


# Belief in Attacks in Epistemic Probabilistic Argumentation\*

Sylwia Polberg <sup>1</sup>, Anthony Hunter<sup>1</sup>, and Matthias Thimm<sup>2</sup>

<sup>1</sup> University College London, London, United Kingdom  
sylwia.polberg@gmail.com

<sup>2</sup> University of Koblenz-Landau, Koblenz, Germany

**Abstract.** The epistemic approach to probabilistic argumentation assigns belief to arguments. This is valuable in dialogical argumentation where one agent can model the beliefs another agent has in the arguments and this can be harnessed to make strategic choices of arguments to present. In this paper, we extend this epistemic approach by also representing the belief in attacks. We investigate properties of this proposal and compare it to the constellations approach showing neither subsumes the other.

**Keywords:** Abstract argumentation · Probabilistic argumentation · Epistemic argumentation

## 1 Introduction

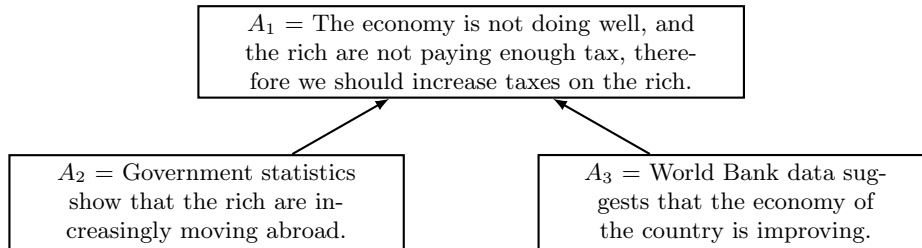
Abstract argumentation as proposed by Dung [8] provides an important formalism for representing and evaluating arguments and counterarguments. Proposals for probabilistic argumentation extend this to address aspects of uncertainty arising in argumentation. The two main approaches to probabilistic argumentation are the constellations and the epistemic approaches [14]. In the **constellations approach**, the uncertainty is in the topology of the graph. This approach is useful when one agent is not sure what arguments and attacks another agent is aware of, and so this can be captured by a probability distribution over the space of possible argument graphs. In the **epistemic approach**, the topology of the argument graph is fixed, but there is uncertainty as to the degree to which each argument is believed.

In this paper, we extend the epistemic approach with a probability distribution over the power set of attacks which we use to represent the uncertainty in each attack. To illustrate, we consider a listener to a political discussion on the radio. This is a situation where the listener acquires all the arguments and attacks that are presented, but does not add or delete arguments or attacks. The argument graph is given in Figure 1. Often the listener would evaluate the

---

\* This research is funded by EPSRC Project EP/N008294/1 “Framework for Computational Persuasion”.

arguments and attacks. For instance, she may have a low belief in  $A_3$  because she has found World Bank predictions to be unreliable in the past, and she may have a high belief in argument  $A_2$ , but a low belief in the attack by  $A_2$  on  $A_1$ . As a result, she may have a high belief in  $A_1$ . Note, in the constellations approach, it is not possible to represent all the arguments and attacks in one graph, and then assign belief to them.



**Fig. 1.** Example of an argument graph acquired by a listener to a debate.

Often uncertainty in attacks arises because “real-world” arguments are normally enthymemes (i. e. some or all of the premises and/or the claim are implicit). When an agent posits an enthymeme, the recipient decodes it to recover the intended argument. This creates a risk that the recipient decodes it differently to the way intended (as illustrated by the attack of  $A_2$  on  $A_1$  in Figure 1).

A potentially valuable role for the extended epistemic approach is in supporting an agent (X) when arguing with another agent (Y). Agent X can model agent Y to reflect the arguments *and* attacks that X thinks Y believes. This extends proposals for using the epistemic approach for user modelling in persuasion dialogues [16, 15].

The contributions of this paper are: (1) a set of constraints on probability distributions that take into account uncertainty of arguments and attacks; (2) results on the constraints showing inter-relationships between them; (3) results showing how with certain combinations of constraints recover and generalize Dung’s dialectical semantics; and (4) a comparison with the constellations approach showing how neither subsumes the other. All proofs are available online<sup>3</sup>.

## 2 Preliminaries

We start with a brief review of abstract argumentation as proposed by [8]. An **argument graph** (or a framework) is a directed graph  $G = (\mathcal{A}, \mathcal{R})$ , where  $\mathcal{A}$  is the set of arguments and  $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$  is the set of attacks. The way we decide which arguments can be accepted or rejected (or neither) is called a

<sup>3</sup> [http://www0.cs.ucl.ac.uk/staff/a.hunter/papers/extended\\_epistemic\\_full.pdf](http://www0.cs.ucl.ac.uk/staff/a.hunter/papers/extended_epistemic_full.pdf)

semantics. We focus on the argument-based approach [5, 1] and the adaptation of the attack-based approach from [2].

An **argument labeling** is a total function  $L : \mathcal{A} \rightarrow \{\text{in}, \text{out}, \text{und}\}$  [1, 5]. By  $\text{in}(L)$ ,  $\text{out}(L)$  and  $\text{und}(L)$  we denote the arguments mapped respectively to in, out and und(ecided) by  $L$ . We will often write a labeling as a triple  $(I, O, U)$ , where  $I$ ,  $O$  and  $U$  are sets of arguments mapped to in, out and und. We say that a set of elements attacks another element if it contains an appropriate attacker. We can now introduce the notion of legality, on which our semantics are based.

**Definition 1.** *An argument  $A \in \mathcal{A}$  is an **attacker** of  $B \in \mathcal{A}$  iff  $(A, B) \in \mathcal{R}$ . Let  $L : \mathcal{A} \rightarrow \{\text{in}, \text{out}, \text{und}\}$  be a labeling:*

- $X \in \text{in}(L)$  is **legally in** iff all its attackers are in  $\text{out}(L)$ .
- $X \in \text{out}(L)$  is **legally out** iff it has an attacker in  $\text{in}(L)$ .
- $X \in \text{und}(L)$  is **legally und** iff not all of its attackers are in  $\text{out}(L)$  and it does not have an attacker in  $\text{in}(L)$ .

**Definition 2.** *Let  $L : \mathcal{A} \rightarrow \{\text{in}, \text{out}, \text{und}\}$  be a labeling:*

- (cf)  $L$  is **conflict-free** iff it holds that if  $A \in \text{in}(L)$ , then none of its attackers is in  $\text{in}(L)$ , and every  $A \in \text{out}(L)$  is legally out
- (ad)  $L$  is **admissible** iff every  $A \in \text{in}(L)$  is legally in and every  $A \in \text{out}(L)$  is legally out.
- (co)  $L$  is **complete** if it is admissible and every  $A \in \text{und}(L)$  is legally und.

*Additionally, a complete labeling is stable (st) if  $\text{und}(L) = \emptyset$ , it is preferred (pr) if  $\text{in}(L)$  is maximal wrt.  $\subseteq$ , and it is grounded (gr) if  $\text{in}(L)$  is minimal wrt.  $\subseteq$ .*

In Dung’s semantics, attacks are seen as secondary to arguments. For example, ensuring that no attack on a given argument is accepted is equivalent to making sure that no argument carrying out an attack is accepted. However, this correspondence does not always hold for various generalizations of Dung’s graph, in which a given conflict may not be successful due to preferences, probabilities, or when it is a target of an attack as well [4]. Therefore, the status assigned to a given attack is not necessarily the same as assigned to its source. We will now adapt the approach from [2] and focus on the **extended labelings**, which are total functions  $L^* : \mathcal{A} \cup \mathcal{R} \rightarrow \{\text{in}, \text{out}, \text{und}\}$ . We introduce the notion of an extended attacker (attacker\*), which is now a conflict, not an argument, and the attackee can be both an argument and a relation.

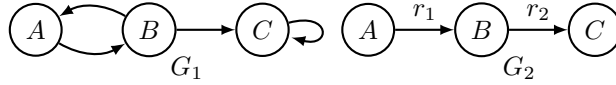
**Definition 3.** *For an attack  $\alpha = (A, B) \in \mathcal{R}$ , the source of  $\alpha$  is  $\text{src}(\alpha) = A$  and the target of  $\alpha$  is  $\text{trg}(\alpha) = B$ . An attack  $\alpha \in \mathcal{R}$  is an **attacker\*** 1) of  $B \in \mathcal{A}$  iff  $B = \text{trg}(\alpha)$ , and 2) of  $\beta \in \mathcal{R}$  iff  $\text{trg}(\alpha) = \text{src}(\beta)$ .*

By replacing attacker with attacker\* in the previous definitions and maximizing/minimizing attacks as well as arguments, we obtain the attack-based semantics, further distinguished with \*. We have the following correspondence between these two families of semantics [2]. Please observe that that extended labelings can, in general, only be projected to their corresponding ordinary labelings if they are at least complete.

**Proposition 1.** *If  $L = (I, O, U)$  is a  $\sigma$ -labeling, where  $\sigma \in \{\text{cf}, \text{ad}, \text{co}, \text{pr}, \text{gr}, \text{st}\}$ , then  $L^* = (I \cup \{\alpha \mid \text{src}(\alpha) \in I\}, O \cup \{\alpha \mid \text{src}(\alpha) \in O\}, U \cup \{\alpha \mid \text{src}(\alpha) \in U\})$  is a  $\sigma^*$ -labeling. If  $L^* = (I^*, O^*, U^*)$  is a  $\delta^*$ -labeling, where  $\delta \in \{\text{co}, \text{pr}, \text{gr}, \text{st}\}$ , then  $L = (I^* \cap \mathcal{A}, O^* \cap \mathcal{A}, U^* \cap \mathcal{A})$  is a  $\delta$ -labeling.*

We use  $\sigma(G)$  to denote the set of labelings of  $G$  according to the semantics  $\sigma \in \{\text{cf}, \text{ad}, \text{co}, \text{pr}, \text{gr}, \text{st}, \text{cf}^*, \text{ad}^*, \text{co}^*, \text{pr}^*, \text{gr}^*, \text{st}^*\}$ . We will say that a set of arguments  $\mathcal{S}$  is a  $\sigma$ -extension iff there exists a  $\sigma$ -labeling  $L$  s. t.  $\text{in}(L) = \mathcal{S}$ .

*Example 1.* Consider the graph  $G_1$  below. The admissible labelings are  $L_1 = (\emptyset, \emptyset, \{A, B, C\})$ ,  $L_2 = (\{A\}, \{B\}, \{C\})$ ,  $L_3 = (\{B\}, \{A\}, \{C\})$  and  $L_4 = (\{B\}, \{A, C\}, \emptyset)$ . Apart from  $L_3$ , all of them are complete.  $L_1$  is grounded,  $L_2$  and  $L_4$  are preferred, and  $L_4$  is stable.



Now consider the graph  $G_2$  above. The admissible labelings are  $L_1 = (\emptyset, \emptyset, \{A, B, C\})$ ,  $L_2 = (\{A\}, \emptyset, \{B, C\})$ ,  $L_3 = (\{A\}, \{B\}, \{C\})$ , and  $L_4 = (\{A, C\}, \{B\}, \emptyset)$ .  $L_4$  is the single complete, preferred, stable and grounded labeling. The admissible\* labelings of  $G_2$  are  $L_1^* = (\emptyset, \emptyset, \{A, B, C, r_1, r_2\})$ ,  $L_2^* = (\{A\}, \emptyset, \{B, C, r_1, r_2\})$ ,  $L_3^* = (\{r_1\}, \emptyset, \{A, B, C, r_2\})$ ,  $L_4^* = (\{r_1\}, \{B\}, \{A, C, r_2\})$ ,  $L_5^* = (\{r_1\}, \{r_2\}, \{A, B, C\})$ ,  $L_6^* = (\{r_1\}, \{B, r_2\}, \{A, C\})$ ,  $L_7^* = (\{A, r_1\}, \emptyset, \{B, C, r_2\})$ ,  $L_8^* = (\{A, r_1\}, \{B\}, \{C, r_2\})$ ,  $L_9^* = (\{A, r_1\}, \{r_2\}, \{B, C\})$ ,  $L_{10}^* = (\{A, r_1\}, \{B, r_2\}, \{A, C\})$ ,  $L_{11}^* = (\{r_1, C\}, \{r_2\}, \{A, B\})$ ,  $L_{12}^* = (\{r_1, C\}, \{B, r_2\}, \{A\})$ ,  $L_{13}^* = (\{A, r_1, C\}, \{r_2\}, \{B\})$  and  $L_{14}^* = (\{A, r_1, C\}, \{B, r_2\}, \emptyset)$ .  $L_{14}^*$  is the single complete\*, preferred\*, stable\* and grounded\* labeling.

We can observe that even though every argument-based labeling has a corresponding extended one, the removal of attacks from the extended labeling does not necessarily give us a standard one (e. g.  $L_{11}^*$  and  $L_{12}^*$ ).

### 3 Constellations Approach

We review the constellations approach [13] which extends [7] and [19]. It allows representation of the uncertainty over the topology of the graph: Each subgraph of the original graph is assigned a probability to represent the chances of it being the actual argument graph of the agent. It can be used to model what arguments and attacks an agent is aware of.

**Definition 4.** *For  $G = (\mathcal{A}, \mathcal{R})$  and  $G' = (\mathcal{A}', \mathcal{R}')$ , the subgraph relation, denoted  $\sqsubseteq$ , is defined as  $G' \sqsubseteq G$  iff  $\mathcal{A}' \subseteq \mathcal{A}$  and  $\mathcal{R}' \subseteq (\mathcal{A}' \times \mathcal{A}') \cap \mathcal{R}$ . The set of subgraphs of  $G$  is  $\text{Sub}(G) = \{G' \mid G' \sqsubseteq G\}$ . A subgraph  $(\mathcal{A}', \mathcal{R}')$  is **full** iff  $\mathcal{A}' \subseteq \mathcal{A}$  and  $\mathcal{R}' = (\mathcal{A}' \times \mathcal{A}') \cap \mathcal{R}$ . A subgraph  $(\mathcal{A}', \mathcal{R}')$  is **spanning** iff  $\mathcal{A}' = \mathcal{A}$  and  $\mathcal{R}' \subseteq \mathcal{R}$ .*

If our uncertainty is about which arguments appear in the graph, then only the full (induced) subgraphs of the argument graph have non-zero probability. If we are only uncertain about which attacks appear, then it is the spanning subgraphs of the argument graph that can have non-zero probability.

**Definition 5.** A **subgraph distribution** is a function  $P^c : \text{Sub}(G) \rightarrow [0, 1]$  with  $\sum_{G' \in \text{Sub}(G)} P^c(G') = 1$ . A subgraph distribution  $P^c$  is a **full subgraph distribution** iff if  $(\mathcal{A}', \mathcal{R}')$  is not a full subgraph, then  $P^c((\mathcal{A}', \mathcal{R}')) = 0$ . A subgraph distribution  $P^c$  is a **spanning subgraph distribution** iff if  $(\mathcal{A}', \mathcal{R}')$  is not a spanning subgraph,  $P^c((\mathcal{A}', \mathcal{R}')) = 0$ .

Determining the probability that a set of arguments is an extension (labeling) of a particular type (e.g. grounded, preferred, etc.) is done by collecting the probabilities of the subgraphs producing the desired labelings. In a similar fashion, we can derive the probability of an argument being accepted in a labeling of a given type.

**Definition 6.** For  $S \subseteq \mathcal{A}$  and  $\sigma \in \{\text{cf}, \text{ad}, \text{co}, \text{pr}, \text{gr}, \text{st}\}$ , the probability that  $L : S \rightarrow \{\text{in}, \text{out}, \text{und}\}$  is a  $\sigma$ -labeling is:

$$P_\sigma(L) = \sum_{G' \in \text{Sub}(G) \text{ s.t. } L \in \sigma(G')} P^c(G')$$

**Definition 7.** Given a semantics  $\sigma \in \{\text{ad}, \text{co}, \text{pr}, \text{gr}, \text{st}\}$ , the probability that  $A \in \mathcal{A}$  is in in a  $\sigma$ -labeling is

$$P_\sigma(A) = \sum_{G' \in \text{Sub}(G) \text{ s.t. } L \in \sigma(G') \text{ and } A \in \text{in}(L)} P(G')$$

*Example 2.* Consider the graph  $G = (\{A, B\}, \{(A, B)\})$ . Its subgraphs are  $G_1 = (\{A, B\}, \{(A, B)\})$ ,  $G_2 = (\{A, B\}, \emptyset)$ ,  $G_3 = (\{A\}, \emptyset)$ ,  $G_4 = (\{B\}, \emptyset)$  and  $G_5 = (\emptyset, \emptyset)$ . Out of them,  $G_1$ ,  $G_3$ ,  $G_4$  and  $G_5$  are full, and  $G_1$  and  $G_2$  are spanning. Consider the following subgraph distribution  $P^c$ :  $P^c(G_1) = 0.09$ ,  $P^c(G_2) = 0.81$ ,  $P^c(G_3) = 0.01$  and  $P^c(G_4) = 0.09$  and  $P^c(G_5) = 0$ . The probability of a given set being a grounded extension is as follows:  $P_{\text{gr}}(\{A, B\}) = P^c(G_2) = 0.81$ ;  $P_{\text{gr}}(\{A\}) = P^c(G_1) + P^c(G_3) = 0.1$ ;  $P_{\text{gr}}(\{B\}) = P^c(G_4) = 0.09$ ; and  $P_{\text{gr}}(\{\}) = P^c(G_5) = 0$ . Therefore  $P_{\text{gr}}(A) = 0.91$  and  $P_{\text{gr}}(B) = 0.9$ .

## 4 Extended Epistemic Approach

In the original version of the epistemic approach [22, 14, 17, 3, 18], an argument graph has an associated probability distribution over the sets of arguments. From this, we derive the probability of a single argument and interpret it as the belief that an agent has in it (i.e. the degree to which the agent believes the premises and the conclusion drawn from those premises). We say that an agent believes an argument  $A$  to some degree when  $P(A) > 0.5$ , disbelieves an argument to some degree when  $P(A) < 0.5$ , and neither believes nor disbelieves an argument when  $P(A) = 0.5$ . Here we extend the approach with uncertainty over attacks. For this, we introduce the probability of attack (i.e. the degree of belief that the attacker does indeed attack the attackee). We use two functions in the definition because we want to investigate the interplay between them.

**Definition 8.** An epistemic bidistribution is a pair  $(P^a, P^r)$  where

- $P^a$  is a function  $P^a : 2^{\mathcal{A}} \rightarrow [0, 1]$  with  $\sum_{S \subseteq \mathcal{A}} P^a(S) = 1$  (**argument belief distribution**).
- $P^r$  is a function  $P^r : 2^{\mathcal{R}} \rightarrow [0, 1]$  with  $\sum_{S \subseteq \mathcal{R}} P^r(S) = 1$  (**attack belief distribution**).

The **probability of an argument**  $A$  is  $P^a(A) = \sum_{S \subseteq \mathcal{A} \text{ s.t. } A \in S} P^a(S)$ . The **probability of an attack**  $\alpha$  is  $P^r(\alpha) = \sum_{S \subseteq \mathcal{R} \text{ s.t. } \alpha \in S} P^r(S)$ . Finally, let  $P^b(X)$  denote  $P^a(X)$  (resp.  $P^r(X)$ ) when  $X \in \mathcal{A}$  (resp.  $X \in \mathcal{R}$ ).

In order to simplify the notation, we drop the brackets for representing the probability of an attack relation, i. e., for  $(A, B) \in \mathcal{R}$ , instead of  $P^r((A, B))$  we write  $P^r(A, B)$ .

The epistemic probability distributions can be constrained by imposing rationality postulates. In what follows we will build up on some of the postulates from [18] and introduce some new ones. The previous results can be retrieved by considering bidistributions in which all attacks are believed. We separate our new approaches into two families of postulates.

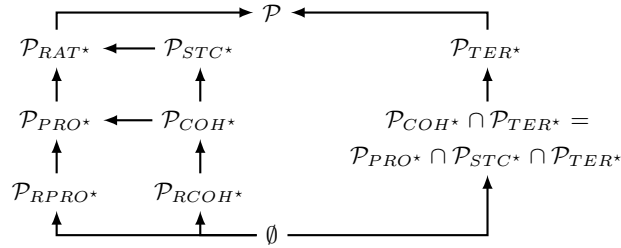
We start with the independent family of postulates in Definition 9 and give results on inter-relationships in Figure 2 where  $\mathcal{P}_\mu$  is the set of bidistributions satisfying postulate  $\mu$  in  $G$ . The family is called independent because there is no dependence imposed between belief in attacks and belief in attackers, i.e. the probabilities assigned to an attack  $\alpha$  and to  $src(\alpha)$  are not necessarily related. RAT\*, TER\*, COH\*, and OPT\* require that both the attacker and the attack itself need to be believed in order to affect the attackee, or that either of them can be disbelieved in order for belief in the target. For RAT\* (resp. STC\*), if an attacker and its attack are believed, then the attackee is not believed (resp. disbelieved). As a dual for STC\*, PRO\* ensures that if an attack and attackee are believed, the attacker is not believed. TRU\* requires that an argument is believed when there is no evidence to the contrary. By DIS\*, an argument can only be disbelieved for a reason. TER\* simply limits beliefs to three values corresponding precisely to the in, out and und statuses from the standard semantics. The ABIN\* postulate prohibits being undecided about beliefs. Finally, while all the previous properties consider belief and disbelief, the COH\* and OPT\* properties give margins for probability assignments—one focuses on the upper, the other on the lower bound. By varying the use of also undecided attacks, we can specialize our axioms further, as seen in the case of RPRO\* and RCOH\*.

**Definition 9.** (The independent family of postulates). An epistemic bidistribution  $(P^a, P^r)$  is:

- (RAT\*) **rational\*** if for all  $A, B \in \mathcal{A}$  s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) > 0.5$ ,  $P^a(A) > 0.5$  implies  $P^a(B) \leq 0.5$ .
- (STC\*) **strict\*** if for all  $A, B \in \mathcal{A}$ , s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) > 0.5$ ,  $P^a(A) > 0.5$  implies  $P^a(B) < 0.5$ .
- (PRO\*) **protective\*** if for all  $A, B \in \mathcal{A}$  s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) > 0.5$ ,  $P^a(B) > 0.5$  implies  $P^a(A) < 0.5$ .

- (**RPRO\***) **restricted protective\*** if for all  $A, B \in \mathcal{A}$  s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) \geq 0.5$ ,  $P^a(B) > 0.5$  implies  $P^a(A) < 0.5$ .
- (**TRU\***) **trusting\*** if for every  $B \in \mathcal{A}$ , it holds that if for all  $C \in \mathcal{A}$  s.t.  $(C, B) \in \mathcal{R}$ , either  $P^a(C) < 0.5$ , or  $P^r(C, B) < 0.5$ , then  $P^a(B) > 0.5$ .
- (**DIS\***) **discharging\*** if for every  $B \in \mathcal{A}$ , if  $P^a(B) < 0.5$ , then there exists  $C \in \mathcal{A}$  s.t.  $(C, B) \in \mathcal{R}$ ,  $P^r(C, B) > 0.5$  and  $P^a(C) > 0.5$ .
- (**TER\***) **ternary\*** if for all  $X \in \mathcal{A} \cup \mathcal{R}$ ,  $P^b(X) \in \{0, 0.5, 1\}$ .
- (**ABIN\***) **attack binary\*** if for all  $X \in \mathcal{R}$ ,  $P^r(X) \neq 0.5$ .
- (**COH\***) **coherent\*** if for all  $A, B \in \mathcal{A}$  s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) > 0.5$ ,  $P^a(A) \leq 1 - P^a(B)$ .
- (**RCOH\***) **restricted coherent\*** if for all  $A, B \in \mathcal{A}$ , s.t.  $(A, B) \in \mathcal{R}$  and  $P^r(A, B) \geq 0.5$ ,  $P^a(A) \leq 1 - P^a(B)$ .
- (**OPT\***) **optimistic\*** if for every  $A \in \mathcal{A}$ , it holds that  $P^a(A) \geq 1 - \sum_{B \text{ s.t. } (B, A) \in \mathcal{R}, P^r(B, A) > 0.5} P^a(B)$ .

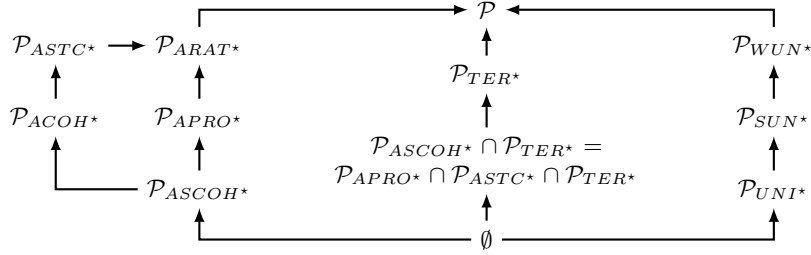
In the independent family, the belief we have in an attacker does not constrain the belief we may have in its attack. We consider it an intuitive modeling, as we do not have to believe two arguments in order to acknowledge a conflict between them. Imagine two people witnessing a robbery, one claiming that the criminal ran away in a car, the other that he used a bike. The statements are clearly conflicting and we can believe the attacks between them independently of the belief we have in the witnesses. Similarly, we do not need to believe a given attack even if we believe the arguments participating in it, as exemplified in the introduction.



**Fig. 2.** Relationships for the independent family of postulates where  $\mathcal{P}_{\mu_1} \rightarrow \mathcal{P}_{\mu_2}$  denotes  $\mathcal{P}_{\mu_1}(G) \subseteq \mathcal{P}_{\mu_2}(G)$ .

We present a second family of postulates in Definition 10, called the dependent family, and give results on inter-relationships in Figure 3. This second family is motivated by the observation that in some situations (e.g. when argument graphs are obtained from logical knowledge bases), it is natural to expect that there is a dependence between belief in an attacker and its attack. Moreover, in many approaches that explicitly include the attacks in extensions and labelings, the conflicts need to conform to the same semantics as the arguments. Consequently, we can demand that the belief in an argument affects

the belief in its attacks and vice versa. For this, we introduce the UNI\*, SUN\* and WUN\* postulates below. Similarly, we also present the attack postulates, which constrain the belief both in the attacked argument and the conflict whose source is attacked, thus implicitly acknowledging the dependency between the two. Moreover, while in the independent family the beliefs in the attack and the attacker had to be mentioned explicitly due to their independence, in this family we consider just the attack itself. This also reflects the intuition behind the attack-based approach.



**Fig. 3.** Relationships for the dependent family of postulates where  $\mathcal{P}_{\mu_1} \rightarrow \mathcal{P}_{\mu_2}$  denotes  $\mathcal{P}_{\mu_1}(G) \subseteq \mathcal{P}_{\mu_2}(G)$ .

**Definition 10.** (*The dependent family of postulates*). An epistemic bidistribution  $(P^a, P^r)$  is:

- (UNI\*) **unified\*** if for all  $(A, B) \in \mathcal{R}$ ,  $P^r(A, B) = P^a(A)$
- (SUN\*) **semi-unified\*** if for all  $(A, B) \in \mathcal{R}$ ,  $P^a(A) > 0.5$  iff  $P^r(A, B) > 0.5$  and  $P^a(A) < 0.5$  iff  $P^r(A, B) < 0.5$
- (WUN\*) **weakly unified\*** if for all  $(A, B) \in \mathcal{R}$ , either both  $P^r(A, B) \geq 0.5$  and  $P^a(A) \geq 0.5$  or both  $P^r(A, B) \leq 0.5$  and  $P^a(A) \leq 0.5$ .
- (ARAT\*) **attack rational\*** iff for every  $\alpha \in \mathcal{R}$ , if  $P^r(\alpha) > 0.5$  and  $\alpha$  is an attacker\* of  $X \in \mathcal{A} \cup \mathcal{R}$ , then  $P^b(X) \leq 0.5$ .
- (ASTC\*) **attack strict\*** iff for every  $\alpha \in \mathcal{R}$ , if  $P^r(\alpha) > 0.5$  and  $\alpha$  is an attacker\* of  $X \in \mathcal{A} \cup \mathcal{R}$ , then  $P^b(X) < 0.5$ .
- (APRO\*) **attack protective\*** iff for every  $X \in \mathcal{A} \cup \mathcal{R}$  and  $\alpha \in \mathcal{R}$  s.t.  $\alpha$  is an attacker\* of  $X$ , if  $P^b(X) > 0.5$ , then  $P^r(\alpha) < 0.5$ .
- (ATRU\*) **attack trusting\*** iff for every  $X \in \mathcal{A} \cup \mathcal{R}$ , it holds that if for every attacker\*  $\beta \in \mathcal{R}$  of  $X$  it is the case that  $P^r(\beta) < 0.5$ , then  $P^b(X) > 0.5$ .
- (ADIS\*) **attack discharging\*** iff for every  $X \in \mathcal{A} \cup \mathcal{R}$ , if  $P^b(X) < 0.5$ , then there exists an attacker\*  $\beta \in \mathcal{R}$  of  $X$  s.t.  $P^r(\beta) > 0.5$ .
- (ACOH\*) **attack coherent\*** iff for every  $X \in \mathcal{A} \cup \mathcal{R}$  and  $\alpha \in \mathcal{R}$  s.t.  $\alpha$  is an attacker\* of  $X$ , if  $P^r(\alpha) > 0.5$ , then  $P^r(\alpha) \leq 1 - P^b(X)$ .
- (ASCOH\*) **attack strongly coherent\*** iff for every  $X \in \mathcal{A} \cup \mathcal{R}$  and  $\alpha \in \mathcal{R}$  s.t.  $\alpha$  is an attacker\* of  $X$ , if  $P^r(\alpha) \geq 0.5$ , then  $P^r(\alpha) \leq 1 - P^b(X)$ .





$\mathcal{A} \mid P^a(A) < 0.5\}$ , and  $U = \{A \in \mathcal{A} \mid P^a(A) = 0.5\}$ . The **extended epistemic labeling** is  $L_{P^a, P^r}^* = (I, O, U)$ , where  $I = \{X \in \mathcal{A} \cup \mathcal{R} \mid P^b(X) > 0.5\}$ ,  $O = \{X \in \mathcal{A} \cup \mathcal{R} \mid P^b(X) < 0.5\}$ , and  $U = \{X \in \mathcal{A} \cup \mathcal{R} \mid P^b(X) = 0.5\}$ .

We will first show how our independent family of postulates (i.e. Def. 9) relates to the classical notions. We can draw a connection between the epistemic bidistributions and the labelings of a subgraph of the original graph, which is obtained by considering only those attacks that are believed:

**Definition 12.** Let  $(P^a, P^r)$  be an epistemic bidistribution. The set of **believed attacks** is  $\text{BAtts}(P^r) = \{(A, B) \in \mathcal{R} \mid P^r(A, B) > 0.5\}$ . The subgraph of  $G$  induced by  $P^r$  is  $G' = (\mathcal{A}, \text{BAtts}(P^r))$ .

**Proposition 2.** Let  $(P^a, P^r)$  be an epistemic bidistribution and  $G' = (\mathcal{A}, \text{BAtts}(P^r))$  the subgraph of  $G$  induced by  $P^r$ .

- $(P^a, P^r) \in \mathcal{P}_{\text{RAT}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G)$  iff  $L_{P^a} \in \text{cf}(G')$ .
- $(P^a, P^r) \in \mathcal{P}_{\text{PRO}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G)$  iff  $L_{P^a} \in \text{ad}(G')$ .
- $(P^a, P^r) \in \mathcal{P}_{\text{PRO}^*}(G) \cap \mathcal{P}_{\text{STC}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G)$  iff  $L_{P^a} \in \text{co}(G')$ .

We can observe that although one attack distribution induces only one subgraph, a single subgraph can be induced by multiple distributions. This is due to the fact that the removal of the attacks depends on whether an attack is believed at all, not on the degree of this belief. Moreover, there can be infinitely many argument distributions associated with a single complete labeling due to the fact that any of the values from  $[0, 0.5)$  (or  $(0.5, 1]$ ) can lead to an out (or in) assignment of a given labeling. Although this is to be expected taking into account the fact that probabilistic semantics carry more information than the classical ones, we would also like to distinguish those probability functions that can be uniquely associated with a given subgraph and its complete labelings. We thus propose the definition of complete probability bidistributions; a given subgraph can be induced only by a single ternary and attack binary distribution, while ternary, trusting, disapproving and coherent postulates lead to a tighter relation with the labelings:

**Definition 13.** An epistemic bidistribution  $(P^a, P^r)$  is **complete** iff  $(P^a, P^r) \in \mathcal{P}_{\text{COH}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G) \cap \mathcal{P}_{\text{TER}^*}(G) \cap \mathcal{P}_{\text{ABIN}^*}(G)$ .

From this, we can further define the preferred, complete and stable bidistributions that lead to appropriate labelings in the associated subgraph by maximizing or minimizing particular assignments, similarly as in the classical semantics. In this case, instead of focusing on in, out and und assignments, we look for probabilities 1, 0 and 0.5.

Although the extended epistemic approach is quite general, the epistemic labelings without any constraints on the attack distributions are connected to the labellings of the subgraphs of a given framework, not necessarily the framework itself. However, if we apply the dependency postulates from the dependent family

(Definition 10)— in particular, the semi-unified one—we can observe that we can focus on the original graph again. The only difference wrt. the previous results is the use of the restricted, not standard protectiveness.

**Proposition 3.** *The following holds:*

- If  $L \in \text{cf}(G)$ , then there exists  $(P^a, P^r) \in \mathcal{P}_{\text{RAT}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$  s.t.  $L = L_{P^a}$ .
- If  $L \in \text{ad}(G)$ , then there exists  $(P^a, P^r) \in \mathcal{P}_{\text{RPRO}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$  s.t.  $L = L_{P^a}$ .
- If  $L \in \text{co}(G)$ , then there exists  $(P^a, P^r) \in \mathcal{P}_{\text{RPRO}^*}(G) \cap \mathcal{P}_{\text{STC}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$  s.t.  $L = L_{P^a}$ .
- If  $(P^a, P^r) \in \mathcal{P}_{\text{RAT}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$ , then  $L_{P^a} \in \text{cf}(G)$ .
- If  $(P^a, P^r) \in \mathcal{P}_{\text{RPRO}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$ , then  $L_{P^a} \in \text{ad}(G)$ .
- If  $(P^a, P^r) \in \mathcal{P}_{\text{RPRO}^*}(G) \cap \mathcal{P}_{\text{STC}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G) \cap \mathcal{P}_{\text{SUN}^*}(G)$ , then  $L_{P^a} \in \text{co}(G)$ .

This leads to the following complete probability bidistribution, which can uniquely describe the complete labelings of the underlying framework. Using this, we can also retrieve the preferred, grounded and stable labellings as for the classical case (as discussed in Section 2).

**Definition 14.** *An epistemic bidistribution  $(P^a, P^r)$  is **jointly complete** iff  $(P^a, P^r) \in \mathcal{P}_{\text{SUN}^*}(G) \cap \mathcal{P}_{\text{RCOH}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G) \cap \mathcal{P}_{\text{TER}^*}(G)$ .*

Let us now focus on the extended classical semantics. As we could have already observed in Example 1, the admissible\* labelings were not necessarily corresponding to the admissible ones. However, we can easily grasp it with our attack epistemic postulates.

**Proposition 4.** *The following holds:*

- $(P^a, P^r) \in \mathcal{P}_{\text{WUN}^*}(G) \cap \mathcal{P}_{\text{ARAT}^*}(G) \cap \mathcal{P}_{\text{ADIS}^*}(G)$  iff  $L_{P^a, P^r}^* \in \text{cf}^*(G)$ .
- $(P^a, P^r) \in \mathcal{P}_{\text{WUN}^*}(G) \cap \mathcal{P}_{\text{APRO}^*}(G) \cap \mathcal{P}_{\text{ADIS}^*}(G)$  iff  $L_{P^a, P^r}^* \in \text{ad}^*(G)$ .
- $(P^a, P^r) \in \mathcal{P}_{\text{SUN}^*}(G) \cap \mathcal{P}_{\text{APRO}^*}(G) \cap \mathcal{P}_{\text{ASTC}^*}(G) \cap \mathcal{P}_{\text{ADIS}^*}(G) \cap \mathcal{P}_{\text{ATRU}^*}(G)$  iff  $L_{P^a, P^r}^* \in \text{co}^*(G)$ .

The fact that the complete\* labelings correspond to bidistributions satisfying the SUN\* postulate gives us one more important result. In particular, under the SUN\* postulate we can replace the other postulates from the dependent family with their counterparts from the independent family. This also means that we can use the jointly complete bidistributions in order to uniquely retrieve the extended labelings of  $G$  that are at least complete\*.

**Theorem 1.** *Let  $(P^a, P^r)$  be an epistemic bidistribution. Then  $L_{P^a, P^r}^* \in \text{co}^*(G)$  iff  $(P^a, P^r) \in \mathcal{P}_{\text{SUN}^*}(G) \cap \mathcal{P}_{\text{RPRO}^*}(G) \cap \mathcal{P}_{\text{STC}^*}(G) \cap \mathcal{P}_{\text{DIS}^*}(G) \cap \mathcal{P}_{\text{TRU}^*}(G)$ .*

These results show that our new proposal for epistemic probabilities can generalize a wider range of argumentation semantics than the original one [18]. Moreover, what we have presented can be easily extended to handle the attack-based semantics from [24] and recursive attacks from [2].

## 6 Comparison with Constellations Approach

The reasoning behind the epistemic and constellations approaches is different, with the former intended to reflect the belief in arguments and attacks, and the latter expressing the uncertainty concerning the topology of the graph, e. g., as to which arguments and attacks are known about or what elements should appear in the graph. Nevertheless, we can still draw some connections between them. We can observe that in a subgraph distribution assigning non-zero probability only to subgraphs without attacks, the grounded extension of each subgraph would consist of all of its arguments. These extensions and their probabilities produce an argument distribution. Thus, the constellations approach can up to some extent mimic the epistemic approach:

**Proposition 5.** *For each argument-belief distribution  $P^a$  over  $G$ , there is a constellations distribution  $P^c$  over  $\text{Sub}(G)$  s.t. for all arguments  $A$  in  $\mathcal{A}$ ,  $P^a(A) = P_{\text{gr}}^c(A)$ .*

In turn, a spanning or full subgraph distribution can be simulated with the attack or argument belief distribution due to the fact that part of a subgraph becomes “fixed” and not directly subject to any uncertainty.

**Proposition 6.** *For each spanning subgraph distribution  $P^c$  over  $G$ , there is an attack belief distribution  $P^r$  s.t. for all subgraphs  $G' \sqsubseteq G$ , and for all sets of attacks  $\mathcal{S} \subseteq \mathcal{R}$ , if  $\mathcal{R}' = \mathcal{S}$ , then  $P^c(G') = P^r(\mathcal{S})$ .*

**Proposition 7.** *For each full subgraph distribution  $P^c$  over  $G$ , there is an argument belief distribution  $P^a$  s.t. for all subgraphs  $G' \sqsubseteq G$ , and for all sets of arguments  $\mathcal{S} \subseteq \mathcal{A}$ , if  $\mathcal{A}' = \mathcal{S}$ , then  $P^c(G') = P^a(\mathcal{S})$ .*

However, we can observe that if a subgraph distribution is neither a full subgraph distribution nor a spanning subgraph distribution, then the constellations approach cannot be captured by the epistemic approach. Moreover, in the constellations approach, the marginal value of a given argument (i. e. the total probability of subgraphs containing this argument) is never less than the marginal for any attack involving that argument. In contrast, the belief in an attacker can be greater than then belief in the attack or attackee. This shows that, in general, the epistemic approach cannot be captured by the constellations method.

**Definition 15.** *Let  $P^c$  be a subgraph distribution. The **argument marginal function** is  $P^m(A) = \sum_{G' \in \text{Sub}(G) \text{ s.t. } A \in \mathcal{A}'} P(G')$ . The **attack marginal function** is  $P^m(A, B) = \sum_{G' \in \text{Sub}(G) \text{ s.t. } (A, B) \in \mathcal{R}'} P(G')$ .*

**Proposition 8.** *Let  $P^c$  be a subgraph distribution. For all  $(A, B) \in \mathcal{R}$ ,  $P^c(A) \geq P^c(A, B)$ .*

*Example 4.* Consider the graph  $G_1 = (\{A, B, C\}, \{(A, B), (B, A), (C, B)\})$  and its subgraphs  $G_2 = (\{A, B, C\}, \{(A, B), (B, A)\})$  and  $G_3 = (\{A, B\}, \{(A, B)\})$ . For this graph, we consider the subgraph distribution  $P^c(G_1) = 0.3$ ,  $P^c(G_2) = 0.5$  and  $P^c(G_3) = 0.2$ , which is neither a full subgraph nor a spanning subgraph distribution. We cannot use  $P^a$  or  $P^r$  to represent  $P^c$ .

We can now consider an epistemic bidistribution  $(P^a, P^r)$  s.t.  $P^r(\{A, B\}) = 1$  and for every set  $\mathcal{S} \subseteq \mathcal{A}$  s.t.  $A \in \mathcal{S}$ ,  $P^a(\mathcal{S}) = 0$  (the remaining assignments are arbitrary as long as we obtain a distribution). Therefore,  $P^a(A) < P^a((A, B))$ . But there cannot be any subgraph distribution  $P^c$  for  $G_1$  s.t.  $P^m(A) < P^m((A, B))$ .

These results show that extended epistemic and constellations approach, although related, do not subsume each other.

## 7 Conclusions

In this paper, we extend the epistemic approach to account for belief in attacks as well as arguments. We do this by introducing the notion of an epistemic bidistribution. We then provide two families of postulates that offer a variety of ways of constraining the bidistributions according to different notions of rational behaviour, give some relationships between these two families, and show how these postulates relate to classical semantics for abstract argumentation, and we show how the extended epistemic and constellations approaches do not subsume each other.

Important dimensions for probabilistic argumentation include the constellations approach to abstract argumentation (e.g. [7, 19, 20, 6, 9]), the equational approach to abstract argumentation [10], and probabilistic structured argumentation (e.g. [21, 7, 23]). The extended epistemic approach is complementary to these existing approaches (see Section 6, for differences with the constellations approach, and see [10], for differences between the epistemic and equational approaches).

The epistemic approach is a promising approach to user modelling in persuasion where a persuader can model the beliefs in arguments of the persuadee and update the model during a dialogue [15, 16, 11], and the user model can be harnessed to make strategic choices of move in a dialogue using decision theory [12]. The extended epistemic approach offers richer user models, and potentially more effective decisions about moves (as indicated by our example in Section 1).

## References

1. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. *The Knowledge Engineering Review* 26(4), 365–410 (2011)
2. Baroni, P., Cerutti, F., Giacomin, M., Guida, G.: AFRA: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning* 52(1), 19–37 (2011)

3. Baroni, P., Giacomin, M., Vicig, P.: On rationality conditions for epistemic probabilities in abstract argumentation. In: Proceedings of COMMA'14. FAIA, vol. 266, pp. 121–132. IOS Press (2014)
4. Brewka, G., Polberg, S., Woltran, S.: Generalizations of Dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems* 29(1), 30–38 (2014)
5. Caminada, M., Gabbay, D.M.: A logical account of formal argumentation. *Studia Logica* 93, 109–145 (2009)
6. Dondio, P.: Multi-valued and probabilistic argumentation frameworks. In: Proceedings of COMMA'14. FAIA, vol. 266, pp. 253–260. IOS Press (2014)
7. Dung, P., Thang, P.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: Proceedings of COMMA'10. FAIA, vol. 216, pp. 171–182. IOS Press (2010)
8. Dung, P.M.: On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence* 77(2), 321–358 (1995)
9. Fazzinga, B., Flesca, S., Parisi, F.: On the complexity of probabilistic abstract argumentation frameworks. *ACM Transactions on Computational Logic* 16(3), 22:1–22:39 (2015)
10. Gabbay, D., Rodrigues, O.: Probabilistic argumentation: An equational approach. *Logica Universalis* 9(3), 345–382 (2015)
11. Hadoux, E., Hunter, A.: Computationally viable handling of beliefs in arguments for persuasion. In: Proceedings of ICTAI'16. pp. 319–326. IEEE (2016)
12. Hadoux, E., Hunter, A.: Strategic sequences of arguments for persuasion using decision trees. In: Proceedings of AAI'17. pp. 1128–1134. AAAI Press (2017)
13. Hunter, A.: Some foundations for probabilistic abstract argumentation. In: Proceedings of COMMA'12. FAIA, vol. 245, pp. 117–128. IOS Press (2012)
14. Hunter, A.: A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning* 54(1), 47–81 (2013)
15. Hunter, A.: Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In: Proceedings of IJCAI'15. pp. 3055–3061. AAAI Press (2015)
16. Hunter, A.: Persuasion dialogues via restricted interfaces using probabilistic argumentation. In: Proceedings of SUM'16. LNCS, vol. 9858, pp. 184–198. Springer (2016)
17. Hunter, A., Thimm, M.: Probabilistic argument graphs for argumentation lotteries. In: Proceedings of COMMA'14. FAIA, vol. 266, pp. 313–324. IOS Press (2014)
18. Hunter, A., Thimm, M.: Probabilistic argumentation with epistemic extensions and incomplete information. Tech. rep., ArXiv (May 2014)
19. Li, H., Oren, N., Norman, T.J.: Probabilistic argumentation frameworks. In: Proceedings of TAFA'11. LNCS, vol. 7132, pp. 1–16. Springer (2011)
20. Polberg, S., Doder, D.: Probabilistic abstract dialectical frameworks. In: Proceedings of JELIA'14. LNCS, vol. 8761, pp. 591–599. Springer (2014)
21. Riveret, R., Rotolo, A., Sartor, G., Prakken, H., Roth, B.: Success chances in argument games: a probabilistic approach to legal disputes. In: Proceedings of JURIX'07. pp. 99–108. IOS Press (2007)
22. Thimm, M.: A probabilistic semantics for abstract argumentation. In: Proceedings of ECAI'12. FAIA, vol. 242, pp. 750–755. IOS Press (2012)
23. Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: Explaining bayesian networks using argumentation. In: Proceedings of ECSQARU'15. LNCS, vol. 9161, pp. 83–92. Springer (2015)
24. Villata, S., Boella, G., van der Torre, L.: Attack semantics for abstract argumentation. In: Proceedings of IJCAI'11. pp. 406–413. AAAI Press (2011)