Proceedings of the Sixteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2018)

Probabilistic Strength of Arguments with Structure

Henry Prakken
Department of Information and Computing Sciences, Utrecht University
Faculty of Law, University of Groningen
The Netherlands

Abstract

This paper investigates the relation between abstract and structured accounts of probabilistic argumentation. The *ASPIC*$^+$ framework is applied to default reasoning with probabilistic generalisations, using the idea that the probability of an argument is the probability of the conjunction of all its premises and conclusions. Based on this idea, two notions of internal and dialectical argument strength are defined and compared. The resulting account is then related to Hunter & Thimm's epistemic approach to abstract probabilistic argumentation.

Introduction

Recently there has been much research on probabilistic abstract argumentation, e.g. (Dung and Thang 2010; Li, Oren, and Norman 2012; Hunter and Thimm 2016; 2017). Its abstract nature makes this work not easy to interpret. For example, there is unclarity about 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. For the same reason statements about whether an argument "is true" or "can be believed" (e.g. (Hunter 2013; Hunter and Thimm 2017)) are in need of clarification. The present paper aims to clarify such issues in terms of the structure of arguments and the nature of attack.

Current approaches to probabilistic argumentation are of two kinds, depending on whether the uncertainty is *in* or *about* the arguments. Sometimes the probabilities are *intrinsic* to an argument in that they express uncertainty concerning the truth of its premises or the reliability of its inferences. An example is default reasoning with probabilistic generalisations, as in *Birds can typically fly, Tweety is a bird, therefore (presumably), Tweety can fly*. This is arguably what (Hunter 2013) calls the 'epistemic' approach to probabilistic argumentation. A second, *extrinsic* use of probability in argumentation (which arguably is what (Hunter 2013) calls the 'constellations' approach) is for expressing uncertainty about whether arguments are accepted as existing by some arguing agent. (Hunter 2014) gives the example of an enthymeme that could leave two alternative premises implicit: if a listener assigns probabilities to these premises, then these translate into probabilities on which argument the speaker meant to construct. This uncertainty is independent of the intrinsic strengths of the two possible arguments: one might be stronger than the other while yet the other is more likely the argument that the speaker had in mind.

This paper focuses on the intrinsic (epistemic) approach. Related work is (Hunter 2013), who defines the strength of classical-logic arguments as the probability of the conjunction of all their premises. While this makes sense when all arguments are deductively valid, it does not apply to cases where additional uncertainty arises from defeasible inferences (as in the above example). This paper therefore generalises Hunter's idea to arguments that apply defeasible inference rules, and studies how the resulting account relates to current abstract models of epistemic probabilistic argumentation, in particular to (Hunter and Thimm 2016; 2017)'s rationality conditions. The problem will be studied in the context of a simple instantiation of the *ASPIC*$^+$ framework for structured argumentation (Prakken 2010; Modgil and Prakken 2013; 2018). An important idea will be that arguments implicitly make probabilistic independence assumptions and that the assumptions of different arguments may be mutually inconsistent.

The proposal will be developed for probabilistic default reasoning, where rules express probabilistic generalisations. In nonmonotonic logic default reasoning is often (though not always) studied in a qualitative way, where the rules are expressed with qualitative modalities, as in 'If X then usually / normally / typically Y'. However, with the rise of big-data machine learning applications, statistical and probabilistic arguments can be increasingly expected in many domains, for example, in legal proof, medical diagnosis, customer acceptance procedures or employee selection procedures. This justifies a quantitative probabilistic study of such arguments.

There have been two earlier attempts to model probabilistic argumentation in *ASPIC*$^+$. (Rienstra 2012) takes the constellations approach and is therefore irrelevant to the present paper. (Timmer et al. 2017) propose a method for explaining forensic Bayesian networks by deriving arguments from them. Their method was by (Prakken 2017) related to (Hunter and Thimm 2017)'s abstract approach. While this work like us takes an epistemic approach, its concerns are different in that explaining forensic Bayesian networks is a

different problem than the one studied in the present paper. In particular, it assumes a single joint probability distribution over the language over which arguments are constructed, while the account developed in the present paper does not make such an assumption, for reasons explained below.

This paper is organised as follows. After presenting the formal preliminaries, we first define a notion of internal argument strength on the basis of only information pertaining to the argument itself. We show that this notion is probabilistically well-defined but can make that conflicting arguments make jointly inconsistent probability assumptions, so that using internal argument strength for resolving conflicts between arguments is problematic. Then we study a notion of dialectical argument strength according to which there are no such probabilistic inconsistencies and which can therefore be used to resolve conflicts between arguments. We will show that dialectical strength better respects (Hunter and Thimm 2017)'s rationality conditions than internal strength but is arguably harder to apply in practice. We end with a discussion of other related research and concluding remarks.

Formal Preliminaries

In this section we summarise the formal frameworks used in this paper. An *abstract argumentation framework* (AF) is a pair $\langle \mathcal{A}, attack \rangle$, where $\mathcal{A}$ is a set of arguments and $attack \subseteq \mathcal{A} \times \mathcal{A}$. The theory of AFs (Dung 1995) identifies sets of arguments (called *extensions*) which are internally coherent and defend themselves against attack. An argument $A \in \mathcal{A}$ is *defended* by a set 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, $E \subseteq \mathcal{A}$ is *admissible* if E is conflict-free and defends all its members; E is a *complete extension* if E is admissible and $A \in E$ iff A is defended by E; 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; and $E \subseteq \mathcal{A}$ is the *grounded extension* if E is the least fixpoint of operator F, where $F(S)$ returns all arguments defended by S. It holds that any preferred, stable or grounded extension is a complete extension. Finally, for $T \in \{$complete, preferred, grounded, stable$\}$, X is *sceptically* or *credulously* justified under the T semantics if X belongs to all, respectively at least one, T extension.

We next specify the present paper's instance of the *ASPIC*$^+$ framework. It defines abstract argumentation systems as structures consisting of a logical language $\mathcal{L}$ with negation and two sets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules defined over $\mathcal{L}$. In the present paper $\mathcal{L}$ is a language of propositional or predicate-logic literals. Arguments are constructed from a knowledge base (a subset of $\mathcal{L}$) by chaining inferences over $\mathcal{L}$ into acyclic graphs. Formally:

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

- $\mathcal{L}$ is a logical language consisting of propositional or ground predicate-logic literals
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a finite set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\{\varphi_1, \ldots, \varphi_n\} \rightarrow \varphi$ and $\{\varphi_1, \ldots, \varphi_n\} \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$. $\varphi_1, \ldots, \varphi_n$ are called the *antecedents* and φ the *consequent* of the rule.[1]

We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$. Note that $-$ is not part of the logical language $\mathcal{L}$ but a metalinguistic function symbol to obtain more concise definitions. Also, for any rule r the *antecedents* and *consequent* are denoted, respectively, with $\texttt{ant}(r)$ and $\texttt{cons}(r)$.

Definition 2 [**Consistency**] For any $S \subseteq \mathcal{L}$, let the *closure of S under strict rules*, denoted $Cl_{R_s}(S)$, be the smallest set containing S and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{R_s}(S)$. Then a set $S \subseteq \mathcal{L}$ is *directly consistent* iff $\nexists\ \psi, \varphi \in S$ such that $\psi = -\varphi$, and *indirectly consistent* iff $Cl_{R_s}(S)$ is directly consistent.

Definition 3 [**Knowledge bases**] A *knowledge base* over an $AS = (\mathcal{L}, \mathcal{R})$ is an indirectly consistent finite set $\mathcal{K} \subseteq \mathcal{L}$.

In this paper $\mathcal{K}$ corresponds to the 'necessary premises' in other *ASPIC*$^+$ publications, which are intuitively certain and therefore not attackable. We will represent what intuitively are uncertain premises φ as defeasible rules $\Rightarrow \varphi$. We also assume that no element of $\mathcal{K}$ occurs in the consequent of any rule in $\mathcal{R}$. The finiteness restrictions on $\mathcal{K}$ and $\mathcal{R}$ are to be in line with (Hunter and Thimm 2017), who assume a finite abstract argumentation framework.

As observed by (Modgil and Prakken 2018), *ASPIC*$^+$ as it has developed over the years is not a single framework but a family of frameworks varying on several elements. Some variations are in the definition of an argument. We adopt a variant in which arguments cannot be circular in that they cannot repeat conclusions of their proper subarguments (a condition adopted by (Grooters and Prakken 2016)), in which the set of all conclusions of an argument has to be indirectly consistent (a condition explored by (Wu and Podlaszewski 2015)) and in which an argument cannot have two different subarguments for the same conclusion. The assumptions of non-circularity and internal consistency are arguably reasonable requirements for any rational notion of an argument. The assumption that an argument cannot have different subarguments for the same conclusion is a pragmatic rationality constraint, expressing that in a single argument one should commit to a single way to support a conclusion. These three assumptions on arguments, besides being reasonable rational constraints, also have technical reasons, as will become apparent in the proof of Theorem 15.

To distinguish between the features of the definition that are present in all work on *ASPIC*$^+$ and the specific features assumed in this paper, we below first define the notion of a 'general' argument and then define arguments as studied in the present paper as general arguments that exhibit the specific features assumed in this paper.

Definition 4 [**(General) arguments**] A *general argument* A on the basis of a knowledge base $\mathcal{K}$ over an argumentation system AS is a structure obtainable by applying one or more of the following steps finitely many times:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{Rules}(A) = \emptyset$; $\texttt{DefRules}(A) = \emptyset$; $\texttt{TopRule}(A) =$ undefined.

[1] Below the brackets around the antecedents will be omitted.

2. $[A_1], \ldots, [A_n] \rightarrow \psi$[2] if $A_1, \ldots, A_n$ are general arguments such that $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi \in \mathcal{R}_s$ with:
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$;
   $\texttt{Conc}(A) = \psi$;
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi\}$;
   $\texttt{DefRules}(A) = \texttt{Rules}(A) \cap \mathcal{R}_d$;
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$.
3. $[A_1], \ldots, [A_n] \Rightarrow \psi$ if $A_1, \ldots, A_n$ are general arguments such that $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi \in \mathcal{R}_d$, with the other notions defined as in (2).

A general argument A is an *argument* iff:

1. $\texttt{Conc}(\texttt{Sub}(A))$ is indirectly consistent; and
2. If A contains subarguments A' and A'' such that $\texttt{Conc}(A') = \texttt{Conc}(A'')$, then $A' = A''$.

An argument A is *strict* if $\texttt{DefRules}(A) = \emptyset$, otherwise A is *defeasible*.

Each of the functions Func in this definition is also defined on sets of arguments $S = \{A_1, \ldots, A_n\}$ as follows: $\texttt{Func}(S) = \texttt{Func}(A_1) \cup \ldots \cup \texttt{Func}(A_n)$. Note that we overload the $\Rightarrow$ symbol to denote an argument while it also denotes defeasible inference rules. This is common practice in argumentation and originates from (Vreeswijk 1997).

Condition (2) of the definition of an argument implies non-circularity, for which reason the formal definition leaves the non-circularity condition implicit. However, the former condition is not implied by the latter, since two different general subarguments for the same conclusion need not have a subargument relation between them. An example is

$$A = [[[p \Rightarrow r], s] \Rightarrow t, [[q \Rightarrow r], u] \Rightarrow v] \rightarrow w$$

General argument A has two different subarguments for r, one applying the rule $p \Rightarrow r$ to p and the other applying the rule $q \Rightarrow r$ to q. Then r is used twice: once in a rule $r, s \Rightarrow t$ and once in a rule $r, u \Rightarrow v$. Then A ends with the top rule $t, v \rightarrow w$. Neither of its two subarguments for r is a subargument of the other.

In general, *ASPIC*$^+$ has three ways of attack: on an argument's uncertain premises (undermining attack), on the conclusion of a defeasible rule (rebutting attack) and on a defeasible rule itself (undercutting attack). However, to keep our present initial explorations as simple as possible, we assume that all premises are certain and that there are no undercutters, and leave the inclusion of undermining and undercutting attack to future research. Thus:

Definition 5 [Attacks] *A attacks B* iff *A rebuts B*, where A *rebuts* argument B (on B') iff $\texttt{Conc}(A) = -\varphi$ 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 6 [**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 arguments on the basis of $\mathcal{K}$ in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

The attack relation tells us which arguments are in conflict with each other. If an argument A *successfully attacks*, i.e., *defeats*, B, then A can be used as a counter-argument to B. Whether a rebutting attack succeeds as a defeat, depends on the argument ordering $\preceq$. In the following definition $A \prec B$ is defined as usual as $A \preceq B$ and $B \not\preceq A$.

Definition 7 [Defeat]. *A defeats B* if A rebuts B on B' and $A \not\prec B'$.

AFs are then generated from $SAFs$ by letting the attacks from an AF be the defeats from a SAF.

Definition 8 [***AFs* corresponding to *SAFs***] 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.

A nonmonotonic consequence notion can then be defined as follows. Let $T \in \{\text{complete, preferred, grounded, stable}\}$ and let $\mathcal{L}$ be from the AT defining SAF. A wff $\varphi \in \mathcal{L}$ is *sceptically T-justified* in SAF if φ is the conclusion of a sceptically T-justified argument, and *credulously T-justified* in SAF if φ is not sceptically T-justified and is the conclusion of a credulously T-justified argument.

Internal Argument Strength

In this section we define a probabilistic notion of 'internal' strength of arguments, where probabilistic generalisations are expressed as rules in $\mathcal{R}$. In doing so we abstract from the distinction between frequentist and subjective interpretations of probabilities; our approach is meant to equally apply to both interpretations. We will assume that premises and strict rules have strength 1 while defeasible rules have a strength less than 1 but greater than 0.5. The premise strengths are meant to be unconditional probabilities while the rule strengths stand for the conditional probability of a rule's consequent given the conjunction of its antecedents. Thus we give a probabilistic semantics to rules in line with the 'probability conditional theory' (cf. [p. 119](Adams 1999)). In philosophy this theory is controversial because of triviality results of (Lewis 1976) but these results assume that probability conditionals can be embedded in a propositional language, which is not the case for *ASPIC*$^+$ rules. The constraint that defeasible rules have strength greater than 0.5 is taken from (Pollock 1995)'s modelling of his statistical syllogism. The constraint agrees with the intuitive reading of defeasible rules $P \Rightarrow Q$ as 'if P then usually/typically Q', which is the reading we want to explore in this paper. Because of this reading, we do not allow that both a rule $S \Rightarrow \varphi$ and a rule $S \Rightarrow -\varphi$ is in $\mathcal{R}_d$.

Definition 9 A *probabilistic argumentation theory* is a tuple $PrAT = (AS, \mathcal{K}, s)$ where $(AS, \mathcal{K})$ is an argumentation theory such that if $S \Rightarrow \varphi \in \mathcal{R}_d$ then $S \rightarrow -\varphi \notin \mathcal{R}_d$ and s is a *rule-premise strength function* assigning a real number r where $0.5 < r < 1$ to all elements of $\mathcal{R}_d$ and $r = 1$ for all

[2] The square brackets make the presentation of examples more concise. They will be omitted if there is no danger for confusion.

elements of $\mathcal{K}$ and $\mathcal{R}_s$. All notions defined above for ATs are also defined for $PrAT$s.

At first sight, it would seem that the probabilistic strength of an argument can, as in (Timmer et al. 2017), be defined as the conditional probability of the argument's conclusion given its premises. However, this definition has some limitations. First, it regards the uncertainty of the premises as irrelevant for the strength of an argument. Second, it ignores the way in which the conclusion was derived from the premises. Consider the following example:

> Bart is Dutch, so Bart presumably likes cycling, so Bart is presumably healthy
> Bart is Dutch, so Bart presumably likes swimming, so Bart is presumably healthy

formalised in *ASPIC*$^+$ as

$$[Dutch \Rightarrow LikesCycling] \Rightarrow IsHealthy$$
$$[Dutch \Rightarrow Swimming] \Rightarrow IsHealthy$$

Since the rules applied in these arguments are defeasible, they may well have different strengths, since the health statistics may be different for swimmers than for runners.

Both limitations can be overcome by defining argument strength as the probability of the conjunction of all premises and conclusions of an argument. This is the definition that will be investigated in this paper. According to probability theory this notion of strength can be calculated with the general version of the chain rule for probability distributions. Consider any argument A with conclusions $C_1, \ldots, C_n$ (which include its premises) and consider any top-down breadth-first order of the argument's conclusions, starting with its final conclusion C_n, when regarding an argument as an acyclic directed graph with as nodes the premises and conclusion of A and as links any application of an inference rule in A. Then the general chain rule amounts to:

$$Pr(C_1 \wedge \ldots \wedge C_n) =$$

$$Pr(C_n|C_1 \wedge \ldots C_{n-1}) \times Pr(C_{n-1}|C_1 \wedge \ldots C_{n-2}) \times \ldots \times Pr(C_1)$$

Interestingly, this definition is a direct generalisation of (Hunter 2013)'s definition of the strength of classical-logic arguments as the probability of the conjunction of all their premises. In (Modgil and Prakken 2013)'s *ASPIC*$^+$ reconstruction of classical argumentation, classically valid arguments only apply strict inference rules, which in the present proposal have strength 1. So for the case of classical argumentation the chain rule definition reduces to Hunter's definition. This holds even if, as in our approach, uncertain premises are expressed as defeasible rules with empty antecedents, since then the *ASPIC*$^+$ reconstruction of classical argumentation still succeeds.

However, we are not yet there, since a practical problem with the chain rule definition of argument strength is that in many cases the conditional probabilities needed for this calculation will not be available. To see this, consider the example argument *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*, abstracted to

$$[A \Rightarrow B] \Rightarrow C$$

The chain rule implies that $Pr(C \wedge B \wedge A)$ equals $Pr(C|A \wedge B) \times Pr(B|A) \times Pr(A)$. However, the arguments were constructed with the rules $A \Rightarrow B$ and $B \Rightarrow C$, so the only probabilities that are given are $Pr(C|B)$ and $Pr(B|A)$ and $Pr(A)$. This will arguably often be the case in practice, where people (or artificial agents) often specify a rule $X \Rightarrow Y$ with strength $Pr(Y|X)$ without also specifying a more specific rule $X, Z \Rightarrow Y$ with strength $Pr(Y|X \wedge Z)$. So in practice the general chain rule will often not be applicable since the required probabilities will not be available.

Fortunately, there is a way out. In many practical contexts, agents will specify a rule $X \Rightarrow Y$ since, as far as they know, X is all that is needed to conclude Y. If they had thought that Z is also needed to conclude Y, they would instead have specified a rule $X, Z \Rightarrow Y$. More generally, it seems reasonable to assume that default reasoning often operates under the general assumption that the conclusion of an inference step in an argument is, conditionally on its premises (the antecedents of the applied rule), statistically independent of anything else. In our abstract example, this assumption amounts to saying that $Pr(C|A \wedge B)$ equals $Pr(C|B)$, which allows us to apply the chain rule. Now according to probability theory this assumption amounts to saying that the general chain rule is equivalent to the following version (where $\texttt{ant}(C_i)$ denotes the set of antecedents of the top rule of the subargument with conclusion C_i, except that for every $\varphi \in \mathcal{K}$, $Pr(\varphi|\texttt{ant}(\varphi)) = Pr(\varphi)$):

$$Pr(C_1 \wedge \ldots \wedge C_n) = \prod_{i=1}^{n} Pr(C_i|\texttt{ant}(C_i))$$

And this means that the probabilistic strength of an argument can be calculated by simply multiplying all premise- and rule strengths of any premise and rule in the argument.

Interestingly, the reduced version of the chain rule equals the chain rule for Bayesian networks (Jensen and Nielsen 2007). This also means that to be probabilistically correct in the present formal setting, we have to make counterparts of the assumptions made for Bayesian networks (BN) that each probabilistic variable occurs only once in the BN and that the BN is acyclic. The corresponding assumptions in the present setup are that no argument repeats a conclusion of one of its proper subarguments, that the set of all conclusions of an argument is assumed to be consistent, and that if two subarguments of an argument have the same conclusion, these arguments are the same. This justifies regarding arguments in the present setup as partial Bayesian networks, that is, as directed acyclic graphs with the nodes corresponding to the argument's conclusions (which includes its premises), with the links corresponding to the applications of the inference rules in the argument and with partially specified conditional probability tables (only partially, since in general if, say, we have a rule $X \Rightarrow Y$ in R_d, the corresponding rule $\neg X \Rightarrow Y$ does not have to be in R_d, in which case the conditional probability table for node Y cannot be fully specified).

This is not yet all there is to say, since the independence assumption that justifies the use of the reduced chain rule is clearly invalid in general. Consider the following example. *People who live in Denmark (D) usually speak Danish (S) but English-speaking university employees (E) who*

live in Denmark usually do not speak Danish. The implicit assumption of an argument using the first rule that $Pr(S|D) = Pr(S|D \wedge E)$ is contradicted by the second rule strength, since all rules are assumed to have strength greater than 0.5. Nevertheless, it seems reasonable to say that the assumption holds in the absence of information to the contrary. On this account, the independence assumption is defeasible and its defeasibility is captured by the possibility of an attacking argument. In our example we have, for a given English-speaking academic who lives in Denmark, the following two arguments which rebut each other.

$A_1 = \quad D \Rightarrow S$
$A_2 = \quad D, E \Rightarrow \neg S$

A variant of this example is the following well-known example: *Adults are usually married but students are usually adults and students are usually not married*. For a given student this yields the following arguments:

$A_1 = \quad [Student \Rightarrow Adult] \Rightarrow Married$
$A_2 = \quad Student \Rightarrow \neg Married$

Argument A_1 assumes that

$Pr(Married|Adult) =$
$Pr(Married|Adult \wedge Student)$

while argument A_2 assumes that

$Pr(\neg Married|Student) =$
$Pr(\neg Married|Adult \wedge Student)$.

Since all rules are assumed to have strength greater than 0.5, these two assumptions are jointly inconsistent.

There are several reasons to study the formal implications of the independence assumption. First, without this assumption, it seems hard to rationally justify the chaining of defeasible rules in arguments, as happens in our above examples and as happens all the time in everyday default reasoning. Furthermore, the assumption is the only way to utilise the given rule strengths in calculating argument strength, and, as argued above, these rule strengths are in many cases the only available probabilities. For these reasons we believe that the formal implications of the assumption are worth exploring, even if it might not apply in all contexts.

In light of all this, the internal strength of an argument ('internal' since it only depends on the argument itself) is now defined as follows. (The definition deliberately overloads the symbol s.)

Definition 10 [Internal argument strength] For any argument A on the basis of a $PrAT$ its *internal strength* $s(A)$ is defined as follows:

1. If $A \in \mathcal{K}$ then $s(A) = 1$
2. If $A = A_1, \ldots, A_n \Rightarrow \psi$ or $A_1, \ldots A_n \rightarrow, \psi$ then $s(A) = s(\texttt{Toprule}(A)) \times s(A_1) \times \ldots \times s(A_n)$

We now formally relate this notion of argument strength, which is defined over the set of arguments $\mathcal{A}$ of an argumentation theory defined by an argumentation system AS, to the notion of a probability distribution over models of the propositional language $\mathcal{L}^{pl}$ composed from the propositional atoms in the language $\mathcal{L}$ of AS. As noted by (Hunter 2013), such a probability distribution over models of $\mathcal{L}^{pl}$ is equivalent to a probability distribution over $\mathcal{L}^{pl}$. Hence we will from now on speak of a probability distribution over $\mathcal{L}^{pl}$, assuming that it is consistent in that it satisfies the axioms of probability theory. We denote classical entailment over $\mathcal{L}^{pl}$ with $\models$. Let furthermore $\mathcal{L}^{pt}$ stand for the set of all well-formulas formulas in probability theory over $\mathcal{L}^{pl}$. Then we denote deductive consequence over $\mathcal{L}^{pt}$ according to probability theory with $\vdash$. For any set of arguments S, we define the set $\Pi(S)$ of its probabilistic assumptions as follows:

For any set of arguments S the set $\Pi(S)$ is defined as $\{Pr(\texttt{cons}(r)|\bigwedge \texttt{ant}(r)) = x \mid r \in \texttt{Rules}(S)$ and $s(r) = x\} \cup \{Pr(\varphi) = y \mid \varphi \in \texttt{Prem}(S)$ and $s(\varphi) = y\} \cup \{C\}$, where C is the specific chain rule defined over $\texttt{Conc}(S)$.

The following proposition, which immediately follows from Definition 10 and the definition of Π, means that the strength definition for arguments is probabilistically well defined.

Proposition 11 For any argument A it holds that $s(A) = x$ iff $\Pi(\{A\}) \vdash Pr(\bigwedge \texttt{Conc}(\texttt{Sub}(A))) = x$.

Corollary 12 For any argument A it holds that $Pr(\texttt{Conc}(A)) \geq s(A)$.

Interestingly, if an argument has at most one defeasible rule, then the definition of internal strength equates (Pollock 1995)'s weakest-link principle:

Proposition 13 If A contains at most one subargument B with a defeasible top rule, then $s(A) = s(B)$.

However, for arguments with more than one defeasible rule this weakest-link principle does not satisfy Proposition 11.

Argument Strength and Conflict Resolution

We next study whether internal argument strength can be used to resolve conflicts between arguments by defining $A \preceq B$ iff $s(A) \leq s(B)$. At first sight, this would seem to be a natural idea. After all, many approaches to preference-based nonmonotonic logic and argumentation suggest that probabilistic strength of rules or premises can be a source of preferences. However, it turns out that using our notion of internal argument strength for this purpose is problematic. As illustrated in the previous section, internal argument strength has an important feature: different arguments can make mutually inconsistent probabilistic assumptions. For instance, in the speaking-Danish example the two arguments make contradictory assumptions about the probability of speaking Danish given that one is an English-speaking academic living in Denmark, and in the student-adult example the two arguments make contradictory assumptions about the probability of being married given that one is an adult who is a student. Argument A_1 implicitly assumes that $Pr(E|A) = Pr(E|A \wedge S)$ while argument A_2 implicitly assumes that $Pr(\neg E|S) = Pr(\neg E|E \wedge S)$. Given that rule strengths exceed 0.5, these two assumptions are jointly inconsistent. More generally the following can be shown:

Proposition 14 For any set S of arguments that is not conflict-free, it holds that $\Pi(E) \vdash \bot$.

Proof: Suppose A and B directly rebut each other and consider their top rules $S \rightarrow \varphi$ and $S' \rightarrow \neg\varphi$. Then $\Pi(S)$ contains both $Pr(\varphi|\bigwedge S) = Pr(\varphi|\bigwedge S \cup S')$ and $Pr(\neg\varphi|\bigwedge S') = Pr(\neg\varphi|\bigwedge S \cup S')$. But this contradicts that rule strengths exceed 0.5. □

The converse does not hold. For a counterexample, consider the following arguments:

A_1: $[p \rightarrow q] \Rightarrow r$
A_2: $p, [p \rightarrow q] \Rightarrow r$

Here A_1 uses defeasible rule $q \Rightarrow r$ while A_2 uses defeasible rule $p, q \Rightarrow r$. If the two defeasible rules have different strengths s_1 and s_2, then $\Pi(A_1)$ implies $Pr(r|p \wedge q) = s_1$ while $\Pi(A_2)$ implies $Pr(r|p \wedge q) = s_2 \neq s_1$. Yet the arguments do not attack each other.

All this implies that resolving conflicts between defeasible arguments by just resorting to their internal argument strengths is problematic, since the arguments assume different and jointly inconsistent probability distributions. Consider again the speaking-Danish example and assume that $s(D \Rightarrow S) > s(D, E \Rightarrow \neg S)$. For instance, 90% of the people who live in Denmark speak Danish while only 75% of the English-speaking academics who live in Denmark does not speak Danish. Do we then want to accept argument A_1 that our English-speaking academic who lives in Denmark presumably speaks Danish? Of course not, since the second statistic is about a more specific class than the first one, so the generally accepted principle to adopt the statistic about the more specific reference class requires that we instead conclude that our person presumably does not speak Danish. For this reason, probability-based comparisons between arguments should, either explicitly or implicitly, involve a kind of specificity principle, which implies that conflicts between defeasible arguments cannot in general be resolved by just resorting to their probabilistic argument strengths. For example, (Pollock 1995)'s so-called subproperty defeater of his statistical syllogism is such a specificity principle.

These observations can be extended to cases in which the attacking arguments 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*. This yields the following rebutting arguments.

A_1: $Quaker \Rightarrow Pacifist$
A_2: $Republican \Rightarrow \neg Pacifist$

Here argument A_1 implicitly assumes that

$Pr(Pacifist|Quaker) =$
$Pr(Pacifist|Quaker \wedge Republican)$

while argument A_2 implicitly assumes that

$Pr(\neg Pacifist|Republican) =$
$Pr(\neg Pacifist|Quaker \wedge Republican)$

which is inconsistent given that rule strengths exceed 0.5. Suppose, furthermore, that $Pr(Pacifist|Quaker) < Pr(\neg Pacifist|Republican)$. Are we then forced to accept argument A_2? No, since what we want to know is $Pr(Pacifist|Quaker \wedge Republican)$ and this probability may be assumed independent of the former probabilities. So in the absence of information about $Pr(Pacifist|Quaker \wedge Republican)$ it seems just as rational to regard both arguments as defensible as preferring A_2 over A_1. For this reason we will below explore an approach which allows this example to have multiple extensions. This approach does not preclude that argument strengths are used to derive argument preferences (this might still be a useful heuristic in some practical cases) but it does not force this either.

Our analysis thus far suggests that conflict resolution in the setting of probabilistic default reasoning may be seen as the *adjustment* of probabilities to obtain a consistent probability distribution over $\mathcal{L}^{pl}$ that can be used to resolve conflicts between arguments. We will explore this approach in the penultimate section. In the present section we study a different question, namely, whether any rational way to resolve the conflict between arguments while leaving their internal strengths as they are yields a consistent set of probabilistic assumptions in that for any resulting extension the set of rule and premise strengths of any argument in the extension is probabilistically consistent. This does not hold in general, as illustrated by the above counterexample to the converse of Proposition 14. However, the consistency result can be proven on the assumption that $\mathcal{R}$ does not contain two different rules with the same consequent, or formally: if $r, r' \in \mathcal{R}$ and $\texttt{cons}(r) = \texttt{cons}(r')$ then $r = r'$. This assumption will below be called the **accrual assumption**.

We now consider any argumentation theory AT that defines a structured argumentation framework SAF that satisfies the rationality postulate of indirect consistency in that for all complete extensions E of the AF corresponding to SAF the set $\texttt{Conc}(E)$ is indirectly consistent (which implies that it is directly consistent). That is, we abstract from the specific way the argument ordering is defined, as long as it makes the SAF satisfy indirect consistency. The argument ordering may in part be based on a notion of specificity, as recommended above, but we also abstract from this. Then for such SAFs the following can be proven.

Theorem 15 Let AT be any argumentation theory satisfying the accrual assumption and let SAF be a structured argumentation framework defined by AT satisfying indirect consistency. For any complete extension $E = \{A_1, \ldots, A_n\}$ of SAF it holds that $\Pi(E) \nvdash \bot$.

Proof: Note first that any complete extension E corresponds to a partial Bayesian network (BN) in the manner explained above, with as nodes all formulas occurring as antecedent or consequent in $\texttt{Rules}(E)$, the links corresponding to applications of any rule in $\texttt{Rules}(E)$ and the conditional probability tables filled in as far as possible with the premise strengths of nodes that are elements of $\mathcal{K}$ and the rule strengths for nodes that are consequents of rules in $\mathcal{R}$. That the network is indeed a well-defined partial BN follows from our assumptions about Definitions 1 and 3 that no element of $\mathcal{K}$ occurs in the consequent of any rule in $\mathcal{R}$, from the assumptions in Definition 4 that arguments have indirectly consistent conclusion sets and do not have different subarguments for the same conclusion and from the present assumptions that AT satisfies the accrual assumption and that

E is indirectly consistent.

Then recall that the theory of Bayesian networks (Jensen and Nielsen 2007) implies that the BN expresses a probability distribution over its set of variables (which equals $\texttt{Conc}(E)$) and, moreover, that for any variable V with value v in the BN the probability $Pr(V = v)$ can be calculated with the specific version of the chain rule, where for any variable V the set $\texttt{ant}(C)$ is the set of all parents of V. Since $\texttt{Conc}(E)$ is indirectly consistent and we only calculate for expressions $v = \mathit{true}$ (where $V \in \mathcal{L}$), such expressions can without danger of confusion be shortened to V.

We then prove with induction on the graph structure of the BN that the probability for a given variable V computed by the chain rule for Bayesian networks equals the strength of the argument corresponding to the subgraph consisting of V and all its ancestors in BN (henceforth denoted as $\texttt{arg}(V)$). Then the result follows from the fact that each BN expresses a probability distribution.

For the base case, note that for any node V in BN without ancestors it holds that $V \in \texttt{Prem}(E)$, so $V \in \mathcal{K}$ so $s(V) = Pr(V)$ as specified in the probability table for V in the BN. Let $\texttt{anc}(V)$ for any BN node V be the set of all ancestors of V in the BN.

The induction hypothesis is that for any given node V of the BN and any parent V' of V in BN the probability $Pr(\bigwedge\{V'\} \cup \texttt{anc}(V'))$ equals $s(\texttt{arg}(V'))$.

Then for the induction step consider any node V in the BN. According to the chain rule $Pr(\bigwedge\{V\} \cup \texttt{anc}(V))$ equals the product of $Pr(V|\bigwedge \texttt{par}(V))$ as specified in the conditional probability table for V and all probabilities $Pr(\bigwedge\{V'\} \cup \texttt{anc}(V'))$. But given the induction hypothesis that the latter probabilities correspond to $s(\texttt{arg}(V'))$ and the fact that $Pr(V|\bigwedge \texttt{par}(V)) = s(\texttt{par}(V) \Rightarrow V)$, it follows that $Pr(\bigwedge\{V\} \cup \texttt{anc}(V))$ equals $s(\texttt{arg}(V)$, from which the result follows. □

Internal Argument Strength and Hunter & Thimm's Rationality Conditions

We next investigate how internal argument strength fares with (Hunter and Thimm 2017)'s rationality conditions for epistemic abstract probabilistic frameworks. Given an abstract argumentation framework $\langle \mathcal{A}, \mathit{attack} \rangle$ where $\mathcal{A}$ is finite, they assume a probability distribution which to each subset of $\mathcal{A}$ assigns a real number between 0 and 1. They then define the probability of an argument $A \in \mathcal{A}$ as

$$Pr(A) = \sum_{A \in S \subseteq \mathcal{A}} Pr(S)$$

At first sight, this definition would seem to prevent instantiation with the above-made proposal, since we do not define probabilities of sets of arguments. However, according to Proposition 12 of (Hunter and Thimm 2017) any assignment of numerical strengths between 0 and 1 to arguments in a finite set S can be extended to a probability function on the powerset of S. Moreover, Hunter & Thimm formulate all their rationality conditions in terms of the probabilities of individual arguments. For these reasons (and since we assume that $\mathcal{R}$ and $\mathcal{K}$ are finite) our above proposal is fully within Hunter & Thimm's formal framework.

(Hunter and Thimm 2017)'s rationality conditions are as follows (below A^- denotes the set of all attackers of A):

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

RAT Pr is *rational* if for every $A, B \in \mathcal{A}$, if A attacks B and $Pr(A) > 0.5$, then $Pr(B) \leq 0.5$.

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 all unattacked $A \in \mathcal{A}$.

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 $A \in \mathcal{A}$.

As observed by (Prakken 2017), these conditions are, when applied to *ASPIC*$^+$ instantiations, ambiguous between *ASPIC*$^+$'s notions of attack and defeat and also between direct attack on an argument's final conclusion and indirect attack on a proper subargument. So for *ASPIC*$^+$ these properties must be independently verified for attack and defeat and for both their direct and general versions.

It turns out that all properties fail in general for all cases. A counterexample to COH and RAT is A: $q \Rightarrow p$ and B: $r \Rightarrow \neg p$ (note that both top rules are assumed to have strength > 0.5). A counterexample to INV, SOPT and OPT is A: $[e_1 \Rightarrow p] \Rightarrow q$ and B: $e_2 \Rightarrow \neg p$ where all three rules have strength 0.6. Finally, a counterexample to SFOU and FOU is an unattacked argument chaining two rules that both have strength 0.7. For some special cases positive results hold. INV and FOU hold for strict arguments, since strict arguments have no attackers. SFOU, SOFT and OPT hold if both A and B apply at most one defeasible rule, since then their strengths exceed 0.5. One reason for the failures is that the strength function s is not based on a single probability distribution on $\mathcal{L}^{pl}$ since different arguments may make jointly inconsistent probabilistic assumptions. Another reason is that several properties implicitly make assumptions on the nature of arguments. For example, INV and FOU are arguably meant for deductively valid arguments only.

The question arises whether the negative results indicate weaknesses of the present approach or instead of (Hunter and Thimm 2017)'s rationality conditions. An answer to this question will be postponed to the concluding section, except for the failure of SFOU, which might at first sight seem counterintuitive. However, note that $s(A) < 0.5$ does not imply that $Pr(\texttt{Conc}(A)) < 0.5$, since the probability of a conjunct may exceed the probability of the conjunction in which it appears. Moreover, in practical applications an argument can be very uncertain even if no attacker can be constructed, for instance, if it chains several defeasible rules. In such cases the probabilistic strength of an argument gives useful additional information to an argument evaluator. For example, the evaluator might in the end decide to only accept the arguments in extensions that have strength greater than

0.5. At the end of the next section we will further explore this idea.

Dialectical Argument Strength based on a Single Probability Distribution

In this section we study the case of a single consistent probability distribution Pr over $\mathcal{L}^{pl}$, which can be used to resolve conflicts between arguments by defining $A \preceq B$ as $Pr(\texttt{Conc}(A)) \leq Pr(\texttt{Conc}(B))$. There are two reasons for studying this case. First, it allows us to see more clearly some of the assumptions underlying Hunter & Thimm's rationality conditions. Second, there may be contexts in which it is feasible to adjust the internal arguments strengths in a way compliant with a single probability distribution over $\mathcal{L}^{pl}$ (although the feasibility of this is doubtful in general, since as argued above, in many applications only the given rule strengths will be available).

Before studying the idea of consistency adjustments, it should be noted that this idea is not the same as (Hunter and Thimm 2017)'s study of restoring consistency of what they call 'contradictory' probability assessments, since they define contradictory probability assessments as probability functions on the powerset A^2 of $\mathcal{A}$ in a $PrAF$ that do not satisfy a given subset of their rationality conditions. Here it is also relevant that our notion of (in)consistent probability functions is not at the level of $\mathcal{A}^2$ of a $PrAF$ but at the level of the propositional language $\mathcal{L}^{pl}$ generated by the language $\mathcal{L}$ of an AT. A probability assessment at the latter level may well be consistent (according to probability theory) while the induced probability assessment at the former level is inconsistent (in Hunter & Thimm's sense) and vice versa. Below we assume that the probability assessment at the level of $\mathcal{L}^{pl}$ is consistent according to probability theory and investigate the consequences at the level of $\mathcal{A}^2$.

The new notion of argument strength generated by a single probability distribution over $\mathcal{L}^{pl}$ will henceforth be called *dialectical argument strength* and will be denoted with d. As with $s(A)$ above we assume that $d(A)$ equals $Pr(\bigwedge \texttt{Conc}(A))$ but we initially abstract from specific ways to define Pr: all we initially assume is that Pr is a probability distribution over $\mathcal{L}^{pl}$.

Proposition 16 For any $PrAT = (AS, \mathcal{K}, s)$ where $AS = (\mathcal{L}, \mathcal{R})$, let $\mathcal{L}^{pl}$ be a propositional language defined from the atoms in $\mathcal{L}$ and let Pr be a probability distribution over $\mathcal{L}$. Let $d(A)$ for any $A \in \mathcal{A}$ be defined as $Pr(\bigwedge \texttt{Conc}(A))$. Then:

1. If A rebuts B then $d(A) \leq 1 - d(B)$.
2. If $A \in \texttt{Sub}(B)$ then $d(B) \leq d(A)$.
3. d satisfies COH and RAT for direct and indirect attack and defeat.

Proof: (1) follows since $\texttt{Conc}(A) \cup \texttt{Conc}(B)$ is propositionally inconsistent by definition of rebutting attack. (2) follows since if $A \in \texttt{Sub}(B)$ then $\bigwedge \texttt{Conc}(B) \models \bigwedge \texttt{Conc}(A)$. Furthermore, (3) follows for direct attack and defeat by (1) and then for indirect attack and defeat by (2). □

We next consider probability distributions on $\mathcal{L}^{pl}$ that assign probabilities in a way similar to in Definition 9.

Proposition 17 Let in addition to everything stated in Proposition 16 Pr be such that $Pr(\varphi) = 1$ for all $\varphi \in \mathcal{K}$, that $Pr(\varphi | \bigwedge S) = 1$ whenever $S \rightarrow \varphi \in \mathcal{R}_s$ and that $Pr(\varphi | \bigwedge S) < 1$ whenever $S \rightarrow \varphi \in \mathcal{R}_d$. Then:

1. $d(A) = 1$ iff A is strict.
2. d satisfies SFOU, FOU, INV for strict arguments for direct and indirect attack and defeat.

Proof: For the only-if part of (1), consider a strict rule $r = S \rightarrow \varphi$ such that $S \subseteq \mathcal{K}$. Since $s(r) = 1$ we have by definition of conditional probabilities that $Pr(\bigwedge S \wedge \varphi)/Pr(\bigwedge S) = 1$. Moreover, $Pr(\bigwedge S) = 1$, so $Pr(\bigwedge S \wedge \varphi) = 1$. Then the general result follows by induction on the structure of a strict argument. The if-part of (1) follows since if A is defeasible, it contains a defeasible rule $r = S \Rightarrow \varphi$ and since $s(r) < 1$ we have that $Pr(\bigwedge S \wedge \varphi)/Pr(\bigwedge S) < 1$, so $Pr(\bigwedge S \wedge \varphi) < 1$. Property (2) follows for SFOU and FOU from property (1), for INV from property (1) and the fact that strict arguments have no attackers and for OPT since strict arguments have no attackers. □

For all other rationality conditions the counterexamples given in the previous section still hold.

Next we informally note two results given further assumptions. First, if every unattacked argument has the same dialectical as internal strength, then SFOU holds for arguments with at most one defeasible rule, since the rule's strength exceeds 0.5. Second, if Pr satisfies the independence assumption underlying internal strength, so that dialectical strength is computed as internal strength (although defeasible rules can now have strength 0.5 or less), then the weakest-link principle holds for the same special case as for internal strength (see Proposition 13). Further properties may be provable under further assumptions but we leave this for future research.

We can conclude that whether Hunter & Thimm's rationality conditions hold depends on various assumptions on the underlying probability assignment to $\mathcal{L}^{pl}$ and the nature of the arguments and attacks.

Finally, we come back to the suggestion at the end of the previous section to only accept arguments in an extension with strength greater than 0.5. When argument strength is based on a single probability distribution over $\mathcal{L}^{pl}$, then (Hunter 2013)'s notion of an *epistemic extension* of a probabilistic abstract argumentation framework can be used, which is defined as $\{A \in \mathcal{A} | Pr(A) > 0.5\}$. The following can be shown for dialectical strength:

Proposition 18 For any $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ with $A \preceq B$ iff $d(A) \leq d(B)$, any $A \in \mathcal{A}$ such that $d(A) > 0.5$ is in E.

Proof: Consider any A such that $d(A) > 0.5$ and consider any rebuttal B of A on its subargument A'. Then $d(A') \geq d(A)$ so by satisfaction of COH it holds that $d(B) < d(A)$, so B does not defeat A'. But then B does not defeat A. So A is admissible with respect to E, so $A \in E$. □

So the policy to accept only those arguments that have dialectical strength > 0.5 only accepts arguments that are in

all extensions. Note that the same result does not hold for internal strength, since then arguments with internal strength > 0.5 might not be in some or all extensions.

Conclusion and Related Work

This paper aimed to clarify the epistemic approach to probabilistic abstract argumentation by relating it to an account of probabilistic structured argumentation. In addressing these aims, we have established relations between probabilistic argumentation and probability theory, including the theory of Bayesian networks. Our account is based on the ideas that the probability of an argument is the probability of the conjunction of all its premises and conclusions and that arguments implicitly make probabilistic independence assumptions. Together these ideas imply that the probabilistic assumptions of conflicting arguments are jointly inconsistent. This in turn implies that resolving conflicts between arguments in terms of their internal strengths is problematic.

The latter observation gives one reason why internal argument strength fails to satisfy several of (Hunter and Thimm 2017)'s rationality conditions for epistemic probabilistic argumentation, since these conditions arguably assume a single probability distribution over the language over which arguments are constructed. To also capture this assumption we distinguished between an argument's internal and dialectical strength and we showed that dialectical strength better respects Hunter & Thimm's rationality conditions than internal strength (though not fully). However, it is arguably harder to apply in practice, since it cannot utilise the given rule strengths in the way internal argument strength does.

Our approach is 'bottom up' in that the formalism is based on an arguably sensible account of probabilistic structured argumentation, which satisfies properties that arguably indicate that the formalism is well-behaved. If these claims about sensibility and well-behavedness are justified, then the partly negative results on the satisfaction of Hunter & Thimm rationality conditions indicate that these conditions cannot all be regarded as minimum rationality conditions for epistemic probabilistic argumentation. More generally, we have shown that modelling probabilistic argumentation at the level of structured argumentation can yield insights that may not be obtained when remaining at the abstract level.

Two topics for future research are generalising our approach to the case where the strict rules encode classical logic, and including undermining and undercutting attacks. Note that with undercutters several results proven above (e.g. on satisfaction of RAT and COH by d) cease to hold, since the conclusion sets of an undercutter and its target may be jointly consistent. This illustrates another reason for the partly negative results on Hunter & Thimm's rationality conditions, namely, that these conditions implicitly make assumptions on the nature of the arguments and attacks.

Some related work was discussed throughout this paper. As noted above, our idea to equate argument strength with the probability of the conjunctions of all premises and conclusions of an argument generalises (Hunter 2013)'s idea to define argument strength for classical-logic argumentation as the probability of the conjunction of the premises of an argument. Apart from Hunter's work and (Prakken 2017), we do not know of any other (epistemic) approach that relates structured to abstract probabilistic argumentation. (Dung and Thang 2010) propose an extension of (Dung 1995)'s abstract frameworks with probability and then instantiate it with assumption-based argumentation. They add to AFs a set W of worlds, where each world is a set of arguments. In each world an argument A has a probability that it is accepted (according to a given semantics). The overall probability that A is accepted is then the sum of the probabilities in each world that A is accepted. In the structured part, each world is a set of assumptions, determining which assumption-based arguments can be constructed in that world. The probabilities are then defined on the assumptions. Interesting as this work is, it concerns the constellation approach, since worlds may contain different AFs. So this work is irrelevant for the present paper.

There is also related work in structured argumentation using alternatives to standard probability theory. (Pollock 1995) uses his 'nomic' theory of probability to assign strengths to inference rules and he defines the strength of arguments with a weakest-link principle. He then uses argument strength to resolve attacks into defeats in a way compliant with Dung's theory of abstract argumentation frameworks. (Verheij 2014) revisits this approach in the light of probability theory and classical logic. (Pollock 2002) deviates from Dung's theory by using the rule strengths for defining a gradual notion of argument justification. (Chesñevar et al. 2004) use possibilistic logic in the context of defeasible logic programming (Garcia and Simari 2004). Possibilistic strengths are added to rules, which are propagated through arguments according to possibilistic logic. Then the propagated strengths are used to resolve attacks into defeats. This idea is not related to standard or probabilistic abstract argumentation. Neither (Pollock 1995) nor (Chesñevar et al. 2004) address the possibility of mutually inconsistent probability assumptions of conflicting arguments.

Recent work on gradual argumentation semantics (reviewed by (Baroni, Rago, and Toni 2018)) may also be relevant to our approach. Some of this work distinguishes between notions of 'base' and 'dialectical' strength of arguments, where only the latter depends on an argument's (support or attack) relations with other arguments. It would, for instance, be interesting to study the extent to which the "basic ideas" discussed by (Baroni, Rago, and Toni 2018) (e.g. that basic strength equates dialectical strength just in case an argument is not attacked) are satisfied by particular ways to adjust internal to dialectical probabilistic argument strength.

Relevant work outside argumentation is the body of work on nonmonotonic probabilistic semantics for defeasible conditionals (reviewed in (Beierle 2016)). This work yields many interesting insights but, unlike argumentation approaches, conceals the interaction between reasons for and against a conclusion in the semantics. As noted earlier by (Caminada 2004, p. 96), this makes conclusions sometimes hard to explain. Moreover, to the best of our knowledge, none of this work deals with inconsistent underlying probability assessments, as in our notion of internal argument strength. In future research it would be interesting to investigate how both approaches can contribute to each other.

References

Adams, E. 1999. *A Primer of Probability Logic*. Stanford, CA: CSLI Publications.

Baroni, P.; Rago, A.; and Toni, F. 2018. How many properties do we need for gradual argumentation? In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*, 1736–1743.

Beierle, C. 2016. Systems and implementations for solving reasoning problems in conditional logics. In Gyssens, M., and Simari, G., eds., *Proceedings of the 9th International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2016)*, number 9616 in Springer Lecture Notes in Computer Science, 83–94. Berlin: Springer Verlag.

Caminada, M. 2004. *For the Sake of the Argument. Explorations into Argument-based Reasoning*. Doctoral dissertation Free University Amsterdam.

Chesñevar, C.; Simari, G.; Alsinet, T.; and Godo, L. 2004. A logic programming framework for possibilistic argumentation with vague knowledge. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 76–84.

Dung, P., and Thang, P. 2010. Towards (probabilistic) argumentation for jury-based dispute resolution. In Baroni, P.; Cerutti, F.; Giacomin, M.; and Simari, G., eds., *Computational Models of Argument. Proceedings of COMMA 2010*. Amsterdam etc: IOS Press. 171–182.

Dung, P. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence* 77:321–357.

Garcia, A., and Simari, G. 2004. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming* 4:95–138.

Grooters, D., and Prakken, H. 2016. Two aspects of relevance in structured argumentation: minimality and paraconsistency. *Journal of Artificial Intelligence Research* 56:197–245.

Hunter, A., and Thimm, M. 2016. On partial information and contradictions in probabilistic abstract argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference*, 53–62. AAAI Press.

Hunter, A., and Thimm, M. 2017. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research* 59:565–611.

Hunter, A. 2013. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning* 54:47–81.

Hunter, A. 2014. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning* 55:607–638.

Jensen, F., and Nielsen, P. 2007. *Bayesian Networks and Decision Graphs*. New York: Springer Verlag, second edition.

Lewis, D. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review* 85:297–315.

Li, H.; Oren, N.; and Norman, T. 2012. Probabilistic argumentation frameworks. In Modgil, S.; Oren, N.; and Toni, F., eds., *Theory and Applications of Formal Argumentation. First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, number 7132 in Springer Lecture Notes in AI, 1–16. Berlin: Springer Verlag.

Modgil, S., and Prakken, H. 2013. A general account of argumentation with preferences. *Artificial Intelligence* 195:361–397.

Modgil, S., and Prakken, H. 2018. Abstract rule-based argumentation. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*, volume 1. London: College Publications. 73–141.

Pollock, J. 1995. *Cognitive Carpentry. A Blueprint for How to Build a Person*. Cambridge, MA: MIT Press.

Pollock, J. 2002. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence* 133:233–282.

Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1:93–124.

Prakken, H. 2017. On relating abstract and structured probabilistic argumentation: a case study. In *Proceedings of the 14th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 17)*, number 10369 in Springer Lecture Notes in AI, 69–79. Berlin: Springer Verlag.

Rienstra, T. 2012. Towards a probabilistic Dung-style argumentation system. In *Proceedings of the First International Conference on Agreement Technologies*, 138–152.

Timmer, S.; Meyer, J.-J.; Prakken, H.; Renooij, S.; and Verheij, B. 2017. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning* 80:475–494.

Verheij, B. 2014. Arguments and their strength: Revisiting Pollock's anti-probabilistic starting points. In Parsons, S.; Oren, N.; Reed, C.; and Cerutti, F., eds., *Computational Models of Argument. Proceedings of COMMA 2014*. Amsterdam etc: IOS Press. 433–444.

Vreeswijk, G. 1997. Abstract argumentation systems. *Artificial Intelligence* 90:225–279.

Wu, Y., and Podlaszewski, M. 2015. Implementing crash-resistence and non-interference in logic-based argumentation. *Journal of Logic and Computation* 25:303–333.