Epistemic Attack Semantics

Matthias THIMM [a], Sylwia POLBERG [b], and Anthony HUNTER [b]
[a] *University of Koblenz-Landau, Germany*
[b] *University College London, UK*

Abstract. We present a probabilistic interpretation of the plausibility of attacks in abstract argumentation frameworks by extending the epistemic approach to probabilistic argumentation with probabilities on attacks. By doing so we also generalise the previously proposed *attack semantics* by Villata et al. to the probabilistic setting and provide a fine-grained assessment of the plausibility of attacks. We also consider the setting where partial probabilistic information on arguments and/or attacks is given and missing probabilities have to be derived.

Keywords. abstract argumentation, probabilistic argumentation, epistemic argumentation

1. Introduction

The probabilistic approaches to computational models of argumentation combine probabilistic reasoning capabilities and formalisms for dealing with the interaction between arguments and counterarguments, which is an essential approach to obtain a realistic model of argumentation [12]. Probabilistic approaches to abstract argumentation can be divided into two groups: the constellations approach [11,8] and the epistemic approach [14,10]. The former is concerned with an extrinsic interpretation of probability and the probabilities on arguments and attacks are used to determine with what likelihood these components actually appear in the argument graphs. In contrast, the epistemic approach considers an intrinsic interpretation of probability and defines the probabilities of arguments as *degrees of belief*. The idea behind the epistemic approach is to allow for a more fine-grained assessment of the acceptability status of an argument by allowing for the whole spectrum of real values between 0 (complete rejection of an argument) and 1 (complete acceptance of an argument). This motivation is similar to *graded semantics* for abstract argumentation, see e. g., [3,1], but the framework of epistemic argumentation explicitly focuses on probabilistic interpretations of numbers rather than leaving the meaning of the values abstract, which results in significantly different semantics.

In this work, we focus on the epistemic approach to probabilistic argumentation and, in particular, the question of deriving degrees of belief from partially specified information [9,10]. We also extend our previous works by considering probabilistic beliefs on attacks with the intuition that a high belief in an attack makes the attack *effective* while low belief in attack means that the attack could almost be ignored. In doing so, we provide an alternative to the approach presented in [13], which treated the belief we have in an attack either as a signal whether the attacker should be considered or as a proxy for the belief in the attacker. However, the work in [13] did not consider how the belief in an attack can provide continuous modulation to the effect that the source argument of the conflict may have on the target, and does not handle scenarios such as in Example 1.

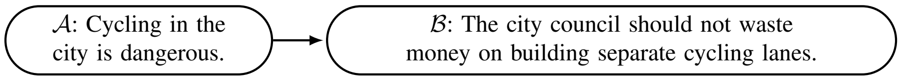

Figure 1. A simple argumentation framework

Example 1. Consider the framework depicted in Figure 1 and imagine a group of colleagues discussing cycling lanes. We observe that arguments $\mathcal{A}$ and $\mathcal{B}$ are enthymemes, hence participants may decode and agree with them differently. John strongly agrees with $\mathcal{A}$ because he believes that cars pose serious threats to cyclists. He strongly agrees with $\mathcal{A}$ being a good counterargument for $\mathcal{B}$, and opposes $\mathcal{B}$. Robin, although also strongly agreeing with $\mathcal{A}$, considers this problem to be more complex and being a combination of both the threat of being hit by a car and breathing in fumes. Consequently, building cycling lanes addresses only one of Robin's concerns, and although Robin still considers $\mathcal{A}$ to be a counterargument for $\mathcal{B}$, she is not as adamant as John and only slightly disagrees with $\mathcal{B}$. Finally, Morgan decodes $\mathcal{A}$ as "Cycling in the city is dangerous and therefore one should not cycle the city". Consequently, $\mathcal{A}$, although strongly agreed to, is no longer really perceived as much of a counterargument for $\mathcal{B}$, and Morgan strongly believes that the city would be wasting money by building lanes. We observe that while all colleagues agree with $\mathcal{A}$ to the same degree, they have divergent opinions on the effect it has on $\mathcal{B}$.

The current paper addresses this issue. Our method is partly based on the work [15] which defined *attack semantics* for Dung's abstract argumentation frameworks and proposed a crisp notion of acceptability of an attack. We investigate some simple rationality postulates for our setting, i. e., properties that describe how the beliefs of arguments and attacks should be (reasonably) linked to each other in the global view of an abstract argumentation framework. The central paradigm here is that high belief in an argument and high belief in an attack originating from that argument, lead to low belief in the argument pointed to by this attack. In formalising this and other intuitions, we obtain a notion of justifiable probability functions that give reasonable degrees of beliefs to arguments and attacks. In the next step, we consider the computational question of completing partial probabilistic beliefs, thus extending [9,10] to our generalized setting. More precisely, given degrees of beliefs in some arguments and attacks, we ask what reasonable inferences for the degrees of beliefs in the remaining arguments and attacks can be made.

In summary, the contributions of this paper are as follows:

1. We develop a notion of reasonable probability assignments by formalising desirable properties through rationality postulates, in particular extending previous (crisp) attack semantics in a principled way (Section 4).
2. We study the problem of completing incomplete probabilistic information from an inferential perspective (Section 5).

Furthermore, Sections 2 and 3 recall needed background information, Section 6 discusses related works, and Section 7 concludes our paper.

2. Preliminaries

Abstract argumentation frameworks do not presuppose any internal structure of an argument and focus on the interactions between them, particularly in context of attack [5]:

Definition 1. A (Dung's) *abstract argumentation framework* AF is a pair $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ where Arg is a set of arguments and $\rightarrow$ is an attack relation $\rightarrow \subseteq \mathsf{Arg} \times \mathsf{Arg}$.

If $(\mathcal{A},\mathcal{B}) \in \rightarrow$, we usually simply write $\mathcal{A} \rightarrow \mathcal{B}$ and read it as "argument $\mathcal{A}$ attacks argument $\mathcal{B}$". For $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$, we use $\mathcal{A} \in \mathsf{AF}$ and $(\mathcal{A},\mathcal{B}) \in \mathsf{AF}$ as shorthands for $\mathcal{A} \in \mathsf{Arg}$ and $(\mathcal{A},\mathcal{B}) \in \rightarrow$, respectively. Note that we only consider finite argumentation frameworks here, i. e., such that the set Arg is finite. We represent frameworks with directed graphs, where nodes stand for arguments and edges model the attack relation.

Semantics of abstract argumentation frameworks are often defined in terms of extensions [5,4,2]. An *extension* E of an argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ is a set of arguments $E \subseteq \mathsf{Arg}$ that gives some coherent view on the argumentation underlying a given AF. Here, we focus only on the four classical semantics of [5], namely grounded, complete, preferred, and stable semantics. For a set of arguments $S \subseteq \mathsf{Arg}$ let $S^- = \{\mathcal{B} \mid \exists \mathcal{A} \in S : \mathcal{B} \rightarrow \mathcal{A}\}$ denote the set of attackers of S and let $S^+ = \{\mathcal{B} \mid \exists \mathcal{A} \in S : \mathcal{A} \rightarrow \mathcal{B}\}$ denote the set of attacked arguments of S.

Definition 2. Let $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ be an argumentation framework. An argument $\mathcal{A} \in \mathsf{Arg}$ is *acceptable* with respect to a set of arguments $F \subseteq \mathsf{Arg}$ iff for every $\mathcal{B} \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$ there is $\mathcal{A}' \in F$ with $\mathcal{A}' \rightarrow \mathcal{B}$. A set of arguments $E \subseteq \mathsf{Arg}$ is:

1. *conflict-free* iff there there are no $\mathcal{A},\mathcal{B} \in E$ with $\mathcal{A} \rightarrow \mathcal{B}$.
2. *admissible* iff it is conflict-free and every $\mathcal{A} \in E$ is acceptable w.r.t. E.
3. a *complete extension* (CO) iff it is admissible and there is no $\mathcal{A} \in \mathsf{Arg} \setminus E$ which is acceptable w.r.t. E.
4. a *grounded extension* (GR) iff it is complete and E is minimal w.r.t. $\subseteq$.
5. a *preferred extension* (PR) iff it is complete and E is maximal w.r.t. $\subseteq$.
6. a *stable extension* (ST) iff it is complete and $E \cup E^+ = \mathsf{Arg}$.

If E is some extension we say that each $\mathcal{A}$ is accepted wrt. E. We observe that with the exception of the grounded semantics, the extensions are not necessarily uniquely determined. Furthermore, a stable extension might not necessarily exist [5]. For the remainder of the paper we use σ to denote any semantics of GR, CO, PR, ST.

3. The Epistemic Approach to Probabilistic Argumentation

In this section we briefly recall the standard approach to epistemic argumentation that only considers beliefs in arguments [7,9,12,14]. The epistemic approach is centered around the probability distributions over the sets of arguments:

Definition 3. A *probability function* P on AF is a function $P : 2^{\mathsf{Arg}} \rightarrow [0,1]$ with $\sum_{M \in 2^{\mathsf{Arg}}} P(M) = 1$. Let $\mathcal{P}_{\mathsf{AF}}$ denote the set of all such functions.

From this, we derive the probability of a single argument and interpret it as the belief that an agent has in it, which can, for example, be seen as the degree to which the agent believes the premises, the conclusion and that the conclusion follows from these premises. Formally speaking, the *marginal belief in an argument* $\mathcal{A} \in \mathsf{AF}$, denoted as $P(\mathcal{A})$, is defined via:

$$P(\mathcal{A}) = \sum_{\mathcal{A} \in M \in 2^{\mathsf{Arg}}} P(M)$$

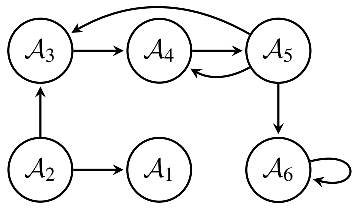

Figure 2. A simple argumentation framework

	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$
P_1	0	1	0	0.5	0.5	0.5
P_2	0	1	0	1	0	0.5
P_3	0	1	0	0	1	0

Table 1. Probability distributions from Example 2.

We say that an agent believes an argument $\mathcal{A}$ to some degree when $P(\mathcal{A}) > 0.5$, disbelieves an argument to some degree when $P(\mathcal{A}) < 0.5$, and neither believes nor disbelieves an argument when $P(\mathcal{A}) = 0.5$.

The epistemic approach provides a wide range of so–called epistemic postulates, which represent reasonable properties that can be used to restrict the probability distributions if desired. In the context of this work, we recall three such postulates:

Coherence P is coherent in AF if $P(\mathcal{B}) \leq 1 - P(\mathcal{A})$ for every $\mathcal{A} \to \mathcal{B} \in \mathsf{AF}$.
Optimism P is optimistic in AF if $P(\mathcal{A}) \geq 1 - \sum_{\mathcal{B} \to \mathcal{A}} P(\mathcal{B})$ for all $\mathcal{A} \in \mathsf{AF}$.
Justifiability P is justifiable in AF if it is both coherent and optimistic in AF.

Coherence states that a high degree of belief in an argument should result in a low degree of belief in any argument attacked by the first argument. *Optimism* states that if all attackers of an argument have a low degree of belief the argument should have a high degree of belief. These two properties formalise the basic intuitions behind Dung's original approach to abstract argumentation [5] in a probabilistic context.

We denote the collection of all coherent (resp. optimistic, justifiable) probability distributions with $\mathcal{P}^C_{\mathsf{AF}}$ (resp. $\mathcal{P}^O_{\mathsf{AF}}$, $\mathcal{P}^J_{\mathsf{AF}}$).

Example 2. Consider the argumentation framework AF in Figure 2. A probability distribution assigning to every argument a probability of 1 would be optimistic, but not coherent. In contrast, a probability distribution assigning to every argument a probability of 0 would be coherent, but not optimistic. Among the justifiable probability distributions of this framework we can find distributions P_1 to P_3 shown in Table 1. We can observe that the provided examples reflect the complete extensions of our graph.

4. Extending the Epistemic Approach

In this section we discuss extending the standard epistemic approach to account for beliefs in arguments as well as in attacks. This calls for extended probability functions that assign degrees of belief to both arguments and attacks. In [13], we have considered pairs of separate argument and attack distributions rather than a single distribution over both arguments and relations. Although the postulates we have previously considered can be easily adapted to both forms, the separate one, due to its structure, may contain less information on how arguments and attacks are jointly believed. For example, we cannot ask for the belief in the set $\{\mathcal{A}, (\mathcal{A}, \mathcal{B})\}$, as we could only consider the sum of probabilities of $\mathcal{A}$ and $(\mathcal{A}, \mathcal{B})$ separately and the obtained number can be as high as 2. Hence, in this paper, we consider the single distribution approach. Let $M_{\mathsf{Arg}} = 2^{\mathsf{Arg} \cup (\mathsf{Arg} \times \mathsf{Arg})}$ denote the set of all possible sets composed of arguments and pairs of arguments.

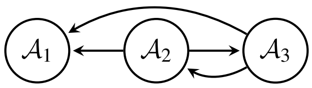

Figure 3. The argumentation framework from Example 3

Definition 4. An *extended probability function* P on AF is a function $P : M_{\mathsf{Arg}} \to [0,1]$ with $\sum_{M \in M_{\mathsf{Arg}}} P(M) = 1$. Let $\mathcal{EP}_{\mathsf{AF}}$ denote the set of all such functions.

In the following, we drop the term "extended" when there is no risk of confusion. Given a probability function P on AF the *marginal belief in an argument* $\mathcal{A} \in \mathsf{AF}$, denoted as $P(\mathcal{A})$, is defined similarly as in the standard approach. The *marginal belief in an attack* $(\mathcal{A},\mathcal{B}) \in \mathsf{AF}$, denoted as $P(\mathcal{A} \to \mathcal{B})$, is defined analogously:

$$P(\mathcal{A}) = \sum_{\mathcal{A} \in M \in M_{\mathsf{Arg}}} P(M) \qquad\qquad P(\mathcal{A} \to \mathcal{B}) = \sum_{(\mathcal{A},\mathcal{B}) \in M \in M_{\mathsf{Arg}}} P(M)$$

Example 3. Consider the argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \to)$ depicted in Figure 3 and define a probability function P_1 on AF via

$$P_1(\{\mathcal{A}_1, \mathcal{A}_2, (\mathcal{A}_3, \mathcal{A}_1)\}) = 0.2 \qquad P_1(\{\mathcal{A}_1, \mathcal{A}_3, (\mathcal{A}_2, \mathcal{A}_3)\}) = 0.3$$
$$P_1(\{\mathcal{A}_3, (\mathcal{A}_3, \mathcal{A}_2), (\mathcal{A}_2, \mathcal{A}_1)\}) = 0.4 \qquad P_1(\{(\mathcal{A}_2, \mathcal{A}_3), (\mathcal{A}_3, \mathcal{A}_1)\}) = 0.1$$

and $P_1(M) = 0$ for all remaining $M \in M_{\mathsf{Arg}}$. Then we have

$$P_1(\mathcal{A}_3 \to \mathcal{A}_1) = 0.3 \quad P_1(\mathcal{A}_2 \to \mathcal{A}_1) = 0.4 \quad P_1(\mathcal{A}_1) = 0.5 \quad P_1(\mathcal{A}_2) = 0.2$$
$$P_1(\mathcal{A}_3 \to \mathcal{A}_2) = 0.4 \quad P_1(\mathcal{A}_2 \to \mathcal{A}_3) = 0.4 \quad P_1(\mathcal{A}_3) = 0.7$$

Note that the epistemic probability of an attack may be higher than the probabilities of the arguments involved in that attack (which is in contrast to the constellations approach, see Section 6 for more discussion). For instance, compare $P(\mathcal{A}_2)$ and $P(\mathcal{A}_2 \to \mathcal{A}_3)$ in the example above. The reason for that is that we interpret these probabilities as *degrees of beliefs* in the validity or strength of the components. For example, we may have a rather low degree of belief in an argument such as "camera footage showed that Jack was at the crime scene at the time of the murder" (because of the low quality of the camera) and equally low degree of belief in an argument such as "camera footage showed that Jack was at a public lecture at the time of the murder". However, even if we do not acknowledge the validity of these arguments, we may have a strong belief in both arguments attacking each other.

In order to compare our new approach to the epistemic approach without probabilities on attacks, we will sometimes consider the fragment of our approach with attacks we are certain of. This allows us to model every probability function in the classical epistemic argumentation as a fully attack-aware extended probability function.

Definition 5. $P \in \mathcal{EP}_{\mathsf{AF}}$ is *fully attack-aware* if for all $(\mathcal{A},\mathcal{B}) \in \mathsf{AF}$, $P(\mathcal{A} \to \mathcal{B}) = 1$. Let $\mathcal{EP}^{\mathrm{faw}}_{\mathsf{AF}}$ denote the set of all fully attack-aware probability functions.

In what follows we will focus on introducing two groups of postulates for the extended epistemic approach, one reflecting the intuitions behind classical epistemic properties [14] and one focusing on the successful attack semantics from [15].

4.1. Basic Rationality Postulates

In the following, we investigate our epistemic framework in terms of desirable relationships between the various degrees of belief of the components involved. We do this by discussing *rationality postulates* in the spirit of our previous works [14,9,10,13].

In order to distinguish the following properties from their counterparts in the existing epistemic approaches, we annotate our new proposals with a superscripted †. Three of the most important properties for the epistemic approach [14] are *coherence*, *optimism*, and *justifiability*, cf. Section 3. Taking probabilities of attacks into account these properties can be formalised as follows. Let $P \in \mathcal{EP}_{\mathsf{AF}}$.

Coherence† P is coherent† in AF if $P(\mathcal{B}) \leq 1 - P(\mathcal{A} \to \mathcal{B})P(\mathcal{A})$ for every $\mathcal{A} \to \mathcal{B} \in \mathsf{AF}$.
Optimism† P is optimistic† in AF if $P(\mathcal{A}) \geq 1 - \sum_{\mathcal{B}\to\mathcal{A}} P(\mathcal{B} \to \mathcal{A})P(\mathcal{B})$ for all $\mathcal{A} \in \mathsf{AF}$.
Justifiability† P is justifiable† in AF if it is both coherent† and optimistic† in AF.

In general, the above properties state how belief in an argument should be bounded given the beliefs in its attackers and the associated attacks. Unlike in [13], the current approach considers how belief in an attack can provide continuous modulation to the effect that the source argument of the conflict may have on the target. In particular, *coherence*† states that the probability of argument should be bounded from above by the inverse of the probability of the attacker, weighted by the probability of the attack. The motivation for this is as follows. If the attack between $\mathcal{A}$ and $\mathcal{B}$ is not acceptable, i. e., $P(\mathcal{A} \to \mathcal{B}) = 0$, the requirement trivialises to $P(\mathcal{B}) \leq 1$, meaning that $\mathcal{A}$ has no influence on $P(\mathcal{B})$ through this attack. On the other hand, if the attack is fully believed, i. e., $P(\mathcal{A} \to \mathcal{B}) = 1$, the requirement becomes $P(\mathcal{B}) \leq 1 - P(\mathcal{A})$ as in the classical epistemic case. Moreover, for different levels of acceptability of the attack, the acceptability of $\mathcal{A}$ weighs more or less into the acceptability of $\mathcal{B}$. The property *optimism*† takes the probability of attacks similarly into account when defining a lower bound. A probability function P is *justifiable*† if it is both coherent† and optimistic†. Let $\mathcal{EP}^J_{\mathsf{AF}} \subseteq \mathcal{EP}_{\mathsf{AF}}$ (resp. $\mathcal{EP}^C_{\mathsf{AF}} \subseteq \mathcal{EP}_{\mathsf{AF}}$, $\mathcal{EP}^O_{\mathsf{AF}} \subseteq \mathcal{EP}_{\mathsf{AF}}$) be the set of all justifiable† (resp. coherent†, optimistic†) probability functions.

Example 4. We continue Example 3 and consider again the framework in Figure 3 and the probability function P_1 with

$$P_1(\mathcal{A}_3 \to \mathcal{A}_1) = 0.3 \quad P_1(\mathcal{A}_2 \to \mathcal{A}_1) = 0.4 \quad P_1(\mathcal{A}_1) = 0.5 \quad P_1(\mathcal{A}_2) = 0.2$$
$$P_1(\mathcal{A}_3 \to \mathcal{A}_2) = 0.4 \quad P_1(\mathcal{A}_2 \to \mathcal{A}_3) = 0.4 \quad P_1(\mathcal{A}_3) = 0.7$$

We first check whether P_1 is coherent†. For that it must hold that

$$P_1(\mathcal{A}_1) \leq 1 - P_1(\mathcal{A}_3 \to \mathcal{A}_1)P_1(\mathcal{A}_3) \quad \Leftrightarrow \quad 0.5 \leq 1 - 0.3 * 0.7$$
$$P_1(\mathcal{A}_1) \leq 1 - P_1(\mathcal{A}_2 \to \mathcal{A}_1)P_1(\mathcal{A}_2) \quad \Leftrightarrow \quad 0.5 \leq 1 - 0.4 * 0.2$$
$$P_1(\mathcal{A}_2) \leq 1 - P_1(\mathcal{A}_3 \to \mathcal{A}_2)P_1(\mathcal{A}_3) \quad \Leftrightarrow \quad 0.2 \leq 1 - 0.4 * 0.7$$
$$P_1(\mathcal{A}_3) \leq 1 - P_1(\mathcal{A}_2 \to \mathcal{A}_3)P_1(\mathcal{A}_2) \quad \Leftrightarrow \quad 0.7 \leq 1 - 0.4 * 0.2$$

which is indeed the case. Now we check whether P_1 is optimistic†. For that it must hold (among others) that

$$P_1(\mathcal{A}_1) \geq 1 - P(\mathcal{A}_3 \to \mathcal{A}_1)P_1(\mathcal{A}_3) - P(\mathcal{A}_2 \to \mathcal{A}_1)P_1(\mathcal{A}_2)$$

$$\Leftrightarrow \qquad 0.5 \geq 1 - 0.3 * 0.7 - 0.4 * 0.2 \qquad \Leftrightarrow \qquad 0.5 \leq 1 - 0.29$$

which is not the case. This is quite clear from observing that $\mathcal{A}_1$ has a roughly medium degree of belief but is only attacked through two weak attacks. Indeed, if the degree of belief of $\mathcal{A}_1$ would be 0.8 (everything else being the same) the above constraint would be satisfied while still being coherent†. However, optimism† is still not satisfied as the corresponding constraints for the arguments $\mathcal{A}_2$ and $\mathcal{A}_3$ are violated as well. An example of a probability function P_2 that is both coherent† and optimistic† is given via

$$P_2(\mathcal{A}_2 \to \mathcal{A}_1) = 0.4 \quad P_2(\mathcal{A}_3 \to \mathcal{A}_2) = 0.5 \quad P_2(\mathcal{A}_1) = 0.7 \quad P_2(\mathcal{A}_3) = 2/3$$

$$P_2(\mathcal{A}_3 \to \mathcal{A}_1) = 0.3 \quad P_2(\mathcal{A}_2 \to \mathcal{A}_3) = 0.5 \quad P_2(\mathcal{A}_2) = 2/3$$

A first observation is that the above properties are indeed faithful generalisations of the analogous properties for the classical epistemic approach [10], i. e., for fully attack-aware probability functions both versions coincide, respectively. The proof of the following theorem is straightforward and omitted.

Theorem 1. *Let $P \in \mathcal{EP}_{\mathsf{AF}}$ and $P' \in \mathcal{P}_{\mathsf{AF}}$ s.t. for every $E \subseteq \mathsf{Arg}$ it holds that $P'(E) = \sum_{X \in M_{\mathsf{Arg}}, E = X \cap \mathsf{Arg}} P(X)$.*

1. *If $P \in \mathcal{EP}^{faw}_{\mathsf{AF}}$, then P is coherent† iff P' is coherent*
2. *If $P \in \mathcal{EP}^{faw}_{\mathsf{AF}}$, then P is optimistic† iff P' is optimistic*
3. *If P' is coherent then P is coherent†*
4. *If P is optimistic† then P' is optimistic*

As our extended approach generalizes the classical epistemic approach, we also obtain several correspondences to classical abstract argumentation generalising those from the literature [14,12]. For $E \subseteq \mathsf{Arg}$ and $P \in \mathcal{P}_{\mathsf{AF}}$ we say that E and P are congruent, denoted $E \sim P$, if for all $\mathcal{A} \in \mathsf{Arg}$, $\mathcal{A} \in E$ implies $P(\mathcal{A}) = 1$, $\mathcal{A} \in E^+$ implies $P(\mathcal{A}) = 0$, and $\mathcal{A} \notin E \cup E^+$ implies $P(\mathcal{A}) = 0.5$. Let $H(P)$ be the *entropy* of P, i. e., $H(P) = -\sum_{M \in M_{\mathsf{Arg}}} P(M) \log P(M)$. The following then straightforwardly generalises results from [14] and is given without proof.

Theorem 2. *Let $E \subseteq \mathsf{Arg}$ and $P \in \mathcal{EP}^{faw}_{\mathsf{AF}}$ with $E \sim P$.*

1. *If E is complete then $P \in \mathcal{EP}^{J}_{\mathsf{AF}}$.*
2. *E is grounded iff $\{P\} = \arg\max_{Q \in \mathcal{EP}^{faw}_{\mathsf{AF}} \cap \mathcal{EP}^{J}_{\mathsf{AF}}} H(Q)$*
3. *If E is stable then $P \in \arg\min_{Q \in \mathcal{EP}^{faw}_{\mathsf{AF}} \cap \mathcal{EP}^{J}_{\mathsf{AF}}} H(Q)$*

Before turning to further rationality postulates, we close the discussion on justifiable probability functions with a brief look on their computational properties. As reasoning with the epistemic approach comes down to solving constraint satisfaction problems on the set of probability functions, it is desirable that the set of feasible solutions (here justifiable probability functions) enjoys some nice topological features. Indeed, in our setting the following holds.

Theorem 3. *The set $\mathcal{EP}^J_{\mathsf{AF}}$ is non-empty and closed.*

Proof. Any function P with $P(\mathcal{A} \to \mathcal{B}) = 0$ and $P(\mathcal{C}) = 1$ for all $\mathcal{A},\mathcal{B},\mathcal{C} \in \mathsf{Arg}$ is in $\mathcal{EP}^J_{\mathsf{AF}}$. As $\mathcal{EP}^J_{\mathsf{AF}}$ is defined using non-strict inequalities it is also closed. □

In contrast to the classical epistemic approach the set $\mathcal{EP}^J_{\mathsf{AF}}$ is not necessarily convex and not even connected.

Example 5. Consider $\mathsf{AF} = (\{\mathcal{A},\mathcal{B}\},\{(\mathcal{A},\mathcal{B})\})$ and P_1, P_2 with

- $P_1(\mathcal{A}) = P_1(\mathcal{B}) = 1$ and $P_1(\mathcal{A} \to \mathcal{B}) = 0$,
- $P_2(\mathcal{A}) = P_2(\mathcal{A} \to \mathcal{B}) = 1$ and $P_2(\mathcal{B}) = 0$,

We have $P_1, P_2 \in \mathcal{EP}^J_{\mathsf{AF}}$ but for the convex combination $P_3 = 0.5 * P_1 + 0.5 * P_2$—yielding $P_3(\mathcal{A}) = 1$, $P_3(\mathcal{A} \to \mathcal{B}) = 0.5$, $P_3(\mathcal{B}) = 0.5$—we have $P_3 \notin \mathcal{EP}^J_{\mathsf{AF}}$, showing that $\mathcal{EP}^J_{\mathsf{AF}}$ is not convex. In fact, in this example $\mathcal{EP}^J_{\mathsf{AF}}$ is also not connected.

Not being convex and connected makes reasoning with $\mathcal{EP}^J_{\mathsf{AF}}$ quite complicated for algorithmic approaches. We will investigate this issue further in future work.

4.2. Epistemic Attack Semantics

We now turn to another, alternative point of view on probabilities of attacks. In [15] a novel perspective on semantics of abstract argumentation frameworks is introduced that focuses on the acceptability of attacks rather than the acceptability of arguments. A particular contribution of that work is a set of rationality postulates that describe how the acceptance of attacks relates to acceptance of arguments and of other attacks. While [15] treats this issue in a qualitative manner—i. e. distinguishing attacks into *successful* and *unsuccessful* attacks—we can lift them to the quantitative case as follows. Let $P \in \mathcal{EP}_{\mathsf{AF}}$.

Attacker Dependence P is attacker-dependent in AF if for all $\mathcal{A},\mathcal{B},\mathcal{C} \in \mathsf{AF}$, if $\mathcal{A} \to \mathcal{B}$ and $\mathcal{A} \to \mathcal{C}$ then $P(\mathcal{A} \to \mathcal{B}) = P(\mathcal{A} \to \mathcal{C})$.

Attack Success P is attack-successful in AF if for all $\mathcal{A} \to \mathcal{B} \in \mathsf{AF}$, $P(\mathcal{A} \to \mathcal{B}) \geq P(\mathcal{A})$.

Attack Failure P is attack-failing in AF all $\mathcal{A},\mathcal{B},\mathcal{C} \in \mathsf{AF}$, if $\mathcal{A} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{C}$ then $P(\mathcal{B} \to \mathcal{C}) \leq 1 - P(\mathcal{A})$.

Attack Defense P is attack-defensive in AF if for all $\mathcal{B} \to \mathcal{C} \in \mathsf{AF}$, $P(\mathcal{B} \to \mathcal{C}) \geq 1 - \sum_{\mathcal{A} \to \mathcal{B}} P(\mathcal{A})$

The above generalizations faithfully extend the corresponding qualitative versions from [15]. More precisely, if we define $P(\mathcal{A} \to \mathcal{B}) = 1$ to mean "The attack $\mathcal{A} \to \mathcal{B}$ is successful", $P(\mathcal{A} \to \mathcal{B}) = 0$ to mean "The attack $\mathcal{A} \to \mathcal{B}$ is unsuccessful", $P(\mathcal{A}) = 1$ to mean "Argument $\mathcal{A}$ is accepted", and $P(\mathcal{A}) = 0$ to mean "Argument $\mathcal{A}$ is not accepted" then the constraints of Definition 4 in [15] are special cases of the above constraints.[1]

Note, however, that imposing the above postulates brings about a slightly different interpretation of the probability $P(\mathcal{A} \to \mathcal{B})$. As outlined before, $P(\mathcal{A} \to \mathcal{B})$ is supposed to represent a degree of belief in the attack $\mathcal{A} \to \mathcal{B}$, i. e., whether the two arguments $\mathcal{A}$ and $\mathcal{B}$ are conflicting and to what degree. In our general setting, this degree of belief

[1] Note that we did not translate the constraints *Qualified* and *Success in terms of acceptance* from [15] as the former is tailored towards the specific formalism of that paper and the latter basically combines the four other properties.

can be high while still having a high degree of belief in $\mathcal{B}$. Imposing the postulates from above interprets "acceptability" as "success". In particular, one central postulate in this view (which is implicit in [15]) is the following.

Attack Certainty P is attack-certain in AF if for $\mathcal{A} \to \mathcal{B}$, $P(\mathcal{B}) \leq 1 - P(\mathcal{A} \to \mathcal{B})$.

This postulate models the basic intuition behind the approach of [15], i. e., the higher the degree of belief in an attack the lower the degree of belief in the attacked argument. This postulate is a special case of coherence[†] when the attacking argument is fully believed. However, imposing *Attack Certainty* also demands that the belief in an argument is bounded even if the belief in the attacking arguments is low.

Assuming both *Attack Success* and *Attack Certainty* implies coherence in the basic epistemic setting.

Proposition 1. *Let $P \in \mathcal{EP}_{\mathsf{AF}}$ and $P' \in \mathcal{P}_{\mathsf{AF}}$ be s.t. for every $E \subseteq \mathsf{Arg}$, $P'(E) = \sum_{X \in M_{\mathsf{Arg}}, E = X \cap \mathsf{Arg}} P(X)$. If P is attack-successful and attack-certain then P' is coherent.*

Proof. Let $\mathcal{A}, \mathcal{B} \in \mathsf{Arg}$ with $\mathcal{A} \to \mathcal{B}$. Due to *Attack Certainty* we have

$$P(\mathcal{B}) \leq 1 - P(\mathcal{A} \to \mathcal{B})$$

and due to *Attack Success* we have $P(\mathcal{A} \to \mathcal{B}) \geq P(\mathcal{A})$ and therefore

$$P(\mathcal{B}) \leq 1 - P(\mathcal{A})$$

showing that P is coherent. □

Assuming *Attack Success*, *Attack Failure* and *Attack Certainty* yields an interesting relationship between arguments in a path of length 2.

Proposition 2. *Let $P \in \mathcal{EP}_{\mathsf{AF}}$. If P satisfies* Attack Success, Attack Failure *and* Attack Certainty *then for all $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathsf{AF}$ with $\mathcal{A} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{C}$*

$$P(\mathcal{B}) \leq 1 - \max\{P(\mathcal{A}), P(\mathcal{C})\}$$

Proof. Let $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathsf{Arg}$ with $\mathcal{A} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{C}$. Due to *Attack Failure* we have

$$P(\mathcal{B} \to \mathcal{C}) \leq 1 - P(\mathcal{A})$$

and due to *Attack Certainty* we have

$$P(\mathcal{C}) \leq 1 - P(\mathcal{B} \to \mathcal{C}) \Leftrightarrow P(\mathcal{B} \to \mathcal{C}) \leq 1 - P(\mathcal{C})$$

and together

$$P(\mathcal{B} \to \mathcal{C}) \leq \min\{1 - P(\mathcal{A}), 1 - P(\mathcal{C})\} \Leftrightarrow P(\mathcal{B} \to \mathcal{C}) \leq 1 - \max\{P(\mathcal{A}), P(\mathcal{C})\}$$

and due to $P(\mathcal{B} \to \mathcal{C}) \geq P(\mathcal{B})$ (*Attack Success*) finally $P(\mathcal{B}) \leq 1 - \max\{P(\mathcal{A}), P(\mathcal{C})\}$. □

So strong belief in either an attacker of $\mathcal{B}$ or an argument attacked by $\mathcal{B}$ yields low belief on $\mathcal{B}$ itself, independent of the beliefs in the attacks.

As we can see, the epistemic approach allows for different points of view on how probabilities of attacks can be interpreted and it depends on the application scenario as to which rationality postulates should be adopted.

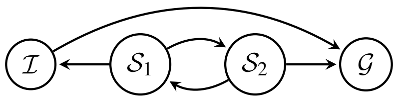

Figure 4. The argumentation framework from Example 6

5. Partial Probability Assignments

We now consider the issue of probabilistic reasoning within our framework, i. e., the challenge of deriving probabilities of all arguments and attacks given partial probabilistic information, cf. [9].

Definition 6. A *partial probability assignment* β on $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ is a partial function $\beta : \mathsf{Arg} \cup (\mathsf{Arg} \times \mathsf{Arg}) \rightarrow [0,1]$.

Let $\mathsf{dom}\,\beta \subseteq \mathsf{Arg} \cup (\mathsf{Arg} \times \mathsf{Arg})$ be the domain of β. A partial probability assignment β provides probabilistic constraints on the subset $\mathsf{dom}\,\beta$ of the components of an argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$. The question then is how can the probabilities of the remaining components be derived by taking the topology of the argumentation framework and rationality postulates into account?

Example 6. We adapt an example from [9] and interpret probabilities of attacks as in Section 4.1. John is either innocent ($\mathcal{I}$) or guilty ($\mathcal{G}$) to have committed the murder of Frank. Footage from a surveillance camera at the crime scene ($\mathcal{S}_1$) shows that someone looking like John was present at the time of the crime, giving a reason that John is not innocent. However, footage from another surveillance camera far away from the crime scene ($\mathcal{S}_2$) gives evidence that a person looking like John was not present at the time of the crime, giving a reason that John is not guilty. This scenario can be modeled with the argumentation framework depicted in Figure 4. Now the footage from the camera $\mathcal{S}_1$ is examined by a lab which assesses that the probability of the person in the pictures is indeed John is 0.7. Moreover, we know that John has a twin brother so the conflict between arguments $\mathcal{S}_1$ and $\mathcal{S}_2$ is uncertain, i. e., only assessed to a degree of belief of 0.6 (both directions). However, we are certain of the validity of the remaining attacks. So given $\beta(\mathcal{S}_1) = 0.7$, $\beta(\mathcal{S}_1 \rightarrow \mathcal{S}_2) = 0.6$, $\beta(\mathcal{S}_2 \rightarrow \mathcal{S}_1) = 0.6$, $\beta(\mathcal{S}_2 \rightarrow \mathcal{G}) = \beta(\mathcal{S}_1 \rightarrow \mathcal{I}) = \beta(\mathcal{I} \rightarrow \mathcal{G}) = 1$ what are now adequate probabilities for the remaining arguments?

A probability function $P \in \mathcal{EP}_{\mathsf{AF}}$ is β*-compliant* if $P(x) = \beta(x)$ for all $x \in \mathsf{dom}\,\beta$. Let $\mathcal{EP}^{\beta}_{\mathsf{AF}} \subseteq \mathcal{EP}_{\mathsf{AF}}$ be the set of β-compliant probability functions. Every $P \in \mathcal{EP}^{\beta}_{\mathsf{AF}}$ is a completion of a partial β and thus provides complete probabilistic information. Let T be some set of rationality postulates from the previous section (such as *coherence*† and *attack success*). Then we can define probabilistic inference as follows.

Definition 7. Let $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ be an abstract argumentation framework and β a partial probability assignment on AF. Then we write $\mathsf{AF}, \beta \mid\!\sim_T C[l, r]$ iff

1. $l = \inf\{P(C) \mid P \in \mathcal{EP}^{\beta}_{\mathsf{AF}}, P \text{ satisfies } T\}$
2. $r = \sup\{P(C) \mid P \in \mathcal{EP}^{\beta}_{\mathsf{AF}}, P \text{ satisfies } T\}$

for all $C \in \mathsf{Arg} \cup \rightarrow$.

In other words, if $\mathsf{AF}, \beta \mid\!\sim_T C[l,r]$ then given the partial probabilistic information of β the probability of the component C (either an argument or an attack) is constrained within the interval $[l,r]$ wrt. all probability functions that satisfy the rationality postulates in T.

Example 7. We continue Example 6 and assume $T = \{\text{coherence}^\dagger\}$, i. e., we are only considering probability functions that are coherent†. Given β defined via

$$\beta(\mathcal{S}_1 \to \mathcal{S}_2) = 0.6 \qquad \beta(\mathcal{S}_2 \to \mathcal{S}_1) = 0.6 \qquad \beta(\mathcal{S}_1) = 0.7$$
$$\beta(\mathcal{S}_1 \to \mathcal{I}) = 1 \qquad \beta(\mathcal{S}_2 \to \mathcal{G}) = 1 \qquad \beta(\mathcal{I} \to \mathcal{G}) = 1$$

we get

$$\mathsf{AF}, \beta \mid\!\sim_T \mathcal{S}_2[0,0.5] \qquad \mathsf{AF}, \beta \mid\!\sim_T \mathcal{G}[0,0.3] \qquad \mathsf{AF}, \beta \mid\!\sim_T \mathcal{I}[0,0.5]$$

For example, the probability that John is innocent lies in the interval $[0,0.5]$.

We leave a deeper analysis of this approach for future work.

6. Related Work

This paper can be seen as orthogonal to the previous work in [13], where the focus was put on exploring the issue of dependent and independent beliefs in attacks and arguments associated with them, showing the interrelationships between these two settings and how they can be used to retrieve the extension–based semantics for abstract argumentation frameworks. In the proposed postulates, the belief in an attack was more seen as a trigger that would allow or prevent the attacker to affect the attackee, but did not vary the degree of this effect. The belief distributions were also always assumed to be complete and the issue of dealing with partially defined function was not explored. In this paper, we have addressed these issues through the introduction of the † properties and a study on partial probability assignments. Additionally, we have provided new postulates concerning the relations between attacks and arguments involved in them and analyzed it in the context of the proposals made in [15].

In the constellations approach [7,11,8,6], the probabilities associated with attacks and arguments are used in order to obtain a probability distribution over the subframeworks of a given framework[2]. The uncertainty associated with a subframework is then seen as the probability of it being the "real" framework of an agent. This modelling means that the constellation probability of an attack is interpreted as a *conditional* probability of the attack being present, given that both the attacker and the attacked argument are present. Therefore, the unconditioned probability of an attack, i. e., the sum of the probabilities of all subframeworks where this attack is present, is always less or equal than the probability of both the attacked argument and the attacker. Due to the fact that epistemic probabilities are interpreted as degrees of beliefs, not as chances of arguments or attacks appearing, the marginal probability of an attack can be greater than the ones of the arguments involved. Although technically speaking a subframework distribution can be mimicked by an extended probability distribution (i.e. the sets of attacks and ar-

[2]For a framework $\mathsf{AF} = (\mathsf{Arg}, \to)$, a subframework is a framework $(\mathsf{Arg}', \to')$ s.t. $\mathsf{Arg}' \subseteq \mathsf{Arg}$ and $\to' \subseteq \to \cap (\mathsf{Arg}' \times \mathsf{Arg}')$.

guments where attacks refer to arguments not in the set have to have a probability of 0), their usage and interpretation is different, which means we can obtain little correlation between the answers provided by the two approaches [13].

7. Summary

We investigated an extension of the epistemic approach to epistemic argumentation by considering probabilities on the attacks between arguments. By interpreting this probability as a degree of belief in the the validity we rephrased previous postulates from the epistemic approach for the new setting and showed several correspondences. We also took an alternative point of view by interpreting probabilities as in the attack semantics of Villata et al. [15], therefore generalising their postulates to the probabilistic setting. Finally, we considered the inferential question of completing incomplete information.

Acknowledgements The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1686/3-1) and the EPSRC funded project "Framework for Computational Persuasion" EP/N008294/1.

References

[1] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In *Proceedings of IJCAI'17*, pages 56–62, 2017.
[2] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.
[3] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of AAAI'16*, pages 914–920, 2016.
[4] M. Caminada. Semi-stable semantics. In *Proceedings of COMMA'06*, pages 121–130,2006.
[5] P.M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–358, 1995.
[6] B. Fazzinga, S. Flesca, and F. Parisi. On the complexity of probabilistic abstract argumentation frameworks. *ACM Transactions on Computational Logic*, 16(3):22:1–22:39, 2015.
[7] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.
[8] A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55:607–638, 2014.
[9] A. Hunter and M. Thimm. On partial information and contradictions in probabilistic abstract argumentation. In *Proceedings of KR'16*, pages 53–62, 2016.
[10] A. Hunter and M. Thimm. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, 59:565–611, August 2017.
[11] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings of TAFA'11*, pages 1–16, 2011.
[12] S. Polberg and A. Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *International Journal of Approximate Reasoning*, 93:487–543, 2018.
[13] S. Polberg, A. Hunter, and M. Thimm. Belief in attacks in epistemic probabilistic argumentation. In *Proceedings of SUM'17*, pages 223–236, 2017.
[14] M. Thimm. A probabilistic semantics for abstract argumentation. In *Proceedings of ECAI'12*, pages 750–755, 2012.
[15] S. Villata, G. Boella, and L. van der Torre. Attack semantics for abstract argumentation. In *Proceedings of IJCAI'11*, pages 406–413, 2011.