

## Admissibility in Probabilistic Argumentation

Christel Baier<sup>1</sup>, Martin Diller<sup>1</sup>, Clemens Dubschlaff<sup>1</sup>  
Sarah Alice Gaggl<sup>1</sup>, Holger Hermanns<sup>2,3</sup>, Nikolai Käfer<sup>1</sup>

<sup>1</sup>Technische Universität Dresden, Faculty of Computer Science, Dresden, Germany

<sup>2</sup>Saarland University, Saarland Informatics Campus, Saarbrücken, Germany

<sup>3</sup>Institute of Intelligent Software, Guangzhou, China

{christel.baier, martin.diller, clemens.dubschlaff, sarah.gaggl, nikolai.kaefer}@tu-dresden.de,  
hermanns@cs.uni-saarland.de

### Abstract

Abstract argumentation is a prominent reasoning framework. It comes with a variety of semantics, and has lately been enhanced by probabilities to enable a quantitative treatment of argumentation. While admissibility is a fundamental notion in the classical setting, it has been merely reflected so far in the probabilistic setting. In this paper, we address the quantitative treatment of argumentation based on probabilistic notions of admissibility in a way that they form fully conservative extensions of classical notions. In particular, our building blocks are not the beliefs regarding single arguments. Instead we start from the fairly natural idea that whatever argumentation semantics is to be considered, semantics systematically induces constraints on the joint probability distribution on the sets of arguments. In some cases there might be many such distributions, even infinitely many ones, in other cases there may be one or none. Standard semantic notions are shown to induce such sets of constraints, and so do their probabilistic extensions. This allows them to be tackled by SMT solvers, as we demonstrate by a proof-of-concept implementation. We present a taxonomy of semantic notions, also in relation to published work, together with a running example illustrating our achievements.

### 1 Introduction

In its basic form, an *abstract argumentation framework* (AF) (Dung 1995) consists of a set of abstract arguments together with a binary relation that represent conflicts between arguments, the so-called *attack relation*. AFs are popular to describe contentious information and draw conclusions from it using formalized arguments. The popularity of the AF concept has led to a variety of extensions like notions to handle preferences and values on arguments (Amgoud and Cayrol 2002; Bench-Capon 2003), weights (Dunne et al. 2011), probabilities (Li, Oren, and Norman 2011; Thimm 2012; Hunter 2013) or introducing a positive influence relation between arguments, so-called *supports* (Amgoud et al. 2008; Nouioua and Risch 2011). Furthermore, *abstract dialectical frameworks* (ADFs) as a powerful generalization of Dung’s framework have been introduced (Brewka and Woltran 2010; Brewka et al. 2013; Straß and Wallner 2015; Gaggl, Rudolph, and Straß 2021), which also allow to handle probabilities (Polberg and Doder 2014).

In this paper, we focus on the emerging field of AFs in the probabilistic setting. As a concrete example to ground

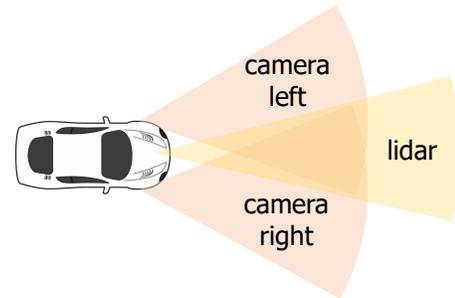


Figure 1: Exemplary sensor layout of a semi-autonomous vehicle

our discussion, we consider an argumentation-based decision framework for a semi-autonomous vehicle as depicted in Figure 1. Here, a central decision entity (the “supervisor” (Faqeh et al. 2020)) has access to possibly conflicting information from several sensors (left/right camera, lidar sensor) with overlapping sensing areas in front of the vehicle. The sensor values are assumed to be of a Boolean nature indicating whether or not an obstacle is detected. They together induce reasons to assume that there is – or is not – an obstacle in some specific area in front of the vehicle. The supervisor aggregates this information in order to decide whether to continue moving forward.

We translate this scenario into a structured argumentation setting as follows. We use literals  $cl$  and  $\bar{cl}$  to denote arguments expressing that the left camera has detected an obstacle or not, and similar for the right camera ( $cr$  and  $\bar{cr}$ ) and for the lidar sensor ( $ld$  and  $\bar{ld}$ ). Slightly more complex arguments represent reasons for and against an obstacle being either on the left ( $l$  and  $\bar{l}$ ), the right ( $r$  and  $\bar{r}$ ), or in the middle ( $m$  and  $\bar{m}$ ) of the area ahead of the vehicle. For example, the argument  $\bar{cl} \Rightarrow \bar{l}$  expresses that a “silent” left camera sensor backs the conclusion that there is no obstacle on the left. If there is no reason to assume that there is an obstacle in the middle, the vehicle will continue driving ( $ct$ ). Otherwise, it will need to stop ( $st$ ). To model the interplay of arguments, we use possibly non-symmetric negation, defeasible rules, and the obvious dualities (Modgil and Prakken 2018). In particular, sensor values might be attacked, reflecting that

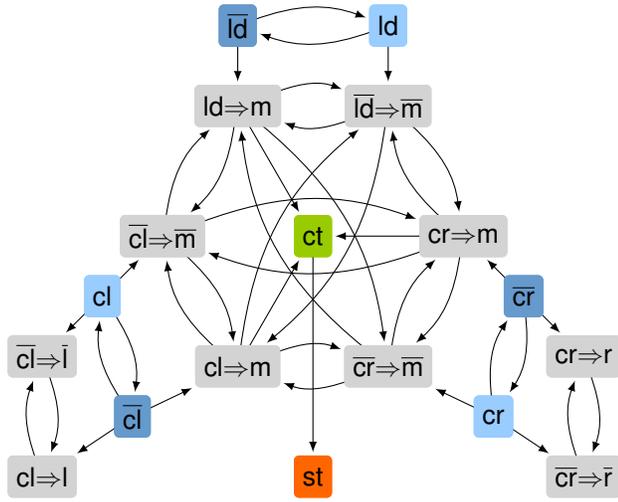


Figure 2: Argument graph for the vehicle example

sensor readings might be erroneous in practice.

The resulting argument graph is depicted in Figure 2, with nodes representing arguments and directed edges representing attacks. In particular, we see that arguments about sensors can “undermine” arguments about the location of obstacles (e.g.,  $\bar{cl}$  undermines  $cl \Rightarrow l$ ) while arguments about the location can “rebut” each other when they make contradictory claims (e.g.,  $cl \Rightarrow m$  and  $\bar{ld} \Rightarrow \bar{m}$  rebut each other).

In this paper, we are interested in the situation where the degree of acceptance of arguments lies on a continuum. For instance, object detection by the right camera might fail with some probability (false negatives), spurious detections may be possible (false positives), and the spatial layout of the overlapping camera views may give quantitative information where to expect an object, albeit the information provided may not be clear-cut. This implies that there will be different degrees of uncertainty as to whether to continue moving forward or not.

This motivation helps to understand why there is a growing body of inspiring research on *probabilistic abstract argumentation frameworks* (PrAFs). The present paper aims at adding to this research spectrum a probability-theoretic perspective, based on the *epistemic* approach (see Section 4 for an overview). We start off from the idea that a PrAF induces probability distributions over various arguments holding or not. Just like classical AFs can induce potentially multiple valid interpretations (called extensions<sup>1</sup>), a PrAF can induce multiple such distributions. We embark on extending the basic notions of classical argumentation theory to the probabilistic setting in a conservative manner. In this, we take the fairly natural view that no argument and its attackers may hold at the same time. This does not mean that both views on an argument and its attacker cannot have non-zero probability, but simply that the probability of an argument and its attacker holding at the same time is zero by construction.

<sup>1</sup>We avoid using “extension” in this sense, and instead reserve the word for discussing *conservative extensions* of AF semantics.

Similarly, we would like to ensure that arguments that hold are defended, meaning that each of its attackers is attacked with probability 1. This is indeed needed for making sense of scenarios like the vehicle example above, but is in contrast to earlier work on PrAFs (Hunter and Thimm 2017) where the belief in single arguments was put in focus.

When lifting the classical concepts from AFs to PrAFs, especially the notion of admissibility gives rise to a hierarchy of different interpretations and lead to an entire taxonomy of semantics. Along this discussion, it becomes apparent that each lifted semantics concept imposes a set of constraints on the joint probability distributions of arguments to hold and not to hold. We will show in this paper that, with one exception, all these sets of constraints we consider are linear, i.e., they belong to the linear arithmetic theory of the reals. Among others, this gives us decidability and bridges to the world of SMT solvers, which are nowadays well capable of handling large sets of linear and even non-linear constraints. Indeed, we present a prototypical tool for studying a variety of questions arising in a PrAF. Specifically, we apply the tool to the above vehicle example to pinpoint some fine details that help to understand our contribution as well as getting an impression of practical relevance. For example, the tool can compute a distribution maximizing the value of  $ct$  assuming that  $cl$  and  $ld$  hold almost surely with additional constraints on the rate of false positives for the sensors, while satisfying a particular semantics (or even sets thereof). The tool and all experimental data are publicly available at

<https://www.perspicuous-computing.science/cpraa/>.

In summary, our contribution is fourfold: We (i) provide a profound study of admissibility and completeness in a probability-theoretic approach to abstract argumentation, (ii) discuss a hierarchy of resulting semantics in the context of earlier work, (iii) present prototypical tool support for experimenting with these notions and further context-specific constraints, and (iv) explicate our contributions by means of the vehicle example introduced above.

## 2 Preliminaries

In this section, we introduce the basics of abstract argumentation along with classical argumentation semantics on which we will base our investigations in the probabilistic setting. We start with a definition of abstract argumentation frameworks following (Dung 1995). For a detailed discussion on argumentation semantics we refer to (Baroni, Caminada, and Giacomin 2011).

**Definition 1.** An abstract argumentation framework (AF) is a pair  $\mathcal{F} = \langle Arg, Att \rangle$  where  $Arg$  is a finite set of arguments and  $Att \subseteq Arg \times Arg$  an attack relation.

A pair  $(A, B) \in Att$  means that argument  $A$  attacks argument  $B$ . We denote by  $\vec{A} := \{B \in Arg : (B, A) \in Att\}$  the set of *attackers* of  $A$  and by  $A^\rightarrow := \{B \in Arg : (A, B) \in Att\}$  the set of  $A$ ’s *attackees*.  $A$  is called *initial* if it has no attackers, i.e., if  $\vec{A}$  is empty. These notions naturally extend to sets of arguments  $\mathcal{S} \subseteq Arg$  by  $\vec{\mathcal{S}} := \bigcup_{A \in \mathcal{S}} \vec{A}$  and  $\mathcal{S}^\rightarrow := \bigcup_{A \in \mathcal{S}} A^\rightarrow$ . Two arguments  $A$  and  $B$  are in *conflict* if  $(A, B) \in Att$  or  $(B, A) \in Att$ , i.e., either one is attacking the other.  $\mathcal{S}$  *defeats* an argument  $B$  if  $B \in \mathcal{S}^\rightarrow$ , i.e., at least

one argument in  $S$  attacks  $B$ . An argument  $C$  is *defended* by  $S$  if  $S$  defeats all attackers  $B \in \neg C$ . The set of all arguments defended by  $S$  is thus given by

$$\text{Defend}(S) := \{C \in \text{Arg} : B \in S \rightarrow \text{for each } B \in \neg C\}.$$

**Classical Argumentation Semantics.** Semantics for AFs are given by collections of argument sets that do not exhibit conflicts.

**Definition 2.** For an AF  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$ , a set  $S \subseteq \text{Arg}$  is said to be conflict-free (CF) if  $(A, B) \notin \text{Att}$  for all  $A, B \in S$ . A conflict-free argument set  $S$  is

(ST) stable if  $S \cup S^\rightarrow = \text{Arg}$ ,

(ADM) admissible if  $S \subseteq \text{Defend}(S)$ ,

(CMP) complete if  $S = \text{Defend}(S)$ ,

(GR) grounded if there is no complete  $T \subseteq \text{Arg}$  with  $T \subset S$ ,

(PRF) preferred if there is no complete  $T \subseteq \text{Arg}$  with  $T \supset S$ .

The classical argumentation semantics  $\sigma$  for  $\mathcal{F}$  with  $\sigma \in \{\text{CF}, \text{ST}, \text{ADM}, \text{CMP}, \text{GR}, \text{PRF}\}$  is the set  $[\mathcal{F}]_\sigma$  of all argument sets  $S \subseteq \text{Arg}$  where condition  $\sigma$  as above holds.

**Assignments.** In the probabilistic setting, each argument  $A$  of a given AF  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  is treated as a Boolean random variable with the same name. An *assignment* is a function  $\beta: \text{Arg} \rightarrow \{\text{T}, \text{F}\}$  that determines for each variable, and thus each argument, whether it holds (T) or not (F). The set of all assignments for  $\text{Arg}$  is denoted by  $\text{Asg}(\text{Arg})$ . An *event*  $\phi$  is a set of assignments, i.e.,  $\phi \subseteq \text{Asg}(\text{Arg})$ . There exists a straight-forward one-to-one correspondence between argument sets and assignments: Given an argument set  $S \subseteq \text{Arg}$ , the corresponding assignment is given by its *characteristic function*  $\text{id}_S$  defined by  $\text{id}_S(A) := \text{T}$  if  $A \in S$  and  $\text{id}_S(A) := \text{F}$  otherwise. Conversely, each assignment  $\beta$  for  $\text{Arg}$  naturally induces an argument set  $\text{Arg}_\beta := \{A \in \text{Arg} : \beta(A) = \text{T}\}$ . The switch between argument sets and assignments is thus merely of syntactic nature and we may see, e.g., any classical semantics  $\sigma$  as a subset of  $\text{Asg}(\text{Arg})$ . Further, we use propositional logic formulas over arguments to specify sets of assignments, e.g., we write  $A \wedge \neg B$  with  $A, B \in \text{Arg}$  for the set of assignments  $\phi \subseteq \text{Asg}(\text{Arg})$  where  $\phi = \{\beta \in \text{Asg}(\text{Arg}) : \beta(A)=\text{T}, \beta(B)=\text{F}\}$ . To this end, we denote by

$$\Delta(C) := \bigwedge_{B \in \neg C} \bigvee_{A \in \neg B} A$$

the set of assignments whose corresponding argument sets defend an argument  $C \in \text{Arg}$ . Note that as usual, empty conjunctions stand for T and empty disjunctions for F.

**Distributions.** A probability distribution over a set  $X$  is a function  $\mu: X \rightarrow [0, 1]$  where  $\sum_{\beta \in X} \mu(\beta) = 1$ . The set of all distributions over  $X$  is denoted by  $\text{Distr}(X)$ . The *support* of a distribution  $\mu$  is defined by  $\text{Supp}(\mu) := \{\beta : \mu(\beta) > 0\}$ .  $\mu$  is a *Dirac distribution* if  $\text{Supp}(\mu)$  is a singleton. For a fixed  $\beta \in X$ , we write  $\text{Dirac}_\beta$  for the uniquely defined Dirac distribution where  $\text{Dirac}_\beta(\beta) = 1$ .

In the following, we are only concerned with distributions over the set of assignments  $\text{Asg}(\text{Arg})$ . For brevity, we write  $\text{Distr}(\mathcal{F})$  for  $\text{Distr}(\text{Asg}(\text{Arg}))$ . Each  $\mu \in \text{Distr}(\mathcal{F})$  induces a

probability measure over events, denoted by  $\mu$  as well, i.e., a function  $\mu: 2^{\text{Asg}(\text{Arg})} \rightarrow [0, 1]$  with  $\mu(\phi) := \sum_{\beta \in \phi} \mu(\beta)$  for events  $\phi \subseteq \text{Asg}(\text{Arg})$ . Note that  $\Delta(C)$  is an event for any argument  $C \in \text{Arg}$ . As usual, we say that  $\mu$  has outcome  $\phi$  *almost surely* in case  $\mu(\phi) = 1$ . For argument sets  $S \subseteq \text{Arg}$ , we use  $\text{Dirac}_S$  as a short form for  $\text{Dirac}_{\text{id}_S}$ . Finally, for two events  $\phi, \psi \subseteq \text{Asg}(\text{Arg})$  with  $\mu(\psi) > 0$ , the *conditional probability* of  $\phi$  given  $\psi$  is defined as

$$\mu(\phi \mid \psi) := \frac{\mu(\phi \wedge \psi)}{\mu(\psi)}.$$

### 3 Probabilistic Argumentation Semantics

A *probabilistic argumentation semantics*  $\rho$  assigns to each AF  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  a subset  $[\mathcal{F}]_\rho$  of  $\text{Distr}(\mathcal{F})$ . While in classical argumentation the argument sets in  $[\mathcal{F}]_\sigma$  only classify the arguments that hold or not hold, distributions in  $[\mathcal{F}]_\rho$  give rise to a probabilistic interpretation of the arguments, e.g., to specify the belief in or acceptance of arguments.

**Definition 3** (Likelihood of arguments). For an AF  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$ , a distribution  $\mu \in \text{Distr}(\mathcal{F})$ , and an argument  $A \in \text{Arg}$ , we refer to  $\mu(A)$  as the  $\mu$ -likelihood of  $A$ .

We denote by  $\text{Arg}_\mu := \{A \in \text{Arg} : \mu(A) = 1\}$  the set of arguments that hold almost surely. Note that here,  $A$  stands for a (basic) propositional logic formula over arguments that specifies the set of all events  $\beta$  where  $\beta(A)=\text{T}$ .

Classical semantics for AFs can be naturally lifted to the probabilistic setting by Dirac distributions corresponding to argument sets. A probabilistic semantics where all induced Dirac distributions agree with the Dirac distributions of a classical semantics is called a *conservative extension*:

**Definition 4** (Conservative extension). A *probabilistic argumentation semantics*  $\rho$  is said to be a conservative extension of a classical argumentation semantics  $\sigma$  if for all AFs  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  and argument sets  $S \subseteq \text{Arg}$ :

$$S \in [\mathcal{F}]_\sigma \text{ iff } \text{Dirac}_S \in [\mathcal{F}]_\rho.$$

As a basic instance of a conservative extension, we present the element-wise lifting of classical argumentation semantics which was first introduced by (Thimm et al. 2017).

**Definition 5** (Element-wise lifting). Let  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  be an AF, and  $\mu \in \text{Distr}(\mathcal{F})$ . For a classical argumentation semantics  $\sigma$ , the *element-wise- $\sigma$  semantics*  $\text{ELM-}\sigma$  is defined as follows:

$$[\mathcal{F}]_{\text{ELM-}\sigma} := \{\mu : \text{Arg}_\beta \in [\mathcal{F}]_\sigma \text{ for all } \beta \in \text{Supp}(\mu)\}.$$

So,  $\text{ELM-}\sigma$  semantics enforces  $\mu(\beta_S) = 0$  for all argument sets  $S \in 2^{\text{Arg}} \setminus [\mathcal{F}]_\sigma$  that are not part of the classical argumentation semantics  $\sigma$ . For instance,  $[\mathcal{F}]_{\text{ELM-ADM}}$  comprises exactly those distributions where the assignments corresponding to the admissible sets of  $\mathcal{F}$  have positive probability.

**Lemma 6.** *Element-wise lifting of classical semantics yields conservative extensions.*

The converse does not hold, i.e., the notion of conservative extension is more liberal.

### 3.1 Assignment Distribution Properties

Before we discuss probabilistic extensions of admissibility and completeness, we define the notions of conflict-freeness and defense for assignments in the probabilistic setting.

**Almost-sure Conflict-freeness.** Recall that a set of arguments  $\mathcal{S}$  is said to be conflict-free if there is no attack  $(A, B) \in Att$  for all  $A, B \in \mathcal{S}$ . By the one-to-one correspondence to assignments, an assignment  $\beta \in \text{Asg}(Arg)$  is thus conflict-free iff there is no pair  $(A, B) \in Att$  with  $\beta(A) = \beta(B) = \top$ . This notion of conflict-freeness naturally transfers to distributions over assignments.

**Definition 7** (Almost-sure conflict-freeness). *For an AF  $\mathcal{F} = \langle Arg, Att \rangle$ , a distribution  $\mu \in \text{Distr}(\mathcal{F})$  is almost-sure conflict-free if*

$$\mu(A \wedge B) = 0 \quad \text{for all } (A, B) \in Att. \quad (\text{ASCF})$$

Notably, ASCF does by no means imply that attacker and attackee cannot both have nonzero  $\mu$ -likelihood. However, the probability of both holding together (constituting a classical conflict) is zero.

Almost-sure conflict-freeness coincides with element-wise lifting of conflict-free argument sets:

**Lemma 8** (ELM-CF  $\leftrightarrow$  ASCF). *For an AF  $\mathcal{F} = \langle Arg, Att \rangle$  a distribution  $\mu \in \text{Distr}(\mathcal{F})$  is almost-sure conflict-free iff  $Arg_\beta$  is conflict-free for each assignment  $\beta \in \text{Supp}(\mu)$ .*

By Lemma 6 and 8, the set of almost-sure conflict-free distributions conservatively extends the set of conflict-free argument sets.

**Almost-sure Defense.** In the classical setting, an argument is defended by defeating each of its attackers. Likewise, we say that a distribution almost surely defends an argument if all attackers are in turn attacked with probability one:

**Definition 9** (Almost-sure defense). *For  $\mathcal{F} = \langle Arg, Att \rangle$  an AF, a distribution  $\mu \in \text{Distr}(\mathcal{F})$  almost surely defends an argument  $C \in Arg$  if*

$$\mu\left(\bigvee_{A \in \neg C} A\right) = 1 \quad \text{for all } B \in \neg C.$$

That is,  $\mu$  almost surely defends the argument  $C$  provided each attacker  $B$  of  $C$  is attacked with probability 1 under  $\mu$ . We let the set  $\text{asDefend}(\mu)$  collect all arguments that are almost surely defended by distribution  $\mu$ . For an argument set  $Arg_\beta$  we write  $\text{Defend}(\beta)$  instead of  $\text{Defend}(Arg_\beta)$  for the set of arguments defended by  $\beta$ .

**Lemma 10.** *Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . Then:*

$$\text{asDefend}(\mu) = \bigcap_{\beta \in \text{Supp}(\mu)} \text{Defend}(\beta).$$

**Corollary 11.** *For any AF  $\mathcal{F} = \langle Arg, Att \rangle$  and distribution  $\mu \in \text{Distr}(\mathcal{F})$ :*

- (a)  $\text{asDefend}(\mu)$  is conflict-free if  $\mu$  is almost-sure conflict-free
- (b)  $\text{asDefend}(\text{Dirac}_S) = \text{Defend}(S)$  for each  $S \subseteq Arg$ .

In other words, for each argument  $C$  and each argument set  $\mathcal{S}$ ,  $\mathcal{S}$  defends  $C$  iff  $\text{Dirac}_S$  almost surely defends  $C$ .

**Relative Defense Constraint.** For a comparison relation  $\sim \in \{\leq, =, \geq\}$ , we define the *relative defense constraint*

$$\mu(C) \sim \mu(\Delta(C)) \quad \text{for all } C \in Arg. \quad (\sim_\mu \Delta)$$

For example, the constraint  $(\leq_\mu \Delta)$  is satisfied if for all arguments the probability that the argument is defended is at least as high as the probability for the argument itself.

### 3.2 Admissibility

Recall that for any AF  $\mathcal{F} = \langle Arg, Att \rangle$  a set of arguments  $\mathcal{S} \subseteq Arg$  is called admissible if  $\mathcal{S}$  is conflict-free and defends all its elements. We say that an assignment  $\beta$  is admissible if  $Arg_\beta$  is admissible. Based on almost-sure conflict-freeness and defense as defined in the last section, we now provide several notions of admissibility in the probabilistic setting.

**Definition 12.** *Let  $\mathcal{F} = \langle Arg, Att \rangle$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$  be almost-sure conflict-free. Then  $\mu$  is called*

- (ELM-ADM) element-wise admissible iff  $\mu \in \llbracket \mathcal{F} \rrbracket_{\text{ELM-ADM}}$ ,
- (PRADM) probabilistically admissible iff for all  $B \in Arg$

$$\mu\left(\bigvee_{C \in B^\rightarrow} C\right) \leq \mu\left(\bigvee_{A \in \neg B} A\right),$$

- (JNTADM) joint-attack admissible iff  $(\leq_\mu \Delta)$  holds,
- (MINADM) min-admissible iff for all  $C \in Arg$

$$\mu(C) \leq \min_{B \in \neg C} \mu\left(\bigvee_{A \in \neg B} A\right),$$

- (WADM) weakly admissible iff  $Arg_\mu \subseteq \text{asDefend}(\mu)$ .

Let us explain the definition of these five notions of admissibility in greater detail. Element-wise admissibility of  $\mu$  requires admissibility of all assignments in the support of  $\mu$ . This can be rephrased by requiring for each argument  $C$  with positive  $\mu$ -likelihood that the conditional probability of assignments defending  $C$  equals one when conditioned on the event where  $C$  holds.

**Lemma 13.** *Let  $\mathcal{F} = \langle Arg, Att \rangle$  and  $\mu \in \text{Distr}(\mathcal{F})$ . Then  $\mu$  is element-wise admissible iff  $\mu$  is almost-sure conflict-free and for each argument  $C \in Arg$  with  $\mu(C) > 0$ , it holds*

$$\mu(\Delta(C) \mid C) = 1.$$

For probabilistic admissibility, recall that an argument set  $\mathcal{S}$  is admissible if for all arguments  $B$  attacking  $\mathcal{S}$ ,  $B$  is attacked by some argument in  $\mathcal{S}$ . Likewise, PRADM requires that for each argument  $B$ , the probability for attacks by  $B$  on arguments with assigned value  $\top$  is bounded by the probability for  $B$  being attacked by arguments with assigned value  $\top$ . Note that this is a tautology for the arguments  $B \in Arg_\mu$ , because then

$$\mu\left(\bigvee_{C \in B^\rightarrow} C\right) = \mu\left(\bigvee_{A \in \neg B} A\right) = 0.$$

Joint-attack and min-admissibility impose upper bounds on the likelihood of arguments. MINADM requires  $\mu(C)$  to be bounded by the probabilities for attacks on  $C$ 's attackers. Conditions MINADM and JNTADM are tautologic for each

$C$  where  $\mu(C) = 0$ . Finally,  $\mu$  is weakly admissible<sup>2</sup> iff  $\mu$  almost surely defends all arguments  $A$  with  $\mu(A) = 1$ .

All five admissibility notions of Definition 12 yield conservative extensions of the non-probabilistic notion of admissibility, as stated in the following lemma:

**Lemma 14** (Conservative extension). *Let  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  be an AF and  $\mathcal{S} \subseteq \text{Arg}$  be conflict-free. Then,  $\mathcal{S}$  is admissible iff  $\text{Dirac}_{\mathcal{S}}$  is  $\kappa$  admissible for  $\kappa \in \{\text{element-wise, probabilistically, joint-attack, min-, weakly}\}$ .*

**Relationships.** We now investigate how our notions of admissibility relate to each other. First, we provide implications that hold between the different semantics and then give concrete examples that show these implications to be strict.

Element-wise admissibility is the most restricted variant, i.e., each distribution that satisfies ELM-ADM also satisfies the other notions of admissibility. Exemplarily, for an attack  $(B, C) \in \text{Att}$ , whenever  $\beta(C) = \top$  for some admissible assignment  $\beta$ , there must be some attacker  $A$  of  $B$  with  $\beta(A) = \top$  as well. Thus, all admissible assignments that contribute to the probability on the left-hand side of the PRADM constraint also appear on the right, so all element-wise admissible distributions are also probabilistically admissible.

**Lemma 15** (ELM-ADM  $\rightarrow$  PRADM, ELM-ADM  $\rightarrow$  JNTADM). *Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is element-wise admissible, then  $\mu$  is probabilistically and joint-attack admissible.*

JNTADM is slightly stronger than MINADM since for any argument  $C$ , we have

$$\mu(\Delta(C)) \leq \min_{B \in \neg C} \mu\left(\bigvee_{A \in \neg B} A\right).$$

Thus the probability for joint attacks on  $C$ 's attackers is an upper bound for the likelihood of  $C$ .

**Lemma 16** (PRADM  $\rightarrow$  MINADM, JNTADM  $\rightarrow$  MINADM). *Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is probabilistically or joint-attack admissible, then  $\mu$  is min-admissible.*

Note that the notions joint-attack and min-admissibility collapse in AFs where each argument has at most one attacker.

As weak admissibility only imposes constraints on the likelihood of arguments belonging to  $\text{Arg}_{\mu}$ , it is strictly weaker than min-admissibility, and hence the most liberal admissibility notion from the ones of Definition 12.

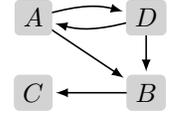
**Lemma 17** (MINADM  $\rightarrow$  WADM). *Let  $\mathcal{F} = \langle \text{Arg}, \text{Att} \rangle$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is min-admissible, then  $\mu$  almost surely defends all arguments in  $\text{Arg}_{\mu}$ .*

These results so far show that the five notions of admissibility for distributions constitute a hierarchy as illustrated in the bottom part of Figure 3. The inclusions indicated by the arrows are strict, as shown by the following examples.

**Example 18** (MINADM  $\not\rightarrow$  PRADM, MINADM  $\not\rightarrow$  JNTADM). *As an example for a min-admissible distribution that is not element-wise admissible and neither joint-attack nor*

*probabilistically admissible, let  $\mu$  be the distribution for the AF on the right where its support is given through*

$$\begin{aligned} \mu(\neg A \wedge \neg B \wedge C \wedge D) &= \frac{1}{3} \\ \mu(A \wedge \neg B \wedge \neg C \wedge \neg D) &= \frac{1}{4} \\ \mu(\neg A \wedge B \wedge \neg C \wedge \neg D) &= \frac{1}{6} \\ \mu(\neg A \wedge \neg B \wedge C \wedge \neg D) &= \frac{1}{4} \end{aligned}$$



*That is,  $\mu(\beta) = 0$  for all other assignments  $\beta$ . Then  $\mu$  is almost-sure conflict-free (since all assignments in its support are conflict-free, cf. Definition 7) and satisfies MINADM:*

- *Argument  $C$  has likelihood  $\mu(C) = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$  and a single attacker  $B$ , which is attacked by  $A$  and  $D$  with probability  $\frac{1}{4} + \frac{1}{3} = \frac{7}{12}$ .*
- *Argument  $B$  has likelihood  $\mu(B) = \frac{1}{6}$ . Its attackers  $A$  and  $D$  are attacked with probability  $\frac{1}{3}$  resp.  $\frac{1}{4}$ .*
- *Argument  $D$  has likelihood  $\mu(D) = \frac{1}{3}$ , and so is the probability for its attacker  $A$  to be attacked. The analogous statement holds for  $A$ .*

*So,  $\mu$  is min-admissible, but  $\mu$  is not element-wise admissible as  $\neg A \wedge \neg B \wedge C \wedge \neg D$  and  $\neg A \wedge B \wedge \neg C \wedge \neg D$  induce the non-admissible argument sets  $\{C\}$  and  $\{B\}$ , respectively, that do not defend their arguments.*

*Further,  $\mu$  is not joint-attack admissible, since the argument  $B$  has no joint attack of its attackers  $A$  and  $D$ :*

$$\mu(B) = \frac{1}{6} \not\leq 0 = \mu(A \wedge D) = \mu(\Delta(B)).$$

*Indeed, in this example there is no admissible argument set containing  $B$ , since  $B$ 's attackers  $A$  and  $D$  attack each other.*

*To see why  $\mu$  is not probabilistically admissible, regard argument  $D$ . We have  $D^{\rightarrow} = \{A, B\}$  and  $\neg D = \{A\}$ . But then*

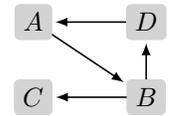
$$\begin{aligned} \mu\left(\bigvee_{C' \in D^{\rightarrow}} C'\right) &= \mu(A \vee B) = \frac{1}{4} + \frac{1}{6} \\ &\not\leq \frac{1}{4} = \mu(A) = \mu\left(\bigvee_{A \in \neg D} A\right). \end{aligned}$$

*So, the constraint PRADM is violated for argument  $D$ . Analogously, PRADM does not hold for  $A$ .*

The following two examples illustrate that probabilistic admissibility and joint-attack admissibility are incomparable.

**Example 19** (JNTADM  $\not\rightarrow$  PRADM). *Consider the AF on the right below and distribution  $\mu$  with the following support:*

$$\begin{aligned} \mu(A \wedge \neg B \wedge C \wedge \neg D) &= \frac{1}{3} \\ \mu(\neg A \wedge B \wedge \neg C \wedge \neg D) &= \frac{1}{3} \\ \mu(\neg A \wedge \neg B \wedge \neg C \wedge D) &= \frac{1}{3} \end{aligned}$$



*$\mu$  obviously is almost-sure conflict-free and the  $\mu$ -likelihood of all four arguments is  $\frac{1}{3}$ . As each argument has exactly one attacker,  $\mu$  is joint-attack admissible: For example, the JNTADM constraint is satisfied for argument  $C$  since*

<sup>2</sup>Not to be confused with the recent notion of *weak admissibility* in the non-probabilistic setting by (Baumann, Brewka, and Ulbricht 2020).

$\neg C = \{B\}$ ,  $\neg B = \{A\}$ , and

$$\mu(C) = \frac{1}{3} \leq \frac{1}{3} = \mu(A) = \mu\left(\bigvee_{A' \in \neg B} A'\right) = \mu(\Delta(C)).$$

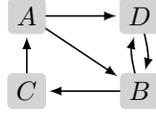
However,  $\mu$  does not satisfy the constraint PRADM for argument  $B$ :  $B^\rightarrow = \{C, D\}$ ,  $\neg B = \{A\}$ , but

$$\mu\left(\bigvee_{C' \in B^\rightarrow} C'\right) = \mu(C \vee D) = \frac{1}{3} + \frac{1}{3} \not\leq \frac{1}{3} = \mu\left(\bigvee_{A' \in \neg B} A'\right).$$

Hence,  $\mu$  is not probabilistically admissible.

**Example 20** (PRADM  $\not\rightarrow$  JNTADM). Consider the following almost-sure conflict-free distribution for the AF on the right:

$$\begin{aligned} \mu(A \wedge \neg B \wedge \neg C \wedge \neg D) &= \frac{1}{4} \\ \mu(\neg A \wedge B \wedge \neg C \wedge \neg D) &= \frac{1}{4} \\ \mu(\neg A \wedge \neg B \wedge C \wedge D) &= \frac{1}{4} \\ \mu(\neg A \wedge \neg B \wedge C \wedge \neg D) &= \frac{1}{4} \end{aligned}$$



We have  $\mu(A) = \mu(B) = \mu(D) = \frac{1}{4}$ , and  $\mu(C) = \frac{1}{2}$ . This implies that  $Arg_\mu$  is empty, such we can immediately infer that  $\mu$  is weakly admissible. To see why  $\mu$  is probabilistically admissible, we have to check that  $\mu$  satisfies PRADM for all arguments:

- For argument  $A$ , we have  $A^\rightarrow = \{B, D\}$  and  $\neg A = \{C\}$ , such that  $\mu(B \vee D) \leq \mu(C)$  has to hold. This is in fact the case due to  $\mu(B \vee D) = \mu(C) = \frac{1}{2}$ .
- For argument  $B$ , we have  $B^\rightarrow = \{C, D\}$ ,  $\neg B = \{A, D\}$ , and  $\mu(C \vee D) = \frac{1}{2} \leq \frac{1}{2} = \mu(A \vee D)$ .
- For argument  $C$ , we have  $C^\rightarrow = \{A\}$ ,  $\neg C = \{B\}$ , and  $\mu(A) = \frac{1}{4} \leq \frac{1}{4} = \mu(B)$ .
- For argument  $D$ , we have  $D^\rightarrow = \{B\}$ ,  $\neg D = \{A, B\}$ , and  $\mu(B) = \frac{1}{4} \leq \frac{1}{2} = \mu(A \vee B)$ .

Thus,  $\mu$  is probabilistically admissible.

To see why  $\mu$  is not joint-attack admissible, we observe that the  $\mu$ -likelihood of argument  $B$  is positive, but there is no joint attack on  $B$ 's attackers  $A$  and  $D$ . More precisely:

$$\mu(B) = \frac{1}{4} \not\leq 0 = \mu(C \wedge (A \vee B)) = \mu(\Delta(B)).$$

Here,  $C$  stands for the attacks on  $A$ , and  $A \vee B$  for the attacks on  $D$ . This shows that  $\mu$  violates the constraint JNTADM for argument  $B$ .

### 3.3 Completeness

In the non-probabilistic setting, completeness is stronger than admissibility, as it additionally requires that all arguments defended by a set  $S \subseteq Arg$  are contained in  $S$ . An assignment  $\beta \in \text{Asg}(Arg)$  is said to be complete if the induced argument set  $Arg_\beta$  is complete. Based on the notions of admissibility for distributions from Definition 12, we now extend completeness towards five notions in the probabilistic setting.

**Definition 21.** Let  $\mathcal{F} = \langle Arg, Att \rangle$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$  be almost-sure conflict-free. Then  $\mu$  is called

- (ELM-CMP) element-wise complete iff  $\mu \in \llbracket \mathcal{F} \rrbracket_{\text{ELM-CMP}}$ ,
- (PRCMP) probabilistically complete iff  $\mu$  satisfies PRADM and  $(\geq_\mu \Delta)$ ,
- (JNTCMP) joint-attack complete iff  $\mu$  satisfies  $(=_\mu \Delta)$ ,
- (MINCMP) min-complete iff  $\mu$  satisfies MINADM and  $(\geq_\mu \Delta)$ , and
- (WCMP) weakly complete iff  $Arg_\mu = \text{asDefend}(\mu)$ .

**Lemma 22.** Let  $\mathcal{F} = \langle Arg, Att \rangle$  and  $\mu \in \text{Distr}(\mathcal{F})$ . Then  $\mu$  is element-wise complete iff  $\mu$  is element-wise admissible and for all arguments  $C \in Arg$  where event  $\Delta(C)$  has positive probability under  $\mu$ , it holds:

$$\mu(C \mid \Delta(C)) = 1.$$

All five completeness notions as of Definition 21 are conservative extensions of the non-probabilistic notion of completeness for argument sets, as stated in the following lemma.

**Lemma 23** (Conservative extension). Let  $\mathcal{F} = \langle Arg, Att \rangle$  be an AF and  $S \subseteq Arg$ . Then,  $S$  is complete iff  $\text{Dirac}_S$  is  $\kappa$  complete for  $\kappa \in \{\text{element-wise, weakly, probabilistically, min-, joint-attack}\}$ .

**Relationships.** Similar as for the notions of admissibility, we draw relationships between notions of completeness introduced in Definition 21. This yields connections between them analogously to the case of admissibility (cf. Figure 3).

**Lemma 24** (ELM-CMP  $\rightarrow$  PRCMP, ELM-CMP  $\rightarrow$  JNTCMP). Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is element-wise complete, then  $\mu$  is probabilistically and joint-attack complete.

**Lemma 25** (PRCMP  $\rightarrow$  MINCMP, JNTCMP  $\rightarrow$  MINCMP). Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is probabilistically or joint-attack complete, then  $\mu$  is min-complete.

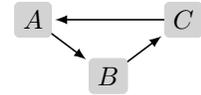
**Lemma 26** (MINCMP  $\rightarrow$  WCMP). Let  $\mathcal{F}$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . If  $\mu$  is min-complete, then  $\mu$  is weakly complete.

We now provide examples illustrating that the implications between the different notions of completeness are strict.

**Example 27.** An example for a distribution that is both joint-attack and probabilistically complete, but without being element-wise complete, is a distribution  $\mu$  that assigns probability  $\frac{1}{3}$  to the assignments  $\text{id}_{\{A\}}$ ,  $\text{id}_{\{B\}}$ , and  $\text{id}_{\{C\}}$  for the simple odd-length cycle AF on the right below: Neither  $\{A\}$ ,  $\{B\}$ , nor  $\{C\}$  are complete sets, so  $\mu$  is not element-wise complete. Because  $A$ ,  $B$ , and  $C$  are isomorphic in the AF under  $\mu$ , for all constraints imposed by PRCMP and JNTCMP the left-hand side equals the right-hand side, so  $\mu$  is probabilistically and joint-attack complete.

For an example of a weakly complete distribution that is not complete w.r.t. any of the other four completeness notions, consider the distribution  $\nu$  with the following support:

$$\begin{aligned} \nu(A \wedge \neg B \wedge \neg C) &= \frac{1}{2} \\ \nu(\neg A \wedge B \wedge \neg C) &= \frac{1}{3} \\ \nu(\neg A \wedge \neg B \wedge C) &= \frac{1}{6} \end{aligned}$$



Then  $Arg_\nu = \emptyset = \text{asDefend}(\nu)$ , so  $\nu$  is weakly complete. However,

$$\nu(A) = \frac{1}{2} \not\leq \frac{1}{6} = \nu(B) = \min_{C' \in \neg A} \nu\left(\bigvee_{B' \in \neg C'} B'\right)$$

and thus, for argument  $A$  the MINADM constraint is violated. To this end,  $\nu$  is not min-complete and thereby not probabilistically, joint-attack, or element-wise complete.

The distribution of Example 19 is joint-attack complete, but not probabilistically admissible, and therefore not probabilistically complete. Vice versa, the distribution  $\mu$  of Example 20 is probabilistically complete, but not joint-attack admissible, and therefore not joint-attack complete. Note that  $\mu$  satisfies the JNTCMP constraint for the arguments  $A$ ,  $C$ , and  $D$  but not for  $B$ , since JNTADM is violated due to  $\mu(B) = \frac{1}{4} > 0 = \mu(\Delta(B))$ . However, the latter inequation also implies that  $\mu$  satisfies PRCMP for  $B$  and together with PRADM holding for  $A$ ,  $C$ , and  $D$ ,  $\mu$  satisfies PRCMP for all arguments.

**Complexity.** With the exception of weak admissibility and weak completeness, the semantics introduced in this section share the characteristics that each of them imposes a certain set of linear constraints on the joint probability distributions over assignments. Checking the existence of a distribution satisfying these constraints is tractable in exponential time (as the feasibility of linear constraint systems can be checked in polynomial time and the number of variables grows exponentially in the number of arguments). An exponential time bound for semantics with the weak admissibility/completeness constraints can be obtained by encountering all subsets of  $Arg$  as candidates for the arguments with a likelihood of one and checking the feasibility of the resulting linear constraint system for each of them.

## 4 Taxonomy and Related Work

There is a large body of work on probabilistic extensions of argumentation frameworks. In general, one distinguishes between the *constellations* approach (Dung and Thang 2010; Li, Oren, and Norman 2011; Hunter 2012; Fazzinga, Flesca, and Parisi 2015) where uncertainty pertains the topology of the framework, and the *epistemic* approach (Hunter and Thimm 2017; Potyka 2019) where the framework is fixed and uncertainty revolves around the acceptance of arguments. This paper falls into the latter category.

In (Baroni, Giacomin, and Vicig 2014), epistemic probabilities are approached from the angle of de Finetti's theory of subjective probabilities (de Finetti 1974). They consider rationality conditions based on the notions of defense and reinstatement, which are closely related to admissibility and completeness.

An investigation of variants of semantics giving uniform distributions over the complete, preferred, and semi-stable labelings of an AF is given in (Rienstra et al. 2018). They show that the schemes investigated produce semantics which are founded, rational, and coherent (see Definition 28). The authors introduce new principles for probabilistic semantics based on SCC-decomposability and SCC-factorability.

In (Thimm et al. 2017), the authors consider probability distributions over argument sets that satisfy classical semantics. This approach is equivalent to our notion of element-wise lifting (see Definition 5).

Most closely related to our approach is the work by Hunter and Thimm (Hunter and Thimm 2017) which in turn is based

on earlier works (Thimm 2012; Hunter 2013), and in which the authors extend AFs towards a probabilistic setting by attributing a degree of belief to arguments. Besides others, the authors consider the case where only for some arguments the degree of belief is a priori known and present probabilistic reasoning techniques to determine possible degrees of belief for the remaining arguments. They further introduce several probabilistic semantics that can be compared to our semantics. Adapted to our notation, we recall these notions below for the sake of self-containedness.

**Definition 28.** Let  $\mathcal{F} = \langle Arg, Att \rangle$  be an AF and  $\mu \in \text{Distr}(\mathcal{F})$ . Then  $\mu$  is called

- (FOU) founded iff  $\mu(A) = 1$  for all initial  $A \in Arg$ ,
- (SFOU) semi-founded iff  $\mu(A) \geq \frac{1}{2}$  for all initial  $A \in Arg$ ,
- (OPT) optimistic iff  $\mu(\neg A) \leq \sum_{B \in \neg A} \mu(B)$  holds for all  $A \in Arg$ ,
- (SOPT) semi-optimistic iff (OPT) holds for all non-initial  $A \in Arg$ ,
- (COH) coherent iff  $\mu(A) \leq \mu(\neg B)$  for all  $(A, B) \in Att$ ,
- (INV) involutory iff  $\mu(A) = \mu(\neg B)$  for all  $(A, B) \in Att$ ,
- (JUS) justifiable iff  $\mu$  is coherent and optimistic,
- (RAT) rational iff  $\mu(A) > \frac{1}{2}$  implies  $\mu(B) \leq \frac{1}{2}$  for all  $(A, B) \in Att$ ,
- (MIN) minimal iff  $\mu(A) = 0$  for all  $A \in Arg$ ,
- (NEU) neutral iff  $\mu(A) = \frac{1}{2}$  for all  $A \in Arg$ , and
- (MAX) maximal iff  $\mu(A) = 1$  for all  $A \in Arg$ .

Figure 3 gives an overview of all semantics introduced in this paper in perspective to the semantic notions by (Hunter and Thimm 2017). An arrow from one semantics to another, e.g.,  $\text{WCMP} \rightarrow \text{JUS}$ , indicates that  $\llbracket \mathcal{F} \rrbracket_{\text{WCMP}} \subseteq \llbracket \mathcal{F} \rrbracket_{\text{JUS}}$  for all arbitrary AFs  $\mathcal{F}$ . Further, there is at least one AF for each arrow such that the set inclusion is strict, and no other arrows (except for the transitive closure) exist. Proofs for  $\text{ELM-CF} \rightarrow \text{COH}$  and  $\text{ELM-ST} \rightarrow \text{JUS}$  are given in (Thimm et al. 2017). All other probabilistic semantics introduced in this paper entail or require almost-sure conflict-freeness of distributions as well, so they all imply coherency. However, apart from ELM-ST, no other semantics entails JUS as shown in the following example.

**Example 29.** Consider again the odd cycle AF  $\mathcal{F}$  appearing in Example 27 and the assignment  $\beta = \{A=F, B=F, C=F\}$ . The distribution  $\text{Dirac}_\beta$  is element-wise preferred and grounded as the corresponding empty argument set is the only element in both  $\llbracket \mathcal{F} \rrbracket_{\text{ST}}$  and  $\llbracket \mathcal{F} \rrbracket_{\text{GR}}$ . However,  $\text{Dirac}_\beta$  is not optimistic (and thus not justifiable) as, e.g.,  $\text{Dirac}_\beta(A) = 1 \not\leq \text{Dirac}_\beta(B) = 0$ .

In turn, all our notions of complete semantics are founded: They all imply weak completeness, and initial arguments have maximal likelihood in weakly complete distributions as initial arguments are always almost surely defended.

In the setting of (Hunter and Thimm 2017), COH and JUS are generalizations of conflict-free argument sets and the complete semantics, respectively. By our definition (cf. Definition 4), coherence is a conservative extension of CF, though this is not the case for JUS and CMP: The assignment  $\beta$  from Example 29 is complete but  $\text{Dirac}_\beta$  is not justifiable.

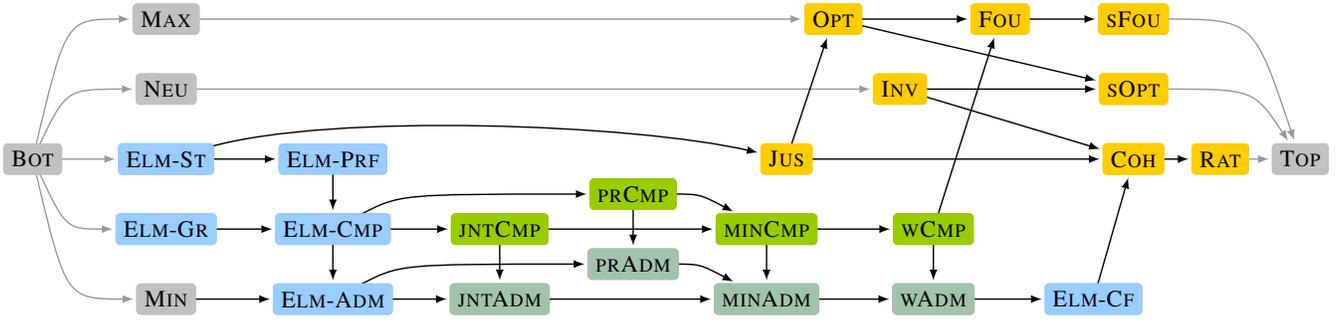
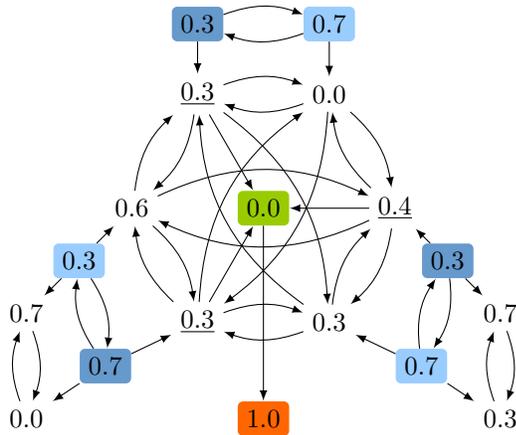


Figure 3: Hierarchy of probabilistic argumentation semantics. Gray boxes indicate trivial semantics, blue boxes stand for element-wise lifted classical semantics, yellow boxes refer to probabilistic semantics introduced by (Hunter and Thimm 2017), and green boxes to the notions of admissibility (Definition 12), respectively completeness (Definition 21).

**Example 30.** To illustrate some of the differences between our semantics and, in particular, justifiability semantics, we return to our motivating example from Section 1. Consider the following distribution  $\mu$  given below as the probabilities of all events in its support, where  $\mu(S)$  abbreviates  $\mu(\text{id}_S)$ .

$$\begin{aligned} \mu(\{\text{st}, \bar{\text{cl}}, \bar{\text{cl}} \Rightarrow \bar{\text{l}}, \bar{\text{cl}} \Rightarrow \bar{\text{m}}, \bar{\text{cr}}, \bar{\text{cr}} \Rightarrow \bar{\text{m}}, \bar{\text{cr}} \Rightarrow \bar{\text{r}}, \bar{\text{ld}}\}) &= 0.2 \\ \mu(\{\text{st}, \text{ld}, \bar{\text{cl}}, \bar{\text{cl}} \Rightarrow \bar{\text{l}}, \bar{\text{cl}} \Rightarrow \bar{\text{m}}, \bar{\text{cr}}, \bar{\text{cr}} \Rightarrow \bar{\text{m}}, \bar{\text{cr}} \Rightarrow \bar{\text{r}}\}) &= 0.1 \\ \mu(\{\text{st}, \text{cr}, \text{cr} \Rightarrow \text{r}, \text{ld}, \bar{\text{cl}}, \bar{\text{cl}} \Rightarrow \bar{\text{l}}, \bar{\text{cl}} \Rightarrow \bar{\text{m}}\}) &= 0.3 \\ \mu(\{\text{st}, \text{cr}, \text{cr} \Rightarrow \text{m}, \text{cr} \Rightarrow \text{r}, \text{ld}, \text{ld} \Rightarrow \text{m}, \bar{\text{cl}}, \bar{\text{cl}} \Rightarrow \bar{\text{l}}\}) &= 0.1 \\ \mu(\{\text{st}, \text{cl}, \text{cl} \Rightarrow \text{m}, \text{cr}, \text{cr} \Rightarrow \text{m}, \text{cr} \Rightarrow \text{r}, \bar{\text{ld}}\}) &= 0.1 \\ \mu(\{\text{st}, \text{cl}, \text{cl} \Rightarrow \text{m}, \text{cr}, \text{cr} \Rightarrow \text{m}, \text{cr} \Rightarrow \text{r}, \text{ld}, \text{ld} \Rightarrow \text{m}\}) &= 0.2 \end{aligned}$$

Based on Figure 2, we depict the resulting  $\mu$ -likelihood of each argument in the figure below:



Though not visible when looking only at the likelihoods, argument **st** (bottom center, red) with  $\mu(\text{st}) = 1$  is only defended by the three (underlined) attackers of **ct** (center, green) with a total probability of

$$\mu(\text{cl} \Rightarrow \text{m} \vee \text{ld} \Rightarrow \text{m} \vee \text{cr} \Rightarrow \text{m}) = 0.4.$$

That is, **st** is not almost surely defended. However, this is required by weak admissibility (and thus by all of our notions of admissibility and completeness) for arguments like **st** that hold almost surely.

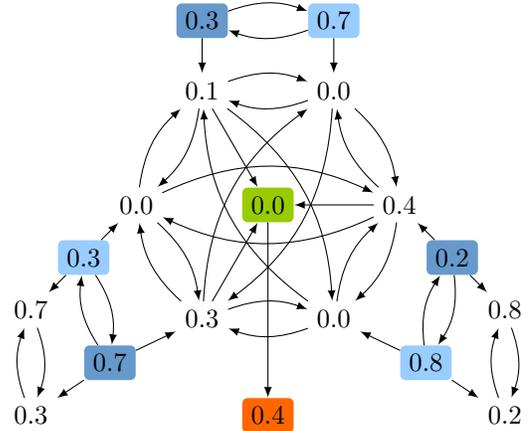
Justifiability, instead, considers solely the likelihoods of the immediate attackers. For the arguments in question, the optimism constraints for argument **st** and argument **ct** hold:

$$\mu(\neg \text{st}) = 0 \leq 0 = \mu(\text{ct})$$

$$\begin{aligned} \mu(\neg \text{ct}) = 1 &\leq 0.3 + 0.3 + 0.4 \\ &= \mu(\text{cl} \Rightarrow \text{m}) + \mu(\text{ld} \Rightarrow \text{m}) + \mu(\text{cr} \Rightarrow \text{m}) \end{aligned}$$

In fact, **OPT** holds for all arguments, so  $\mu$  is justifiable as coherency is given as well.

**Example 31.** In the figure below, we show the argument likelihoods induced by a distribution  $\mu$  that is joint-attack complete but not justifiable.



We see  $\mu$  is not justifiable as, e.g., for argument **st**, the **OPT** constraint is violated:  $\mu(\neg \text{st}) = 0.6 \not\leq 0 = \mu(\text{ct})$ . Note that **JNTCMP** cannot be verified from the likelihoods alone. We refer to Example 33 for details how  $\mu$  was constructed.

**The Bigger Picture.** Apart from the technical relations between these notions, there is a foundational difference between the semantical notions from (Hunter and Thimm 2017) and the notions we propose. The former only impose constraints on the likelihood of single arguments (which arise as marginal distributions  $\mu(A)$  in our setting) and therefore tend to be more coarse than our semantics from Section 3. The latter crucially exploit the possibility to impose constraints

across the *joint probability distributions*, and this makes it possible to express dependencies as needed already to spell out admissibility in an adequate manner.

Specifically, when restricting to constraints over the marginal probabilities, dependencies between the truth values of arguments cannot be expressed: Whenever a system of constraints for  $\mu(A)$  can be satisfied by at least one distribution, then the solution space contains at least (i) one distribution  $\mu$  where the arguments are pairwise independent and (ii) one distribution  $\mu$  such that for all arguments  $A, B$  where  $\mu(A) \geq \mu(B)$ , it is implied that  $B \rightarrow A$  holds for every assignment in the support of  $\mu$ . If  $B$  attacks  $A$  and the likelihood of  $B$  is positive then both (i) and (ii) are in contrast to standard argumentation semantics. In particular, linear constraints for  $\mu(A)$  cannot express that arguments are complements of each other and thus mutually exclusive, which is needed in our vehicle example. This observation should not be read as a critique at the earlier work by Hunter and Thimm as their focus is on modeling the belief of arguments and the induced three-values labelings, rather than a conservative extension of standard concepts (such as conflict-freeness or admissibility) on the level of distributions.

**Epistemic Approach to Abstract Argumentation.** In (Hunter, Polberg, and Thimm 2020) the authors extended the approach by (Hunter and Thimm 2017) and proposed *epistemic graphs*. Besides the notion of *support* of arguments that complements attacks (Boella et al. 2010), they augment argument graphs by constraints that restrict the degree of beliefs in arguments as well as how these beliefs influence each other. These constraints can be formulated as Boolean combinations of polynomial inequalities over terms denoting the probability of acceptance of argument sets (represented with propositional atoms). The semantics considered in (Hunter, Polberg, and Thimm 2020) associates sets of distributions over the powerset of arguments with each epistemic graph. The authors consider three forms of semantics or types of constraints on the distributions, the simplest being the satisfaction semantics, which simply returns the distributions consistent with the constraints of an argument graph. A central impulse for the development of epistemic graphs appears to be work on using argumentation for persuasion (Hadoux, Hunter, and Polberg 2021), which was already partly addressed also in (Hunter, Polberg, and Thimm 2020). Epistemic graphs are a very general framework. As a matter of fact, all constraints appearing in our work can be cast into the setting of epistemic graphs and its satisfaction semantics.

## 5 Implementation and Evaluation

To support the understanding and evaluation of probabilistic abstract argumentation frameworks, we have developed a prototypical implementation that is capable of answering a variety of questions arising naturally in this context.

**Tool Architecture.** Each semantics is implemented as a function taking a representation of an AF as input and returning a set of constraints on the induced joint distribution. Further, context specific constraints can be added right away. This approach enables the combination of semantics as needed, including the option to consider complement

semantics, and supports easy addition of new probabilistic semantics. The tool is implemented in Python, but provides integrated support for constraint analysis via a number of external solvers for the purpose of identifying satisfying distributions. Two kinds of solver back-ends are available to tackle different tasks: SMT solvers like Z3 are able to handle arbitrary polynomial constraints in the existential theory of the reals that may arise when checking for the existence of a distribution contained in one or more semantics. On the other hand, linear-optimization solvers can be used for tasks like synthesis of a distribution with maximal likelihood of a certain argument or to enumerate the distributions at the corners of the convex polytope which forms the solution space. The latter class of solvers is only applicable to semantics that are induced by linear constraints, so all of our notions except weak admissibility and weak completeness.

**Supported Functionality.** The tool provides support for

- synthesizing distributions satisfying the probabilistic argumentation semantics presented in this paper and (Hunter and Thimm 2017),
- maximizing or minimizing the marginal probability of certain user-defined arguments,
- exporting the corner distributions of the convex solution polytope,
- checking for credulous or skeptical acceptance of arguments with respect to thresholds,
- incorporating context-specific constraints in SMT-LIB format (Barrett, Fontaine, and Tinelli 2016),
- utilizing several state-of-the-art SMT solvers as back-end via the pySMT interface (Gario and Micheli 2015) and linear solvers via CVXOPT (Andersen, Dahl, and Vandenberghe 2014), and
- generating labelings according to a variety of labeling schemes.

The tasks involving acceptance checking and labeling schemes are described in more detail in (Käfer 2020).

**Practical Applicability.** The exponentiality of the number of constraints induced by admissibility and completeness might put doubts on the applicability of a tool on examples with growing sizes of argument sets. Our tool relies on state-of-the-art SMT solvers and linear constraint solvers that are designed to cope with this challenge. We give some impression on the applicability of our tool regarding several instances of the vehicle example.

**Example 32.** *Example 30 was produced by synthesizing a distribution that satisfies all constraints of the semantics JUS and ELM-CF, as well as the complement semantics of wCMP. Additionally, we imposed the following constraints on the likelihoods of some of the arguments:  $\mu(\overline{\text{cl}}) = 0.7$ ,  $\mu(\text{ld}) = 0.7$ ,  $\mu(\text{cr} \Rightarrow \text{r}) \in [0.7, 1]$ , and  $\mu(\text{cr} \Rightarrow \text{m}) \in [0.4, 1]$ . Intuitively, e.g., the latter means that the probability of camera right truly detecting an object in the middle in front of the car is between 0.4 and 1. Such constraints on marginal probabilities are introduced by annotating the specification of the AF's nodes and edges.*

instance	#nodes	distribution synthesis		maximizing probability of ct		all corners of solution polytope	
cl-ld-cr	6	0.017	(0.004)	0.006	(0.003)	0.022	(0.003)
cl-ld-cr	8	0.055	(0.018)	0.030	(0.018)	1.076	(0.018)
cl-ld-cr	12	1.286	(0.651)	2.736	(0.762)	> 600	timeout
cl-ld-cr	14	7.423	(3.913)	43.093	(15.887)	> 600	timeout
cl-ld-cr	18	361.033	(119.800)	> 600	timeout	> 600	timeout

Table 1: Running times (in seconds) on subgraphs of the vehicle example with increasing size for different tasks. cl-ld-cr is the full example, otherwise the arguments in grey stand for missing sensors and are dropped including related edges. The time to generate the constraints is given in parenthesis.

**Example 33.** For Example 31, the semantics JNTCMP and the complement of JUS were enforced, along with the same constraints on marginals as above. Furthermore, some context-specific constraints were imposed in an additional file in SMT-LIB format:

**Complementation** In the example, the arguments for the three sensors come with a complement, e.g., ld and  $\bar{ld}$ . We added constraints such that these arguments are in fact their respective inverse, i.e.,  $\mu(\text{ld}) = 1 - \mu(\bar{\text{ld}})$ .

**Conditional probability** A maximum 2% risk of a false positive detection was enforced via the conditional probability constraint  $\mu(\text{cr} \Rightarrow \text{r} \vee \text{cr} \Rightarrow \text{m} \mid \text{cr}) \geq 0.98$ .

**Dependency** In line with the sensor arrangement visualized in Section 1, we enforced  $\mu(\text{cr} \Rightarrow \text{r}) = 2 \cdot \mu(\text{cr} \Rightarrow \text{m})$ . This means that one third of the right camera view angle is monitoring the (overlapping) middle.

**Running Times.** All experiments were run on an Intel i9-10900K machine with 64GB of RAM, running Ubuntu 20.10 and Python 3.8.6. Computing solutions according to Example 32 took seventy minutes, and less than five minutes for Example 33. Table 1 provides further statistics for typical tasks applied to the vehicle example under PRCMP semantics. In a nutshell, the results are very encouraging. The tool and all experimental data are publicly available at <https://www.perspicuous-computing.science/cpraa/>.

## 6 Conclusion

In this paper we contributed to the quest for quantitative abstract argumentation frameworks from a probability-theoretic perspective. We have provided a profound study of admissibility and completeness and have discussed a hierarchy of resulting semantics, also in relation to earlier work. At the core of our approach, we are viewing each semantics as inducing sets of constraints on the joint distribution over argument sets.

We experimented with these semantic notions on a probabilistic abstract argumentation framework inspired by an autonomous driving scenario. For this, we implemented a tool based on SMT solvers to harvest present and future advances. In particular, by providing generic support for including additional constraints, it is capable of addressing a variety of adapted semantic notions well beyond the notions of admissibility and completeness spelled out in this paper, as well as adapting them towards context-specific needs. Our

tool is a research prototype, built for the ease of experimentation with semantic notions across the wider spectrum of probabilistic abstract argumentation frameworks. Indeed it turned out very helpful for the authors of this paper to sharpen their intuition.

**Future Work.** The tool already offers some elementary optimization tasks like maximizing the likelihood of selected arguments. Expanding on this functionality, we plan to investigate quantifications of how close distributions are to satisfy a certain semantics. This could allow to find, e.g., the “most” min-admissible distribution even if no distribution exists that completely satisfies the MINADM constraints. Likewise, basic support for *labeling schemes* is already available, connecting probabilistic argumentation semantics back to classical semantics by discretizing distributions. Different schemes exist and more are imaginable, calling for a thorough investigation of their properties and relationships.

While we gave membership results for the complexity of several tasks in probabilistic argumentation, establishing hardness (if possible) remains to be done.

Finally, we identified two further directions that might be worthwhile to explore under the lens of our probabilistic approach. In *dynamic abstract argumentation*, nodes and edges can be added or removed from an initial AF, giving rise to the question how these dynamic interventions affect, e.g., the accepted argument sets (see, e.g., (Diller et al. 2018) or (Doutre and Mailly 2018) for a survey). In the probabilistic setting, changes to the resulting distributions could be quantified. Secondly, several logical languages for abstract argumentation have been developed (see, e.g., YALLA (Dupin de Saint-Cyr et al. 2016)). Such languages allow to encode and subsequently reason about the basic notions of abstract argumentation and extensions in the probabilistic domain could prove useful.

## Acknowledgments

This research was partially funded by the DFG through the Collaborative Research Center TRR 248 (see <https://perspicuous-computing.science>, project ID 389792660), the Cluster of Excellence EXC 2050/1 (CeTI, project ID 390696704, as part of Germany’s Excellence Strategy), the Research Training Groups QuantLA (GRK 1763) and RoSI (GRK 1907), and the projects BA-1679/11-1 and BA-1679/12-1, by the Bundesministerium für Bildung und Forschung (BMBF) Förderkennzeichen 01IS20056\_NAVAS,

by the European Research Council through the Advanced Investigators Grant 695614 (POWVER), and by the Key-Area Research and Development Program Grant 2018B010107004 of Guangdong Province.

## References

- Amgoud, L., and Cayrol, C. 2002. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence* 34(1-3):197–215.
- Amgoud, L.; Cayrol, C.; Lagasque-Schiex, M.; and Livet, P. 2008. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems* 23(10):1–32.
- Andersen, M.; Dahl, J.; and Vandenberghe, L. 2014. Cvxopt – python software for convex optimization. <http://cvxopt.org/>.
- Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An introduction to argumentation semantics. *Knowledge Engineering Review* 26(4):365–410.
- Baroni, P.; Giacomin, M.; and Vicig, P. 2014. On rationality conditions for epistemic probabilities in abstract argumentation. In Parsons, S.; Oren, N.; Reed, C.; and Cerutti, F., eds., *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA 2014)*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, 121–132. IOS Press.
- Barrett, C.; Fontaine, P.; and Tinelli, C. 2016. The Satisfiability Modulo Theories Library (SMT-LIB). [www.SMT-LIB.org](http://www.SMT-LIB.org).
- Baumann, R.; Brewka, G.; and Ulbricht, M. 2020. Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, 2742–2749. AAAI Press.
- Bench-Capon, T. J. M. 2003. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation* 13(3):429–448.
- Boella, G.; Gabbay, D. M.; van der Torre, L. W. N.; and Villata, S. 2010. Support in abstract argumentation. In Baroni, P.; Cerutti, F.; Giacomin, M.; and Simari, G. R., eds., *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA 2010)*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, 111–122. IOS Press.
- Brewka, G., and Woltran, S. 2010. Abstract Dialectical Frameworks. In Lin, F.; Sattler, U.; and Truszczyński, M., eds., *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, 102–111. AAAI Press.
- Brewka, G.; Straß, H.; Ellmauthaler, S.; Wallner, J. P.; and Woltran, S. 2013. Abstract Dialectical Frameworks Revisited. In Rossi, F., ed., *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, 803–809. IJCAI/AAAI.
- de Finetti, B. 1974. *Theory of Probability*. Wiley.
- Diller, M.; Haret, A.; Linsbichler, T.; Rümmele, S.; and Woltran, S. 2018. An extension-based approach to belief revision in abstract argumentation. *Int. J. Approx. Reason.* 93:395–423.
- Doutre, S., and Maily, J. 2018. Constraints and changes: A survey of abstract argumentation dynamics. *Argument Comput.* 9(3):223–248.
- Dung, P. M., and Thang, P. M. 2010. Towards (probabilistic) argumentation for jury-based dispute resolution. In Baroni, P.; Cerutti, F.; Giacomin, M.; and Simari, G. R., eds., *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA 2010)*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, 171–182. IOS Press.
- Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–358.
- Dunne, P. E.; Hunter, A.; McBurney, P.; Parsons, S.; and Wooldridge, M. 2011. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence* 175(2):457–486.
- Dupin de Saint-Cyr, F.; Bisquert, P.; Cayrol, C.; and Lagasque-Schiex, M.-C. 2016. Argumentation update in yalla (yet another logic language for argumentation). *International Journal of Approximate Reasoning* 75:57–92.
- Faqeh, R.; Fetzer, C.; Hermanns, H.; Hoffmann, J.; Klauck, M.; Köhl, M. A.; Steinmetz, M.; and Weidenbach, C. 2020. Towards dynamic dependable systems through evidence-based continuous certification. In Margaria, T., and Steffen, B., eds., *Proceedings of the 9th International Symposium on Leveraging Applications of Formal Methods, ISoLA 2020, Verification and Validation: Engineering Principles, Part II*, volume 12477 of *Lecture Notes in Computer Science*, 416–439. Springer.
- Fazzinga, B.; Flesca, S.; and Parisi, F. 2015. On the complexity of probabilistic abstract argumentation frameworks. *ACM Transactions on Computational Logic* 16(3):22:1–22:39.
- Gaggl, S. A.; Rudolph, S.; and Straß, H. 2021. On the decomposition of abstract dialectical frameworks and the complexity of naive-based semantics. *Journal of Artificial Intelligence Research* 70:1–64.
- Gario, M., and Micheli, A. 2015. Pysmt: a solver-agnostic library for fast prototyping of smt-based algorithms. In *SMT Workshop 2015*.
- Hadoux, E.; Hunter, A.; and Polberg, S. 2021. Strategic argumentation dialogues for persuasion: Framework and experiments based on modelling the beliefs and concerns of the persuadee. *CoRR* abs/2101.11870.
- Hunter, A., and Thimm, M. 2017. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research* 59:565–611.
- Hunter, A.; Polberg, S.; and Thimm, M. 2020. Epistemic graphs for representing and reasoning with positive

and negative influences of arguments. *Artificial Intelligence* 281:103236.

Hunter, A. 2012. Some foundations for probabilistic abstract argumentation. In Verheij, B.; Szeider, S.; and Woltran, S., eds., *Proceedings of the International Conference on Computational Models of Argument (COMMA 2012)*, volume 245 of *Frontiers in Artificial Intelligence and Applications*, 117–128. IOS Press.

Hunter, A. 2013. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning* 54(1):47–81.

Käfer, N. 2020. A general framework for probabilistic abstract argumentation. Master's thesis, Saarland University.

Li, H.; Oren, N.; and Norman, T. J. 2011. Probabilistic argumentation frameworks. In Modgil, S.; Oren, N.; and Toni, F., eds., *Revised Selected Papers of the 1st International Workshop on Theory and Applications of Formal Argumentation (TAFAs 2011)*, volume 7132 of *Lecture Notes in Computer Science*, 1–16. Springer.

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*. College Publications. chapter 6.

Nouioua, F., and Risch, V. 2011. Argumentation frameworks with necessities. In Benferhat, S., and Grant, J., eds., *Proceedings of the 5th International Conference on Scalable Uncertainty Management (SUM 2011)*, volume 6929 of *Lecture Notes in Computer Science*, 163–176. Springer.

Polberg, S., and Doder, D. 2014. Probabilistic abstract dialectical frameworks. In Fermé, E., and Leite, J., eds., *Proceedings of the 14th European Conference on Logics in Artificial Intelligence (JELIA 2014)*, volume 8761 of *Lecture Notes in Computer Science*, 591–599. Springer.

Potyka, N. 2019. A polynomial-time fragment of epistemic probabilistic argumentation. *International Journal of Approximate Reasoning* 115:265–289.

Rienstra, T.; Thimm, M.; Liao, B.; and van der Torre, L. W. N. 2018. Probabilistic abstract argumentation based on SCC decomposability. In Thielscher, M.; Toni, F.; and Wolter, F., eds., *Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018)*, 168–177. AAAI Press.

Straß, H., and Wallner, J. P. 2015. Analyzing the Computational Complexity of Abstract Dialectical Frameworks via Approximation Fixpoint Theory. *Artificial Intelligence* 226:34–74.

Thimm, M.; Baroni, P.; Giacomin, M.; and Vicig, P. 2017. Probabilities on extensions in abstract argumentation. In Black, E.; Modgil, S.; and Oren, N., eds., *Proceedings of the 4th International Workshop on Theory and Applications of Formal Argumentation (TAFAs 2017)*, volume 10757 of *Lecture Notes in Computer Science*, 102–119. Springer.

Thimm, M. 2012. A probabilistic semantics for abstract argumentation. In Raedt, L. D.; Bessière, C.; Dubois, D.; Doherty, P.; Frasconi, P.; Heintz, F.; and Lucas, P. J. F., eds., *Proceedings of the 20th European Conference on Artificial*

*Intelligence (ECAI 2012). Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, 750–755. IOS Press.