

Encompassing Attacks to Attacks in Abstract Argumentation Frameworks

Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida¹

Dipartimento di Elettronica per l'Automazione, Università di Brescia, Via Branze 38,
I-25123 Brescia, Italy

Abstract. In the traditional definition of Dung's abstract argumentation framework (AF), the notion of attack is understood as a relation between arguments, thus bounding attacks to start from and be directed to arguments. This paper introduces a generalized definition of abstract argumentation framework called $AFRA$ (Argumentation Framework with Recursive Attacks), where an attack is allowed to be directed towards another attack. From a conceptual point of view, we claim that this generalization supports a straightforward representation of reasoning situations which are not easily accommodated within the traditional framework. From the technical side, we first investigate the extension to the generalized framework of the basic notions of conflict-free set, acceptable argument, admissible set and of Dung's fundamental lemma. Then we propose a correspondence from the $AFRA$ to the AF formalism, showing that it satisfies some basic desirable properties. Finally we analyze the relationships between $AFRA$ and a similar extension of Dung's abstract argumentation framework, called $EAF+$ and derived from the recently proposed formalism EAF .

1 Introduction

An argumentation framework (AF in the following), as introduced in a seminal paper by Dung [1], is an abstract entity consisting of a set of elements, called *arguments*, whose origin, nature and possible internal structure is not specified and by a binary relation of *attack* on the set of arguments, whose meaning is not specified either. This abstract formalism has been shown to be able to encompass a large variety of more specific formalisms in areas ranging from nonmonotonic reasoning to logic programming and game theory, and, as such, is widely regarded as a powerful tool for theoretical analysis. Several variations of the original AF formalism have been proposed in the literature. On one hand, some approaches introduce new elements in the basic scheme in order to encompass explicitly some additional conceptual notions, useful for a “natural” representation of some reasoning situations. This is the case for instance of value-based argumentation frameworks [2], where a notion of value is associated to arguments, and of bipolar argumentation frameworks [3], where a relation of support between arguments is considered besides the one of attack. On the other hand, one may investigate generalized versions of the original AF definition (in particular, of the notion

of attack) while not introducing additional concepts within the basic scheme, as in [4–6]. This paper lies in the latter line of investigation and pursues the goal of generalizing the AF notion of attack by allowing an attack, starting from an argument, to be directed not just towards an argument but also towards any other attack. This will be achieved by a recursive definition of the attack relation leading to the introduction and preliminary investigation of a formalism called $AFRA$ (Argumentation Framework with Recursive Attacks).

The paper is organised as follows. In Section 2 motivations for a recursive notion of attack will be discussed, leading in Section 3 to the formal definition of $AFRA$ and of the necessary companion notions. Section 4 proposes a translation procedure from $AFRA$ to AF , able to ensure a full correspondence between the notions of conflict-free set, acceptable argument and admissible set. In section 5 we compare $AFRA$ with an alternative way to encompass attacks to attacks, called $EAF+$ (in turn based on the EAF formalism, proposed in [4–6]). Section 6 concludes the paper.

2 Background and Motivations

In Dung’s theory an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ is a pair where \mathcal{A} is a set of arguments (whatever this may mean) and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ a binary relation on it. The terse intuition behind this formalism is that arguments may attack each other and useful formal definitions and theoretical investigations may be built on this simple basis. In particular, the related fundamental notions of conflict-free set, acceptable argument and admissible set are recalled in Definition 1.

Definition 1. *Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$:*

- a set $\mathcal{S} \subseteq \mathcal{A}$ is conflict-free if $\nexists A, B \in \mathcal{S}$ s.t. $(A, B) \in \mathcal{R}$;
- an argument $A \in \mathcal{A}$ is acceptable with respect to a set $\mathcal{S} \subseteq \mathcal{A}$ if $\forall B \in \mathcal{A}$ s.t. $(B, A) \in \mathcal{R}$, $\exists C \in \mathcal{S}$ s.t. $(C, B) \in \mathcal{R}$;
- a set $\mathcal{S} \subseteq \mathcal{A}$ is admissible if \mathcal{S} is conflict-free and every element of \mathcal{S} is acceptable with respect to \mathcal{S} .

The notions recalled in Definition 1 lie at the heart of the definitions of Dung’s argumentation semantics, a topic which is only marginally covered by this paper. For our purposes it is sufficient to recall that an argumentation semantics identifies for an argumentation framework, a set of *extensions*, namely sets of arguments which are “collectively acceptable”, or, in other words, are able to survive together the conflict represented by the attack relation.

Even from this quick review it emerges that the main role of the notion of attack in Dung’s theory is supporting the identification of “surviving” arguments, on which the definition of extensions is exclusively focused. On this basis, one might say that attacks are necessary but accessory in this theory.

At a merely abstract level, one might then envisage an alternative approach where attacks are ascribed an extended (in a sense, empowered) role. A simple way to achieve this is allowing an attack to be directed also towards another

attack. From a general point of view, this amounts to conceive an attack as an entity able to affect any other entity (be it an argument or an attack) rather than just a by-product of how arguments relate each other. As a further consequence, this opens the way to include attacks as first-class elements in the definitions of all the basic notions we have seen before, from conflict-free sets to extensions.

However, before proceeding with what could be regarded as a sort of technical exercise, one might wonder whether there are practical motivations and concrete intuitions backing this kind of investigation. We provide an affirmative answer by means of an example in the area of modeling decision processes. Suppose Bob is deciding about his Christmas holidays and, as a general rule of thumb, he always buys cheap last minute offers. Suppose two such offers are available, one for a week in Gstaad, another for a week in Cuba. Then, using his behavioral rule, Bob can build two arguments, one, let say G , whose premise is “There is a last minute offer for Gstaad” and whose conclusion is “I should go to Gstaad”, the other, let say C , whose premise is “There is a last minute offer for Cuba” and whose conclusion is “I should go to Cuba” (note that if more last minute offers were available, more arguments of the same kind would be constructed). As the two choices are incompatible, G and C attack each other, a situation giving rise to an undetermined choice. Suppose however that Bob has a preference P for skiing and knows that Gstaad is a ski resort, how can we represent this fact?

P might be represented implicitly by suppressing the attack from C to G , but this is unsatisfactory, since it would prevent, in particular, further reasoning on P , as described below. So let us consider P as an argument whose premise is “Bob likes skiing” and whose conclusion is “When it is possible, Bob prefers to go to a ski resort”. P might attack C , but this does not seem sound since P is not actually in contrast with the existence of a good last minute offer for Cuba and the fact that, according to Bob’s general behavioral rule, this gives him a reason for going to Cuba. Thus, it seems more reasonable to represent P as attacking the attack from C to G , causing G to prevail. Note that the attack from C to G is not suppressed, but only made ineffective, in the specific situation at hand, due to the attack of P .

Assume now that Bob learns that there have been no snowfalls in Gstaad since one month and from this fact he derives that it might not be possible to ski in Gstaad. This argument (N), whose premise is “The weather report informs that in Gstaad there were no snowfalls since one month” and whose conclusion is “It is not possible to ski in Gstaad”, does not affect neither the existence of last minute offers for Gstaad nor Bob’s general preference for ski, rather it affects the ability of this preference to affect the choice between Gstaad and Cuba. Thus argument N attacks the attack originated from P .

Suppose finally that Bob is informed that in Gstaad it is anyway possible to ski, thanks to a good amount of artificial snow. This allows to build an argument, let say A , which attacks N , thus in turn reinstating the attack originated from P and intuitively supporting the choice of Gstaad. A graphical illustration of this example is provided in Figure 1.

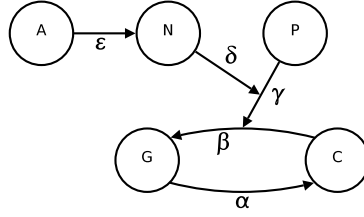


Fig. 1. Bob's last minute dilemma.

While the quick formalization adopted for this largely informal example is clearly not the only possible one and might even be questionable, we believe that the example anyway provides the kind of intuitive backing we sought for the notion of attacks towards attacks in abstract argumentation and the necessity of extending concepts like acceptability, admissibility and reinstatement to attacks too. A formal counterpart to this intuition will be provided in the next section.

3 AFRA: Argumentation Framework with Recursive Attacks

An Argumentation Framework with Recursive Attacks (*AFRA*) is defined, similarly to Dung's argumentation framework, as a pair consisting of a set of arguments and a set of attacks. Unlike the original definition, every attack is defined recursively as a pair where the first member is an argument and the second is another attack or an argument (base case).

Definition 2 (AFRA). *An Argumentation Framework with Recursive Attacks (AFRA) is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$ where:*

- \mathcal{A} is a set of arguments;
- \mathcal{R} is a set of attacks, namely pairs (A, \mathcal{X}) s.t. $A \in \mathcal{A}$ and $(\mathcal{X} \in \mathcal{R}$ or $\mathcal{X} \in \mathcal{A})$.

Given an attack $\alpha = (A, \mathcal{X}) \in \mathcal{R}$, we will say that A is the source of α , denoted as $\text{src}(\alpha) = A$ and \mathcal{X} is the target of α , denoted as $\text{trg}(\alpha) = \mathcal{X}$.

We start substantiating the role played by attacks by introducing a notion of defeat which regards attacks, rather than their source arguments, as the subjects able to defeat arguments or other attacks, as encompassed by Definition 3.

Definition 3 (Direct Defeat). *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{V} \in \mathcal{R}$, $\mathcal{W} \in \mathcal{A} \cup \mathcal{R}$, then \mathcal{V} directly defeats \mathcal{W} iff $\mathcal{W} = \text{trg}(\mathcal{V})$.*

Moreover, as we are interested also in how attacks are affected by other attacks, we introduce a notion of indirect defeat for an attack, corresponding to the situation where its source receives a direct defeat.

Definition 4 (Indirect Defeat). Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{V} \in \mathcal{R}$, $\mathcal{W} \in \mathcal{A}$, if \mathcal{V} directly defeats \mathcal{W} then $\forall \alpha \in \mathcal{R}$ s.t. $\text{src}(\alpha) = \mathcal{W}$, \mathcal{V} indirectly defeats α .

Therefore an element \mathcal{V} of a AFRA defeats another element \mathcal{W} if there is a direct or an indirect defeat from \mathcal{V} to \mathcal{W} .

Definition 5 (Defeat). Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{V} \in \mathcal{R}$, $\mathcal{W} \in \mathcal{A} \cup \mathcal{R}$, then \mathcal{V} defeats \mathcal{W} , denoted as $\mathcal{V} \rightarrow_R \mathcal{W}$, iff \mathcal{V} directly or indirectly defeats \mathcal{W} .

To exemplify, the case of Fig. 1 can be represented by an AFRA $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A} = \{C, G, P, N, A\}$ and $\mathcal{R} = \{\alpha, \beta, \gamma, \delta, \epsilon\}$ with $\alpha = (G, C)$, $\beta = (C, G)$, $\gamma = (P, \beta)$, $\delta = (N, \gamma)$, $\epsilon = (A, N)$. There are five direct defeats, namely $\epsilon \rightarrow_R N$, $\delta \rightarrow_R \gamma$, $\gamma \rightarrow_R \beta$, $\beta \rightarrow_R G$, $\alpha \rightarrow_R C$, and three indirect defeats: $\epsilon \rightarrow_R \delta$, $\beta \rightarrow_R \alpha$, $\alpha \rightarrow_R \beta$.

The definition of conflict-free set follows directly.

Definition 6 (Conflict-free). Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ is conflict-free iff $\nexists \mathcal{V}, \mathcal{W} \in \mathcal{S}$ s.t. $\mathcal{V} \rightarrow_R \mathcal{W}$.

Note that, while the definition of conflict-free set for AFRA is formally almost identical to the corresponding one in AF, actually they feature substantial differences, related to the underlying notion of defeat. In fact in AFRA every set of arguments $\mathcal{S} \subseteq \mathcal{A}$ is conflict-free, since only the explicit consideration of attacks gives rise to conflict in this approach. For instance if $A, B \in \mathcal{A}$ and $\alpha = (A, B) \in \mathcal{R}$, the set $\{A, B\}$ is conflict-free, while the set $\{A, B, \alpha\}$ is not.

Continuing the above example, the maximal, w.r.t. inclusion, conflict free sets of Γ are $\{C, P, A, \epsilon, \beta\}$, $\{C, G, P, N, A, \delta\}$, $\{C, G, P, N, A, \gamma\}$, $\{C, G, A, P, \epsilon, \gamma\}$, $\{C, P, N, A, \delta, \beta\}$, $\{G, P, N, A, \delta, \alpha\}$, $\{G, P, N, A, \alpha, \gamma\}$, $\{G, A, P, \epsilon, \gamma, \alpha\}$.

Also the definition of acceptability is formally very similar to the traditional one, but now it is applied to both arguments and attacks.

Definition 7 (Acceptability). Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$, $\mathcal{W} \in \mathcal{A} \cup \mathcal{R}$, \mathcal{W} is acceptable w.r.t. \mathcal{S} iff $\forall \mathcal{Z} \in \mathcal{R}$ s.t. $\mathcal{Z} \rightarrow_R \mathcal{W} \exists \mathcal{V} \in \mathcal{S}$ s.t. $\mathcal{V} \rightarrow_R \mathcal{Z}$.

On this basis, the definition of admissibility is formally identical to the one proposed by Dung. As a consequence, it is also possible to directly introduce the notion of *preferred extension* (which is simply a maximal admissible set) lying at the basis of Dung's well-known *preferred semantics* [1].

Definition 8 (Admissibility). Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ is admissible iff it is conflict-free and each element of \mathcal{S} is acceptable w.r.t. \mathcal{S} .

Definition 9 (Preferred extension). A preferred extension of an AFRA is a maximal (w.r.t. set inclusion) admissible set.

Referring again to the example of Fig. 1, the only preferred extension of Γ is $\{A, P, G, \epsilon, \gamma, \alpha\}$.

It is also possible to prove that a straightforward transposition of Dung's fundamental lemma holds in the context of AFRA.

Lemma 1 (Fundamental lemma). *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ an admissible set and $\mathcal{A}, \mathcal{A}' \in \mathcal{A} \cup \mathcal{R}$ elements acceptable w.r.t. \mathcal{S} . Then:*

1. $\mathcal{S}' = \mathcal{S} \cup \{\mathcal{A}'\}$ is admissible; and
2. \mathcal{A}' is acceptable w.r.t. \mathcal{S}' .

Proof.

1. \mathcal{A} is acceptable w.r.t. \mathcal{S} so each element of \mathcal{S}' is acceptable w.r.t. \mathcal{S}' . Suppose \mathcal{S}' is not conflict-free; therefore there exists an element $\mathcal{B} \in \mathcal{S}$ such that either $\mathcal{A} \rightarrow_R \mathcal{B}$ or $\mathcal{B} \rightarrow_R \mathcal{A}$. From the admissibility of \mathcal{S} and the acceptability of \mathcal{A} there exists an element $\tilde{\mathcal{B}} \in \mathcal{S}$ such that $\tilde{\mathcal{B}} \rightarrow_R \mathcal{B}$ or $\tilde{\mathcal{B}} \rightarrow_R \mathcal{A}$. Since \mathcal{S} is conflict-free it follows that $\tilde{\mathcal{B}} \rightarrow_R \mathcal{A}$. But then there must exist an element $\hat{\mathcal{B}} \in \mathcal{S}$ such that $\hat{\mathcal{B}} \rightarrow_R \tilde{\mathcal{B}}$. Contradiction.
2. Obvious.

4 A correspondence between AFRA and AF

We consider now the issue of expressing an AFRA in terms of a traditional AF and drawing the relevant correspondences between the basic notions introduced in Section 3. This kind of correspondence provides a very useful basis for further investigations as it allows one to reuse or adapt, in the context of AFRA, the many theoretical results available in Dung's framework.

Definition 10. *Let $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$ be an AFRA, the corresponding AF $\tilde{\Gamma} = \langle \tilde{\mathcal{A}}, \tilde{\mathcal{R}} \rangle$ is defined as follows:*

- $\tilde{\mathcal{A}} = \mathcal{A} \cup \mathcal{R}$;
- $\tilde{\mathcal{R}} = \{(\mathcal{V}, \mathcal{W}) \mid \mathcal{V}, \mathcal{W} \in \mathcal{A} \cup \mathcal{R} \text{ and } \mathcal{V} \rightarrow_R \mathcal{W}\}$.

In words both arguments and attacks of the original AFRA Γ become arguments of its corresponding AF version $\tilde{\Gamma}$ while the defeat relations in AF correspond to all direct and indirect defeats in the original AFRA. We can now examine the relationships between the relevant basic notions in Γ and $\tilde{\Gamma}$.

Lemma 2. *Let $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$ an AFRA and $\tilde{\Gamma} = \langle \tilde{\mathcal{A}}, \tilde{\mathcal{R}} \rangle$ its corresponding AF:*

1. \mathcal{S} is a conflict-free set for Γ iff \mathcal{S} is a conflict-free set for $\tilde{\Gamma}$;
2. \mathcal{A} is acceptable w.r.t. $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ in Γ iff \mathcal{A} is acceptable w.r.t. \mathcal{S} in $\tilde{\Gamma}$;
3. \mathcal{S} is an admissible set for Γ iff \mathcal{S} is an admissible set for $\tilde{\Gamma}$.

Proof.

1. The conclusion follows directly from Def. 10.
2. *Right to left half:* Let $\mathcal{A} \in \mathcal{A} \cup \mathcal{R}$ be acceptable w.r.t. $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ in Γ and suppose \mathcal{A} is not acceptable w.r.t. \mathcal{S} of $\tilde{\Gamma}$. So, there exists $\mathcal{B} \in \tilde{\mathcal{A}} = \mathcal{A} \cup \mathcal{R}$ s.t. $(\mathcal{B}, \mathcal{A}) \in \tilde{\mathcal{R}}$ and $\nexists \mathcal{C} \in \mathcal{S}$ s.t. $(\mathcal{C}, \mathcal{B}) \in \tilde{\mathcal{R}}$. From Def. 10, $(\mathcal{B}, \mathcal{A}) \in \tilde{\mathcal{R}}$ iff $\mathcal{B} \rightarrow_R \mathcal{A}$ and $(\mathcal{C}, \mathcal{B}) \in \tilde{\mathcal{R}}$ iff $\mathcal{C} \rightarrow_R \mathcal{B}$. Then $\exists \mathcal{A} \in \mathcal{S}$, $\exists \mathcal{B} \in \mathcal{A} \cup \mathcal{R}$ s.t. $\mathcal{B} \rightarrow_R \mathcal{A}$ and $\nexists \mathcal{C} \in \mathcal{S}$ s.t. $\mathcal{C} \rightarrow_R \mathcal{B}$. Therefore \mathcal{A} is not acceptable w.r.t. \mathcal{S} in Γ . Contradiction.

Left to right half: Follows the same reasoning line with obvious modifications.

3. Follows directly from 1 and 2.

5 Comparing *AFRA* with *EAF+*

AFRA provides a formally quite simple way to generalize the notion of attack of Dung’s framework, such simplicity partially hiding some substantial underlying differences, concerning in particular the notion of conflict-free set. A more restricted, but similar in spirit, generalization of Dung’s framework has recently been proposed in [4–6], with the name of Extended Argumentation Framework (*EAF*). This approach is motivated by the need to express preferences between arguments and supports a very interesting form of meta-level argumentation about the values that arguments promote [2]. In *EAF* a limited notion of attacks to attacks is encompassed: only attacks whose target is an argument (i.e. the “traditional” ones) can be attacked, while attacks whose target is another attack can not be in turn attacked. Referring to Figure 1, only the attack originated from P could be represented, while the one originated from N could not. On the other hand, the notion of conflict-free set introduced in [5] for *EAF* is somehow closer to Dung’s original one. To compare this kind of approach with *AFRA* we investigate in this section an extension of *EAF* (called *EAF+*) which allows for recursive attacks, while attempting to follow as close as possible the original *EAF* definitions provided in [5].

The definition of *EAF+* keeps the original elements of Dung’s definition, adding, as a separate entity, a relation of attack between attacks.

Definition 11 (*EAF+*). *An EAF Plus ($EAF+$) is a tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ s.t.:*

1. \mathcal{A} is a set of arguments;
2. $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$;
3. $\mathcal{D}+$ is a set of pairs (A, δ) s.t. $A \in \mathcal{A}$ and $(\delta \in \mathcal{R}$ or $\delta \in \mathcal{D}+)$.

We extend in the obvious way the notions of source and target of an attack to the pairs both in \mathcal{R} and in $\mathcal{D}+$. The notion of defeat for *EAF+* turns out to be articulated into four cases as it has to encompass the roles of both arguments and attacks.

Definition 12 (Defeat). *Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, $\mathcal{V}, \mathcal{W} \in \mathcal{A} \cup \mathcal{E}$, \mathcal{V} defeats \mathcal{W} (denoted $\mathcal{V} \rightarrow_E \mathcal{W}$) if any of the following conditions holds:*

1. $\mathcal{V}, \mathcal{W} \in \mathcal{A}, (\mathcal{V}, \mathcal{W}) \in \mathcal{R}$;
2. $\mathcal{V} \in \mathcal{A}, \mathcal{W} \in \mathcal{E}$ and $(\mathcal{V}, \mathcal{W}) \in \mathcal{E}$;
3. $\mathcal{V}, \mathcal{W} \in \mathcal{E}$ and $\mathcal{W} = \text{trg}(\mathcal{V})$;
4. $\mathcal{V} \in \mathcal{E}, \mathcal{W} \in \mathcal{A}$ and $\mathcal{W} = \text{trg}(\mathcal{V})$.

Keeping again distinct the treatment of arguments and attacks, the property of being conflict-free has to be introduced for pairs consisting of a set of arguments and a set of attacks. Moreover we need to constrain the sets of arguments and attacks in the pair, by restricting our attention on *self-contained* pairs, where the presence of an attack implies also the presence of its source.

Definition 13 (Self-contained pair). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, $\mathcal{S} \subseteq \mathcal{A}$, $\mathcal{T} \subseteq \mathcal{E}$, a pair $\mathcal{W} = (\mathcal{S}, \mathcal{T})$ is self-contained iff $\forall \alpha \in \mathcal{T}$, $src(\alpha) \in \mathcal{S}$. We will denote a self-contained pair as $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$.

Definition 14 (Conflict-free). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, a self-contained pair $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ is conflict-free iff $\forall A, B \in \mathcal{S}$ s.t. $A \rightarrow_E B$ (so there exists $\alpha = (A, B) \in \mathcal{R}$), $\exists \beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \alpha$ and $\nexists \gamma \in \mathcal{T}$, $\nexists \mathcal{D} \in \mathcal{S} \cup \mathcal{T}$ s.t. $\mathcal{D} = trg(\gamma)$.

In words, any attack between the arguments in \mathcal{S} has to be attacked in turn by \mathcal{T} (and thus made ineffective), while attacks in \mathcal{T} directed against any other element in the pair are simply not allowed. Note that, differently from the definition provided for AF , there may be a conflict even in absence of attacks in \mathcal{T} , so a pair $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \emptyset)}$ may not be conflict-free.

In $EAF+$, like in Dung's AF , an element (argument or attack) is acceptable with respect to a self-contained pair when it is defended against any attack it receives. Defense may consist of a defeat against the attack or against its source.

Definition 15 (Acceptability). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, and $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ a self-contained pair :

- $A \in \mathcal{A}$ is acceptable w.r.t. $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ iff $\forall \beta \in \mathcal{R}$ s.t. $A = trg(\beta)$, $\exists \alpha \in \mathcal{T}$ s.t. $\alpha \rightarrow_E src(\beta)$ or $\alpha \rightarrow_E \beta$;
- $\alpha \in \mathcal{E}$ is acceptable w.r.t. $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ iff $src(\alpha)$ is acceptable w.r.t. $\widehat{\mathcal{W}}$ and $\forall \beta \in \mathcal{D}+$ s.t. $\beta \rightarrow_E \alpha$, $\exists \gamma \in \mathcal{T}$ s.t. $\gamma \rightarrow_E src(\beta)$ or $\gamma \rightarrow_E \beta$.

Following [1, 5], we consider a self-contained pair admissible if and only if it is conflict-free and any of its elements is acceptable with respect to it.

Definition 16 (Admissibility). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, a self-contained pair $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ is admissible iff it is conflict-free, $\forall A \in \mathcal{S}$, A is acceptable w.r.t. $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ and $\forall \alpha \in \mathcal{T}$, α is acceptable w.r.t. $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$.

In order to introduce the notion of preferred extension for $EAF+$ we have to define an inclusion relation between self-contained pairs.

Definition 17 (Inclusion of self-contained pair). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, $\mathcal{S} \subseteq \mathcal{A}$, $\mathcal{T} \subseteq \mathcal{E}$, $\widehat{\mathcal{W}}' = \widehat{(\mathcal{S}', \mathcal{T}')}$ is included in $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ iff $\mathcal{S}' \subseteq \mathcal{S}$ and $\mathcal{T}' \subseteq \mathcal{T}$.

Definition 18 (Preferred extension). A preferred extension of an $EAF+$ is a maximal (according to the inclusion relation introduced in Definition 17) admissible self-contained pair of $EAF+$.

It is also possible to prove Dung's fundamental Lemma for the case of $EAF+$.

Lemma 3 (Fundamental lemma). Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an $EAF+$, $\mathcal{E} = \mathcal{R} \cup \mathcal{D}+$, $\widehat{\mathcal{W}} = \widehat{(\mathcal{S}, \mathcal{T})}$ an admissible self-contained pair and $\mathcal{A}, \mathcal{A}' \in \mathcal{A} \cup \mathcal{E}$ acceptable w.r.t. $\widehat{\mathcal{W}}$. Then:

1. $\widehat{\mathcal{W}}' = \begin{cases} (\widehat{\mathcal{S}'}, \widehat{\mathcal{T}}), \mathcal{S}' = \mathcal{S} \cup \{\mathcal{A}\} & \text{if } \mathcal{A} \in \mathcal{A} \\ (\widehat{\mathcal{S}}, \widehat{\mathcal{T}'}) , \mathcal{T}' = \mathcal{T} \cup \{\mathcal{A}\} & \text{if } \mathcal{A} \in \mathcal{E} \end{cases}$ is admissible; and
2. \mathcal{A}' is acceptable w.r.t. $\widehat{\mathcal{W}}'$.

Proof. To prove this lemma, we have to consider four different cases:

- A. $\mathcal{A}, \mathcal{A}' \in \mathcal{A}$;
- B. $\mathcal{A} \in \mathcal{A}, \mathcal{A}' \in \mathcal{E}$;
- C. $\mathcal{A} \in \mathcal{E}, \mathcal{A}' \in \mathcal{A}$;
- D. $\mathcal{A}, \mathcal{A}' \in \mathcal{E}$.

Case A.

1. \mathcal{A} is acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ therefore in $\widehat{\mathcal{W}}' = (\widehat{\mathcal{S}'}, \widehat{\mathcal{T}})$, with $\mathcal{S}' = \mathcal{S} \cup \{\mathcal{A}\}$, $\forall A \in \mathcal{S}'$, A is acceptable w.r.t. $\widehat{\mathcal{W}}'$ and $\forall \alpha \in \mathcal{T}$, α is acceptable w.r.t. $\widehat{\mathcal{W}}'$. Suppose $\widehat{\mathcal{W}}'$ is not conflict-free. There are two possible cases: (a) $\exists A, B \in \mathcal{S}'$ s.t. $\alpha = (A, B) \in \mathcal{R}$ and $\nexists \beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \alpha$; (b) $\exists \alpha \in \mathcal{T}, \exists \mathcal{D} \in \mathcal{S}' \cup \mathcal{T}$ s.t. $\mathcal{D} = \text{trg}(\alpha)$.

Considering case (b), from the admissibility of $\widehat{\mathcal{W}}$ we have $\mathcal{D} = \mathcal{A} = \text{trg}(\alpha)$; from the acceptability of \mathcal{A} w.r.t. $\widehat{\mathcal{W}}$ and Def. 15 two cases are possible but $\alpha \in \mathcal{T}$, so both cases imply that $\widehat{\mathcal{W}}$ is not conflict-free: contradiction.

Considering case (a), from the admissibility of $\widehat{\mathcal{W}}$, $A = \mathcal{A}$ or $B = \mathcal{A}$.

Suppose $A = \mathcal{A}$, i.e. $\alpha = \mathcal{A} \rightarrow_E B$. Since B is acceptable w.r.t. $\widehat{\mathcal{W}}$ from Def. 15 two cases are in turn possible: (i) $\exists \beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \mathcal{A}$ and $\text{src}(\beta) \in \mathcal{S}$; (ii) $\exists \beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \alpha$.

In case (i) from the acceptability of \mathcal{A} with respect to $\widehat{\mathcal{W}}$ we would have that β or its source (both belonging to $\widehat{\mathcal{W}}$) should be defeated by an element of $\widehat{\mathcal{W}}$ thus contradicting the fact that $\widehat{\mathcal{W}}$ is conflict free. In case (ii) we contradict the assumption (a).

Suppose now $B = \mathcal{A}$; so $\text{trg}(\alpha) = \mathcal{A}$. From the acceptability of \mathcal{A} w.r.t. $\widehat{\mathcal{W}}$ there exists $\beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \alpha$ or $\beta \rightarrow_E \text{src}(\alpha)$, but again in both cases we contradict the fact that $\widehat{\mathcal{W}}$ is conflict free.

2. Obvious.

Case B.

1. Same as Case A, item 1.
2. Suppose \mathcal{A}' is acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ but not w.r.t. $\widehat{\mathcal{W}}' = (\widehat{\mathcal{S}'}, \widehat{\mathcal{T}})$; then (i) $\text{src}(\mathcal{A}')$ is not acceptable w.r.t. $\widehat{\mathcal{W}}'$ or (ii) $\exists \beta \in \mathcal{D}^+$ s.t. $\beta \rightarrow_E \mathcal{A}'$ and $\nexists \gamma \in \mathcal{T}$ s.t. $\gamma \rightarrow_E \text{src}(\beta)$ or $\gamma \rightarrow_E \beta$. In case (i), noting that $\text{src}(\mathcal{A}')$ is acceptable w.r.t. $\widehat{\mathcal{W}}$, we contradict what we have proved in Case A, item 1. Case (ii) contradicts the fact that \mathcal{A}' is acceptable w.r.t. $\widehat{\mathcal{W}}$.

Case C.

1. \mathcal{A} is acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ so in $\widehat{\mathcal{W}}' = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}'})$, with $\mathcal{T}' = \mathcal{T} \cup \{\mathcal{A}\}$, $\forall A \in \mathcal{S}$, A is acceptable w.r.t. $\widehat{\mathcal{W}}'$ and $\forall \alpha \in \mathcal{T}'$, α is acceptable w.r.t. $\widehat{\mathcal{W}}'$. Suppose $\widehat{\mathcal{W}}'$ is not conflict-free. There are two possible cases: (a) $\exists A, B \in \mathcal{S}$

s.t. $\alpha = (A, B) \in \mathcal{R}$ and $\nexists \beta \in \mathcal{T}'$ s.t. $\beta \rightarrow_E \alpha$; (b) $\exists \alpha \in \mathcal{T}', \exists \mathcal{D} \in \mathcal{S} \cup \mathcal{T}'$ s.t. $\mathcal{D} = \text{trg}(\alpha)$.

Case (a) is impossible because $\mathcal{T}' \supseteq \mathcal{T}$ and $\widehat{\mathcal{W}}$ is admissible.

Let us consider case (b): from the admissibility of $\widehat{\mathcal{W}}$, $\alpha = \mathcal{A}$ or $\mathcal{D} = \mathcal{A}$.

Suppose $\alpha = \mathcal{A}$; so $\mathcal{A} \rightarrow_E \mathcal{D}$ and $\text{src}(\mathcal{A})$ is acceptable w.r.t. $\widehat{\mathcal{W}}$ (because \mathcal{A} is acceptable w.r.t. $\widehat{\mathcal{W}}$). From the acceptability of \mathcal{D} w.r.t. $\widehat{\mathcal{W}}$, there exists $\gamma \in \mathcal{T}$, $\text{src}(\gamma) \in \mathcal{S}$ s.t. (i) $\gamma \rightarrow_E \text{src}(\mathcal{A})$ or (ii) $\gamma \rightarrow_E \mathcal{A}$. In both cases, from the acceptability of $\text{src}(\mathcal{A})$ w.r.t. $\widehat{\mathcal{W}}$, there exists $\delta \in \mathcal{T}$ s.t. $\delta \rightarrow_E \text{src}(\gamma)$ or $\delta \rightarrow_E \gamma$. Either case implies that $\widehat{\mathcal{W}}$ is not conflict-free: contradiction.

Suppose $\mathcal{D} = \mathcal{A}$; so $\alpha \rightarrow_E \mathcal{A}$. Using the acceptability of \mathcal{A} w.r.t. $\widehat{\mathcal{W}}$, we can then apply the same reasoning as above leading to contradict the fact that $\widehat{\mathcal{W}}$ is conflict-free.

2. Obvious.

Case D.

1. Same of Case C, item 1.
2. Same of Case B, item 2.

Let us now analyse the relation between *EAF+* and *AFRA*: first it is possible to draw a direct correspondence between the two formalisms.

Definition 19 (AFRA–EAF+ correspondence). For any *EAF+* $\Delta = \langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ we define the corresponding *AFRA* as $\Delta_R = \langle \mathcal{A}, \mathcal{R} \cup \mathcal{D}+ \rangle$. For any *AFRA* $\Gamma = \langle \mathcal{A}, \mathcal{R}' \rangle$ we define the corresponding *EAF+* as $\Gamma_E = \langle \mathcal{A}, \mathcal{R}' \cap (\mathcal{A} \times \mathcal{A}), \mathcal{R}' \setminus (\mathcal{A} \times \mathcal{A}) \rangle$.

The correspondence between the notions of defeat is drawn in Lemma 4.

Lemma 4. Let $\Delta = \langle \mathcal{A}, \mathcal{R}, \mathcal{D}+ \rangle$ be an *EAF+* and Δ_R its corresponding *AFRA*. For any $\mathcal{V}, \mathcal{W} \in (\mathcal{A} \cup \mathcal{R} \cup \mathcal{D}+)$, $\mathcal{V} \rightarrow_E \mathcal{W}$ in Δ iff the disjunction of the following conditions holds:

- (a) \mathcal{W} is directly defeated by \mathcal{V} in Δ_R ;
- (b) $\exists \mathcal{Z} \in (\mathcal{R} \cup \mathcal{D}+)$ s.t. $\mathcal{V} = \text{src}(\mathcal{Z})$ and \mathcal{Z} directly defeats \mathcal{W} in Δ_R .

Proof. Consider the four cases of defeat for *EAF+* of Def. 12. The disjunction of cases 1 and 2 is equivalent to condition (b). In fact, in these cases $\exists \mathcal{Z} \in (\mathcal{R} \cup \mathcal{D}+)$ s.t. $\mathcal{V} = \text{src}(\mathcal{Z})$ and $\mathcal{W} = \text{trg}(\mathcal{Z})$. It follows that \mathcal{Z} directly defeats \mathcal{W} in Δ_R . Conversely, if $\exists \mathcal{Z} \in (\mathcal{R} \cup \mathcal{D}+)$ s.t. $\mathcal{V} = \text{src}(\mathcal{Z})$ and \mathcal{Z} directly defeats \mathcal{W} in Δ_R it follows that $\mathcal{W} = \text{trg}(\mathcal{Z})$ which implies that case 1 or 2 holds in Δ . The disjunction of cases 3 and 4 is equivalent to condition (a): the fact that either 3 or 4 implies (a) follows directly from Definition 3, the converse implication is immediate too.

From the proof of Lemma 4 we can observe that the four-cases definition of defeat in *EAF+* is somewhat redundant and might be summarized referring to the more synthetic formulation used in *AFRA*. We can now analyse the relationship involving the notions of conflict-free set, acceptability and admissibility: those defined for *EAF+* turn out to be a specialization of the ones in *AFRA*.

Lemma 5. Let $\Delta = \langle \mathcal{A}, \mathcal{R}, \mathcal{D} \rangle$ an *EAF+* and $\Gamma = \langle \mathcal{A}, \mathcal{R}' \rangle = \Delta_R$:

1. if $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ is conflict-free in Δ then $\mathcal{S} \cup \mathcal{T}$ is a conflict-free set for Γ ;
2. if \mathcal{V} is acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ in Δ then \mathcal{V} is acceptable w.r.t. $\mathcal{S} \cup \mathcal{T}$ in Γ ;
3. if $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ is admissible for Δ then $\mathcal{S} \cup \mathcal{T}$ is an admissible set for Γ .

Proof.

1. Let $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ a conflict-free pair for Δ . From Def. 14 we have: (a) $\forall A, B \in \mathcal{S}, \alpha = (A, B) \in \mathcal{R}, \exists \beta \in \mathcal{T}$ s.t. $\beta \rightarrow_E \alpha$; and (b) $\nexists \gamma \in \mathcal{T}, \nexists \mathcal{D} \in \mathcal{S} \cup \mathcal{T}$ s.t. $\mathcal{D} = \text{trg}(\gamma)$.
Let $\mathcal{U} = \mathcal{S} \cup \mathcal{T} \subseteq \mathcal{A} \cup \mathcal{R}'$. According to Def. 6 we have to show that $\nexists \mathcal{V}, \mathcal{W} \in \mathcal{U}$ s.t. $\mathcal{V} \rightarrow_R \mathcal{W}$ which in turn amounts to require: (i) $\nexists \mathcal{V} \in \mathcal{U} \cap \mathcal{R}', \nexists \mathcal{W} \in \mathcal{U}$ s.t. $\mathcal{W} = \text{trg}(\mathcal{V})$; (ii) $\nexists \mathcal{V}, \mathcal{Z} \in \mathcal{U} \cap \mathcal{R}'$, s.t. $\text{src}(\mathcal{Z}) = \text{trg}(\mathcal{V})$. From (b) it follows that $\nexists \mathcal{V} \in \mathcal{U} \cap \mathcal{R}', \nexists \mathcal{W} \in \mathcal{U}$ s.t. $\mathcal{W} = \text{trg}(\mathcal{V})$, i.e. condition (i). To prove (ii), assume that $\exists \mathcal{V}, \mathcal{Z} \in \mathcal{U} \cap \mathcal{R}'$, s.t. $\text{src}(\mathcal{Z}) = \text{trg}(\mathcal{V})$. To avoid contradiction with (b) we have to assume $\text{src}(\mathcal{Z}) \notin \mathcal{S}$, but this contradicts in turn the fact that $\widehat{\mathcal{W}}$ is self-contained.
2. Let \mathcal{V} be acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$. By Def. 7, we have to show that $\forall \mathcal{Z} \in \mathcal{R}'$ s.t. $\mathcal{Z} \rightarrow_R \mathcal{V}$ (i.e. (i) $\mathcal{V} = \text{trg}(\mathcal{Z})$ or, if $\mathcal{V} \in \mathcal{R}'$, (ii) $\text{src}(\mathcal{V}) = \text{trg}(\mathcal{Z})$), $\exists \mathcal{W} \in \mathcal{U} = \mathcal{S} \cup \mathcal{T}$ s.t. $\mathcal{W} \rightarrow_R \mathcal{Z}$ (i.e. (iii) $\mathcal{Z} = \text{trg}(\mathcal{W})$ or (iv) $\text{src}(\mathcal{Z}) = \text{trg}(\mathcal{W})$). Assume first (i) and $\mathcal{V} \in \mathcal{A}$: then (iii) or (iv) follows directly from the first part of Def. 15. Assuming (i) and $\mathcal{V} \in \mathcal{R}'$, according to the second part of Def. 15 we obtain again that (iii) or (iv) holds. Assuming (ii), we have $\mathcal{V} \in \mathcal{R}'$ and, by Def. 15, $\text{src}(\mathcal{V})$ must be acceptable w.r.t. $\widehat{\mathcal{W}} = (\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$ and (iii) or (iv) follows again from the first part of Def. 15.
3. Follows directly from 1 and 2.

6 Discussion and Conclusions

We have proposed a preliminary investigation about *AFRA*, a generalization of Dung’s argumentation framework where attacks to attacks are recursively encompassed without restriction. An intuitive justification for this kind of formalism has been provided in relation with the representation of decision processes.

The idea of encompassing attacks to attacks in abstract argumentation framework has been first considered in [7], in the context of an extended framework encompassing argument strengths and their propagation. In this quite different context, deserving further development, Dung style semantics issues have not been considered.

Focusing on approaches closer to “traditional” Dung’s framework, attacks to attacks have been considered in the context of reasoning about preferences [4–6] and reasoning about coalitions [8]. In both cases only attacks to attacks between arguments are covered, i.e. only one level of recursion is allowed. A detailed conceptual analysis motivating *EAF* with respect to a variety of reference

domains is provided in particular in [6]. While developing this kind of analysis for *AFRA* and the relevant comparison with *EAF* is beyond the scope of this paper, one can note that *EAF* adopts some specific assumptions, for instance a limited level of recursion and a constraint on some attacks to be symmetric, when the involved arguments represent conflicting preferences. These assumptions are fully justified in the context of reasoning about preferences but may not be necessary in general. The *AFRA* formalism, though originally conceived to support some intuitive forms of reasoning in the context of decision making, addresses the general need to reason about conflicts which may be themselves defeasible. In order to satisfy this need, it seems reasonable to provide attacks with an ontological status encompassing defeasibility, differently than in *AF*.

To complete the comparison from a more technical point of view, we have considered in Section 5 the *EAF+* formalism, namely a possible extension of *EAF* aimed at overcoming these restrictive assumptions, and we have drawn correspondences between *EAF+* and *AFRA*. It turns out that the *AFRA* formalism supports more compact (and in some cases also more general) definitions of the fundamental notions of defeat, conflict-free set, acceptability and admissibility. The “translation” from *AFRA* to Dung’s *AF* proposed in Section 4 opens the way to one of the main future work directions, namely enlarging the theoretical bases of *AFRA* and investigating the definition of argumentation semantics in this context, possibly exploiting the rich corpus of results available for the traditional framework.

References

1. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming, and n-person games. *Artificial Intelligence* **77**(2) (1995) 321–357
2. Bench-Capon, T.J.M.: Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation* **13**(3) (2003) 429–448
3. Amgoud, L., Cayrol, C., Lagasque-Schiex, M.C., Livet, P.: On bipolarity in argumentation frameworks. *Int. Journal of Intelligent Systems* **23** (2008) 1062–1093
4. Modgil, S.: An abstract theory of argumentation that accommodates defeasible reasoning about preferences. In: *Proc. 9th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, Hammamet, TN (2007) 648–659
5. Bench-Capon, T.J.M., Mogdil, S.: Integrating object and meta-level value based argumentation. In: *Proc. 2nd Int. Conf. on Computational Models of Argument (COMMA 2008)*, Toulouse, F (2008) 240–251
6. Modgil, S.: Reasoning about preferences in argumentation frameworks. *Artificial Intelligence* **In press** (2009)
7. Barringer, H., Gabbay, D.M., Woods, J.V.: Temporal dynamics of support and attack networks: From argumentation to zoology. In Hutter, D., Stephan, W., eds.: *Mechanizing Mathematical Reasoning*. Springer (2005) 59–98
8. Boella, G., van der Torre, L., Villata, S.: Social viewpoints for arguing about coalitions. In: *Proc. 11th Pacific Rim Int. Conf. on Multi-Agents (PRIMA’08)*. (2008) 66–77