On Relating Abstract and Structured Probabilistic Argumentation: a Case Study (corrected version)

Henry Prakken[1,2]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands,
h.prakken@uu.nl
[2] Faculty of Law, University of Groningen, Groningen, The Netherlands

Abstract. This paper investigates the relations between Timmer et al.'s proposal for explaining Bayesian networks with structured argumentation and abstract models of probabilistic argumentation. First some challenges are identified for incorporating probabilistic notions of argument strength in structured models of argumentation. Then it is investigated to what extent Timmer et al's approach meets these challenges and satisfies semantics and rationality conditions for probabilistic argumentation frameworks proposed in the literature. The results are used to draw conclusions about the strengths and limitations of both approaches.

1 Introduction and Motivation

There is a recent increase in interest in models of probabilistic argumentation. In argumentation theory, Hahn and others have advocated a probabilistic interpretation of argument schemes (e.g. [3]). A limitation of this work is that it does not deal with several crucial features of argumentation-based inference, such as attacks and combinations of arguments. Recent AI research on abstract models of probabilistic argumentation, e.g. [7, 6], addresses the first limitation. However, since it says nothing about the structure of arguments and the nature of attack, the proposed models have so far been hard to interpret. For example, it is not easy to understand what the probability of an argument means, since 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. What is required here is a precise account in terms of the structure of arguments and the nature of attack. The present paper aims to offer such an account.

In the literature two different uses of probability theory in argumentation can be seen, depending on whether the uncertainty is *in* or *about* the arguments. In the first use, probabilities are *intrinsic* to an argument in that they are used for capturing the strength of an argument given uncertainty concerning the truth of its premises or the reliability of its inferences. An example is default reasoning with probabilistic generalisations, as in *The large majority of Belgian people speak French, Mathieu is Belgian, therefore (presumably) Mathieu speaks French*. Clearly, if all premises of an argument are certain and it only makes deductive inferences, the argument should be given maximum probabilistic strength. [4] calls this use of probability the *epistemic* approach.

A second, *extrinsic* use of probability in argumentation (by [4] called the *constellations* approach) is for expressing uncertainty about whether arguments are accepted as

existing by some arguing agent. [5] gives the example of a dialogue participant who utters an enthymeme and where the listener can imagine two reasonable premises that the speaker had in mind: the listener can then assign probabilities to these options, which translate into probabilities on which argument the speaker meant to construct. This uncertainty has nothing to do with the intrinsic strengths of the two candidate completed arguments: one might be stronger than the other while yet the other is more likely the argument that the speaker had in mind. Note that in this approach even deductive arguments from certain premises can have less than maximal strength.

This paper focuses on the first use of probability theory, unlike most recent work on probabilistic abstract argumentation, which instead focuses on the second use (cf. the overview in [6]). An exception is [4], who formally distinguishes and develops both approaches, followed-up in e.g. [6]. For its epistemic approach [4] instantiates probabilistic argumentation frameworks with classical argumentation and defines the strength of an argument as the probability of the conjunction of all its premises. However, while for classical argumentation (where all arguments are deductive) this makes sense, this is not the case for accounts where arguments make defeasible inferences from certain premises (as in the above example of default reasoning, where it is both certain that the large majority of Belgian people speaks French and that Mathieu is Belgian but where the conclusion does not deductively follow from them): here all arguments should according to [4] be given strength 1, which is clearly undesirable.

Accordingly, the problem studied in this paper is how to instantiate abstract probabilistic frameworks with an account of intrinsic probabilistic strength of structured arguments, where the premises of all arguments are certain but their inferences can be defeasible. The problem will be studied in the context of a simple instantiation of the *ASPIC*$^+$ framework [8]. In particular, [10]'s recent proposal will be studied to explain forensic Bayesian networks in terms of *ASPIC*$^+$-style structured argumentation frameworks ($SAFs$) with probabilistic argument strengths. Since $SAFs$ are an instance of [1]'s abstract argumentation frameworks (AFs), Timmer's probabilistic $SAFs$ are a suitable candidate for being related to abstract probabilistic frameworks.

This paper is organised as follows. Section 2 presents the formal preliminaries. Section 3 gives a conceptual analysis of the problem of defining probabilistic strengths of structured arguments. Section 4 summarises [10]'s structured model of probabilistic argumentation. Section 5 then formally investigates its relation with abstract probabilistic argumentation frameworks, while Section 6 draws some general conclusions on the relation between abstract and structured models of probabilistic argumentation.

2 Formal Preliminaries

An *abstract argumentation framework* (AF) is a pair $\langle \mathcal{A}, attack \rangle$, where $\mathcal{A}$ is a set arguments and $attack \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation. The theory of AFs addresses how sets of arguments (called *extensions*) can be identified which are internally coherent and defend themselves against attack. A key notion here is that of an argument being *acceptable with respect to*, or *defended by* a set of arguments: $A \in \mathcal{A}$ is defended by $S \subseteq \mathcal{A}$ if for all $B \in \mathcal{A}$: if B attacks A, then some $C \in S$ attacks B. Then relative to a given AF various types of extensions can be defined.

- E is *admissible* if E is conflict-free and defends all its members;
- E is a *preferred extension* if E is a $\subseteq$-maximal admissible set;
- E is a *stable extension* if E is admissible and attacks all arguments outside it;
- $E \subseteq \mathcal{A}$ is the *grounded extension* if E is the least fixpoint of operator F, where $F(S)$ returns all arguments defended by S.

Various proposals for extending abstract argumentation frameworks with probabilities exist. Here we focus on one of the simplest proposals, the one of [7] as adapted by [4]. A *probabilistic argumentation framework* ($PrAF$) is a triple $\langle \mathcal{A}, attack, Pr \rangle$ where $\langle \mathcal{A}, attack \rangle$ is an abstract argumentation framework and $Pr : \mathcal{A} \mapsto [0, 1]$. Further notions concerning $PrAFs$ will be discussed in Section 5 below.

ASPIC$^+$ [8] is a general framework for structured argumentation. It abstracts from the logical language $\mathcal{L}$ except that it assumes a binary contrariness relation defined over $\mathcal{L}$. In the present paper $\mathcal{L}$ will be a language of propositional or predicate-logic atoms. Arguments are constructed from a knowledge base expressed in $\mathcal{L}$ by chaining inference rules defined over $\mathcal{L}$ into graphs (which are trees if no premise is used more than once). For present purposes only certain (non-attackable) premises and defeasible (attackable) inference rules are needed. All this reduces to the following definitions:

Definition 1 (Argumentation System). *An* argumentation system *(AS) is a tuple* $AS = (\mathcal{L}, \bar{}, \mathcal{R})$ *where:*

- $\mathcal{L}$ *is a logical language consisting of propositional or predicate-logic atoms*
- $\bar{}: \mathcal{L} \mapsto Pow(\mathcal{L})$ *is a contrariness function over* $\mathcal{L}$
- $\mathcal{R}$ *is a set of (defeasible) inference rules of the form* $\phi_1, \ldots, \phi_n \Rightarrow \phi$ *(where* ϕ, ϕ_i *are meta-variables ranging over wff in* $\mathcal{L}$*).*

Definition 2. ***[Knowledge Bases and Arguments]*** *An* argument A *on the basis of a* knowledge base $\mathcal{K} \subseteq \mathcal{L}$ *in an argumentation system* AS *is:*

1. φ *if* $\varphi \in \mathcal{K}$ *with:* $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$;
2. $A_1, \ldots A_n \Rightarrow \psi$ *if* $A_1, \ldots, A_n$ *are arguments such that* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi \in \mathcal{R}$ *with:* $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$, $\texttt{Conc}(A) = \psi$, $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;

An argument A is said to *attack* an argument B iff A *rebuts* B, where A *rebuts* B (on B') iff $\texttt{Conc}(A) = \overline{\varphi}$ for some $B' \in \texttt{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \varphi$.

The *ASPIC*$^+$ counterpart of an abstract argumentation framework is a structured argumentation framework.

Definition 3. ***[Structured Argumentation Frameworks]*** *Let* AT *be an argumentation theory* $(AS, \mathcal{K})$*. A structured argumentation framework (*SAF*) defined by* AT*, is a triple* $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ *where* $\mathcal{A}$ *is the set of all finite arguments constructed from* $\mathcal{K}$ *in* AS*,* $\preceq$ *is an ordering on* $\mathcal{A}$*, and* $(X, Y) \in \mathcal{C}$ *iff* X *attacks* Y*.*

A relation of *defeat* is then defined as follows ($A \prec B$ is defined as usual as $A \preceq B$ and $B \not\preceq A$). A *defeats* B iff A rebuts B on B' and $A \not\prec B'$. Abstract argumentation frameworks are then generated from $SAFs$ by letting the attacks from an AF be the defeats from a SAF.

Definition 4 (Argumentation frameworks). *An* abstract argumentation framework (AF) corresponding to a SAF = $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ *(where* $\mathcal{C}$ *is* ASPIC$^+$*'s attack relation) is a pair* $(\mathcal{A}, attack)$ *such that* $attack$ *is the defeat relation on* $\mathcal{A}$ *determined by* SAF.

3 Probabilistic Argument Strength: a Conceptual Analysis

We can now make the problem studied in this paper even more specific. The problem is: in the context of the just-described simple instantiation of *ASPIC*$^+$, how can a probabilistic notion of argument strength be defined such that for two arguments A and B we have that $A \preceq B$ just in case $strength(A) \leq strength(B)$? In other words: how can probabilistic argument strength be used to resolve attacks into defeats?

A first challenge here is that a higher internal strength does not necessarily make an argument dialectically stronger. Suppose 90% of the birds can fly, and 80% of penguins cannot fly. Should the argument *Tweety can fly since it is a penguin so it is a bird* be stronger than the argument *Tweety cannot fly since it is a penguin*? Of course not, since probability theory requires that all evidence is taken into account, and since penguins are a special kind of bird, we should (defeasibly) conclude that Tweety cannot fly. More formally, if we have $Pr(q|p) = x$ and $Pr(q|p \wedge r) = y$ and both p and r are given, then we should base our inference on $Pr(q|p \wedge r) = y$. For this reason, probability-based comparisons between arguments should, either explicitly or implicitly, involve a kind of specificity principle. For example, [2] do so in their notion of specificity defeat for probabilistic assumption-based argumentation. More generally this shows that the probabilistic strength of an argument cannot be calculated independent of its attackers.

This issue arises in a different way in case of attacks between arguments that do not have a specificity relation. Consider the following well-known example from nonmonotonic logic: *Quakers are usually pacifists, Republicans are usually not pacifists, Nixon was a quaker and a republican*. It is wrong to compare $Pr(P|Q)$ with $Pr(\neg P|R)$. What counts is $Pr(P|Q \wedge R)$ and in general the latter probability is independent of the former probabilities (although in special cases this may be different).

A third challenge arises from the step-by-step nature of arguments. Consider *I see smoke, my observations are usually correct, therefore (presumably) there is smoke. Where there is smoke, there usually is fire, therefore (presumably) there is fire*. Can a recursive definition of argument strength be given where the strength of the entire argument depends on the strength of its subargument for *there is smoke* and the strength of its final step? This is not a trivial problem, since $Pr(fire|seesmoke)$ does in general not follow from $Pr(fire|smoke)$ and $Pr(smoke|seesmoke)$.

4 Explaining Bayesian Networks with Argumentation

In this section we summarise [10]'s method for explaining (forensic) Bayesian networks with argumentation. A Bayesian network is a graphical representation of a joint probability distribution. Formally, it is a pair (G, Pr) where G is a directed acyclic graph $(\mathbf{V}, \mathbf{E})$, with a finite set of variables $\mathbf{V}$ connected by edges $\mathbf{E}$ from $\mathbf{V} \times \mathbf{V}$, and Pr

is a probability function which specifies for every variable V_i the probability distribution $Pr(V_i|parents(V_i))$ of its outcomes conditioned on its parents $parents(V_i)$ in the graph. [10] assume that all variables are boolean.

[10] first set the language $\mathcal{L}$ of the *ASPIC*$^+$ argumentation system to the set of all $V = v$ expressions where $V \in \mathbf{V}$ and v is a possible value of V. Then $\varphi \in \overline{\psi}$ iff φ and ψ assign different values to the same variable. [10] then derive an *ASPIC*$^+$ SAF from a BN plus a set of instantiated variables (the *evidence*) in terms of an intermediate structure called a *support graph*, which for a given variable of interest from $\mathbf{V}$ captures the potential reasoning paths through the BN. Entering evidence in a BN prunes all branches of the support graph that do not end in evidence. Arguments can be constructed from the support graph by making its non-premise nodes either true or false.

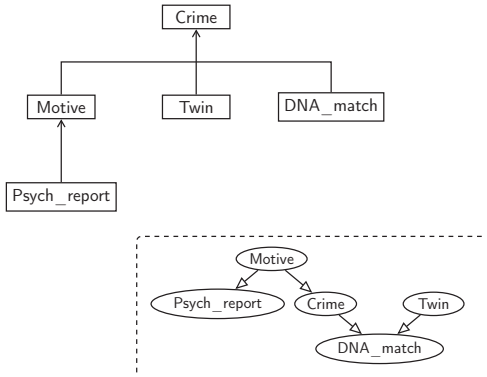

Fig. 1. A Bayesian network (below) and a support graph (above).

The basic idea is illustrated with Figures 1 and 2, displaying an example from [10]. The variable of interest in the BN is whether the suspect committed the *Crime*. Evidence for this can be a *DNA_match* between the suspect's DNA and DNA found at the crime scene. Such a DNA match may also be explained by the existence of a *Twin* brother. The existence of a *Motive* makes the crime more likely. Evidence for a motive may be given in a psychological report (*Psych_report*). After the evidence *Psych_report* = *True* and *DNA_match* = *True* is entered into the BN, the chain in the support graph from *Twin* to *Crime* is pruned away. The arguments generated from this are all inference trees corresponding to the pruned graph or a subgraph with all non-evidence nodes instantiated in any possible way (i.e., with either *True* or *False*). Figure 2 displays the arguments when *Psych_report* and *Crime* are both true. Formally (with contrariness for convenience encoded with negation, even though $\neg$ is strictly speaking not part of $\mathcal{L}$) both these pieces of evidence are arguments and moreover:

$A_1 =$ *Psych_report* $\Rightarrow$ *Motive* $\quad A_2 = A_1,$ *DNA_match* $\Rightarrow$ *Crime*

But, for instance, the following arguments can also be constructed:

$$B_1 = \textit{Psych\_report} \Rightarrow \neg \textit{Motive} \quad B_2 = B_1, \textit{DNA\_match} \Rightarrow \neg \textit{Crime}$$
$$C = A_1, \textit{DNA\_match} \Rightarrow \neg \textit{Crime}$$

Suppose that now *Twin* also becomes available as evidence. At first sight, one might expect that this gives rise to rebuttal of A_2 of the form $\textit{Twin} \Rightarrow \neg \textit{Crime}$. However, this is not how the method works. Instead, it captures all variables relevant to a conclusion in a single argument concerning that conclusion. So, A_2, B_2 and C are modified to:

$$A_2' = A_1, \textit{DNA\_match}, \textit{Twin} \Rightarrow \textit{Crime} \qquad B_2' = B_1, \textit{DNA\_match}, \textit{Twin} \Rightarrow \neg \textit{Crime}$$
$$C' = A_1, \textit{DNA\_match}, \textit{Twin} \Rightarrow \neg \textit{Crime}$$

So for every constructable argument there is a rebuttal with the same 'skeleton' but with some truth values of non-premises flipped, and there are no other (direct) rebuttals.

Space limitations prevent listing the formal definitions of the support graph and the induced SAF. Essentially, the knowledge base consists of the evidence while the set of defeasible rules corresponds to links in the support graph. For present purposes all that is relevant is that the following definition of argument strength, when used to resolve attacks into defeats, gives the induced SAF a number of 'good' properties, which means that [10]'s way to define probabilistic structured argumentation as an instantiation of Dung-style AFs makes sense (with some simplified notation compared to [10]):

$$\textit{strength}(A) = Pr(\texttt{Conc}(A)|\mathcal{K})$$

(where $\mathcal{K}$ is the knowledge base of the AT induced by the BN-with-evidence). Thus the

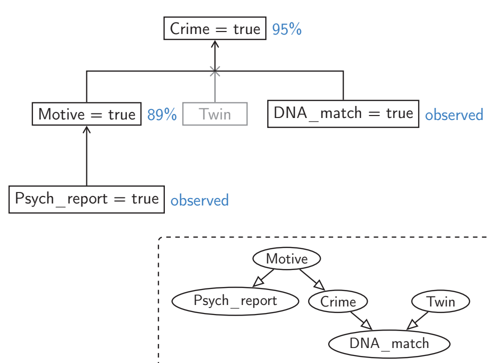

Fig. 2. Deriving arguments from the pruned support graph after entering evidence.

strength of an argument A equals the posterior probability of $\texttt{Conc}(A)$ in the BN-with-evidence inducing the SAF. This definition implies that the strengths of two directly rebutting arguments always adds up to 1, since $Pr(Q|P) = 1 - Pr(\neg Q|P)$.

Figure 2 displays the strengths of the non-premise arguments in the figure, based on the probability tables in [10]. The strength of B_1 is $1 - \textit{strength}(A_1) = 11\%$ and

the strength of B_2 and C is $1 - strength(A_2) = 5\%$. In fact, to obtain the results listed below, for arguments for the variable of interest the definition of strength can be reduced to the equivalent definition $strength(A) = Pr(\texttt{Conc}(A)|\texttt{Prem}(A))$. However, for subarguments the inclusion of all of $\mathcal{K}$ is needed. For the reasons why see [10]. It should be noted that a strength of 1 of a non-premise argument does not mean that it should be modelled as applying strict rules, since $Pr(P|Q) = 1$ does not imply $Pr(P|Q \wedge R) = 1$. So even a non-premise argument of strength 1 is defeasible.

The 'good' properties (as proven by or easily following from [10]) are as follows.

1. The grounded extension equals the set of undefeated arguments.
2. The grounded extension satisfies subargument closure, direct and indirect consistency and (trivially) strict closure. For the definitions of these properties see [8].
3. If A is in the grounded extension, then A is the strongest argument for $\texttt{Conc}(A)$.
4. If A is in the grounded extension, then $strength(A) > 0.5$.

Interestingly, an argument can be stronger than some of its subarguments, An example is Figure 2, where argument A_2 for *Crime* is stronger than its subargument A_2 for motive. This can happen since by combining arguments A_1 and *DNA_match*, argument A_2 aggregates the support given to its conclusion by two pieces of evidence. This reflects a general feature of probabilistic reasoning, namely, that the combination of pieces of evidence that each have weak probative force can have strong probative force.

Let us see how [10]'s approach deals with the challenges discussed in Section 3. The first two challenges are dealt with since, firstly, all arguments contain all variables from the BN that are relevant for their conclusions and, secondly, argument strength is defined relative to all (relevant) evidence. While this is good, it also has an obvious limitation, namely, that two arguments with the same premises and conclusion but different internal structure have the same strength. Thus the third challenge is not fully met. On the other hand, it is still partially met since two such arguments can have subarguments of different strengths, and this can be reported to those to which the BN is explained. Another possible limitation is that in [10] the reasons pro and con a conclusion are not, as usual in argumentation, distributed over conflicting arguments but are all contained in a single argument for or against the conclusion, which is not according to the conceptual idea underlying argumentation-based inference. Nevertheless, for the purpose of explaining forensic Bayesian networks to judges, prosecutors and defence lawyers this may be perfectly adequate.

5 Relating the Abstract and Structured Accounts

We now investigate whether the work of [10] can be seen as an instantiation of epistemically interpreted probabilistic abstract frameworks in the sense of [4]. The first step is obvious, namely, equating the probability of an argument in a $PrAF$ with the argument's strength according to [10]. However, the next step, instantiating the *attack* relation of $PrAFs$, is less obvious: should it be instantiated with *ASPIC*$^+$'s $\mathcal{C}$ relation of attack or with its defeat relation? Let us first assume that the probability of an argument is to be used to resolve attacks: can this be modelled at the abstract level or should this be modelled while taking the structure of arguments into account? [8] provide reasons

for the latter approach, since the former approach cannot distinguish between direct and indirect attack relations and therefore runs the risk of applying the wrong probabilities to an attack. Consider an argument $A_3 = A_1, A_2 \Rightarrow \varphi$ and an argument B rebutting A_3 on A_1. [8] show that with a last-link ordering in terms of rule priorities it may be that $A_1 \prec B$ while $B \prec A_3$. Then resolving the attacks at the abstract level by saying that an argument A attacks (i.e., ***ASPIC***$^+$-defeats) an argument B iff A rebuts B and $A \not\prec B$ results in a grounded/preferred/stable/complete extension $\{A_2, A_3, B\}$, which is not closed under the subargument relation. This problem arises since the rebuttal of B on A_3 is incorrectly resolved with $B \prec A_3$ (so B does not attack A_3 in the AF) while it should be resolved with $A_1 \prec B$ (so B does attack A_3 in the AF), resulting in a grounded extension $\{A_2, B\}$. As shown by [8], the same problem can make conclusion sets of extensions violate consistency. Since, as noted above, in [10]'s approach an argument can also be stronger than some of its subarguments, all these problems also arise if the Pr function is used at the abstract level of $PrAFs$ to resolve attacks. This is an important lesson that can be learned from the present analysis.

In [4], which instantiates $PrAFs$ with classical argumentation, the argument probabilities are not used in defining the attack relation between arguments, so (in terms of ***ASPIC***$^+$), [4] instantiates the *attack* relation of $PrAFs$ with ***ASPIC***$^+$'s $\mathcal{C}$ relation. Yet his approach does not necessarily suffer from the just-sketched problems, since [4] does not use the Pr functions to resolve the attacks in $PrAFs$. Instead, he defines the following notions. The *epistemic extension* of a $PrAF$ is $\{A \in \mathcal{A} | Pr(A) > 0.5\}$. A probability function is *rational* iff for every pair of arguments A and B such that A attacks B, if $Pr(A) > 0.5$ then $Pr(B) \leq 0.5$. A $PrAF$ is called rational if its Pr is rational. [4] then proves that for every rational $PrAF$ its epistemic extension is conflict-free with respect to the *attack* relation.

Yet the notion of an epistemic extension combined with the rationality constraint on Pr is not a good abstraction of [10]'s approach. To start with, the grounded extension in [10]'s approach does not always equal the epistemic extension. Consider a support graph $E \longrightarrow H_1 \longrightarrow H_2$ and consider the arguments $(E \Rightarrow H_1) \Rightarrow H_2$ and $(E \Rightarrow \neg H_1) \Rightarrow H_2$. Both have the same strength. Suppose their strength exceeds 0.5. Then they cannot be both in the grounded extension, since they have rebutting subarguments. Yet both are in the epistemic extension. This can happen since in [10]'s approach an argument can be stronger than some of its subarguments.

Next, [10]'s probability function is not guaranteed to be rational. Consider the same support graph, the above arguments and their respective subarguments $E \Rightarrow H_1$ and $E \Rightarrow H_2$, and suppose *strength*$((E \Rightarrow H_1) \Rightarrow H_2) = 0.6$ and *strength*$((E_1 \Rightarrow H_1)) = 0.4$ and *strength*$((E_1 \Rightarrow \neg H_1)) = 0.6$. The argument for $\neg H_1$ indirectly attacks the argument for H_2 but both have strength > 0.5. This can happen since the attack is indirect. In $PrAFs$ the distinction between direct and indirect attack cannot be modelled. However, if the rationality constraint is confined to direct attacks, then it holds, since the strengths of two directly rebutting arguments always add up to 1. This is a first indication of the importance of taking the structure of arguments into account.

Further indications follow from an analysis of the other "rationality conditions" on $PrAFs$ proposed in [6]. (Below for any $A \in \mathcal{A}$, $A^- = \{B | B \text{ attacks } A\}$).

COH Pr is *coherent* if for every $A, B \in \mathcal{A}$, if A attacks B then $Pr(A) \leq 1 - Pr(B)$.

INV Pr is *involutary* if for every $A, B \in \mathcal{A}$, if A attacks B then $Pr(A) = 1 - Pr(B)$.

SFOU Pr is *semi-founded* if $Pr(A) \geq 0.5$ for every unattacked $A \in \mathcal{A}$.

FOU Pr is *founded* if $Pr(A) = 1$ for every $A \in \mathcal{A}$ with $A^- = \emptyset$.

SOPT Pr is *semi-optimistic* if $Pr(A) \geq 1 - \Sigma_{B \in A^-} Pr(B)$ whenever $A^- \neq \emptyset$.

OPT Pr is *optimistic* if $Pr(A) \geq 1 - \Sigma_{B \in A^-} Pr(B)$ for every $A \in \mathcal{A}$.

We now investigate whether these properties hold for [10]'s approach, for both the *ASPIC*$^+$ $\mathcal{C}$ relation and its defeat relation. In doing so, we will use the support graph $E \longrightarrow H_1 \longrightarrow H_2$ and the various arguments it generates with evidence E, assuming that both $Pr(H_1|\mathcal{K}) > 0.5$ and $Pr(H_2|\mathcal{K}) > 0.5$. (For space limitations we omit a proof that a BN that generates such a support graph, arguments and strengths exists).

COH in general neither holds for $\mathcal{C}$ nor for defeat, since these relations can be indirect. For example, argument $(E \Rightarrow H_1) \Rightarrow H_2$ rebuts $(E \Rightarrow \neg H_1) \Rightarrow H_2$ on $E \Rightarrow \neg H_1$ but both arguments have strength > 0.5 (even though the latter's subargument for $\neg H_1$ has strength < 0.5). However, COH does hold when restricted to direct $\mathcal{C}$ or defeat relations, since for every pair of direct rebuttals their strengths add up to 1. All these observations also hold for INV.

SFOU holds in general for both $\mathcal{C}$ and defeat. For $\mathcal{C}$, note that the only non-rebutted arguments are elements of $\mathcal{K}$, which by definition have strength 1. For defeat, if B unsuccessfully directly rebuts A, then $strength(B) < strength(A)$ so since these strengths add up to 1, $strength(A) > 0.5$. Note further that every non-premise argument has at least one rebuttal, so every non-defeated argument has strength > 0.5.

FOU holds in general for $\mathcal{C}$ but not for defeat. For $\mathcal{C}$ FOU holds for the same reason as why SFOU holds. Our above example provides a counterexample for defeat if the strength of the argument for H_2 does not equal 1. This is also a counterexample for FOU restricted to direct defeats.

SOPT neither holds for $\mathcal{C}$ nor for defeat, and neither for the direct nor for the indirect relations. In our example, $(E \Rightarrow H_1) \Rightarrow H_2$ has two direct rebuttals, namely, $(E \Rightarrow H_1) \Rightarrow \neg H_2$ and $(E \Rightarrow \neg H_1) \Rightarrow H_2$. If the strength of the argument for H_2 is higher than 0.5 but below 0.75, then the strengths of its two rebuttals add up to above 0.5. This is also a counterexample to OPT.

SOPT holds in general for both $\mathcal{C}$ and defeat. For direct attack and direct defeat it holds since if A directly attacks B, then by definition of the strength of arguments it holds that $Pr(A) + Pr(B) = 1$. Furthermore, by construction of $\mathcal{A}$ every argument has at least one direct attacker. If an argument A only has indirect defeaters, then $Pr(A) > 0.5$ since it has no direct defeaters. Then the property holds for defeat in general since if B indirectly defeats A then B directly attacks a subargument of A so $Pr(B) \geq 0.5$.

Finally, OPT holds for in general for attack since an argument without attackers has probability 1. However, it holds neither for direct defeat nor for defeat in general since an argument without defeaters can still have probability less than 1.

These results are summarised in the following Table.

	Direct attack	General attack	Direct defeat	General defeat
RAT	x		x	
COH	x		x	
INV	x		x	
SFOU	x	x	x	x
FOU	x	x		
SOPT	x	x	x	x
OPT	x	x		

The negative results do not indicate flaws of [10]'s approach, since they are due to two of its features which both are reasonable for probabilistic argumentation: the distinction between direct and indirect attack and the fact that an argument can be stronger than some of its subarguments. It can therefore be concluded that [6]'s set of rationality conditions cannot be seen as minimum conditions for the well-behavedness of $PrAFs$.

6 Conclusion

In this paper we have investigated to what extent [10]'s probabilistic version of *ASPIC*$^+$, proposed for explaining Bayesian networks, satisfies semantics and rationality conditions for probabilistic argumentation frameworks proposed in the literature. Some results were positive but other results were negative. The negative results do not seem to point at flaws of [10]'s approach but instead at limitations of current abstract models of probabilistic argumentation, in particular their failure to distinguish between direct and indirect relations of attack and defeat. One conclusion is that to make this distinction in a proper way, the structure of arguments and the nature of attack and defeat must be made explicit. Another conclusion is that not all rationality conditions for probabilistic models of argumentation proposed in the literature can be regarded as minimum requirements for the well-behavedness of these models.

This paper has also identified several challenges for attempts to use probabilistic argument strength for resolving attacks into defeats. These challenges arise from the difference between the global nature of Bayesian probabilistic reasoning (where all evidence has to be taken into account) and the local nature of argumentation (where particular conflicting arguments are compared). [10] found a way to meet these challenges but with some limitations. While [10]'s approach may suffice for its intended application of explaining forensic Bayesian networks, future research should study whether more general solutions are possible without these limitations.

Some related work was already discussed throughout this paper. In addition, [2] propose an extension of [1]'s abstract frameworks with probability and then extend assumption-based argumentation with the means to label literals in rules with probabilities. The abstract approach models more than what is of present concern, namely, some aspects of multi-agent argumentation, while the abstract and assumption-based parts are not formally related. In future research it would be interesting to investigate [2]'s probabilistic version of assumption-based argumentation in the same way as we did for [10]'s probabilistic version of *ASPIC*$^+$.

In this paper we studied the use of probability for two things: for resolving attacks into defeats within *ASPIC*$^+$ and for identifying epistemic extensions in the sense of [4]. In future research it would be interesting to investigate the use of probability to define graded notions of argument acceptability, as in e.g. [9]. We conjecture that here, too, it is important to take the structure of arguments and the nature of attack into account.

References

1. P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence*, 77:321–357, 1995.
2. 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, Amsterdam etc, 2010.
3. U. Hahn and J. Hornikx. A normative framework for argument quality: argumentation schemes with a Bayesian foundation. *Synthese*, 193:1833–1873, 2016.
4. A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54:47–81, 2013.
5. A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55:607–638, 2014.
6. A. Hunter and M. Thimm. On partial information and contradictions in probabilistic abstract argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference (KR-16)*, pages 53–62. AAAI Press, 2016.
7. H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings first Workshop on the Theory and Applications of Formal Argument*, pages 1–16, 2011.
8. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
9. M. Thimm. A probabilistic semantics for abstract argumentation. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 750–755, 2012.
10. S. Timmer, J.-J.Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.