN. In this approach, the initial BN graph is semi-automatically constructed in a manner similar to the approach of Fenz [2012] by considering the concepts and relations in a given ontology. The CPTs of the BN under construction are then calculated by applying text mining approaches to the Wikipedia pages corresponding to the concepts under consideration, among other things by calculating the frequencies with which a given concept appears on the Wikipedia page of a related concept.

8.2.4 Concluding remarks

In this section we compared our approach for constructing BN graphs from IGs of Chapter 5 to BN construction methods from the literature. Compared to idiom-based approaches to BN graph construction (Section 8.2.1), in which instantiations of proposed fragments can be gradually incorporated in the BN under construction, our BN construction approach can be applied to IGs that are constructed without having to conform to any predefined pattern or configuration of generalisations and conflicts. Arguably this allows our BN construction approach to be applied more flexibly in practice, a claim that should be empirically evaluated in future work. In Sections 8.2.2 and 8.2.3 approaches for constructing BN graphs from information specified in arguments and ontologies were discussed. As mentioned earlier, we do not wish to use existing argumentation formalisms such as ASPIC$^+$ as a starting point in BN construction in this thesis, among other things because existing argumentation formalisms only allow for the construction of either deductive or abductive arguments and do not consider the constraints on performing inference we argue should be imposed (see also Sections 3.2 and 8.1.3). In contrast with ontologies, which capture generic relations between concepts, the IG-formalism is a formal account of reasoning about evidence that captures the interplay between the different types of inferences and generalisations; hence, the IG-formalism is more closely related to informal sense-making tools such as mind maps that are under consideration in this thesis. Moreover, compared to ontologies our IG-formalism allows for actually performing inference instead of only representing knowledge and reasoning. Accordingly, we wish to refrain from using ontologies as an intermediary formalism between analyses performed using sense-making tools and formal AI systems.

8.3 Probabilistic argumentation

In this thesis, we have focussed on facilitating the construction of formal representations within two types of AI systems, namely probabilistic models, more specifically BNs, and computational models of argument. In the literature, approaches to so-called *probabilistic argumentation* have also been proposed, which combine argumentation and probabilities. In this section, a number of different approaches to probabilistic argumentation are reviewed. We then conclude this section by comparing our IG-formalism to these approaches, after which we discuss ways in which the construction of formal representations within probabilistic argumentation systems can possibly be facilitated in future work.

Approaches to probabilistic argumentation can be distinguished based on whether they consider uncertainty to be in or about the arguments [Hunter, 2013]. In describing the general differences between these two approaches, we follow Prakken [2018b]. In the first type of approach, probabilities are intrinsic to an argument in that they express uncertainty concerning the truth of the argument's (ordinary) premises or the reliability of its inferences (e.g. Hunter [2013]; Prakken [2018b]; Hunter and Thimm [2017]; Hunter et al. [2020]); as noted by Prakken [2018b], this is arguably what Hunter [2013] calls the 'epistemic' approach to probabilistic argumentation. An example is default reasoning with probabilistic generalisations. In the second type of approach (which arguably is what Hunter [2013] calls the 'constellations' approach to probabilistic argumentation), probabilities are extrinsic to arguments and are used to express grades of uncertainty about whether (elements of) arguments or attacks and defeats are accepted as existing by some arguing agent (e.g. Li et al. [2012]; Dung and Thang [2010]; Rienstra [2012]). Prakken [2018b] provides the following example of extrinsic argument uncertainty, adapted from Hunter [2014]. Suppose that an enthymeme is posed that leaves two alternative premises implicit. A listener could then assign probabilities to these premises, which translate into probabilities on which argument the speaker intended to construct. This uncertainty is independent of the intrinsic strengths of the two possible arguments: one argument may be stronger than the other while the other is more likely to be the argument that the speaker had in mind.

8.3.1 Constellation approaches

We first consider constellations approaches to probabilistic argumentation in more detail. As discussed by Timmer [2017, Section 6.2], Dung and Thang [2010] proposed an argumentation framework for jury-based dispute resolution by incorporating probabilistic reasoning into AFs. A set of worlds is added to a given AF, where each world is a set of arguments. These possible worlds are shared among the jurors but each juror can assign their own probabilities to arguments in these worlds. In each world, an argument A then has a probability that it is accepted according to a given semantics (with respect to one juror), where the overall probability that A is accepted (according to that juror) is calculated by taking the sum of the probabilities in each world that A is accepted. Dung and Thang [2010] then instantiate their approach with assumption-based argumentation [Dung et al., 2009], where each world is a set of assumptions. Based on these assumptions arguments can be constructed in that world, where probabilities are defined on the assumptions. This work concerns the constellations approach to probabilistic argumentation as worlds may contain different AFs. Li and colleagues [2012] proposed a similar approach by associating probabilities with arguments and defeats in an AF, where these probabilities represent the likelihood of existence of an argument or a defeat. This results in a probability distribution over possible Dung frameworks. For a set of arguments, the probability that this set is a subset of a specific Dung extension can then be computed. This probability is used as the degree of justification of those arguments.

The incorporation of extrinsic uncertainty in structured arguments was investigated by Rienstra [2012], who proposed an instantiation of ASPIC$^+$ in which defeasible and strict inference rules are annotated with a number between 0 and 1 indicating the probability with which the rule is 'active' and may be applied in constructing an argument. This uncertainty may, for instance, be based on not knowing whether a given rule is a valid constituent of an argument. Rienstra then proposed a method to calculate the uncertainty in the existence of arguments, and subsequently uses these probabilities in calculating bounds on the probability that a given argument has a given dialectical status.

8.3.2 Epistemic approaches

Next, epistemic approaches to probabilistic argumentation are considered. These approaches can be distinguished based on whether they consider uncertainty at the abstract or at the structured level. Most approaches concern abstract argumentation. In the epistemic approach by Hunter [2013], probabilities associated with arguments in an AF represent the degree to which arguments are believed, in contrast to the approach of Li and colleagues [2012], in which probabilities represent the likelihood of existence of an argument. So-called epistemic extensions are then defined, subsets of arguments A for which $\Pr(A) > 0.5$. By adopting appropriate constraints on the probability distribution, it is then shown that the epistemic extensions correspond to complete, grounded, preferred, stable, or semi-stable extensions [Hunter and Thimm, 2017]. Building on this approach, probabilistic models of argumentation have been introduced that can also deal with incomplete and inconsistent probabilistic information [Hunter and Thimm, 2017]. The epistemic abstract approach has been generalised in recent work [Hunter et al., 2020], among other things to allow for modelling both defeat and support, context-sensitivity, and different perspectives.

An issue with the aforementioned epistemic approaches is that their abstract nature makes these approaches not easy to interpret. For instance, in these approaches there is unclarity about what the probability of an argument means, as in probability theory probabilities are assigned to the truth of statements or to outcomes of events, and an argument is in general neither a statement nor an event. Hence, it has been investigated how the probability of an argument can be specified in terms of its structure. Hunter [2013] instantiates his abstract epistemic approach with classical-logic argumentation, providing an account of epistemic structured probabilistic argumentation. In his approach, the strength of an argument is defined as the probability of the conjunction of all its premises. As noted by Prakken [2018b], while this makes sense when all argument are strict, it does not apply when arguments are constructed using defeasible inferences. Prakken [2018b] accordingly generalises Hunter's approach to arguments that apply defeasible inferences constructed in a simple instantiation of ASPIC$^+$, where the strength of an argument is defined as the probability of the conjunctions of all premises and conclusions of an argument. He then relates his account to abstract models of epistemic probabilistic argumentation [Hunter and Thimm, 2017]. An important idea behind Prakken's approach is that arguments implicitly make probabilistic independence assumptions, which implies that the probabilistic assumptions of conflicting arguments are jointly inconsistent. In contrast to the aforementioned epistemic approaches to structured argumentation, in which arguments are constructed from an available knowledge base and are then assigned probabilities, in the approach of Hunter [2020] structured arguments are instead constructed directly from a probability distribution. Methods are then proposed to select arguments and counterarguments to present in a Dung framework. This approach among other things allows for argument-based explanations of probability distributions.

Approaches to structured argumentation using alternatives to standard probability theory have also been proposed. Pollock [1995] formulated argument strength in terms of quantitative degrees of belief, where against Bayesian approaches he argued that degrees of belief and justification do not conform to the laws of probability theory. In his approach, defeasible inference rules are assigned a finite positive strength and strict rules are assigned infinite strength. The strength of arguments is then defined using a weakest-link principle, where an argument is at most as strong as its weakest sub-argument and its weakest rule. Argument strength is used to resolve attack into defeat in a manner compliant with Dung's [1995] abstract approach, where defeat is defined in an all-or-nothing manner in that defeaters weaker than their target cannot affect the status of their target. This approach is revisited by Verheij [2014] in the light of probability theory and classical logic. In later work, Pollock [2001] deviates from Dung's [1995] original approach by using rule strengths for defining degrees of argument justification, which allows defeaters to weaken the justification status of their stronger targets.

Verheij [2017] proposed an entirely different approach to all aforementioned approaches in which arguments, scenarios and probabilities are combined in a single framework. More specifically, in his approach arguments to and from different, mutually incompatible scenarios are viewed in the context of classical probability theory. Scenarios are then compared on their strength. An advantage of the formalism is that, in contrast to BNs, which require the specification of a full probability distribution, Verheij's formalism does not require more numbers than are available. In particular, it is not assumed that all arguments are assigned a probabilistic strength. The formalism allows for a qualitative and a quantitative interpretation, where the quantitative interpretation uses probability distributions and the qualitative interpretation uses total preorderings. As discussed by Verheij [2020], limitations of this approach include that it pays less attention to the internal structure of arguments than other argumentation approaches, that it does not study the roles of causal and evidential reasoning, and that it does not consider knowledge representation aspects.

The approaches discussed in Section 8.2.2 for constructing BNs from information specified in arguments do not combine argumentation and probabilities and therefore cannot be considered approaches to probabilistic argumentation in the same sense as the epistemic approaches discussed in this section. Instead, the problem of BN construction is considered, which is typically considered to be a difficult and error-prone process, where methods are proposed to facilitate that process.

8.3.3 Discussion

We now compare our IG-formalism to the approaches discussed in this section. Note that the IG-formalism is not an argumentation formalism and that we have currently opted not to account for probabilities in our IG-formalism as (numerical) probabilities are typically not indicated using sense-making tools. In comparing our IG-formalism to the approaches discussed in this section, we instead consider whether intrinsic (non-numeric) uncertainty is incorporated in IGs or whether there is uncertainty regarding the existence of an IG's elements. First, we note that the IG-formalism is designed to serve as an intermediary formalism between analyses performed using sense-making tools and formal AI systems, where the elements of IGs, similar to the elements of analyses performed using tools, are considered to be fixed. In particular, while the inferences represented in these analyses and that can be read from IGs may be uncertain, the existence of the elements represented in analyses and IGs is undisputed. Hence, work that concerns the constellations approach to probabilistic argumentation is irrelevant for current purposes.

Work that concerns the epistemic approach to probabilistic argumentation is relevant to us, as in IGs only intrinsic and not extrinsic uncertainty is incorporated, where this uncertainty concerns the reliability of its defeasible generalisations and inferences. As mentioned earlier, most epistemic approaches to probabilistic argumentation concern abstract argumentation; these are therefore irrelevant for current purposes (cf. Section 8.3.2 above and Prakken [2018b]). Epistemic approaches to structured probabilistic argumentation such as proposed by Hunter [2013] and Prakken [2018b] may serve as a source of inspiration for defining an epistemic approach to structured probabilistic argumentation based on IGs in future work. In particular, our IG-formalism may be extended with epistemic probabilities by allowing strengths of generalisations to be specified, which may then serve to guide the construction of formal representations within accounts of epistemic structured probabilistic argumentation. Compared to the approaches of Hunter [2013] and Prakken [2018b], which assign probabilistic strengths to deductive classical-logic arguments and ASPIC$^+$-style arguments, respectively, such an approach based on probabilistic IGs would allow probabilistic strengths to be assigned to both deductive and abductive arguments, where probabilistic arguments constructed using such an approach would adhere to the constraints on performing inference we argue should be imposed. In addition, such probabilistic IGs may also allow for deriving more probabilistic constraints on BNs.

8.4 Explaining Bayesian networks using argumentation

Approaches for explaining the reasoning patterns captured in BNs in terms of argumentation have been proposed [Vreeswijk, 2004; Keppens, 2012; Timmer et al., 2017], which are intended to allow domain experts who are not familiar with BNs but are accustomed to argumentation to understand the probabilistic reasoning captured in a BN. More specifically, these approaches can be used to summarise reasoning patterns from a given evidence set to a conclusion variable in a given BN in terms of argumentation. These approaches differ from the approach proposed in Chapter 7 for capturing and resolving disagreements about BN elements using argumentation in that they do not allow for any argumentative discussion about the construction of BNs. Furthermore, compared to the work on constructing BNs from information specified in arguments (Section 8.2.2), the approaches of Vreeswijk [2004], Keppens [2012] and Timmer and colleagues [2017] are in the reverse direction, namely from

BNs to arguments. In particular, Timmer and colleagues introduced support graphs as an intermediary formalism that captures general reasoning patterns represented by a BN for the purpose of explaining such patterns in terms of ASPIC$^+$-style arguments; hence, this work is in the direction opposite of the work by Bex and Renooij [2016] and Wieten and colleagues [2018a] on constructing BNs from information specified in ASPIC$^+$-style arguments. Similar to IGs, support graphs serve to guide the construction of arguments, but in contrast with IGs support graphs are designed to capture reasoning patterns represented by a BN while the IG-formalism is designed to model the process of reasoning about evidence.

Chapter 9

Conclusions and future research

The main motivation for this research has been to guide the construction of formal representations of evidential knowledge and inference by exploiting knowledge specified by domain experts about the domain using informal sense-making tools they are familiar with. More specifically, we have focussed on guiding the construction of formal representations within two types of AI systems, namely argumentation frameworks and BNs. In addition, we have studied how argumentation can be used to argue about a BN under construction instead of about the domain.

In the following sections, we restate the research questions from the introduction of this thesis and summarise our answers to these questions, after which we discuss possible avenues for future research.

9.1 Constructing formal representations within AI systems from informal sense-making tools

In this thesis we have addressed the following main research question:

Research question 1 How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of formal representations within AI systems?

To answer research question 1, we have studied examples of analyses performed using such tools, namely Wigmore charts and mind maps, in the light of our conceptual analysis of reasoning about evidence. We have observed that when performing analyses domain experts naturally mix deductive and abductive inference with the various types of generalisations distinguished in our conceptual analysis, where the used generalisations and the inference type are left implicit. Furthermore, the manner of conflict is typically not precisely specified and the assumptions of domain experts underlying their analyses are typically not explicitly stated; hence, we have concluded that informal sense-making tools do not directly allow for guiding the construction of formal representations within AI systems.

Accordingly, we have set out to formalise and disambiguate analyses performed using informal sense-making tools in a manner that (1) allows for guiding the construction of formal representations within AI systems and that (2) is in line with our conceptual analysis of reasoning about evidence, while (3) allowing inference to be performed and visualised in a manner that is closely related to the way in which inference is performed and visualised by domain experts using such tools. In particular, we have proposed the IG-formalism in Chapter 3, which formalises analyses performed using such tools in a manner that makes the used generalisations and conflicts explicit. Our IG-formalism is tailored to model the process of reasoning about evidence in that it provides a precise account of the interplay between deductive and abductive inference and causal, evidential, abstractions, and other types of generalisations. In particular, we haven taken into account important constraints we argue should be imposed on the types of inferences that may be performed with the different types of generalisations. In designing the IG-formalism, we have opted for a graph-based formalism instead of a logic-based formalism to remain closely related to the manner analyses are visualised using aforementioned graph-based tools as well as the BN-formalism. Finally, inspired by formal-logical models of abductive reasoning we have accounted for conflicts between competing alternative explanations.

Our IG-formalism can be used to guide the construction of formal representations within AI systems by serving as an intermediary formalism between analyses performed using informal sense-making tools and formalisms that allow for formal evaluation. In Chapters 4 and 5 we have demonstrated the use of our IG-formalism in guiding the construction of formal representations within two types of AI systems, namely computational models of argument and BNs. In the following two sections, we summarise our research regarding these two applications of our IG-formalism and thereby provide an answer to research question 1.

9.1.1 Guiding the construction of argumentation frameworks from information graphs

In Chapter 4 we have addressed the following subquestion of research question 1:

Research question 1a How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of argumentation frameworks?

We have answered this question by defining a framework for structured argumentation based on IGs that allows for the construction and evaluation of deductive and abductive arguments in one unifying framework. Arguments based on IGs are defined as sequences of deductive and abductive inference applications than can be read from the IG given the evidence. We have defined several types of attacks between arguments based on IGs, based on the different types of conflict considered in our IG-formalism. In particular, a new notion of attack, namely alternative attack, is defined between arguments whose conclusions are abductively inferred competing explanations, inspired by formal-logical models of abduction. Our approach generates a Dung-style AF which allows argument frameworks based on IGs to be formally evaluated. More specifically, while the initially constructed AF may be further specified if desired, for instance by supplementing it with preferences, it may also be directly evaluated using Dung's argumentation semantics. Hence, by formalising analyses performed by domain experts using informal sense-making tools as IGs as an intermediary step, this allows for the construction and evaluation of AFs on the basis of IGs.

Our approach is designed to ensure that AFs constructed on the basis of IGs automatically satisfy a number of desirable properties. We have proven that arguments constructed on the basis of IGs adhere to the constraints we argue should be imposed on performing inferences with different types of information. Moreover, we have shown that instantiations of our argumentation formalism satisfy key rationality postulates [Caminada and Amgoud, 2007], which are widely accepted as important desiderata for structured argumentation formalisms. Satisfaction of these postulates warrants instantiations of our argumentation formalism to be of good quality, as it implies that unintuitive and undesirable results regarding inconsistency and non-closure as identified by [Caminada and Amgoud, 2007] are avoided.

9.1.2 Guiding the construction of Bayesian networks from information graphs

In Chapter 5 we have considered BN construction and addressed the following subquestion of research question 1:

Research question 1b How can domain knowledge expressed by experts in analyses performed using informal sense-making tools be exploited to guide the construction of Bayesian networks?

To answer this question, we have investigated the application of our IG-formalism in guiding the construction of BN graphs. In manual BN graph construction, the notion of causality is typically used as a guiding principle in directing arcs [Fenton and Neil, 2012; Jensen and Nielsen, 2007]. Accordingly, we have proposed a BN graph construction approach that exploits a notion of causality as expressed by the knowledge captured in an IG, namely by its causal and evidential generalisations and conflicts. Our approach serves for automatically constructing a directed BN graph from an IG. Moreover, we have demonstrated that the inferences that can be read from an IG allow us to derive some qualitative probabilistic constraints on the BN under construction. These qualitative probabilistic constraints may serve as input for a subsequent elicitation procedure for obtaining a fully specified QPN or BN for (qualitative) probabilistic inference. Hence, our IG-formalism, together with our BN construction approach, allows us to construct an initial BN from a domain expert's initial analysis performed using an informal tool; it thereby facilitates the BN elicitation process.

Our BN construction approach is designed to ensure that initial BN graphs constructed from IGs automatically satisfy a number of desirable properties. In particular, we have formally proven that BN graphs constructed by our approach capture reasoning patterns similar to those that can be read from the original IG, and we have identified conditions on the IG under which the fully automatically constructed initial graph is guaranteed to be acyclic. Arguably, satisfaction of these properties partly simplifies the (manual) validation step involved in BN graph construction, in which it is verified that the initially constructed graph is acyclic and correctly captures the (conditional) independencies.

9.2 Exploiting arguments about Bayesian networks to facilitate their construction

As IGs only express qualitative and not quantitative (probabilistic) information, our BN construction approach from Chapter 5 can only serve for constructing a partially specified initial BN. Accordingly, we have investigated how the construction of BNs can be further facilitated. In constructing BNs in practice, disagreements about BN elements may arise among experts; as such disagreements are inherently argumentative, for research question 2 we have investigated how argumentation techniques can be used to express such disagreements and how constructed arguments about BNs may subsequently be exploited to facilitate their construction:

Research question 2 How can Bayesian network construction be facilitated by exploiting expert knowledge expressed as arguments about BN elements?

To answer this question, in Chapter 7 we have studied an actual disagreement about a forensic BN, where we have analysed where disagreements about BNs typically arise and how such disagreements are typically expressed and resolved manually by experts in practice. Based on our argument-based analysis of this disagreement, we have provided an answer to research question 2 by proposing a method that allows experts to explicitly express their reasons pro and con modelling decisions regarding the structure and parameterisation of a (fully or partially specified) BN using formal argumentation. Disagreements are then resolved as much as possible by utilising the dialectical status of the constructed arguments to derive probabilistic and structural constraints on the BN. Besides supporting experts in resolving conflicts about BNs, another important aspect of our method is that it allows for structurally documenting reasons pro and con BN modelling decisions, which can play a vital role in allowing experts to understand, use, and accept BNs.

While disagreements may also arise about analyses performed using informal sense-making tools or IGs, we have proposed a method that allows for capturing and resolving disagreements about BNs instead of about analyses or IGs, which allows our method to be applied to both BNs constructed from IGs and BNs otherwise constructed. Another advantage of specifying our method at the BN level is that it allows for argumentation about probabilities, as probabilities are typically not expressed using sense-making tools and are currently not accounted for in our IG-formalism (see also Section 8.3.3).

9.3 Future research

We now discuss ways in which the research presented in this thesis may be extended in future work. Note that possible ways in which the IG-formalism and our approaches for constructing argumentation frameworks and BNs from IGs may be extended, adjusted or evaluated in future work were already discussed in Chapters 3, 4, and 5, respectively. Similarly, future work regarding our method for capturing and resolving disagreements about BNs using argumentation proposed in Chapter 7 was already discussed in the conclusion of that chapter.

In Section 9.3.1 we discuss guiding the construction of formal representations within formal systems other than AFs and BNs. In Section 9.3.2 we discuss ways in which our construction approaches can be used to (theoretically) compare BNs and AFs in future work, which may help gain a better understanding of relations between argumentation and probabilistic approaches. In Section 9.3.3 we provide suggestions by which the construction of AFs and BNs can be further facilitated.

9.3.1 Guiding the construction of representations within other types of formal systems using information graphs

In this thesis, we have focussed on the construction of formal representations within two types of AI systems using IGs, namely computational models of argument and BNs. In future work, ways in which the construction of representations within other types of formal systems can be facilitated may be investigated. In Section 8.3 we have discussed guiding the construction of instantiations of probabilistic argumentation systems using (adjustments to) the IG-formalism. Other formal representations whose construction may be guided using IGs include instantiations of formal-logical models of abduction (see Section 8.1.1) and instantiations of Bex' [2011] hybrid theory (see Section 8.1.2). From a theoretical perspective, developing such construction approaches is interesting as it allows for comparing the way in which evaluation is performed using these different formal systems. More specifically, the similarities and differences obtained from evaluating different formal representations within these systems constructed from the same IG may be compared, which may then serve to gain a better understanding of relations between these systems.

From a practical perspective, the construction of formal representations within systems other than argumentation frameworks and BNs may be useful for specific application contexts. As discussed in Section 8.1.1, formal-logical models of abduction only allow for abduction and only allow for performing inference using causal rules. However, for small and pre-defined domains (for which these models were originally intended), these models may be adequate, as they allow for obtaining a relatively simple global overview of a case by providing different explanations of observed evidence. As noted by Bex [2011, p. 253], these approaches essentially model the story-based approach to reasoning about evidence and hence these models have the advantages and disadvantages of this approach (see Section 1.2.3). Another option is to construct instantiations of Bex' [2011] hybrid theory from IGs. An advantage of Bex' hybrid theory is that it combines the story-based approach to reasoning about evidence with a framework for structured argumentation, where stories are used to explain the evidence and to judge the global coherence of a case while arguments constructed from the evidence are used to attack or support specific elements of stories. Disadvantages of using Bex' hybrid theory compared to using our argumentation formalism based on IGs were discussed in Section 8.1.2.

9.3.2 Further investigating relations between arguments and BNs

The two main formal AI systems under consideration in this thesis are argumentation frameworks and BNs. In previous work, it was investigated how the central concepts of these two systems are connected to help gain a better understanding of relations between them; this includes work on deriving constraints on BNs given information specified in arguments (see Section 8.2.2) and work on explaining BNs using argumentation (see Section 8.4). In future work, this work may be extended upon, for instance by considering how the dialectical status of arguments or preferences over arguments may be used in deriving constraints on BNs, elements of argumentation not considered by the approaches discussed in Section 8.2.2. Another option is to apply the approach of Bex and Renooij [2016] for constructing BNs from information specified in arguments (see Section 8.2.2) and the argumentation-based explanation method for BNs of Timmer and colleagues [2017] (see Section 8.4) in sequence for a given ASPIC$^+$-style structured argumentation framework (see Section 2.2.2), where the differences between the original arguments used as input for BN construction and the arguments returned by explanation methods can be compared; this was preliminarily investigated by us in [Wieten et al., 2018a].

A way in which the approaches from Chapters 4 and 5 for constructing AFs and BNs from IGs can aid in theoretically investigating relations between BNs and AFs is by comparing the manner in which a BN and an AF constructed from the same IG using our approaches are formally evaluated (see also Section 9.3.1). As discussed in Section 5.5, the defeasible inferences that can be read from an IG given the evidence may possibly be interpreted as qualitative influences, which allows for constructing a fully specified QPN from an IG that can be used for performing qualitative probabilistic inference. Hence, this interpretation would allow us to compare the manner in which AFs constructed from IGs are evaluated using Dung's semantics to the way QPNs constructed from IGs are evaluated; in particular, it allows us to compare the dialectical status of conclusions to the signs that are assigned to variables upon performing qualitative probabilistic inference.

9.3.3 Further facilitating the construction of argumentation frameworks and BNs

In this thesis we have proposed approaches and methods that aid the construction of AFs and BNs. We now provide a number of suggestions by which the construction of AFs and BNs can be further facilitated. In our case study of Chapter 6 we constructed an IG corresponding to a Wigmore chart according to a number of general heuristics. For instance, in establishing which generalisations could have been used in constructing the chart we among other things determined whether sequences of described events could be interpreted as instances of story schemes (see Section 6.2). In future work general guidelines for IG construction may be formulated. For instance, a database of schemes that capture general patterns of defeasible reasoning (including argument and story schemes) may be composed, instantiations of which can be used as building blocks in facilitating IG construction. Such an approach would in turn facilitate AF and BN construction. In the context of BNs such an approach is comparable to the idiom-based approaches to BN construction discussed in Section 8.2.1.

In Chapter 6 we have also illustrated that depending on the modelling choices made in constructing an IG different representations within AI systems may be obtained upon applying our approaches from Chapters 4 and 5. From a practical perspective we do not consider this to be undesirable as our approaches serve for constructing an initial AF or BN that may be adjusted or further specified. From a theoretical perspective it may be investigated under which conditions local adjustments to the original IG result in the same AF or BN, or under which conditions a fragment of the obtained AF or BN is identical. Instead of considering how local adjustments to IGs can be incorporated in a BN under construction, we have proposed a method in Chapter 7 that allows domain experts to use argumentation to argue about the BN under construction instead of about the domain, where the dialectical status of the constructed arguments is used to derive constraints on the BN. By comparison, in the field of argumentation approaches that allow for meta-argumentation [Wooldridge et al., 2006] about AFs [Modgil and Bench-Capon, 2010] and ASPIC$^+$-style arguments [Müller et al., 2013] have been proposed that allow one to adjust one's AF or structured arguments, respectively. Our argumentation formalism based on IGs may similarly be extended in future work by allowing for meta-argumentation. For instance, as causality is a contentious topic, meta-argumentation about labels of generalisations used in constructing arguments may be considered. Similarly, reasoning about the validity of generalisations used in constructing arguments may be accounted for, as considered by e.g. Bex [2011].

9.4 Final remarks

In this thesis we have set out to bridge the gap between informal sense-making tools and formal systems by proposing approaches and methods that facilitate the construction of instantiations of such formal systems. While sense-making tools such as mind maps are suited for creating an initial sketch of a problem, they do not support experts in formally evaluating the problem. The approaches, formalisms, and methods proposed in this thesis are designed to aid the construction of BNs and AFs, thereby allowing experts to formally evaluate their problems in a probabilistic manner by computing probabilities of interest using BNs or to evaluate arguments on their acceptability using AFs.

To help increase the uptake and application of AFs and BNs in practice, it is important not only to consider facilitating their construction but also to consider how the *use* of AFs and BNs can be facilitated for domain experts unfamiliar with them by making instantiations of these systems more explainable. Explainable AI (XAI) has recently received much attention; however, most of this work is directed at explaining decisions of machine learning algorithms [Adadi and Berrada, 2018; Guidotti et al., 2018] instead of explaining instantiations of symbolic knowledge-based systems. In the context of BNs, approaches have been proposed that allow domain experts to understand the elements of a BN model or the reasoning patterns captured in BNs, for instance by explaining BNs using text, visualisations or arguments (see Sections 6.3 and 6.4 of Timmer [2017] for an overview). In the context of argumentation, approaches have been proposed that allow for explaining why a conclusion or argument is accepted under a given semantics using (elements of) arguments (see Borg and Bex [2021] for an overview) which may be used in the context of argumentation frameworks constructed on the basis of IGs to explain outcomes to domain experts.

While we focussed on facilitating the construction of instantiations of formal systems, the formalisms and approaches we proposed are designed in a way that hopefully makes them understandable for domain experts. For BNs documented using our approach from Chapter 7, the reasons behind BN modelling decisions may be returned to experts, which may aid them in better understanding BNs. Moreover, our IG-formalism is designed to allow inference to be performed and visualised in a way closely related to sense-making tools familiar to domain experts. Arguably, this makes it more straightforward for domain experts to understand, construct and use IGs rather than directly constructing and using instantiations of formal systems, a claim that should be empirically evaluated in future work.

Bibliography

A. Adadi and M. Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018.

L. Amgoud and C. Cayrol. A model of reasoning based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–215, 2002.

T. J. Anderson, D. A. Schum, and W. L. Twining. *Analysis of Evidence*. Cambridge University Press, 2nd edition, 2005.

P. Antal, T. Mészáros, B. de Moor, and T. Dobrowiecki. Annotated Bayesian networks: a tool to integrate textual and probabilistic medical knowledge. In A. Jacobs, editor, *Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems*, pages 177–182. IEEE, 2001.

D. Batens. A universal logic approach to adaptive logics. *Logica Universalis*, 1(1): 221–242, 2007.

M. Beirlaen, J. Heyninck, P. Pardo, and C. Straßer. Argument strength in formal argumentation. *Journal of Applied Logics - IfCoLog Journal of Logics and their Applications*, 5(3):629–676, 2018.

T. J. M. Bench-Capon and H. Prakken. Justifying actions by accruing arguments. In P. E. Dunne and T. J. M. Bench-Capon, editors, *Computational Models of Arguments: Proceedings of COMMA 2006*, pages 247–258. IOS Press, 2006.

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.

P. Besnard, A. Garcia, A. Hunter, S. Modgil, H. Prakken, G. Simari, and F. Toni. Introduction to structured argumentation. *Argument & Computation*, 5(1):1–4, 2014.

F. Bex. *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Springer, 2011.

F. Bex. An integrated theory of causal stories and evidential arguments. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law*, pages 13–22. ACM Press, 2015.

F. Bex and S. Renooij. From arguments to constraints on a Bayesian network. In P. Baroni, T. F. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument: Proceedings of COMMA 2016*, pages 95–106. IOS Press, 2016.

F. Bex and B. Verheij. Solving a murder case by asking critical questions: an approach to fact-finding in terms of argumentation and story schemes. *Argumentation*, 26:325–353, 2012.

F. Bex, H. Prakken, C. A. Reed, and D. Walton. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, 11(2-3):125–165, 2003.

F. Bex, S. Modgil, H. Prakken, and C. A. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, 23(5):951–989, 2013.

T. Boneh, A. E. Nicholson, and E. A. Sonenberg. Matilda: a visual tool for modeling with Bayesian networks. *International Journal of Intelligent Systems*, 21(11): 1127–1150, 2006.

R. Booth, D. Gabbay, S. Kaci, T. Rienstra, and L. van der Torre. Abduction and dialogical proof in argumentation and logic programming. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *Proceedings of the Twenty-First European Conference on Artificial Intelligence*, volume 263, pages 117–122. IOS Press, 2014.

A. Borg. Assumptive sequent-based argumentation. *Journal of Applied Logics*, 2631 (3):227–294, 2020.

A. Borg and F. Bex. A basic framework for explanations in argumentation. *IEEE Intelligent Systems*, 36(2):25–35, 2021.

S. J. Buckingham Shum. The roots of computer supported argument visualization. In P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, editors, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, pages 3–24. Springer, 2003.

M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, volume 4160 of *LNAI*, pages 111–123. Springer, 2006.

M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation. In L. Godo, editor, *Proceedings of the Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 3571, pages 378–389. Springer, 2005.

L. A. Chalaguine and A. Hunter. A persuasive chatbot using a crowd-sourced argument graph and concerns. In H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi, editors, *Computational Models of Argument. Proceedings of COMMA 2020*, pages 9–20. IOS Press, 2020.

P. W. Cheng and L. R. Novick. Causes versus enabling conditions. *Cognition*, 40: 83–120, 1990.

L. Console and D. T. Dupré. Abductive reasoning with abstraction axioms. In G. Lakemeyer and B. Nebel, editors, *Foundations of Knowledge Representation and Reasoning*, volume 810, pages 98–112. Springer, 1994.

L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 7(3):133–141, 1991.

F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.

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.

C. P. de Campos and F. Cozman. Belief updating and learning in semi-qualitative probabilistic networks. In F. Bacchus and T. Jaakkola, editors, *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 153–160. AUAI Press, 2005.

T. de Leeuw. Formulating guidelines for the manual construction of BN graphs in data-poor domains. Master's thesis, Master's in Artificial Intelligence, Utrecht University, The Netherlands, 2020.

A. M. Dingley. The ballpoint case: a Wigmorean analysis. In M. Malsch and J. F. Nijboer, editors, *Complex Cases: Perspectives on the Netherlands Criminal Justice System*, pages 179–188. Thela Thesis, 1999.

G. Doekhie. A Bayesian network for assigning probabilities on which finger left a mark. Master's thesis, Master's in Forensic Science, University of Amsterdam, The Netherlands, 2012.

V. Doshi. Validation of Bayesian networks with a case study on fingerprint general patterns. Master's thesis, Master's in Computing Science, Utrecht University, The Netherlands, 2013.

M. J. Druzdzel and M. Henrion. Intercausal reasoning with uninstantiated ancestor nodes. In D. Heckerman and E. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 317–325. Morgan Kaufmann, 1993.

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?" Guest editors' introduction. *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 P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari, editors, *Computational Models of Argument: Proceedings of COMMA 2010*, pages 171–182. IOS Press, 2010.

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.

P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175:457–486, 2011.

N. Fenton and M. Neil. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2012.

N. 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.

N. Fenton, M. Neil, and D. Berger. Bayes and the law. *Annual Review of Statistics and Its Application*, 3:51–77, 2016.

S. Fenz. An ontology-based approach for constructing Bayesian networks. *Data & Knowledge Engineering*, 73:73–88, 2012.

T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5):93:1–93:42, 2018.

R. Haraksim, D. Meuwly, G. Doekhie, P. Vergeer, and M. Sjerps. Assignment of the evidential value of a fingermark general pattern using a Bayesian network. In A. Brömme and C. Busch, editors, *Proceedings of the Twelfth International Conference of the Biometrics Special Interest Group*, volume 212, pages 99–109. GI, 2012.

E. M. Helsper and L. C. van der Gaag. Building Bayesian networks through ontologies. In F. van Harmelen, editor, *Proceedings of the Fifteenth European Conference on Artificial Intelligence*, volume 77, pages 680–684. IOS Press, 2002.

E. M. Helsper and L. C. van der Gaag. Ontologies for probabilistic networks: a case study in the oesophageal-cancer domain. *The Knowledge Engineering Review*, 22 (1):67–86, 2007.

A. B. Hepler, A. P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1-4):275–293, 2007.

J. Heyninck and C. Straßer. Relations between assumption-based approaches in nonmonotonic logic and formal argumentation. In G. Kern-Isberner and R. Wassermann, editors, *Proceedings of the Sixteenth International Workshop on Non-Monotonic Reasoning (NMR'16)*, pages 65–75. Technische Universität Dortmund, 2016.

A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.

A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55(2):607–638, 2014.

A. Hunter. Generating instantiated argument graphs from probabilistic information. In G. D. Giacomo, A. Catala, B. Dilkina, M. Milano, S. Barro, A. Bugarín, and J. Lang, editors, *Proceedings of the Twenty-Fourth European Conference on Artificial Intelligence*, volume 325, pages 769–776. IOS Press, 2020.

A. Hunter and M. Thimm. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, 59:565–611, 2017.

A. Hunter, S. Polberg, and M. Thimm. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *Artificial Intelligence*, 281:103236, 2020.

F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, 2nd edition, 2007.

J. R. Josephson. On the proof dynamics of inference to the best explanation. In M. MacCrimmon and P. Tillers, editors, *The Dynamics of Judicial Proof. Computation, Logic and Common Sense*, pages 287–306. Physica, 2002.

J. R. Josephson and S. G. Josephson. *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, 1994.

J. B. Kadane and D. A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. John Wiley & Sons Inc, 1996.

A. C. Kakas, R. Kowalski, and F. Toni. Abductive logic programming. *Journal of Logic and Computation*, 2(6):719–770, 1993.

H. A. Kautz. A formal theory of plan recognition and its implementation. In J. F. Allen, H. A. Kautz, R. N. Pelavin, and J. D. Tenenberg, editors, *Reasoning About Plans*, pages 69–124. Morgan Kaufmann, 1991.

J. Keppens. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law*, 20(2):109–143, 2012.

J. Keppens. On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning. *Artificial Intelligence and Law*, 22(3):239–290, 2014.

D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In D. Geiger and P. P. Shenoy, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 302–313. Morgan Kaufmann, 1997.

R. Kowalski and M. Sergot. A logic-based calculus of events. *New Generation Computing*, 4(5):67–94, 1986.

D. A. Lagnado, N. Fenton, and M. Neil. Legal idioms: a framework for evidential reasoning. *Argument & Computation*, 4(1):46–63, 2013.

K. B. Laskey and S. M. Mahoney. Network fragments: representing knowledge for constructing probabilistic models. In D. Geiger and P. P. Shenoy, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 334–341. Morgan Kaufmann, 1997.

H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In S. Modgil, N. Oren, and F. Toni, editors, *Theory and Applications of Formal Argumentation. First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, volume 7132, pages 1–16. Springer, 2012.

P. Lucas. Analysis of notions of diagnosis. *Artificial Intelligence*, 105(1-2):295–343, 1998.

J. Meheus and D. Batens. A formal logic for abductive reasoning. *Logic Journal of the IGPL*, 14(2):221–236, 2006.

S. Modgil and T. J. M. Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation*, 21(6):959–1003, 2010.

S. Modgil and H. Prakken. Resolutions in structured argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument: Proceedings of COMMA 2012*, pages 310–321. IOS Press, 2012.

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 & Computation*, 5(1):31–62, 2014.

S. Modgil and H. Prakken. Abstract rule-based argumentation. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, pages 286–361. College Publications, 2018.

S. Modgil, F. Toni, F. Bex, I. Bratko, C. I. Chesñevar, W. Dvořák, M. A. Falappa, X. Fan, S. A. Gaggl, A. J. García, M. P. Gonzalez, T. F. Gordon, J. Leite, M. Možina, C. A. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer, 2013.

J. Müller, A. Hunter, and P. Taylor. Meta-level argumentation with argument schemes. In W. Liu, V. S. Subrahmanian, and J. Wijsen, editors, *Scalable Uncertainty Management. SUM 2013*, volume 8078 of *LNCS*, pages 92–105. Springer, 2013.

M. Neil, N. Fenton, and L. Nielson. Building large-scale Bayesian networks. *The Knowledge Engineering Review*, 15(3):257–284, 2000.

M. Neil, N. Fenton, D. Lagnado, and R. D. Gill. Modelling competing legal arguments using Bayesian model comparison and averaging. *Artificial Intelligence and law*, 27(4):403–430, 2019.

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.

E. P. Nyberg, A. E. Nicholson, K. B. Korb, M. Wybrow, I. Zukerman, S. Mascaro, S. Thakur, A. O. Alvandi, J. Riley, R. Pearson, S. Morris, M. Herrmann, A. K. M. Azad, F. Bolger, U. Hahn, and D. Lagnado. BARD: a structured technique for group elicitation of Bayesian networks to support analytic reasoning. *arXiv preprint arXiv:2003.01207*, 2020.

D. Odekerken, A. Borg, and F. Bex. Estimating stability for efficient argument-based inquiry. In H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi, editors, *Computational Models of Argument. Proceedings of COMMA 2020*, pages 307–318. IOS Press, 2020.

A. Okada, S. J. Buckingham Shum, and T. Sherborne, editors. *Knowledge Cartography: Software Tools and Mapping Techniques*. Springer, 2nd edition, 2014.

C. L. Ortiz Jr. A commonsense language for reasoning about causation and rational action. *Artificial Intelligence*, 111(1-2):73–130, 1999.

J. Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35(2): 259–271, 1988a.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988b.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

N. Pennington and R. Hastie. Reasoning in explanation-based decision making. *Cognition*, 49(1-2):123–163, 1993.

P. Pirolli and S. K. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the 2005 International Conference on Intelligence Analysis*, 2005.

J. Pitchforth and K. Mengersen. A proposed validation framework for expert elicited Bayesian networks. *Expert Systems with Applications*, 40(1):162–167, 2013.

J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

J. L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, 1995.

J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133(1):233–282, 2001.

D. Poole. Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence*, 5(2):97–110, 1989.

D. Poole. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence*, 11(1-4):33–50, 1994.

H. Prakken. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1(2):93–124, 2010.

H. Prakken. Some reflections on two current trends in formal argumentation. In *Logic Programs, Norms and Action. Essays in Honour of Marek J. Sergot on the Occasion of his 60th Birthday*, pages 249–272. 2012.

H. Prakken. Historical overview of formal argumentation. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, pages 73–141. College Publications, 2018a.

H. Prakken. Probabilistic strength of arguments with structure. In M. Thielscher, F. Toni, and F. Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference (KR2018)*, pages 158–167. AAAI Press, 2018b.

H. Prakken. A new use case for argumentation support tools: supporting discussions of Bayesian analyses of complex criminal cases. *Artificial Intelligence and Law*, (28):27–49, 2020.

H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1-2):25–75, 1997.

H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In R. Goebel and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Springer, 2002.

H. Prakken, F. Bex, and A. R. Mackor. Editors' review and introduction: models of rational proof in criminal law. *Topics in Cognitive Science*, 12(4):1053–1067, 2020.

I. Rahwan and C. A. Reed. The argument interchange format. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 383–402. Springer, 2009.

A. Ramírez-Noriega, R. Juárez-Ramírez, J. J. Tapia, V. H. Castillo, and S. Jiménez. Construction of conditional probability tables of Bayesian networks using ontologies and Wikipedia. *Computación y Sistemas*, 23(4):1275–1289, 2019. doi:10.13053/CyS-23-4-2705.

C. A. Reed and G. Rowe. A pluralist approach to argument diagramming. *Law, Probability and Risk*, 6(1-4):59–85, 2007.

C. A. Reed, D. Walton, and F. Macagno. Argument diagramming in logic, law and artificial intelligence. *The Knowledge Engineering Review*, 22(1):1–22, 2007.

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132, 1980.

S. Renooij. *Qualitative Approaches to Quantifying Probabilistic Networks*. PhD thesis, Utrecht University, The Netherlands, 2001.

S. Renooij and L. C. van der Gaag. From qualitative to quantitative probabilistic networks. In A. Darwiche and N. Friedman, editors, *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pages 422–429. Morgan Kaufmann, 2002.

T. Rienstra. Towards a probabilistic Dung-style argumentation system. In S. Ossowski, F. Toni, and G. A. Vouros, editors, *Proceedings of the First International Conference on Agreement Technologies*, pages 138–152. CEUR, 2012.

C. Sakama. Abduction in argumentation frameworks. *Journal of Applied Non-Classical Logics*, 28(2-3):218–239, 2018.

M. Shanahan. Prediction is deduction but explanation is abduction. In N. S. Sridharan, editor, *Proceedings of International Joint Conference on Artificial Intelligence 89*, pages 1055–1060. Morgan Kaufmann, 1989.

G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157, 1992.

C. Straßer. *Adaptive Logics for Defeasible Reasoning. Applications in Argumentation, Normative Reasoning and Default Reasoning*, volume 38 of *Trends in Logic*. Springer, 2014.

F. Taroni, C. G. Aitken, P. Garbolino, and A. Biedermann. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. Wiley, 2014.

S. T. Timmer. *Designing and understanding Bayesian networks using argumentation*. PhD thesis, Utrecht University, The Netherlands, 2017.

S. T. Timmer, J.-J. C. 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*, volume 279, pages 173–176. IOS Press, 2015.

S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.

N. Timmers. The hybrid theory in practice: a case study at the Dutch police force. Master's thesis, Master's in Business Informatics, Utrecht University, The Netherlands, 2017.

M. Uschold and M. Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 11(2):93–136, 1996.

S. W. van den Braak. *Sensemaking software for crime analysis*. PhD thesis, Utrecht University, The Netherlands, 2010.

S. W. van den Braak, H. van Oostendorp, H. Prakken, and G. A. W. Vreeswijk. Representing narrative and testimonial knowledge in sense-making software for crime analysis. In E. Francesconi, G. Sartor, and D. Tiscornia, editors, *Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference*, pages 160–169. IOS Press, 2008.

L. C. van der Gaag and E. M. Helsper. Experiences with modelling issues in building probabilistic networks. In A. Gómez-Pérez and V. R. Benjamins, editors, *International Conference on Knowledge Engineering and Knowledge Management*, volume 2473, pages 21–26. Springer, 2002.

L. C. van der Gaag and J.-J. C. Meyer. The dynamics of probabilistic structural relevance. In L. C. van der Gaag and J.-J. C. Meyer, editors, *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC '96)*, pages 145–156, 1996.

L. C. van der Gaag and H. J. M. Tabachneck-Schijf. Library-style ontologies to support varying model views. *International Journal of Approximate Reasoning*, 51(2):196–208, 2010.

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, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 647–654. Morgan Kaufmann, 1999.

B. van Gijzel and H. Prakken. Relating Carneades with abstract argumentation via the ASPIC$^+$ framework for structured argumentation. *Argument & Computation*, 3(1):21–47, 2012.

S. P. van Gosliga and I. van de Voorde. Hypothesis management framework: a flexible design pattern for belief networks in decision support systems. In S. Renooij, H. J. M. Tabachneck-Schijf, and S. M. Mahoney, editors, *Proceedings of the Sixth Bayesian Modelling Applications Workshop at the Conference on Uncertainty in Artificial Intelligence*. CEUR, 2008.

R. van Hooff. Abductie als argumentatie. In Dutch (Abduction as argumentation). Master's thesis, Master's in Artificial Intelligence, Utrecht University, The Netherlands, 2004.

B. Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003.

B. Verheij. Arguments and their strength: revisiting Pollock's anti-probabilistic starting points. In S. Parsons, N. Oren, C. A. Reed, and F. Cerutti, editors, *Computational Models of Argument: Proceedings of COMMA 2014*, pages 433–444. IOS Press, 2014.

B. Verheij. Proof with and without probabilities: correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty. *Artificial Intelligence and Law*, 25:127–154, 2017.

B. Verheij. Analyzing the Simonshaven case with and without probabilities. *Topics in Cognitive Science*, 12(4):1175–1199, 2020.

B. Verheij, F. Bex, S. T. Timmer, C. S. Vlek, J.-J. C. Meyer, S. Renooij, and H. Prakken. Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, 15(1): 35–70, 2016.

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.

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. A. Reed, editors, *International Workshop on Argumentation in Multi-Agent Systems*, volume 3366, pages 111–129. Springer, 2004.

W. A. Wagenaar, P. J. van Koppen, and H. F. M. Crombag. *Anchored Narratives: The Psychology of Criminal Evidence*. Harvester Wheatsheaf, 1993.

D. Walton, C. A. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.

M. P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3):257–303, 1990.

R. Wieten, J. de Zoete, B. Blankers, and B. Kokshoorn. The interpretation of traces found on adhesive tapes. *Law, Probability and Risk*, 14(1):305–322, 2015.

R. Wieten, F. Bex, L. C. van der Gaag, H. Prakken, and S. Renooij. Refining a heuristic for constructing Bayesian networks from structured arguments. In B. Verheij and M. Wiering, editors, *Artificial Intelligence. BNAIC 2017. Communications in Computer and Information Science*, volume 823 of *LNCS*, pages 32–45. Springer, 2018a.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Exploiting causality in constructing Bayesian networks from legal arguments. In M. Palmirani, editor, *Legal Knowledge and Information Systems. JURIX 2018: The Thirty-First Annual Conference*, volume 313, pages 151–160. IOS Press, 2018b.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Supporting discussions about forensic Bayesian networks using argumentation. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, pages 143–152. ACM Press, 2019a.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Constructing Bayesian network graphs from labeled arguments. In G. Kern-Isberner and Z. Ognjanović, editors, *Proceedings of the Fifteenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, volume 11726 of *LNAI*, pages 99–110. Springer, 2019b.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Deductive and abductive reasoning with causal and evidential information. In H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi, editors, *Computational Models of Argument. Proceedings of COMMA 2020*, pages 383–394. IOS Press, 2020.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Information graphs and their use for Bayesian network graph construction. *International Journal of Approximate Reasoning*, 136:249–280, 2021a.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Deductive and abductive argumentation based on information graphs. Manuscript accepted for publication in *Argument & Computation*, 2021b. doi:10.3233/AAC-200539.

J. H. Wigmore. *The Principles of Judicial Proof*. Little, Brown and Company, 1913.

M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems: Second International Workshop, ArgMAS 2005*, volume 4049 of *LNAI*, pages 42–56. Springer, 2006.

B. Yet, Z. B. Perkins, N. R. M. Tai, and D. W. R. Marsh. Clinical evidence framework for Bayesian networks. *Knowledge and Information Systems*, 50(1):117–143, 2017.

Samenvatting

Redeneren met onzekerheid en bewijs speelt een belangrijke rol bij het nemen van beslissingen en het oplossen van problemen in vele domeinen, waaronder geneeskunde, engineering, forensisch onderzoek en het recht. Er bestaan verscheidene technieken om domeinexperts te ondersteunen in het uitvoeren van hun taken en te helpen een probleem te begrijpen. Voorbeelden zijn informele graafgebaseerde *sense-making* tools zoals mindmaps, argumentatiediagrammen en Wigmore-diagrammen, waarmee de gebruiker het probleem en de gemaakte redeneerstappen in het oplossen ervan kan structureren en visualiseren. Een beperking van deze tools is dat ze alleen bedoeld zijn voor het visualiseren van de redenering van de gebruiker: ze voorzien niet in het uitvoeren van berekeningen of geautomatiseerd redeneren met de gevisualiseerde informatie. Dus hoewel deze tools geschikt zijn om een initiële schets van het probleem te maken, ondersteunen deze tools experts niet in het formeel evalueren van het probleem. Formele systemen voor redeneren met bewijs zijn voorgesteld in het vakgebied kunstmatige intelligentie (KI), waaronder argumentatieformalismen en kansmodellen zoals Bayesiaanse netwerken (BNs). Deze systemen maken geautomatiseerd redeneren en berekenen mogelijk en ondersteunen experts daarmee wél in het formeel evalueren van een probleem. In de praktijk worden zowel *sense-making* tools als formele representaties binnen KI-systemen gebruikt door domeinexperts, aangezien beiden hun voor- en nadelen hebben.

Formele systemen zijn, in tegenstelling tot bovengenoemde *sense-making* tools, nauwkeurig gedefinieerd wat betreft hun notatie en semantiek en maken daarmee automatische evaluatie en berekeningen mogelijk. De werking van formele systemen is algemeen bekend en de eigenschappen van specifieke toepassingen van deze systemen kunnen worden bestudeerd. Domeinexperts hebben echter doorgaans niet de expertise om formele representaties van een probleem in een KI-systeem te bouwen. Vooral in data-arme domeinen moet hun constructie daarom meestal handmatig worden gedaan door een KI-expert via een kenniselicitatieprocedure in overleg met de domeinexpert, wat een moeilijk en foutgevoelig proces is.

In dit proefschrift willen we de constructie van modelleringen in een KI-systeem ondersteunen zodat domeinexperts hun problemen automatisch kunnen evalueren. Hiertoe bestuderen we hoe domeinkennis die is vastgelegd met behulp van een *sense-*

making tool kan worden gebruikt om de constructie van formele representaties binnen KI-systemen te ondersteunen. We richten ons op twee soorten formele systemen, namelijk kansmodellen, in het bijzonder BNs, en argumentatieraamwerken. Argumentatie is geschikt voor domeinen zoals het recht, waar argumenten voor en tegen beweringen worden opgebouwd vanuit bewijs. Argumenten kunnen dan formeel worden geëvalueerd op hun aanvaardbaarheid. Kansmodellen zoals BNs maken redeneren op basis van statistische en probabilistische informatie mogelijk, waardoor experts hun probleem kunnen evalueren aan de hand van kansberekeningen.

Dit proefschrift behandelt de volgende hoofdonderzoeksvraag:

Onderzoeksvraag 1 Hoe kan domeinkennis uitgedrukt door experts met informele *sense-making* tools worden gebruikt om de constructie van formele representaties binnen KI-systemen te ondersteunen?

Meer specifiek zullen we deze onderzoeksvraag beantwoorden voor de twee genoemde KI-systemen, namelijk argumentatieraamwerken (onderzoeksvraag 1a) en BNs (onderzoeksvraag 1b). Om onderzoeksvraag 1 te beantwoorden hebben we voorbeelden van analyses die gemaakt zijn met behulp van dergelijke *sense-making* tools bestudeerd. We hebben geconstateerd dat domeinexperts van nature verschillende typen informatie en gevolgtrekkingen (deductieve en abductieve) combineren maar dit niet expliciet maken. De gespecificeerde kennis is daarmee niet precies genoeg om direct gebruikt te kunnen worden ter ondersteuning van de constructie van formele representaties binnen KI-systemen.

Om de kloof tussen informele *sense-making* tools en formele systemen te overbruggen introduceren we het *informatiegraafformalisme* (IG-formalisme) waarin analyses uitgevoerd met dergelijke tools gerepresenteerd kunnen worden op een manier die de gebruikte informatie en conflicten expliciet maakt. Ons IG-formalisme is speciaal ontworpen om het proces van redeneren over bewijs te modelleren, door het expliciet maken van de wisselwerking tussen deductieve en abductieve gevolgtrekkingen en verschillende typen informatie. Bij het ontwerpen van het IG-formalisme hebben we gekozen voor een graafgebaseerd formalisme in plaats van een op logica gebaseerd formalisme om nauw verwant te blijven aan de manier waarop analyses worden gevisualiseerd met behulp van bovengenoemde graafgebaseerde tools.

Door analyses die zijn uitgevoerd door domeinexperts met behulp van informele *sense-making* tools te formaliseren als IGs, kan de IG dienen als tussenstap in de constructie van formele KI-systemen. We gebruiken de IGs dan ook om onderzoeksvraag 1 te beantwoorden. We hebben onderzoeksvraag 1a beantwoord door een raamwerk voor gestructureerde argumentatie op basis van IGs te definiëren dat de constructie en evaluatie van deductieve en abductieve argumenten in één verenigd raamwerk mogelijk maakt. Argumenten op basis van IGs worden gedefinieerd als reeksen van deductieve en abductieve gevolgtrekkingstoepassingen die kunnen worden afgelezen uit de IG op basis van het bewijs. We hebben verschillende soorten aanvallen gedefinieerd tussen argumenten op basis van IGs, gebaseerd op de verschillende soorten conflicten die we in ons IG-formalisme onderscheiden. Onze methode genereert een argumentatieraamwerk en maakt het daarmee mogelijk dat argumenten op basis van IGs formeel geëvalueerd kunnen worden met behulp van computationele argumentatie.

Onze methode is ontworpen om te garanderen dat argumentatieraamwerken die zijn geconstrueerd op basis van IGs automatisch voldoen aan een aantal gewenste eigenschappen. In het bijzonder hebben we aangetoond dat instantiaties van ons argumentatieformalisme voldoen aan belangrijke rationaliteitspostulaten, wat garandeert dat instantiaties van ons argumentatieformalisme van goede kwaliteit zijn omdat tegenintuïtieve resultaten en inconsistenties worden vermeden.

Om onderzoeksvraag 1b te beantwoorden hebben we onderzocht hoe ons IG-formalisme kan worden gebruikt om de constructie van BNs te ondersteunen. Een BN bestaat uit een gerichte graaf die de probabilistische onafhankelijkheidsrelatie tussen voor het domein relevante variabelen vastlegt, en een verzameling voorwaardelijke kansverdelingen die tezamen een gezamenlijke kansverdeling beschrijven. In de handmatige constructie van BN-grafen wordt het begrip causaliteit vaak gebruikt als leidraad voor het richten van pijlen. Om deze reden hebben we een methode voor de constructie van BN-grafen voorgesteld die gebruikmaakt van een notie van causaliteit zoals uitgedrukt door de kennis die is vastgelegd in een IG. Onze methode kan worden gebruikt om automatisch een gerichte BN-graaf uit een IG te construeren. Bovendien hebben we laten zien dat de gevolgtrekkingen die uit een IG kunnen worden afgelezen gegeven het bewijs ons in staat stellen enkele randvoorwaarden af te leiden voor de kansverdeling behorende bij de te construeren BN. Deze randvoorwaarden kunnen als input dienen tijdens het eliciteren van de kansen voor het verkrijgen van een volledig gespecificeerde BN. Daarom stelt ons IG-formalisme, samen met onze BN-constructiemethode, ons in staat om een initiële BN te construeren op basis van de initiële analyse van een domeinexpert uitgevoerd met een informele tool; het ondersteunt daarmee het BN-elicitatieproces.

Onze BN-constructiemethode is ontworpen om te garanderen dat initiële BN-grafen die zijn geconstrueerd op basis van IGs automatisch voldoen aan een aantal gewenste eigenschappen. We hebben aangetoond dat BN-grafen die door onze methode zijn geconstrueerd redeneerpatronen vangen die soortgelijk zijn aan degene die kunnen worden afgelezen uit de oorspronkelijke IG. We hebben bovendien voorwaarden aan IGs geïdentificeerd waaronder de volledig automatisch geconstrueerde initiële graaf gegarandeerd acyclisch is.

Aangezien IGs geen kwantitatieve informatie uitdrukken kan onze BN-constructiemethode alleen worden gebruikt om een gedeeltelijk gespecificeerde initiële BN te construeren. Daarom hebben we onderzocht hoe de constructie van BNs verder kan worden ondersteund. Bij het construeren van BNs kan in de praktijk onenigheid ontstaan onder de betrokken experts over de gemaakte keuzes. Aangezien dergelijke onenigheden argumentatief van aard zijn hebben we voor onderzoeksvraag 2 onderzocht hoe argumentatietechnieken kunnen worden gebruikt om dergelijke onenigheden uit te drukken en hoe argumenten over BNs vervolgens gebruikt kunnen worden om hun constructie te ondersteunen:

Onderzoeksvraag 2 Hoe kan de constructie van BNs worden ondersteund door gebruik te maken van expertkennis uitgedrukt als argumenten?

Om deze vraag te beantwoorden hebben we een daadwerkelijke discussie tussen experts over een forensisch BN bestudeerd, waarbij we hebben geanalyseerd waarover onenigheid kan ontstaan en hoe een dergelijke onenigheid in de praktijk wordt uitgedrukt en opgelost. Op basis van onze op argumenten gebaseerde analyse van deze onenigheid hebben we een antwoord gegeven op onderzoeksvraag 2 door een methode voor te stellen die experts in staat stelt expliciet hun redenen voor en tegen modelleringsbeslissingen met betrekking tot de structuur en parameters van een (geheel of gedeeltelijk gespecificeerde) BN met behulp van formele argumentatie vast te leggen. Onenigheden worden dan zoveel mogelijk opgelost door formele evaluatie van de geconstrueerde argumenten. Naast het ondersteunen van experts bij het oplossen van conflicten over BNs is een ander belangrijk aspect van onze methode dat het experts in staat stelt redenen voor en tegen BN-modelleringsbeslissingen te documenteren, wat cruciaal is voor het begrip, gebruik en de acceptatie van BNs.

Acknowledgements

First and foremost, I would like to express my gratitude towards my supervisors. Henry, your feedback never failed to be illuminating. I always left our meetings thinking about the questions you raised, and your suggestions have been invaluable to the completion of this thesis. Floris, I would like to thank you for always taking the time to discuss my work and ideas with me. In particular, I would like to thank you for providing helpful feedback on how to better motivate my work, including its social and scientific relevance. Silja, I would like to thank you for your personal approach to supervising. I am grateful that you were always available to discuss even the smallest of details with me. I would also like to thank Linda for her supervision during the first year of my project and for teaching me how to write more concisely.

I would like to thank the members of my assessment committee - Jan Broersen, Dragan Doder, Anthony Hunter, Martin Neil, Marjan Sjerps, and Bart Verheij - for taking the time to read and evaluate my thesis and for providing valuable feedback.

It would be remiss of me if I did not thank my colleagues from the Intelligent Systems group and the National Police Lab AI for the many interesting research colloquia, the entertaining and fruitful discussions we had over lunch and during coffee breaks, and the fun we had during barbecues, borrels and board game nights. In particular, I would like to thank my office mates Bas and Marijn, with whom I shared many laughs, stories and insights, as well as AnneMarie, who in addition to that provided important commentary on the argumentation part of this thesis.

I would like to thank my dear friends and family, who helped me relax and take my mind of work. I especially would like to thank Kenny, Walter, Leonie, Tim, Tom, Maxim, Maarten and Feben. A special thanks to Freddy, who designed the beautiful cover of this thesis while also planning his own wedding. It could not have been easy! I would like to thank Jan and Remon of my band, Photo Morgana, for their musical companionship. Also, a shout-out to my forensic science friends George, Iris and Guillem: finally, our group of four will be ‘doctors only’! I am very grateful to my parents Tonnie and Marinus, as well as to my in-laws Anny and Klaas, for supporting me unconditionally and for ensuring me that everything will be fine in the end. Finally, I am most grateful and thankful towards Judith. Especially in the final year of the project when the going got rough, you always remained patient and understanding. Now we can close this chapter and the three of us can start building our future together.

Remi

Kampen, July 2021

Curriculum vitae

Education

2016 - 2020
Utrecht University: PhD, direction: Artificial Intelligence

2012 - 2015
University of Amsterdam: Master's in Forensic Science, cum laude

2008 - 2012
Radboud University: Bachelor's in Mathematics

2002 - 2008
Almere College, Kampen: VWO

Experience

Sep 2015 - Nov 2016
EY
Advisor Forensic Technology & Discovery Services

Feb 2014 - Aug 2014
Netherlands Forensic Institute
Research Intern: The Interpretation of Traces Found on Adhesive Tapes

SIKS dissertation series

2011

01	Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*
02	Nick Tinnemeier (UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*
03	Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems*
04	Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference*
05	Bas van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.*
06	Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage*
07	Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction*
08	Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues*
09	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 (VU), *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 (VU), *Analysis and Validation of Models for Trust Dynamics*

26 Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*

27 Aniel Bhulai (VU), *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 (VU), *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 (VU), *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

01 Terry Kakeeto (UvT), *Relationship Marketing for SMEs in Uganda*

02 Muhammad Umair (VU), *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models*

03 Adam Vanya (VU), *Supporting Architecture Evolution by Mining Software Repositories*

04 Jurriaan Souer (UU), *Development of Content Management System-based Web Applications*

05 Marijn Plomp (UU), *Maturing Interorganisational Information Systems*

06 Wolfgang Reinhardt (OU), *Awareness Support for Knowledge Workers in Research Networks*

07 Rianne van Lambalgen (VU), *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*

08 Gerben de Vries (UVA), *Kernel Methods for Vessel Trajectories*

09 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 (VU), *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.*

16 Fiemke Both (VU), *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 (VU), *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 (VU), *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 (VU), *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 Withdrawn

44 Anna Tordai (VU), *On Combining Alignment Techniques*

45 Benedikt Kratz (UvT), *A Model and Language for Business-aware Transactions*

46 Simon Carter (UVA), *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation*

47 Manos Tsagkias (UVA), *Mining Social Media: Tracking Content and Predicting Behavior*

48 Jorn Bakker (TUE), *Handling Abrupt Changes in Evolving Time-series Data*

49 Michael Kaisers (UM), *Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions*

50 Steven van Kervel (TUD), *Ontology driven Enterprise Information Systems Engineering*

51 Jeroen de Jong (TUD), *Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching*

2013

01 Viorel Milea (EUR), *News Analytics for Financial Decision Support*

02 Erietta Liarou (CWI), *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing*

03 Szymon Klarman (VU), *Reasoning with Contexts in Description Logics*

04 Chetan Yadati (TUD), *Coordinating autonomous planning and scheduling*

05 Dulce Pumareja (UT), *Groupware Requirements Evolutions Patterns*

06 Romulo Goncalves (CWI), *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience*

07 Giel van Lankveld (UvT), *Quantifying Individual Player Differences*

08 Robbert-Jan Merk (VU), *Making enemies: cognitive modeling for opponent agents in fighter pilot simulators*

09 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 (VU), *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*

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 (VU), *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 (VU), *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

01 Nicola Barile (UU), *Studies in Learning Monotone Models from Data*

02 Fiona Tuliyano (RUN), *Combining System Dynamics with a Domain Modeling Method*

03 Sergio Raul Duarte Torres (UT), *Information Retrieval for Children: Search Behavior and Solutions*

04 Hanna Jochmann-Mannak (UT), *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation*

05 Jurriaan van Reijsen (UU), *Knowledge Perspectives on Advancing Dynamic Capability*

06 Damian Tamburri (VU), *Supporting Networked Software Development*

07 Arya Adriansyah (TUE), *Aligning Observed and Modeled Behavior*

08 Samur Araujo (TUD), *Data Integration over Distributed and Heterogeneous Data Endpoints*

09 Philip Jackson (UvT), *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*

10 Ivan Salvador Razo Zapata (VU), *Service Value Networks*

11 Janneke van der Zwaan (TUD), *An Empathic Virtual Buddy for Social Support*

12 Willem van Willigen (VU), *Look Ma, No Hands: Aspects of Autonomous Vehicle Control*

13 Arlette van Wissen (VU), *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 (VU), *Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare*

16 Krystyna Milian (VU), *Supporting trial recruitment and design by automatically interpreting eligibility criteria*

17 Kathrin Dentler (VU), *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 (VU), *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 (VU), *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 (VU), *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. Met als ondertitel: 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

01 Niels Netten (UvA), *Machine Learning for Relevance of Information in Crisis Response*

02 Faiza Bukhsh (UvT), *Smart auditing: Innovative Compliance Checking in Customs Controls*

03 Twan van Laarhoven (RUN), *Machine learning for network data*

04 Howard Spoelstra (OUN), *Collaborations in Open Learning Environments*

05 Christoph Bösch (UT), *Cryptographically Enforced Search Pattern Hiding*

06 Farideh Heidari (TUD), *Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes*

07 Maria-Hendrike Peetz (UvA), *Time-Aware Online Reputation Analysis*

08 Jie Jiang (TUD), *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions*

09 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 (VU), *Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks*

13 Giuseppe Procaccianti (VU), *Energy-Efficient Software*

14 Bart van Straalen (UT), *A cognitive approach to modeling bad news conversations*

15 Klaas Andries de Graaf (VU), *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 Lois 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 (VU), *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

01 Syed Saiden Abbas (RUN), *Recognition of Shapes by Humans and Machines*

02 Michiel Christiaan Meulendijk (UU), *Optimizing medication reviews through decision support: prescribing a better pill to swallow*

03 Maya Sappelli (RUN), *Knowledge Work in Context: User Centered Knowledge Worker Support*

04 Laurens Rietveld (VU), *Publishing and Consuming Linked Data*

05 Evgeny Sherkhonov (UVA), *Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers*

06 Michel Wilson (TUD), *Robust scheduling in an uncertain environment*

07 Jeroen de Man (VU), *Measuring and modeling negative emotions for virtual training*

08 Matje van de Camp (TiU), *A Link to the Past: Constructing Historical Social Networks from Unstructured Data*

09 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 (Tue), *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 (TUe), *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), *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), *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

01 Jan-Jaap Oerlemans (UL), *Investigating Cybercrime*

02 Sjoerd Timmer (UU), *Designing and Understanding Forensic Bayesian Networks using Argumentation*

03 Daniël Harold Telgen (UU), *Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines*

04 Mrunal Gawade (CWI), *Multi-core Parallelism in a Column-store*

05 Mahdieh Shadi (UVA), *Collaboration Behavior*

06 Damir Vandic (EUR), *Intelligent Information Systems for Web Product Search*

07 Roel Bertens (UU), *Insight in Information: from Abstract to Anomaly*

08 Rob Konijn (VU), *Detecting Interesting Differences: Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery*

09 Dong Nguyen (UT), *Text as Social and Cultural Data: A Computational Perspective on Variation in Text*

10 Robby van Delden (UT), *(Steering) Interactive Play Behavior*

11 Florian Kunneman (RUN), *Modelling patterns of time and emotion in Twitter #anticipointment*

12 Sander Leemans (TUE), *Robust Process Mining with Guarantees*

13 Gijs Huisman (UT), *Social Touch Technology - Extending the reach of social touch through haptic technology*

14 Shoshannah Tekofsky (UvT), *You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior*

15 Peter Berck (RUN), *Memory-Based Text Correction*

16 Aleksandr Chuklin (UVA), *Understanding and Modeling Users of Modern Search Engines*

17 Daniel Dimov (UL), *Crowdsourced Online Dispute Resolution*

18 Ridho Reinanda (UVA), *Entity Associations for Search*

19 Jeroen Vuurens (UT), *Proximity of Terms, Texts and Semantic Vectors in Information Retrieval*

20 Mohammadbashir Sedighi (TUD), *Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility*

21 Jeroen Linssen (UT), *Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)*

22 Sara Magliacane (VU), *Logics for causal inference under uncertainty*

23 David Graus (UVA), *Entities of Interest — Discovery in Digital Traces*

24 Chang Wang (TUD), *Use of Affordances for Efficient Robot Learning*

25 Veruska Zamborlini (VU), *Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search*

26 Merel Jung (UT), *Socially intelligent robots that understand and respond to human touch*

27 Michiel Joosse (UT), *Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors*

28 John Klein (VU), *Architecture Practices for Complex Contexts*

29 Adel Alhuraibi (UvT), *From IT-Business Strategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT*

30 Wilma Latuny (UvT), *The Power of Facial Expressions*

31 Ben Ruijl (UL), *Advances in computational methods for QFT calculations*

32 Thaer Samar (RUN), *Access to and Retrievability of Content in Web Archives*

33 Brigit van Loggem (OU), *Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity*

34 Maren Scheffel (OU), *The Evaluation Framework for Learning Analytics*

35 Martine de Vos (VU), *Interpreting natural science spreadsheets*

36 Yuanhao Guo (UL), *Shape Analysis for Phenotype Characterisation from High-throughput Imaging*

37 Alejandro Montes Garcia (TUE), *WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy*

38 Alex Kayal (TUD), *Normative Social Applications*

39 Sara Ahmadi (RUN), *Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR*

40 Altaf Hussain Abro (VUA), *Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems*

41 Adnan Manzoor (VUA), *Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle*

42 Elena Sokolova (RUN), *Causal discovery from mixed and missing data with applications on ADHD datasets*

43 Maaike de Boer (RUN), *Semantic Mapping in Video Retrieval*

44 Garm Lucassen (UU), *Understanding User Stories - Computational Linguistics in Agile Requirements Engineering*

45 Bas Testerink (UU), *Decentralized Runtime Norm Enforcement*

46 Jan Schneider (OU), *Sensor-based Learning Support*

47 Jie Yang (TUD), *Crowd Knowledge Creation Acceleration*

48 Angel Suarez (OU), *Collaborative inquiry-based learning*

2018

01 Han van der Aa (VUA), *Comparing and Aligning Process Representations*

02 Felix Mannhardt (TUE), *Multi-perspective Process Mining*

03 Steven Bosems (UT), *Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction*

04 Jordan Janeiro (TUD), *Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks*

05 Hugo Huurdeman (UVA), *Supporting the Complex Dynamics of the Information Seeking Process*

06 Dan Ionita (UT), *Model-Driven Information Security Risk Assessment of Socio-Technical Systems*

07 Jieting Luo (UU), *A formal account of opportunism in multi-agent systems*

08 Rick Smetsers (RUN), *Advances in Model Learning for Software Systems*

09 Xu Xie (TUD), *Data Assimilation in Discrete Event Simulations*

10 Julienka Mollee (VUA), *Moving forward: supporting physical activity behavior change through intelligent technology*

11 Mahdi Sargolzaei (UVA), *Enabling Framework for Service-oriented Collaborative Networks*

12 Xixi Lu (TUE), *Using behavioral context in process mining*

13 Seyed Amin Tabatabaei (VUA), *Computing a Sustainable Future*

14 Bart Joosten (UVT), *Detecting Social Signals with Spatiotemporal Gabor Filters*

15 Naser Davarzani (UM), *Biomarker discovery in heart failure*

16 Jaebok Kim (UT), *Automatic recognition of engagement and emotion in a group of children*

17 Jianpeng Zhang (TUE), *On Graph Sample Clustering*

18 Henriette Nakad (UL), *De Notaris en Private Rechtspraak*

19 Minh Duc Pham (VUA), *Emergent relational schemas for RDF*

20 Manxia Liu (RUN), *Time and Bayesian Networks*

21 Aad Slootmaker (OUN), *EMERGO: a generic platform for authoring and playing scenario-based serious games*

22 Eric Fernandes de Mello Araujo (VUA), *Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks*

23 Kim Schouten (EUR), *Semantics-driven Aspect-Based Sentiment Analysis*

24 Jered Vroon (UT), *Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots*

25 Riste Gligorov (VUA), *Serious Games in Audio-Visual Collections*

26 Roelof Anne Jelle de Vries (UT), *Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology*

27 Maikel Leemans (TUE), *Hierarchical Process Mining for Scalable Software Analysis*

28 Christian Willemse (UT), *Social Touch Technologies: How they feel and how they make you feel*

29 Yu Gu (UVT), *Emotion Recognition from Mandarin Speech*

30 Wouter Beek (VU), *The "K" in "semantic web" stands for "knowledge": scaling semantics to the web*

2019

01 Rob van Eijk (UL), *Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification*

02 Emmanuelle Beauxis Aussalet (CWI, UU), *Statistics and Visualizations for Assessing Class Size Uncertainty*

03 Eduardo Gonzalez Lopez de Murillas (TUE), *Process Mining on Databases: Extracting Event Data from Real Life Data Sources*

04 Ridho Rahmadi (RUN), *Finding stable causal structures from clinical data*

05 Sebastiaan van Zelst (TUE), *Process Mining with Streaming Data*

06 Chris Dijkshoorn (VU), *Nichesourcing for Improving Access to Linked Cultural Heritage Datasets*

07 Soude Fazeli (TUD), *Recommender Systems in Social Learning Platforms*

08 Frits de Nijs (TUD), *Resource-constrained Multi-agent Markov Decision Processes*

09 Fahimeh Alizadeh Moghaddam (UVA), *Self-adaptation for energy efficiency in software systems*

10 Qing Chuan Ye (EUR), *Multi-objective Optimization Methods for Allocation and Prediction*

11 Yue Zhao (TUD), *Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs*

12 Jacqueline Heinerman (VU), *Better Together*

13 Guanliang Chen (TUD), *MOOC Analytics: Learner Modeling and Content Generation*

14 Daniel Davis (TUD), *Large-Scale Learning Analytics Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses*

15 Erwin Walraven (TUD), *Planning under Uncertainty in Constrained and Partially Observable Environments*

16 Guangming Li (TUE), *Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models*

17 Ali Hurriyetoglu (RUN), *Extracting actionable information from microtexts*

18 Gerard Wagenaar (UU), *Artefacts in Agile Team Communication*

19 Vincent Koeman (TUD), *Tools for Developing Cognitive Agents*

20 Chide Groenouwe (UU), *Fostering technically augmented human collective intelligence*

21 Cong Liu (TUE), *Software Data Analytics: Architectural Model Discovery and Design Pattern Detection*

22 Martin van den Berg (VU), *Improving IT Decisions with Enterprise Architecture*

23 Qin Liu (TUD), *Intelligent Control Systems: Learning, Interpreting, Verification*

24 Anca Dumitrache (VU), *Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing*

25 Emiel van Miltenburg (VU), *Pragmatic factors in (automatic) image description*

26 Prince Singh (UT), *An Integration Platform for Synchromodal Transport*

27 Alessandra Antonaci (OUN), *The Gamification Design Process applied to (Massive) Open Online Courses*

28 Esther Kuindersma (UL), *Cleared for take-off: Game-based learning to prepare airline pilots for critical situations*

29 Daniel Formolo (VU), *Using virtual agents for simulation and training of social skills in safety-critical circumstances*

30 Vahid Yazdanpanah (UT), *Multiagent Industrial Symbiosis Systems*

31 Milan Jelisavcic (VU), *Alive and Kicking: Baby Steps in Robotics*

32 Chiara Sironi (UM), *Monte-Carlo Tree Search for Artificial General Intelligence in Games*

33 Anil Yaman (TUE), *Evolution of Biologically Inspired Learning in Artificial Neural Networks*

34 Negar Ahmadi (TUE), *EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES*

35 Lisa Facey-Shaw (OUN), *Gamification with digital badges in learning programming*

36 Kevin Ackermans (OUN), *Designing Video-Enhanced Rubrics to Master Complex Skills*

37 Jian Fang (TUD), *Database Acceleration on FPGAs*

38 Akos Kadar (OUN), *Learning visually grounded and multilingual representations*

2020

01 Armon Toubman (UL), *Calculated Moves: Generating Air Combat Behaviour*

02 Marcos de Paula Bueno (UL), *Unraveling Temporal Processes using Probabilistic Graphical Models*

03 Mostafa Deghani (UvA), *Learning with Imperfect Supervision for Language Understanding*

04 Maarten van Gompel (RUN), *Context as Linguistic Bridges*

05 Yulong Pei (TUE), *On local and global structure mining*

06 Preethu Rose Anish (UT), *Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support*

07	Wim van der Vegt (OUN), *Towards a software architecture for reusable game components*
08	Ali Mirsoleimani (UL), *Structured Parallel Programming for Monte Carlo Tree Search*
09	Myriam Traub (UU), *Measuring Tool Bias and Improving Data Quality for Digital Humanities Research*
10	Alifah Syamsiyah (TUE), *In-database Preprocessing for Process Mining*
11	Sepideh Mesbah (TUD), *Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models*
12	Ward van Breda (VU), *Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment*
13	Marco Virgolin (CWI), *Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming*
14	Mark Raasveldt (CWI/UL), *Integrating Analytics with Relational Databases*
15	Konstantinos Georgiadis (OUN), *Smart CAT: Machine Learning for Configurable Assessments in Serious Games*
16	Ilona Wilmont (RUN), *Cognitive Aspects of Conceptual Modelling*
17	Daniele Di Mitri (OUN), *The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences*
18	Georgios Methenitis (TUD), *Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems*
19	Guido van Capelleveen (UT), *Industrial Symbiosis Recommender Systems*
20	Albert Hankel (VU), *Embedding Green ICT Maturity in Organisations*
21	Karine da Silva Miras de Araujo (VU), *Where is the robot?: Life as it could be*
22	Maryam Masoud Khamis (RUN), *Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar*
23	Rianne Conijn (UT), *The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging*
24	Lenin da Nobrega Medeiros (VUA/RUN), *How are you feeling, human? Towards emotionally supportive chatbots*
25	Xin Du (TUE), *The Uncertainty in Exceptional Model Mining*
26	Krzysztof Leszek Sadowski (UU), *GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization*
27	Ekaterina Muravyeva (TUD), *Personal data and informed consent in an educational context*
28	Bibeg Limbu (TUD), *Multimodal interaction for deliberate practice: Training complex skills with augmented reality*
29	Ioan Gabriel Bucur (RUN), *Being Bayesian about Causal Inference*
30	Bob Zadok Blok (UL), *Creatief, Creatieve, Creatiefst*
31	Gongjin Lan (VU), *Learning better – From Baby to Better*
32	Jason Rhuggenaath (TUE), *Revenue management in online markets: pricing and online advertising*
33	Rick Gilsing (TUE), *Supporting service-dominant business model evaluation in the context of business model innovation*
34	Anna Bon (MU), *Intervention or Collaboration? Redesigning Information and Communication Technologies for Development*
35	Siamak Farshidi (UU), *Multi-Criteria Decision-Making in Software Production*

2021

01	Francisco Xavier Dos Santos Fonseca (TUD), *Location-based Games for Social Interaction in Public Space*
02	Rijk Mercuur (TUD), *Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models*
03	Seyyed Hadi Hashemi (UVA), *Modeling Users Interacting with Smart Devices*
04	Ioana Jivet (OU), *The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning*
05	Davide Dell'Anna (UU), *Data-Driven Supervision of Autonomous Systems*
06	Daniel Davison (UT), *"Hey robot, what do you think?" How children learn with a social robot*
07	Armel Lefebvre (UU), *Research data management for open science*
08	Nardie Fanchamps (OU), *The Influence of Sense-Reason-Act Programming on Computational Thinking*
09	Cristina Zaga (UT), *The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play*
10	Quinten Meertens (UvA), *Misclassification Bias in Statistical Learning*
11	Anne van Rossum (UL), *Nonparametric Bayesian Methods in Robotic Vision*
12	Lei Pi (UL), *External Knowledge Absorption in Chinese SMEs*
13	Bob R. Schadenberg (UT), *Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning*
14	Negin Samaeemofrad (UL), *Business Incubators: The Impact of Their Support*
15	Onat Ege Adali (TU/e), *Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm*
16	Esam A. H. Ghaleb (UM), *Bimodal Emotion Recognition from Audio-Visual Cues*
17	Dario Dotti (UM), *Human Behavior Understanding from motion and bodily cues using deep neural networks*
18	Remi Wieten (UU), *Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks*