Conditional Probabilistic Bipolar Argumentation Framework: Explanations, Complexity and Approximation

Gianvincenzo Alfano, Sergio Greco, Domenico Mandaglio, Francesco Parisi, Irina Trubitsyna

Department of Informatics, Modeling, Electronics and System Engineering, University of Calabria, Italy
{g.alfano, greco, d.mandaglio, fparisi, i.trubitsyna}@dimes.unical.it

Abstract

Recently, there has been an increasing interest in extending Dung's framework with probability theory, leading to the Probabilistic Argumentation Framework (PAF), and with supports in addition to attacks, leading to the Bipolar Argumentation Framework (BAF). In this paper, we introduce the Conditional Probabilistic Bipolar Argumentation Framework (CPBAF), which extends Probabilistic and Bipolar AF by allowing conditional probabilities on arguments, attacks, and on (possibly cyclic) supports. In this setting, we address the problem of computing the probability that a given argument is accepted. This is carried out by introducing the concept of probabilistic explanation for a given (probabilistic) extension. We show that the complexity of the problem is $\text{FP}^{\#\text{P}}$-hard and propose polynomial approximation algorithms with bounded additive error for CPBAF where cycles with an odd number of attacks are forbidden.

1 Introduction

Research on rational discourse and conflict resolution has become increasingly prominent in Artificial Intelligence, prompting the development of the formal argumentation field (Bench-Capon and Dunne 2007; Simari and Rahwan 2009; Atkinson et al. 2017). A foundational model in this domain is Dung's *abstract Argumentation Framework* (AF) (Dung 1995), which captures conflicts among agents using a set of *arguments* and a binary *attack* relation.

To extend its expressive power, various refinements have been proposed, e.g., (Alfano et al. 2023b,c, 2024b,c,e, 2025d,e,c). *Bipolar Argumentation Frameworks* (BAFs) introduce *support* relations in addition to attacks (Nouioua and Risch 2011; Villata et al. 2012; Alfano et al. 2024a), allowing for cooperative and adversarial links between arguments.

Recently, increasing attention has been given to modeling uncertainty in argumentation. Probabilistic approaches, such as the *constellation* one (Dung and Thang 2010; Rienstra 2012; Doder and Woltran 2014; Hunter 2012; Li, Oren, and Norman 2011; Popescu and Wallner 2024), address this by associating probabilities with alternative configurations of arguments—*possible worlds*—thus enabling reasoning under incomplete knowledge. In a *Probabilistic Argumentation Framework* (PAF) (Li, Oren, and Norman 2011; Fazzinga, Flesca, and Parisi 2015, 2016; Fazzinga, Flesca, and Furfaro 2019), and in its extension called *Probabilistic Bipolar Argumentation Framework* (PBAF) (Fazzinga, Flesca, and Furfaro 2018b), a probability distribution function (PDF) on the set of possible worlds is entailed by the probabilities that are associated with arguments, attacks, and supports.

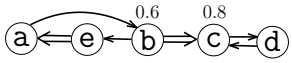

Figure 1: Probabilistic Bipolar AF Δ of Example 1.

Example 1. Consider a situation where a party planner invites Alice, Bob, Carl, David, and Erik to join a party. The party planner knows that: (i) Bob (resp., Carl, David, and Erik) replies that he will join the party if Alice (resp., David, Carl, and Bob) does not, and (ii) to let Carl (resp., Alice) join the party, it is necessary that Bob (resp., Erik) joins the party as well. Moreover, the party planner knows that Bob's and Carl's availability for the party's date is uncertain, with probability 0.6 for Bob and 0.8 for Carl (while the others are certainly available). This situation can be modeled by means of the PBAF Δ whose corresponding graph is shown in Figure 1, where nodes represent arguments, single edges ($\rightarrow$) represent attacks, and double edges ($\Rightarrow$) represent supports, whereas probabilities different from 1 are specified nearby them. Argument x states that "(the person whose name's initial is) x joins the party". □

In this paper we do not address the problem of assigning probabilities to arguments or attacks/supports, as is instead done e.g. in (Hunter 2012, 2013), and assume they are given.

Several argumentation semantics—e.g. *grounded* (gr), *complete* (co), *preferred* (pr), and *stable* (st)—have been defined for Bipolar AFs, leading to the characterization of σ*-extensions*, which intuitively consist of the sets of arguments that can be collectively accepted under semantics σ. Consider for instance the deterministic version of the PBAF in Example 1, obtained by assuming that all arguments are certain (i.e., they have probability 1), the stable extensions are $E_1 = \{\texttt{b}, \texttt{c}\}$, $E_2 = \{\texttt{b}, \texttt{d}\}$, and $E_3 = \{\texttt{a}, \texttt{d}, \texttt{e}\}$.

The semantics of a PBAF is given by considering all possible worlds (i.e., BAFs) obtained by removing consistent subsets of the probabilistic elements. Every possible world has associated a probability value derived from the probabil-

ities of the elements that have been kept or removed. Moreover, every possible world admits a set of σ-extensions.

Example 2 (cont'd)**.** There exist four possible worlds w_i (with $i \in [1, 4]$). w_1 is the BAF obtained from the PBAF by keeping all arguments, attacks, and supports. Moreover, w_2 (resp., w_3, and w_4) is the BAF obtained from the PBAF by removing c (resp., b, both b and c) and, consistently with this, the attacks and supports towards/from the removed argument(s). As it will be clear later (cf. Example 7), the probabilities of the possible worlds w_1, w_2, w_3, and w_4 are 0.48, 0.12, 0.32, and 0.08. Since w_1 coincides with the deterministic version of Δ, its st-extensions are E_1, E_2, and E_3 given earlier. The st-extensions of w_2 (resp., w_3) are E_2 and E_3 (resp., $E_4 = \{\mathtt{a}, \mathtt{e}, \mathtt{c}\}$ and E_3). Finally, w_4 admits only E_3 as an st-extension. □

Notably, PBAFs currently defined in the literature do not allow for encoding i) *cyclic supports* (though the semantics of cyclic Bipolar AFs has been investigated), as well as ii) *conditional probabilities*, that would allow to model more general situations, as shown in the following example.

Example 3 (cont'd)**.** Assume now that the party planner desires to model the fact that the probability that David will attend the party, provided that both Bob and Carl will attend too, is 0.4. This can be modeled by means of a conditional probability $P(\mathtt{d} \mid \mathtt{b} \wedge \mathtt{c}) = 0.4$. □

To this end, we introduce *Conditional Probabilistic Bipolar Argumentation Frameworks* (CPBAF), where conditional probabilities on arguments, attacks, and (possibly cyclic) supports can be easily modeled.

Interesting problems recently investigated in the context of probabilistic argumentation are the *probabilistic credulous and skeptical acceptance* problems (Fazzinga, Flesca, and Furfaro 2018a, 2019), and their generalization to *probabilistic acceptance* (Alfano et al. 2023a), where a probability distribution function over the set of σ-extensions of possible worlds is considered. Given a PAF (i.e., a PBAF without supports) Δ and a goal argument g, the probabilistic acceptance (denoted as $\mathsf{PrA}[\sigma]$) returns the probability that g is accepted under semantics $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{pr}, \mathtt{st}\}$. In more detail, $\mathsf{PrA}[\sigma]$ implicitly assumes that a PDF over the set of σ-extensions of any possible world of PAF Δ is defined (herein a possible world of a PAF is an AF). Thus, a concrete instance of $\mathsf{PrA}[\sigma]$ is obtained after defining such a PDF.

In this paper, we explore an instantiation of $\mathsf{PrA}[\sigma]$ in CPBAF where the PDF over the σ-extensions of a world relies on the concept of *explanation*, firstly introduced in (Alfano et al. 2020) for PAF. This problem is called *Explanation-based Probabilistic Acceptance*, denoted as $\mathsf{PrEA}[\sigma]$. Intuitively, an explanation for a σ-extension E is a *sequence of arguments* occurring in E that 'justify' E. Every explanation is associated with a probability entailed by the possible choices that can be made. These choices must be consistent with an ordering entailed by the strongly connected components of the underlying BAF, and they are used to guide the construction of an extension. The sum of the probabilities of the explanations for an extension E gives the probability of E. Thus, given a CPBAF Δ, we assign (i) to each possible world w of Δ, a probability $I(w)$, and (ii) to each extension E of w (under semantics σ), a probability $Pr(E, w, \sigma)$, whose value is based on the probabilities of the explanations of E.

Contributions. In this paper we tackle a new problem that we call *Probabilistic Acceptance* of an argument.

- We first introduce the *Conditional Probabilistic Bipolar Argumentation Framework* (CPBAF) that extends PBAF by allowing the modeling of conditional probabilities on arguments, attacks, and supports. Moreover, in CPBAF the support relation may be cyclic, encompassing the limitations of previous approaches.
- We define the problem of Probabilistic Acceptance $\mathsf{PrA}[\sigma]$, for some semantics σ, in CPBAF. Given a CPBAF Δ and an argument g, the problem asks for the probability that g is accepted in Δ, by means of some fixed PDF over the σ-extensions of the possible worlds of Δ.
- We introduce our notion of explanation for a BAF Δ, and exploit it to provide a PDF over the σ-extensions of Δ. This leads to an instantiation of $\mathsf{PrA}[\sigma]$, dubbed $\mathsf{PrEA}[\sigma]$.
- To deal with the intractability of $\mathsf{PrA}[\sigma]$ and $\mathsf{PrEA}[\sigma]$, whose complexity is shown to be $\mathrm{FP}^{\#\mathrm{P}}$-hard, we propose an *additive approximation algorithm* for $\mathsf{PrEA}[\sigma]$ for CPBAF without cycles with an odd number of attacks and semantics $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$.

2 Preliminaries

We briefly recall the Bipolar Argumentation Framework and its probabilistic variant (where support cycles are avoided). We assume familiarity with basic complexity theory notions.

2.1 Bipolar Argumentation Frameworks

A *Bipolar Argumentation Framework* (BAF) is a triple $\langle A, R, S\rangle$, where A is a set of *arguments*, $R \subseteq A \times A$ is a set of *attacks*, and $S \subseteq A \times A$ is a set of *supports* (with $R \cap S = \emptyset$). A BAF can be represented by a directed graph with two types of edges: *attacks* and *supports*, denoted by $\rightarrow$ and $\Rightarrow$, respectively. A *support path* $a_0 \Rightarrow a_n$ from argument a_0 to argument a_n is a sequence of $n > 0$ edges $a_{i-1} \Rightarrow a_i$ with $0 < i \leq n$. We use $\vec{a} = \{b \in A \mid b \rightarrow a \vee \exists c \in A.(b \rightarrow c) \wedge c \Rightarrow a\}$ to denote the set of arguments that directly or transitively attack a.

Given a cycle of attacks and supports, we say that the *cycle is odd* (resp., *even*) if it contains an odd (resp., even) number of attacks. A BAF is (i) *acyclic* if it does not have cycles, (ii) *support-acyclic* if it does not have cycles containing only supports, i.e. there is no argument a such that $a \Rightarrow a$, (iii) *odd-cycle-free* if it does not have odd cycles.

Example 4. The BAF encoding the deterministic version of the PBAF of Example 1 is $\Lambda = \langle A = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}, R = \{(\mathtt{a}, \mathtt{b}), (\mathtt{b}, \mathtt{e}), (\mathtt{c}, \mathtt{d}), (\mathtt{d}, \mathtt{c})\}, S = \{(\mathtt{b}, \mathtt{c}), (\mathtt{e}, \mathtt{a})\}\rangle$. □

Different interpretations of the support relation have been proposed in the literature (Simari and Rahwan 2009; Cayrol, Cohen, and Lagasquie-Schiex 2021; Villata et al. 2012). We focus on necessary supports,[1] whose semantics is intended

[1]The results can be easily extended to the deductive interpretation of the support relation by reversing the direction of supports.

to capture the following intuition: if argument a supports argument b, then to have b accepted, it is necessary that a is accepted as well; the non-acceptance of a implies the non-acceptance of b (Nouioua and Risch 2011). Moreover, we consider BAF that could be possibly cyclic, whose semantics is recalled next.

Definition 1 (BAF Semantics (Alfano et al. 2024d))**.** *For any BAF $\langle A, R, S\rangle$ and set of arguments $E \subseteq A$:*

- $\text{DEF}(E) = E^+ = \{a \in A \mid (\exists b \in E \,.\, b \rightarrow a) \vee (\exists b \in \text{DEF}(E) \,.\, b \Rightarrow a) \vee a \Rrightarrow a\}$;
- $\text{ACC}(E) = \{a \in A \mid (\forall b \in A \,.\, b \rightarrow a \textit{ implies } b \in \text{DEF}(E)) \wedge (\forall c \in A \,.\, c \Rightarrow a \textit{ implies } c \in \text{ACC}(E))\}$.

Observe that the set of acceptable arguments $\text{ACC}(E)$ explicitly excludes self-supported arguments (i.e., arguments in a support cycle or supported by an argument in a support cycle), that are assumed to be defeated. For a given set E, $E^* = E \cup E^+$ denotes the set of elements which are either in E or defeated by E (i.e., $E^* = E \cup \text{DEF}(E)$).

Different argumentation semantics have been defined leading to the characterization of collectively acceptable sets of arguments, called *extensions* (Dung 1995).

Given a BAF $\Lambda = \langle A, R, S\rangle$, a set $E \subseteq A$ of arguments is said to be i) *conflict-free* iff $E \cap \text{DEF}(E) = \emptyset$, and ii) *admissible* iff it is conflict-free and $E \subseteq \text{ACC}(E)$. Moreover, E is an extension:

- *complete* iff it is conflict-free and $E = \text{ACC}(E)$;
- *preferred* iff it is a $\subseteq$-maximal complete extension;
- *stable* iff it is a total ($E \cup \text{DEF}(E) = A$) preferred extension;
- *grounded* iff it is the $\subseteq$-smallest complete extension.

The set of σ-extensions (with $\sigma \in \{\texttt{gr}, \texttt{co}, \texttt{st}, \texttt{pr}\}$) of a BAF Λ will be denoted by $\sigma(\Lambda)$. If we consider a generic semantics σ, we refer to a semantics in $\{\texttt{gr}, \texttt{co}, \texttt{pr}, \texttt{st}\}$.

Given an extension E, we say that the *status* of an argument a is true (resp., false, undecided) if $a \in E$ (resp., $a \in E^+$, $a \in A \setminus E^*$). We also say that a is accepted, defeated, or undefined (w.r.t. E), respectively.

Clearly, if the set S of supports is: i) acyclic, then the semantics extends that of AF with Necessities (Nouioua and Risch 2011); ii) empty, then a BAF is an AF (Dung 1995).

Example 5 (cont'd)**.** Consider the cyclic BAF Λ' obtained from Λ by replacing attacks $(\texttt{a}, \texttt{b})$ and $(\texttt{b}, \texttt{e})$ with supports. Then, $\text{DEF}(\{\texttt{d}\}) = \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{e}\}$ and $\text{ACC}(\{\texttt{d}\}) = \{\texttt{d}\}$. Under necessary interpretation of supports, Λ' has a unique complete extension, $\{\texttt{d}\}$, that is grounded, preferred and stable.□

Given a BAF $\Lambda = \langle A, R, S\rangle$ and set $E \subseteq A$, $\Lambda_{\downarrow E} = \langle E, R \cap (E \times E), S \cap (E \times E)\rangle$ denotes the *restriction* of Λ to E. Moreover, $\Lambda_{\uparrow E} = \Lambda_{\downarrow A \setminus E}$ denotes the restriction of Λ to $A \setminus E$. A strongly connected component (SCC) of Λ is a maximal subset $\mathcal{C}$ of A such that, for every pair of arguments $a, b \in \mathcal{C}$, there is a path from a to b along the attack and support relation in the graph representing Λ. Note that, differently from the standard definition, we use SCC to denote a set of nodes (i.e. arguments), not a subgraph. Hereafter, we assume the existence of a linear ordering over the SCCs of Λ that follows the topological ordering of the underlying graph. An SCC is said to be *first* if there is no other SCC preceding it in the ordering. As an example, the two SCCs of the BAF shown in Figure 1 (i.e., the deterministic version of the PBAF in Example 1) are $\{\texttt{a}, \texttt{b}, \texttt{e}\}$ and $\{\texttt{c}, \texttt{d}\}$, where the former precedes the latter.

2.2 Probabilistic Acyclic BAF

Probabilistic acyclic Bipolar Argumentation Framework (PBAF), introduced in (Fazzinga, Flesca, and Furfaro 2018b), extends acyclic BAF with probabilities on arguments, attacks, and supports.

Definition 2. *A Probabilistic Bipolar Argumentation Framework (PBAF) is a quadruple $\langle A, R, S, P\rangle$ where $\langle A, R, S\rangle$ is a BAF, and P is a total function assigning a non-zero marginal probability value to every element in $A \cup R \cup S$, that is, $P : (A \cup R \cup S) \rightarrow (0, 1]$.*

In this section, for the sake of the presentation, we assume that only arguments are uncertain (and attacks and supports are certain, i.e., their probability is 1). This is w.l.o.g. as any PBAF can be rewritten into an equivalent one where only arguments are uncertain (see in Section 3.1). Also, *marginal probabilities equal to* 1 *will not be explicitly defined*.

Example 6. The PBAF underlying the scenario of Example 1 is $\Delta = \langle A, R, S, P\rangle$, where $\langle A, R, S\rangle$ is the BAF of Example 4, while $P = \{P(\texttt{b}) = 0.6, P(\texttt{c}) = 0.8\}\rangle$. □

Intuitively, the value assigned by P to any argument a represents the probability that a actually occurs. The formal meaning of a PBAF is given in terms of *possible worlds*. Given a PBAF $\Delta = \langle A, R, S, P\rangle$, a possible world of Δ is a BAF $w = \langle A', R', S'\rangle$ such that $A' \subseteq A$, $R' = R \cap (A' \times A')$, $S' = S \cap (A' \times A')$, and arguments $a \in A$ occur in A' whenever they are certain (i.e., $P(a) = 1$). We use $pw(\Delta)$ to denote the set of all possible worlds of Δ.

An *interpretation* for a PBAF $\Delta = \langle A, R, S, P\rangle$ is a PDF I over the set $pw(\Delta)$ of the possible worlds. Each $w = \langle A', R', S'\rangle \in pw(\Delta)$ is assigned by I the probability:

$$I(w) = \prod_{a \in A'} P(a) \times \prod_{a \in A \setminus A'} \big(1 - P(a)\big). \qquad (1)$$

Example 7. The possible worlds of the PBAF Δ of Examples 1 and 6 are w_1, w_2, w_3 and w_4 given in Example 2. Then, interpretation I is as follows:[2]

- $I(w_1) = P(\texttt{b}) \cdot P(\texttt{c}) = 0.6 \cdot 0.8 = 0.48$,
- $I(w_2) = P(\texttt{b}) \cdot (1 - P(\texttt{c})) = 0.6 \cdot 0.2 = 0.12$,
- $I(w_3) = (1 - P(\texttt{b})) \cdot P(\texttt{c}) = 0.4 \cdot 0.8 = 0.32$,
- $I(w_4) = (1 - P(\texttt{b})) \cdot (1 - P(\texttt{c})) = 0.4 \cdot 0.2 = 0.08$. □

3 Conditional Probabilistic BAF

Before defining the syntax and semantics of the proposed framework, we introduce some notation.

In defining probabilities for attacks and supports, we will use the symbol α_{ab} (resp., β_{ab}) to denote the attack (resp., support) (a, b). For instance, to say that the probability that the attack (a, b) occurs, provided that the support (b, c) occurred, we write $P(\alpha_{ab} | \beta_{bc})$. Given a set R of attacks and a set S of supports, we introduce the sets $\alpha = \{\alpha_{ab} \mid (a, b) \in$

[2] Probabilities equal to one have been omitted, as they do not affect the final result.

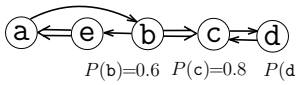

Figure 2: CPBAF Δ of Example 8.

$R\}$ and $\beta = \{\beta_{ab} \mid (a,b) \in S\}$, that is, α and β are alternative ways to denote the sets of attacks and supports, respectively, that will be used in what follows.

Definition 3 (CPBAF)**.** *A Conditional Probabilistic Bipolar Argumentation Framework (CPBAF) is a quadruple $\langle A, R, S, P\rangle$ where $\langle A, R, S\rangle$ is an BAF, and P is a total conditional probability function assigning a non-zero probability value $P(a|C_a)$ to every element $a \in A \cup \alpha \cup \beta$, where C_a is a conditional event consisting of a propositional logic formula whose atoms are taken from $A \cup \alpha \cup \beta$.*

The marginal probability of an element a will be simply denoted by $P(a)$ (instead of $P(a|true)$). *Again, marginal probabilities equal to* 1 *will not be explicitly defined.* We assume that the probabilistic function uniquely defines the probability of an element, i.e., there are no two distinct instances $P(a|C'_a)$ and $P(a|C''_a)$. Clearly, the probability that an attack α_{ab} or a supports β_{ab} occurs is conditional to the occurrence of the related arguments a and b. Thus, whenever we write $P(\alpha_{ab})$ (resp. $P(\alpha_{ab} \mid C_{ab})$) we mean $P(\alpha_{ab}|a \wedge b)$ (resp. $P(\alpha_{ab} \mid a \wedge b \wedge C_{ab})$). Analogous assumptions are made regarding supports' probabilities.

Example 8. The CPBAF of Example 3 is $\Delta = \langle A, R, S, P\rangle$, where $\langle A, R, S\rangle$ is the BAF of Example 4, while $P = \{P(\mathtt{b}) = 0.6, P(\mathtt{c}) = 0.8, P(\mathtt{d} \mid \mathtt{b} \wedge \mathtt{c}) = 0.4\}$, as shown in Figure 2. □

Given a CPBAF $\langle A, R, S, P\rangle$ the *probability dependency graph* is a graph whose nodes are elements in $A \cup \alpha \cup \beta$ and there is an edge from b to a if atom b is conditioning atom a, that is, let $P(a|C_a)$ be the conditional probability of a, b occurs in the formula C_a. We say that a CPBAF is *well defined* if the probability dependency graph is acyclic. *From now on, we assume that CPBAFs are well defined.*

The semantics of CPBAF is given in terms of *possible worlds*; it naturally extends that of PBAF as follows. Given a CPBAF $\Delta = \langle A, R, S, P\rangle$, a possible world of Δ is a BAF $w = \langle A', R', S'\rangle$ such that $A' \subseteq A$, $R' \subseteq R \cap (A' \times A')$, $S' \subseteq S \cap (A' \times A')$, $w \models C_t$ for any $t \in A' \cup R' \cup S'$, and elements $t \in A \cup R \cup S$ s.t. $P(t|C_t) = 1$ and $w \models C_t$ occur in w. Herein, $w \models C_t$ means that the formula C_t is true in w.[3] This condition generalizes that of possible worlds for PBAF where only marginal probabilities are defined and thus C_t is $true$ for each argument t. We use $pw(\Delta)$ to denote the set of all possible worlds of CPBAF Δ.

Definition 4. *An interpretation for a CPBAF Δ is a probability distribution function I over $pw(\Delta)$. I assigns to each $w \in pw(\Delta)$ the probability*

$$I(w) = \prod_{t \in T' \wedge w \models C_t} P(t \mid C_t) \times \prod_{t \in T \setminus T' \wedge w \models C_t} \big(1 - P(t \mid C_t)\big), \quad (2)$$

where $T = A \cup R \cup S$ and $T' = A' \cup R' \cup S'$.

[3] In evaluating $w \models C_t$, an atom a is *true* iff it occurs in w.

Example 9. The possible worlds of the CPBAF $\Delta = \langle A, R, S, P\rangle$ of Example 8 are the following:

- $w_1 = \langle A, R, S\rangle$ with probability $I(w_1) = .6 \times .8 \times .4 = .192$;
- $w_2 = \langle \{\mathtt{a}, \mathtt{e}, \mathtt{b}, \mathtt{c}\}, \{(\mathtt{a}, \mathtt{b}), (\mathtt{b}, \mathtt{e})\}, S\rangle$ with probability $I(w_2) = .6 \times .8 \times (1 - .4) = .288$;
- $w_3 = \langle \{\mathtt{a}, \mathtt{e}, \mathtt{b}\}, \{(\mathtt{a}, \mathtt{b}), (\mathtt{b}, \mathtt{e})\}, \{(\mathtt{e}, \mathtt{a})\}\rangle$ with probability $I(w_3) = .6 \times (1 - .8) = .12$;
- $w_4 = \langle \{\mathtt{a}, \mathtt{e}, \mathtt{c}\}, \emptyset, \{(\mathtt{e}, \mathtt{a})\}\rangle$ with probability $I(w_4) = (1\text{-}.6) \times (.8) = .32$;
- $w_5 = \langle \{\mathtt{a}, \mathtt{e}\}, \emptyset, \{(\mathtt{e}, \mathtt{a})\}\rangle$ with $I(w_4) = (1\text{-}.6) \times (1\text{-}.8) = .08$. □

A relevant problem in the field of formal argumentation is that of *acceptance*, that is determine the probability that a goal argument g of a framework Δ is accepted under a given semantics σ. We next extend the acceptance problem defined for non-conditional PAF in (Alfano et al. 2023a) to CPBAF.

Definition 5 (Probabilistic Acceptance)**.** *Given a CPBAF $\Delta = \langle A, R, S, P\rangle$ and an argument $g \in A$, $\mathsf{PrA}[\sigma]$ is the problem of determining the probability $PrA^{\sigma}_{\Delta}(g)$ that g is acceptable w.r.t. semantics σ, defined as follows:*

$$PrA^{\sigma}_{\Delta}(g) = \sum_{w \in pw(\Delta) \ \wedge \ E \in \sigma(w) \ \wedge \ g \in E} I(w) \cdot Pr(E, w, \sigma)$$

where $Pr(\cdot, w, \sigma)$ is a PDF over the set $\sigma(w)$.

The previous definition generalizes the notion of Credulous Acceptance (CA) for deterministic AFs proposed in (Thimm 2012) (where a PDF over the set of σ-extensions is assumed to be given), as well as the notion of probabilistic CA for PBAF (Fazzinga et al. 2019), and probabilistic acceptance in PAF, where it has also been shown that $\mathsf{PrA}[\sigma]$ encompasses some drawback of the classical CA problem.

3.1 CBAF Mappings

W.l.o.g., we can restrict our attention to arg-CPBAFs, a subclass of CPBAFs in which the probabilities assigned to attacks and supports are equal to 1. Indeed, for any CPBAF, there exists an equivalent arg-CPBAF.

Definition 6. *Given a CPBAF $\Delta = \langle A, R, S, P\rangle$ we denote by $\Delta' = arg(\Delta) = \langle A', R', S', P\rangle$ the arg-CPBAF derived from Δ as follows:*

- $A' = A \cup \{\alpha_{ab} \mid (a,b) \in R\} \cup \{\beta_{ab} \mid (a,b) \in S\}$;
- $R' = \{(\alpha_{ab}, b) \mid (a,b) \in R\}$;
- $S' = \{(a, \alpha_{ab}) | (a,b) \in R\} \cup \{(a, \beta_{ab}), (\beta_{ab}, b) | (a,b) \in S\}$.

Thus, any attack (resp., support) (a,b) is replaced by a support (a, α_{ab}) and an attack (α_{ab}, b) (resp., two supports (a, β_{ab}) and (β_{ab}, b)), where the fresh argument α_{ab} (resp β_{ab} has associated the same probability of (a,b). It is worth noting that, to obtain arg-CPBAFs, certain elements (e.g., attacks (a,b) with $P(\alpha_{ab}|C_{ab}) = 1$ and $C_{ab} \equiv true$) can be ignored by the rewriting operation. Analogous results mapping CPBAF into CPBAF where only attacks/supports are uncertain can be similarly provided.

Example 10. Consider the CPBAF $\Delta = \langle \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}, \{(\mathtt{b}, \mathtt{c})\}, \{(\mathtt{a}, \mathtt{b})\}, P = \{P(\beta_{\mathtt{ab}}) = 0.4, P(\alpha_{\mathtt{bc}} \mid \beta_{\mathtt{ab}}) = 0.6\}\rangle$. The derived (equivalent) arg-CPBAF is $\Delta' = \langle \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \beta_{\mathtt{ab}}, \alpha_{\mathtt{bc}}\}, \{(\alpha_{\mathtt{bc}}, \mathtt{c})\}, \{(\mathtt{a}, \beta_{\mathtt{ab}}), (\beta_{\mathtt{ab}}, \mathtt{b}), (\mathtt{b}, \alpha_{\mathtt{bc}})\}, P\rangle$. □

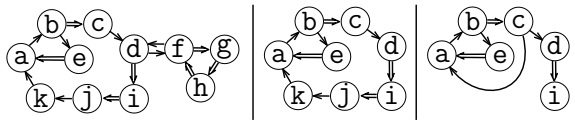

Figure 3: BAFs Λ, $\Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$, and $\widehat{\Lambda_{\mathtt{j}}}$ of Examples 11 and 12.

For any CPBAF $\Delta = \langle A, R, S, P\rangle$, we denote by $\Delta'' = \langle A'', R'', S'', P\rangle$ the CPBAF derived from $\Delta' = arg(\Delta) = \langle A', R', S', P\rangle$ where: A''=$A' \cap A$; R''=$(R' \cap R) \cup \{(a,b) \mid \alpha_{ab} \in A'\}$; and S''=$(S' \cap S) \cup \{(a,b) \mid \beta_{ab} \in A'\}$.

Fact 1. *For any CPBAF Δ, $\Delta'' = arg^{-1}(arg(\Delta)) = \Delta$.*

Thus, the transformation to derive Δ'' from Δ' defines an inverse mapping w.r.t. that used to derive Δ' from Δ.

The next theorem states the 'equivalence' between CPBAF and the derived arg-CPBAF.

Theorem 1. *Let Δ be a CPBAF and $\Delta' = arg(\Delta)$. Then:*
- *there is a bijective mapping m from $pw(\Delta)$ to $pw(\Delta')$;*
- *for each $w \in pw(\Delta)$, $I(w) = I(\mathsf{m}(w))$; and*
- *$\sigma(w) = \{E \cap A \mid E \in \sigma(\mathsf{m}(w))\}$.*

Thus, hereafter we assume that in CPBAFs only arguments may be uncertain, and simply call them CPBAFs.

4 Explanations for Bipolar AF

We introduce the concept of *explanation*, that is a sequence of necessary suggestions (guided by an ordering on SCCs) useful to construct a given extension. The explanation of an extension E is determined recursively for: *i*) a BAF stripped of all arguments determined to be true or false by the grounded extension, and *ii*) an (initially empty) set $\mathcal{F}$ of arguments, whose status is not yet determined but, based on the assumptions made, will have to be determined to be defeated. We next introduce some notations.

Given a BAF $\Lambda = \langle A, R, S\rangle$, a set of arguments $\mathcal{F} \subseteq A$ whose status is to be confirmed as defeated, and an argument $a \in A$ (that will be assumed to be accepted) whose set of (direct or transitive) attackers is $\vec{a}$, we denote by:

- Λ_a the BAF obtained from Λ by deleting attacks and supports targeting a, that is, $\langle A, R\backslash(A\times\{a\}), S\backslash(A\times\{a\})\rangle$;
- $\widehat{\Lambda_a}$ the BAF obtained from $\Lambda_{\uparrow \mathtt{gr}(\Lambda_a)^*}$ by adding the following set of (virtual) attacks:
  $$\vec{a} \times \big(\{b \mid (x,b) \in R \ \wedge \ x \in \mathtt{gr}(\Lambda_a)^+\}\backslash \mathtt{gr}(\Lambda_a)^*\big).$$
  As it will be clarified in what follows, the intuitive meaning of virtual attacks is not to determine the status of attacked arguments (since attackers are defeated), but to preserve the (SCC) topology.

Example 11. Consider the BAF Λ shown in Figure 3(left), extending that of Example 4, and whose grounded extension is $\mathtt{gr}(\Lambda) = \emptyset$. Since $\mathtt{f}, \mathtt{g}$ and $\mathtt{h}$ are derived as defeated (as occurring in a support cycle), $\mathtt{gr}(\Lambda)^* = \mathtt{gr}(\Lambda) \cup \{\mathtt{f}, \mathtt{g}, \mathtt{h}\}$. Taking the BAF $\Lambda = \Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$ reported in Figure 3(center), we have that $\Lambda_{\mathtt{j}}$ is obtained by removing the support $(\mathtt{i}, \mathtt{j})$, and consequently $\mathtt{gr}(\Lambda_{\mathtt{j}}) = \{\mathtt{j}\}$, and $\mathtt{gr}(\Lambda_{\mathtt{j}})^* = \{\mathtt{j}, \mathtt{k}\}$. As $\vec{\mathtt{j}} = \{\mathtt{c}\}$, $\widehat{\Lambda_{\mathtt{j}}}$ is reported on Figure 3(right). □

The next definition recursively identifies when a sequence of arguments is an explanation for an extension E and it can be used to determine an explanation X for E. The computation of X can be carried out considering one SCC at a time following the topological order of the argumentation graph, and for each SCC by alternating 'deterministic computations' (using the grounded semantics) and the choice of an argument to be accepted (or alternatively all arguments in the SCC are undefined). In the first case i) it is possible to determine the status of a further set U of arguments, ii) state that all attackers of the chosen argument must be determined subsequently as defeated (and added to $\mathcal{F}$), and iii) add 'virtual attacks' from the attackers of the arguments attacked by U. If $\mathcal{F}$ is not empty, the choice of the argument is restricted to arguments making some argument in $\mathcal{F}$ defeated.

Definition 7 (Explanation)**.** *Let $\Lambda = \langle A, R, S\rangle$ be a BAF, $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$ a semantics, $E \in \sigma(\Lambda)$, and $\mathcal{F} \subset A$ a set of assumptions about the false status of some arguments. A sequence $X = \langle a_1, \ldots, a_n\rangle$, where $a_i \in E \cup \{\varepsilon\}$ (with $i \in [1...n]$ and ε a fresh symbol not used for arguments names), is an explanation for E (w.r.t. Λ and $\mathcal{F}$) if, letting $\Lambda = \Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$ and $\mathcal{C}$ be the first SCC of Λ, one of the following three conditions hold:*

1. *(Final Step) $X = \langle\rangle$, $\mathcal{F} = \emptyset$, and $\Lambda = \langle\emptyset, \emptyset, \emptyset\rangle$.*
2. *(Choice Step) (i) a_1 occurs in an even cycle of $\mathcal{C}$;*
   (ii) $\big(\exists x \in \mathcal{C}\,.(a_1, x) \in R\big)$;
   (iii) $\big(\mathcal{F} \neq \emptyset\big)$ implies $\big(\mathtt{gr}(\Lambda_{a_1})^+ \cap \mathcal{F} \neq \emptyset\big)$;
   (iv) $\langle a_2, ..., a_n\rangle$ is an explanation for $E' = E \setminus \mathtt{gr}(\Lambda_{a_1})$ w.r.t. $\widehat{\Lambda_{a_1}}$, and $(\mathcal{F} \cup \vec{a}_1) \setminus \mathtt{gr}(\Lambda_{a_1})^+$.
3. *(Skip Step) $a_1 = \varepsilon$, $\mathcal{F} = \emptyset$ and $\langle a_2, \ldots, a_n\rangle$ is an explanation for E w.r.t. $\Lambda_{\uparrow \mathcal{C}}$ and $\mathcal{F}$.*

At each choice step, an argument belonging to the extension (say it a_1) is selected. Intuitively, $\mathcal{F}$ denotes arguments whose status has not yet been determined but, based on previous assumptions, will have to be derived as defeated. After assuming that a_1 is accepted, all arguments in $\vec{a}_1$ that transitively attack a_1 should be derived as defeated and, thus, added to $\mathcal{F}$ (if their status has not been determined during the computation of $\mathtt{gr}(\Lambda_{a_1})$). In assuming a_1 as accepted, $\mathtt{gr}(\Lambda_{a_1})$ is the set of arguments that are derived to be accepted, whereas arguments $\mathtt{gr}(\Lambda_{a_1})^+$ are derived as defeated. Then, the BAF $\Lambda_{\uparrow \mathtt{gr}(\Lambda_{a_1})^*}$ is obtained by removing the arguments (in $\mathtt{gr}(\Lambda_{a_1})^*$) whose status has been determined (either as true of false). Finally, virtual attacks are added to $\Lambda_{\uparrow \mathtt{gr}(\Lambda_{a_1})^*}$, obtaining the BAF $\widehat{\Lambda_{a_1}}$. The meaning of virtual attacks in $\widehat{\Lambda_{a_1}}$ is to ensure that if a_1 belonged to an SCC $\mathcal{C}$, after eliminating the arguments in $\mathtt{gr}(\Lambda_{a_1})^*$ and the related attacks and supports, arguments in $\vec{a}_1$ belonging to $\mathcal{C}$ continue to belong to $\mathcal{C}$. It is worth noting that, as arguments in $\vec{a}_1$ will be determined as defeated, the added attacks do not contribute to determining the status of other arguments.

Example 12. Continuing from Example 11, $\langle \mathtt{j}, \mathtt{a}\rangle$ is an explanation for the complete extension $E = \{\mathtt{a}, \mathtt{d}, \mathtt{e}, \mathtt{i}, \mathtt{j}\}$, obtained as follows. The initial BAF is $\Lambda^{(1)} = \Lambda = \Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$ reported on Figure 3 (center), whereas the initial extension is denoted by $E^{(1)} = E$, and $\mathcal{F}^{(1)} = \emptyset$ — superscripts de-

notes steps. Assume in the first (choice) step that j is selected in the first SCC $\mathcal{C} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{i}, \mathtt{j}, \mathtt{k}\}$ of $\Lambda^{(1)}$. Then, i) $\mathtt{gr}(\Lambda_{\mathtt{j}}) = \{\mathtt{j}\}$, ii) $\Lambda^{(2)} = \widehat{\Lambda_{\mathtt{j}}}$ is obtained (see Figure 3 (right)); iii) $E^{(2)} = E^{(2)} \setminus \{\mathtt{j}\}$; and iv) c (the only attacker of j whose status has not yet been determined) is added to $\mathcal{F}^{(1)}$, obtaining $\mathcal{F}^{(2)}$. In the second (and choice) step, recalling that $\mathcal{F}^{(2)} = \{\mathtt{c}\}$ intuitively means that c must be 'confirmed' as defeated in $\Lambda^{(2)}$, the only choice in the first SCC $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{e}\}$ of $\Lambda^{(2)}$ making c defeated is a. Then, i) $\mathtt{gr}(\Lambda^{(2)}_{\mathtt{a}}) = \{\mathtt{a}, \mathtt{d}, \mathtt{e}, \mathtt{i}\}$, ii) $\Lambda^{(3)} = \widehat{\Lambda^{(2)}_{\mathtt{a}}} = \langle \emptyset, \emptyset, \emptyset \rangle$ is obtained; iii) $E^{(3)} = \mathcal{F}^{(3)} = \emptyset$. In the third and final step, the empty explanation $\langle\rangle$ for $E^{(3)}$ w.r.t. $\Lambda^{(3)}$ and $\mathcal{F}^{(3)}$ is given. □

Observe that: (i) when the computation of an SCC terminates, i.e. the status of all arguments in it has been defined, then $\mathcal{F}$ becomes empty, (ii) under stable semantics Item 3 of Definition 7 is never applied, and (iii) under preferred semantics Item 3 is applied only if Item 2 cannot be applied.

4.1 Properties

Definition 7 gives rise to an *explanation strategy* (Ulbricht and Wallner 2021). We use ξ to denote an explanation strategy. Moreover, $\xi^{\sigma}_{\Lambda}(E)$ denotes the set of ξ-explanations for $E \in \sigma(\Lambda)$, under semantics σ, whereas $\xi^{\sigma}(\Lambda) = \cup_{E \in \sigma(\Lambda)} \xi^{\sigma}_{\Lambda}(E)$ is the set of ξ-explanations for Λ under σ.

Most of the explanation strategies defined in the literature return sets. In contrast, our explanation strategy gives a sequence of arguments choices. Before formalizing a set of fundamental properties for explanation strategies introduced in the literature (Ulbricht and Wallner 2021; Borg and Bex 2024), we define the concept of *explanation-set* that allows us to uniformly state properties for explanations consisting of sets or sequences, thus enabling formal comparison of different types of explanations strategies. For any BAF $\Lambda = \langle A, R, S \rangle$ and explanation $X = \langle x_1, ..., x_n \rangle \in \xi^{\sigma}_{\Lambda}(E)$ (with $E \in \sigma(\Lambda)$), we say that $\tilde{\mathrm{X}} = set(X) \cap A$ is an *explanation-set* for E, where $set(X) = \{x_1, ..., x_n\}$. For explanation strategies where X is a set, $\tilde{\mathrm{X}} = X \cap A$.

Definition 8. *Let Λ be a BAF and $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$ a semantics. Then, we say that an explanation strategy ξ satisfies a property τ, if for every σ-extension $E \in \sigma(\Lambda)$ and for every explanation $X \in \xi^{\sigma}_{\Lambda}(E)$, it holds that:*
τ = σ-basic*, if* $E = \mathtt{gr}(\langle A, R \setminus (A \times \tilde{X}), S \setminus (A \times \tilde{X}) \rangle)$*;*
τ = σ-existence*, if* $\xi^{\sigma}_{\Lambda}(E) \neq \emptyset$*;*
τ = Conflict-freeness*, if $\tilde{X}$ is conflict-free in Λ;*
τ = Relevance*, if* $\tilde{X} \subseteq \cup_{a \in E} \{b \in A \mid b$ *reaches* a *in* $\Lambda\}$*;*
τ = Disjointness*, if* $F \in \sigma(\Lambda) \setminus \{E\}$ *implies* $\xi^{\sigma}_{\Lambda}(E) \cap \xi^{\sigma}_{\Lambda}(F) = \emptyset$*;*
τ = Subset-Minimality*, if* $\nexists\, Y \in \xi^{\sigma}_{\Lambda}(E)$ *with* $\tilde{Y} \subset \tilde{X}$*.*

The σ-basic property, adapted from (Ulbricht and Wallner 2021), intuitively captures the fact that explanation-sets must be consistent sets of arguments that, when assumed to be true (i.e., they receive no attack/support), should allow to derive all arguments in E. An explanation strategy satisfies *σ-existence* if, every extension admits at least one explanation. *Conflict-freeness* requires explanation sets to be conflict-free. *Relevance* ensures that each argument in the extension is reachable via a path in Λ from at least one argument in the explanation set; this is meaningful for semantics σ satisfying the directionality principle, as is the case in this work. *Disjointness* prescribes that distinct extensions must correspond to disjoint explanation sets. *Subset-Minimality* demands subset-minimality of explanation sets. Note that all these properties implicitly assume explanations being sets of arguments, and therefore do not fully align with our notion of explanation as a sequence of arguments. Therefore, we also consider *Sequence-Minimality*, that is, let $X = \langle x_1, ..., x_n \rangle \in \xi^{\sigma}_{\Lambda}(E)$, there is no $\langle x_1, ..., x_k \rangle \in \xi^{\sigma}_{\Lambda}(E)$ with $k < n$. As also stated next, an explanation X for an extension E is a minimal sequence of arguments that, if assumed to be true, allows us to derive all arguments in E. Finally, it is worth considering *Inclusion*: for $E, F \in \sigma(\Lambda)$, if $E \subset F$, then an explanation $X \in \xi^{\sigma}_{\Lambda}(E)$ is a prefix of at least one explanation $Y \in \xi^{\sigma}_{\Lambda}(F)$. Hence, *Inclusion* states that an extension contained in other extensions gives explanations that are prefixes of explanations for the larger extensions. Clearly, this property makes sense for complete extensions only, as for the other semantics σ there cannot be two extensions $E, F \in \sigma(\Lambda)$ such that $E \subset F$.

Theorem 2. *The explanation strategy of Definition 7 satisfies σ-basic, σ-existence, Conflict-freeness, Relevance, Disjointness, Sequence-Minimality, and Inclusion.*

4.2 Inducing Probabilities on Extensions

Since a given extension may have multiple explanations of different length, it is reasonable to assume that some explanations are preferred to others. We now introduce probabilities for explanations. As said before, the grounded semantics has a unique empty explanation which has probability 1. To define probabilities of explanations, we exploit the concepts of *trie* (standard prefix tree data structure) and its probabilistic version (Alfano et al. 2023a).

Definition 9. *Given a BAF $\Lambda = \langle A, R, S \rangle$ and a semantics σ, the probabilistic trie for Λ under semantics σ is the triple $\mathcal{Q}^{\sigma}_{\Lambda} = \langle N, H, \pi \rangle$ of nodes N and edges H where $\langle N, H \rangle$ is the trie of all sequences in $\xi^{\sigma}(\Lambda)$,* [4] *$\pi : N \to (0, 1]$ is the function inductively defined as: $\pi(\langle\rangle) = 1$ and $\pi(x) = \pi(\mathrm{PAR}(x))/|\mathrm{CHI}(\mathrm{PAR}(x))|$ where $\mathrm{PAR}(x)$ denotes the parent of x, whereas $|\mathrm{CHI}(x)|$ denotes the number of children of x.*

Since the set of leaves of the probabilistic trie $\mathcal{Q}^{\sigma}_{\Lambda} = \langle N, H, \pi \rangle$ coincides with $\xi^{\sigma}(\Lambda)$ (i.e. $leaves(\mathcal{Q}^{\sigma}_{\Lambda}) = \xi^{\sigma}(\Lambda)$) hereafter, with a little abuse of notation, we assume that π is a function from $\xi^{\sigma}(\Lambda)$ to $(0, 1]$. By definition, we have that $\sum_{X \in \xi^{\sigma}(\Lambda)} \pi(X) = 1$.

Definition 10. *For any BAF Λ and σ-extension $E \in \sigma(\Lambda)$, $PrE(E, \Lambda, \sigma) = \sum_{X \in \xi^{\sigma}_{\Lambda}(E)} \pi(X)$ denotes the* Explanation-based Probability *associated with E.*

[4] The trie $\langle N, H \rangle$ of all sequences in $\xi^{\sigma}(\Lambda)$ is built in the standard way, by considering an explanation as a sequence of characters, each representing an argument name or symbol ϵ. Thus, the root node represents an empty string, that is, an empty explanation; each node in N represents the addition of a single character from the given sequences, that is, an argument name or symbol ϵ from an explanation in $\xi^{\sigma}(\Lambda)$; edges in H connect nodes that are labeled with argument names or ϵ; and a path from the root to any node represents a prefix of one or more explanations in $\xi^{\sigma}(\Lambda)$.

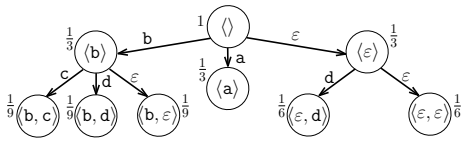

Figure 4: Probabilistic trie for the BAF Λ of Example 13.

Example 13. The explanations for the complete extensions $\mathtt{co}(\Lambda) = \{E_1 = \emptyset, E_2 = \{\mathtt{b}\}, E_3 = \{\mathtt{d}\}, E_4 = \{\mathtt{b}, \mathtt{c}\}, E_5 = \{\mathtt{b}, \mathtt{d}\}, E_6 = \{\mathtt{a}, \mathtt{d}, \mathtt{e}\}\}$ of the BAF Λ of Example 4 are represented by the leaf nodes of the trie shown in Figure 4. The probability of explanation $\langle \mathtt{a} \rangle$ for extension E_6 is $1/3$, that is also the value for $PrE(E_6, \Lambda, \sigma)$. □

As for probabilistic CPBAF Δ, we have to consider all BAFs w in the set $pw(\Delta)$ of possible worlds. This leads to the *Explanation-based Probabilistic Acceptance* problem, that is a specialization of $\mathsf{PrA}[\sigma]$ (cf. Definition 5).

Definition 11. *Given a CPBAF $\Delta = \langle A, R, S, P \rangle$ and an argument $g \in A$, $\mathsf{PrEA}[\sigma]$ is the* Explanation-based Probabilistic Acceptance *problem, that is, the problem of determine the probability $PrEA^{\sigma}_{\Delta}(g)$ that g is acceptable w.r.t. semantics σ, computed as follows:*

$$PrEA^{\sigma}_{\Delta}(g) = \sum_{w \in pw(\Delta) \ \wedge \ E \in \sigma(w) \ \wedge \ g \in E} I(w) \cdot PrE(E, w, \sigma).$$

The intuition behind the previous definition is that, as extensions do not share explanations, the probability of explanations is transferred to the extensions and acceptance of arguments. With a little effort, it can be checked that, for the CPBAF of Example 8, $PrEA^{\mathtt{co}}_{\Delta}(g) = 1/3$ (resp., $1/3$, $1/9$, $4/9$, and $1/3$), with $g = \mathtt{a}$ (resp., $\mathtt{b}$, $\mathtt{c}$, $\mathtt{d}$, and $\mathtt{e}$).

5 Complexity and Approximation

We address the complexity of problems $\mathsf{PrA}[\sigma]$ and $\mathsf{PrEA}[\sigma]$. Recall that $\mathsf{PrA}[\sigma]$ is defined after choosing an arbitrary but fixed PDF over the set of extensions of a BAF, while $\mathsf{PrEA}[\sigma]$ uses the specific PDF $Pr(E, \Lambda, \sigma)$ of Definition 11. As shown next, these problems are intractable.

Theorem 3. *For $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$, $\mathsf{PrA}[\sigma]$ is FP$^{\#P}$-hard, even for acyclic CPBAFs and for any chosen PDF.*

The high computational complexity of $\mathsf{PrA}[\sigma]$ and $\mathsf{PrEA}[\sigma]$, even for the very simple setting of acyclic CPBAF, suggests that one would need to focus on finding efficient algorithms that solve the problem approximately. Next, we present a complete picture of the approximability landscape of our problems, under different semantics and *approximation schemes*, where the definition of the latter is as follows.

Definition 12 (Approximation schemes)**.** Consider a function $f : \{0,1\}^* \to \mathbb{Q}$. A *fully polynomial-time randomized approximation scheme* (FPRAS) for f is a randomized algorithm A that given as input $x \in \{0,1\}^*$, and numbers $\epsilon > 0$, $0 < \delta < 1$, outputs a random variable $A(x, \epsilon, \delta)$ such that:

$$\mathsf{Pr}\left(|A(x, \epsilon, \delta) - f(x)| \leq \epsilon \cdot f(x)\right) \geq 1 - \delta, \qquad (3)$$

and A runs in polynomial time in $|x|$, $1/\epsilon$, and $\ln(1/\delta)$.

Algorithm 1: Apx

Input: A CPBAF Δ, a semantics σ, a goal $g \in A$, error parameter $\epsilon > 0$, and uncertainty parameter $0 < \delta < 1$.
Output: a random number p s.t. $PrEA^{\sigma}_{\Delta}(g) \in [p-\epsilon, p+\epsilon]$ with probability $1 - \delta$.

1: $n = \lceil \frac{1}{2\epsilon^2} \times \ln(\frac{2}{\delta}) \rceil$; $\kappa = 0$;
2: **for** $i \in \{1, \ldots, n\}$ **do**
3: Choose $\Lambda = \langle A, R, S \rangle$ in $pw(\Delta)$ with probability $\mathcal{I}(\Lambda)$;
4: Choose $X \in \xi^{\sigma}(\Lambda)$ with probability $\pi(X)$;
5: **if** $g \in \mathtt{gr}(\langle A, R \setminus (A \times \tilde{\mathrm{X}}), S \setminus (A \times \tilde{\mathrm{X}}) \rangle)$ **then** $\kappa = \kappa + 1$;
6: **return** κ / n;

A *fully polynomial-time additive randomized approximation scheme* (FPARAS) for a function f is defined as in Equation 3, where the inequality $|A(x, \epsilon, \delta) - f(x)| \leq \epsilon \cdot f(x)$ is replaced with $|A(x, \epsilon, \delta) - f(x)| \leq \epsilon$.

The type of error guarantee distinguishes the two schemes: an *additive* error bounded by ϵ in FPARAS, whereas a *relative* error within a factor of $\epsilon \cdot f(x)$ in FPRAS.

As stated next, approximate computation of $\mathsf{PrA}[\sigma]$ via FPRASes is not possible even for acyclic CPBAFs.

Theorem 4. *Unless NP $\subseteq$ BPP, there is no FPRAS for $\mathsf{PrA}[\sigma]$ with $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$, even for acyclic CPBAFs.*

Next, we also show that even approximation algorithms with bounded additive error cannot be devised, for CPBAFs of general shape and preferred and stable semantics.

Theorem 5. *Unless NP $\subseteq$ BPP, there is no FPARAS for $\mathsf{PrA}[\mathtt{st}]$ and $\mathsf{PrA}[\mathtt{pr}]$, for any chosen PDF.*

Our results highlight an intrinsic difficulty in providing efficient procedures (either exact or approximate) for *any* approach assigning a probability to an argument by means of a probability distribution over the extensions. Thus, our efforts should be towards approximation schemes with bounded additive error guarantees, i.e., FPARASes. In particular, in the light of Theorem 5, one could still provide an FPARAS either when $\sigma = \mathtt{gr}$, or when some restriction on the input CPBAF is assumed. The following theorem identifies cases where using explanations for devising a PDF over extensions allows us to construct an FPARAS.

Theorem 6. *$\mathsf{PrEA}[\sigma]$ has an FPARAS if i) $\sigma = \mathtt{gr}$, or ii) $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$ and the input CPBAF is odd-cycle-free.*

Our FPARAS algorithm is presented in Algorithm 1. Consider a CPBAF Δ, a semantics σ and an argument g. The high-level idea is to perform a number of iterations n, and at each iteration sample a world w of Δ and an explanation X in $\xi^{\sigma}(w)$, and count the fraction of iterations for which the given argument g is in the σ-extension E explained by X.

We point out that, besides line 4, all steps of our algorithm can be easily implemented in polynomial time regardless of the shape of the input CPBAF and the semantics. Particularly, to prove that Algorithm 1 leads to an FPARAS in the cases described above, it suffices to prove that i) line 4 can be implemented in polynomial time when either $\sigma = \mathtt{gr}$ or

Algorithm 2

Input: A BAF $\Lambda = \langle A, R, S\rangle$ and a semantics σ.
Output: An explanation for a σ-extension.

1: Let $X = \langle\rangle$; $\mathcal{F} = \mathcal{U} = \emptyset$; $\Lambda = \Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$;
2: **while** $\Lambda \neq \langle\emptyset, \emptyset, \emptyset\rangle$ **do**
3: Let A' be the first SCC of Λ;
4: Let $\mathcal{C} = \{a \in A' \mid a \notin (\mathcal{F} \cup \mathcal{U}) \wedge \big(\exists x \in A'.(a, x) \in R\big) \wedge \big((\mathcal{F} \neq \emptyset) \text{ implies } (\mathtt{gr}(\Lambda_a)^+ \cap \mathcal{F} \neq \emptyset)\big)\}$;
5: **if** $\sigma = \mathtt{gr}$ **then** $\mathcal{C} = \{\varepsilon\}$;
6: **if** $\sigma = \mathtt{co} \wedge \mathcal{F} = \emptyset$ **then** $\mathcal{C} = \mathcal{C} \cup \{\varepsilon\}$;
7: Select $a \in \mathcal{C}$ with probability $\frac{1}{|\mathcal{C}|}$ and append it to X;
8: **if** $a = \varepsilon$ **then**
9: $\mathcal{U} = \mathcal{U} \cup \{x | (y, x) \in ((R \cup S) \cap (A' \times (A \setminus A'))\}$; $\Lambda = \Lambda_{\uparrow A'}$;
10: **else** $\mathcal{F} = \mathcal{F}_a$; and $\Lambda = \widehat{\Lambda_a}$;
11: **return** X

Δ is odd-cycle-free. This is done via Algorithm 2, which operates as follows. At the beginning Λ is updated to $\Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$ by deleting arguments occurring in the grounded extension or defeated by them (line 1). Clearly, if the source BAF is acyclic, we have that $\Lambda_{\uparrow \mathtt{gr}(\Lambda)^*}$ is empty and the algorithm terminates, returning the empty explanation (with probability 1). Then, it iterates until the current BAF Λ becomes empty (line 2) and the following steps are executed iteratively. It determines the set $\mathcal{C}$ of arguments in the first SCC, from which it is possible to choose the next element (lines 3 and 4). Moreover, if $\sigma = \mathtt{gr}$, the only possible choice will be ε (line 5), whereas if $\sigma = \mathtt{co}$ and $\mathcal{F}$ is empty, ε is added to $\mathcal{C}$, as it is also an admissible choice (line 6). Then, an element a is nondeterministically selected from $\mathcal{C}$ with probability $1/|\mathcal{C}|$ (line 7) and appended to X. The next steps depend on the choice made. If $a = \varepsilon$ (line 9) all arguments in the first component are deleted from Λ and all arguments attacked or supported by these arguments are added to $\mathcal{U}$ to remember that their status cannot be true and, thus, they cannot be chosen in the next steps. By deleting the whole components and adding ε to X we are stating that the status of all elements in the component is undefined. If the chosen element a is an argument, the following steps are executed: i) the attackers of a are added to $\mathcal{F}$ as we are assuming that their status must be false, and all elements whose status is determined (i.e. those in $\mathtt{gr}(\Lambda_a)^+$) are deleted from $\mathcal{F}$ and ii) the BAF Λ is updated to $\widehat{\Lambda_a}$ (line 10), so that the resulting component continues to be strongly connected by reconstructing paths in the graph that were broken through the deletion of nodes.

Theorem 7. *Whenever i) $\sigma = \mathtt{gr}$, or ii) $\sigma \in \{\mathtt{gr}, \mathtt{co}, \mathtt{st}, \mathtt{pr}\}$ and the input BAF Λ is odd-cycle-free, then: Algorithm 2 runs in polynomial time and, for each $E \in \sigma(\Lambda)$, it outputs E with probability $Pr(E, \Lambda, \sigma)$.*

Thus, Algorithm 1 enjoys the probabilistic and error guarantees of an FPARAS (Hoeffding 1963).

6 Related Work

Looking for transparent and interpretable models has led to the exploration of several explanation paradigms in eXplainable AI (XAI) (Marques-Silva and Ignatiev 2022; Ignatiev et al. 2022; Malfa et al. 2021; Ignatiev, Narodytska, and Marques-Silva 2019; Alfano et al. 2025a,b), also in the probabilistic setting (Izza et al. 2021; Subercaseaux, Arenas, and Meel 2025; Arenas et al. 2025; Wäldchen et al. 2021).

Integrating explanations in argumentation systems is important for enhancing the argumentation and persuasion capabilities of software agents (Moulin et al. 2002; Bex and Walton 2016; Cyras et al. 2019; Miller 2019). For these reasons, several researchers explored how to deal with explanations in formal argumentation. Significant work in this field includes (Fan and Toni 2015), where a new argumentation semantics is proposed for capturing explanations in AF, and (Craven and Toni 2016) that focuses on ABA framework (Dung, Kowalski, and Toni 2009). They treat an explanation as a semantics to answer why an argument is accepted or not. Thus, an explanation is viewed as a set of arguments, instead of a sequence of arguments, needed for explaining such an extension. In (Fan and Toni 2015), an explanation is a set of arguments justifying a given argument by means of a proponent-opponent dispute-tree (Dung, Mancarella, and Toni 2007). A similar approach based on debate trees as proof procedure for computing grounded, ideal, and preferred semantics is given in (Thang, Dung, and Hung 2009).

Probabilities in argumentation have been widely explored (Dung and Thang 2010; Rienstra 2012; Doder and Woltran 2014; Hunter 2012; Li, Oren, and Norman 2011; Thang, Dung, and Hung 2009; Hunter et al. 2021). Almost all frameworks so far proposed have considered marginal probabilities with either standard AF (Fazzinga, Flesca, and Furfaro 2018a, 2022) or support-acyclic BAF (Fazzinga et al. 2019). A formalization of conditional probabilistic argumentation based on probabilistic conditional logic has been provided in (Hunter and Potyka 2023). A language defining constraints by means of conditional probabilities is used in epistemic graphs (De Bona, Rocha, and Cozman 2021). However, none of those works specifically deals with explanations, which is the focus of our work.

7 Conclusions

We have introduced CPBAF, combining cyclic BAF with conditional probabilities. Then, we have introduced a notion of explanation in BAF yielding a probability distribution function over BAF's extensions. This leads to an instantiation of the acceptance problem in CPBAF ($\mathsf{PrA}[\sigma]$) called explainable acceptance ($\mathsf{PrEA}[\sigma]$). After showing that $\mathsf{PrA}[\sigma]$ and $\mathsf{PrEA}[\sigma]$ are $\mathrm{FP}^{\#\mathrm{P}}$-hard even for acyclic CPBAFs, we have proposed a polynomial time additive approximation algorithm for solving $\mathsf{PrEA}[\sigma]$ for general CPBAFs and either $\sigma = \mathtt{gr}$ or odd-cycle-free CPBAFs under $\sigma \in \{\mathtt{co}, \mathtt{st}, \mathtt{pr}\}$.

Future work will be devoted to the investigation of other ways of defining a PDF over the set of extensions, thus enabling other instantiations of $\mathsf{PrA}[\sigma]$.

Acknowledgments

We acknowledge financial support from PNRR MUR projects FAIR (PE0000013) and SERICS (PE00000014), project Tech4You (ECS0000009), and MUR project PRIN 2022 EPICA (H53D23003660006).

References

Alfano, G.; Calautti, M.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2020. Explainable Acceptance in Probabilistic Abstract Argumentation: Complexity and Approximation. In *Proc. of International Conference on Principles of Knowledge Representation and Reasoning*, 33–43.

Alfano, G.; Calautti, M.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2023a. Explainable acceptance in probabilistic and incomplete abstract argumentation frameworks. *Artif. Intell.*, 323: 103967.

Alfano, G.; Cohen, A.; Gottifredi, S.; Greco, S.; Parisi, F.; and Simari, G. R. 2024a. Credulous acceptance in high-order argumentation frameworks with necessities: An incremental approach. *Artif. Intell.*, 333: 104159.

Alfano, G.; Gould, A.; Leofante, F.; Rago, A.; and Toni, F. 2025a. Counterfactual Explanations Under Model Multiplicity and Their Use in Computational Argumentation. In *Proc. of International Joint Conference on Artificial Intelligence*, 4321–4329.

Alfano, G.; Greco, S.; Mandaglio, D.; Parisi, F.; Shahbazian, R.; and Trubitsyna, I. 2025b. Even-if Explanations: Formal Foundations, Priorities and Complexity. In *Proc. of AAAI*, 15347–15355.

Alfano, G.; Greco, S.; Mandaglio, D.; Parisi, F.; and Trubitsyna, I. 2024b. Abstract argumentation frameworks with strong and weak constraints. *Artif. Intell.*, 336: 104205.

Alfano, G.; Greco, S.; Molinaro, C.; Parisi, F.; and Trubitsyna, I. 2025c. Extending Abstract Argumentation Frameworks with Knowledge Bases. In *Proc. of International Conference on Principles of Knowledge Representation and Reasoning*, 24–35.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2023b. On acceptance conditions in abstract argumentation frameworks. *Inf. Sci.*, 625: 757–779.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2023c. Preferences and Constraints in Abstract Argumentation. In *Proc. of International Joint Conference on Artificial Intelligence*, 3095–3103.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2024c. Counterfactual and Semifactual Explanations in Abstract Argumentation: Formal Foundations, Complexity and Computation. In *Proc. of International Conference on Principles of Knowledge Representation and Reasoning*, 14–26.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2024d. Cyclic Supports in Recursive Bipolar Argumentation Frameworks: Semantics and LP Mapping. *Theory Pract. Log. Program.*, 24(4): 921–941.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2024e. General Epistemic Abstract Argumentation Framework: Semantics and Complexity. In *Proc. of International Joint Conference on Artificial Intelligence*, 3206–3214.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2025d. Constraints and lifting-based (conditional) preferences in abstract argumentation. *Artif. Intell.*, 349: 104437.

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2025e. Featured Argumentation Framework: Semantics and Complexity. In *Proc. of International Joint Conference on Artificial Intelligence*, 4311–4320.

Arenas, M.; Barceló, P.; Kozachinskiy, A.; Romero, M.; and Subercaseaux, B. 2025. On Computing Probabilistic Explanations for Decision Trees. *J. Artif. Intell. Res.*, 83.

Atkinson, K.; Baroni, P.; Giacomin, M.; Hunter, A.; Prakken, H.; Reed, C.; Simari, G. R.; Thimm, M.; and Villata, S. 2017. Towards Artificial Argumentation. *Artificial Intelligence Magazine*, 38(3): 25–36.

Bench-Capon, T.; and Dunne, P. E. 2007. Argumentation in Artificial Intelligence. *Artificial Intelligence*, 171: 619 – 641.

Bex, F.; and Walton, D. 2016. Combining explanation and argumentation in dialogue. *Argument & Computation*, 7(1): 55–68.

Borg, A.; and Bex, F. 2024. Minimality, necessity and sufficiency for argumentation and explanation. *Int. J. Approx. Reason.*, 168: 109143.

Cayrol, C.; Cohen, A.; and Lagasquie-Schiex, M.-C. 2021. Higher-Order Interactions (Bipolar or not) in Abstract Argumentation: A State of the Art. *FLAP*, 8(6): 1339–1436.

Craven, R.; and Toni, F. 2016. Argument graphs and assumption-based argumentation. *Artificial Intelligence*, 233: 1–59.

Cyras, K.; Birch, D.; Guo, Y.; Toni, F.; Dulay, R.; Turvey, S.; Greenberg, D.; and Hapuarachchi, T. 2019. Explanations by arbitrated argumentative dispute. *Expert Systems with Applications*, 127: 141–156.

De Bona, G.; Rocha, V. H. N.; and Cozman, F. 2021. Epistemic Argumentation with Conditional Probabilities and Labeling Constraints. In *Proc. of Int. Symposium on Imprecise Probability*, 100–109. PMLR.

Doder, D.; and Woltran, S. 2014. Probabilistic Argumentation Frameworks-A Logical Approach. In *Proceedings of International Conference on Scalable Uncertainty Management (SUM)*, 134–147.

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: 321–358.

Dung, P. M.; Kowalski, R. A.; and Toni, F. 2009. Assumption-Based Argumentation. In *Argumentation in Artificial Intelligence*, 199–218. Springer.

Dung, P. M.; Mancarella, P.; and Toni, F. 2007. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15): 642–674.

Dung, P. M.; and Thang, P. M. 2010. Towards (Probabilistic) Argumentation for Jury-based Dispute Resolution. In *Proceeding of International Conference on Computational Models of Argument (COMMA)*, 171–182.

Fan, X.; and Toni, F. 2015. On Computing Explanations in Argumentation. In *Proceedings of AAAI Conference on Artificial Intelligence*, 1496–1502.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2018a. Credulous and skeptical acceptability in probabilistic abstract argumentation: complexity results. *Intelligenza Artificiale*, 12(2): 181–191.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2018b. Probabilistic bipolar abstract argumentation frameworks: complexity results. In *Proc. of IJCAI*, 1803–1809.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2019. Complexity of fundamental problems in probabilistic abstract argumentation: Beyond independence. *Artificial Intelligence*, 268: 1–29.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2022. Abstract Argumentation Frameworks with Marginal Probabilities. In Raedt, L. D., ed., *Proc. of IJCAI*, 2613–2619.

Fazzinga, B.; Flesca, S.; Furfaro, F.; and Scala, F. 2019. Efficiently computing extensions' probabilities over probabilistic Bipolar Abstract Argumentation Frameworks. *Intelligenza Artificiale*, 13(2): 189–200.

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.

Fazzinga, B.; Flesca, S.; and Parisi, F. 2016. On efficiently estimating the probability of extensions in abstract argumentation frameworks. *International Journal of Approximate Reasoning*, 69: 106–132.

Hoeffding, W. 1963. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301): 13–30.

Hunter, A. 2012. Some Foundations for Probabilistic Abstract Argumentation. In *Proceeding of International Conference on Computational Models of Argument (COMMA)*, 117–128.

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

Hunter, A.; Polberg, S.; Potyka, N.; Rienstra, T.; and Thimm, M. 2021. Probabilistic argumentation: A survey. *Handbook of Formal Argumentation*, 2: 397–441.

Hunter, A.; and Potyka, N. 2023. Syntactic reasoning with conditional probabilities in deductive argumentation. *Artificial Intelligence*, 321: 103934.

Ignatiev, A.; Izza, Y.; Stuckey, P. J.; and Marques-Silva, J. 2022. Using MaxSAT for Efficient Explanations of Tree Ensembles. In *Proceedings of AAAI Conference on Artificial Intelligence*, 3776–3785.

Ignatiev, A.; Narodytska, N.; and Marques-Silva, J. 2019. Abduction-Based Explanations for Machine Learning Models. In *Proceedings of AAAI Conference on Artificial Intelligence*, 1511–1519.

Izza, Y.; Ignatiev, A.; Narodytska, N.; Cooper, M. C.; and Marques-Silva, J. 2021. Efficient Explanations With Relevant Sets. *CoRR*, abs/2106.00546.

Li, H.; Oren, N.; and Norman, T. J. 2011. Probabilistic Argumentation Frameworks. In *Proceeding of International Workshop on Theorie and Applications of Formal Argumentation (TAFA)*, 1–16.

Malfa, E. L.; Michelmore, R.; Zbrzezny, A. M.; Paoletti, N.; and Kwiatkowska, M. 2021. On Guaranteed Optimal Robust Explanations for NLP Models. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 2658–2665.

Marques-Silva, J.; and Ignatiev, A. 2022. Delivering Trustworthy AI through Formal XAI. In *Proceedings of AAAI Conference on Artificial Intelligence*, 12342–12350.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267: 1–38.

Moulin, B.; Irandoust, H.; Bélanger, M.; and Desbordes, G. 2002. Explanation and Argumentation Capabilities: Towards the Creation of More Persuasive Agents. *Artificial Intelligence Review*, 17(3): 169–222.

Nouioua, F.; and Risch, V. 2011. Argumentation Frameworks with Necessities. In *SUM*, 163–176.

Popescu, A.; and Wallner, J. P. 2024. Advancing Algorithmic Approaches to Probabilistic Argumentation under the Constellation Approach. In *Proc. of KR*, 585–596.

Rienstra, T. 2012. Towards a Probabilistic Dung-style Argumentation System. In *Proceedings of International Conference on Agreement Technologies (AT)*, 138–152.

Simari, G. R.; and Rahwan, I., eds. 2009. *Argumentation in Artificial Intelligence*. Springer.

Subercaseaux, B.; Arenas, M.; and Meel, K. S. 2025. Probabilistic Explanations for Linear Models. In *Proc. of AAAI*, 20655–20662.

Thang, P. M.; Dung, P. M.; and Hung, N. D. 2009. Towards a Common Framework for Dialectical Proof Procedures in Abstract Argumentation. *Journal of Logic and Computation*, 19(6): 1071–1109.

Thimm, M. 2012. A Probabilistic Semantics for abstract Argumentation. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, 750–755.

Ulbricht, M.; and Wallner, J. P. 2021. Strong Explanations in Abstract Argumentation. In *Proc. of AAAI*, 6496–6504.

Villata, S.; Boella, G.; Gabbay, D. M.; and van der Torre, L. W. N. 2012. Modelling defeasible and prioritized support in bipolar argumentation. *Ann. Math. Artif. Intell.*, 66(1-4): 163–197.

Wäldchen, S.; Macdonald, J.; Hauch, S.; and Kutyniok, G. 2021. The computational complexity of understanding binary classifier decisions. *Journal of Artificial Intelligence Research*, 70: 351–387.