

# Argumentation Frameworks with Attack Classification

Alexandros Vassiliades<sup>1,2</sup>      Giorgos Flouris<sup>2</sup>  
Theodore Patkos<sup>2</sup>      Antonis Bikakis<sup>3</sup>      Nick Bassiliades<sup>1</sup>  
Dimitris Plexousakis<sup>2</sup>

<sup>1</sup> School of Informatics, Aristotle University of Thessaloniki, Greece

<sup>2</sup> Institute of Computer Science, Foundation for Research and  
Technology - Hellas, Greece

<sup>3</sup> Department of Information Studies, University College London, UK

valexande@csd.auth.gr, fgeo@ics.forth.gr, patkos@ics.forth.gr,  
a.bikakis@ucl.ac.uk, nbassili@csd.auth.gr, dp@ics.forth.gr

## Abstract

Abstract Argumentation Frameworks (AAFs), introduced by Dung in [11], enabled a new way of reasoning with arguments, which does not take into account the internal structure of arguments, but only how they are related to each other. The only form of relation considered in AAFs is a binary attack relation on the set of arguments. From the definitions of acceptability semantics of AAFs, it is obvious that attacks actually have a dual role: on the one hand, they generate conflicts, on the other hand, they can defend other arguments from attacks. In this paper, we propose a framework, where the modeler can explicitly specify the role of each attack. For this purpose, we define a set of conflict-generating attacks  $\mathcal{R}_c$  and a set of defending attacks  $\mathcal{R}_d$ , as well as a family of semantics that considers the role of each attack while determining which arguments are attacked, which are defended, and which will be included in each extension. We study the formal properties of the proposed framework and semantics, show that our framework is a generalization of AAFs, and assess its semantics against a set of principles. Finally, we present a web application that provides an interface for creating custom argumentation frameworks and uses ASP to compute their extensions.

**Keywords**— Argumentation, Abstract Argumentation Framework, Attack Classification, Defending Attack, Conflict-Generating Attack

# 1 Introduction

Abstract Argumentation Frameworks (AAFs), introduced by Dung in [11], enabled a new way of reasoning with arguments, which does not take into account the internal structure of arguments, but only how they are related to each other. The simplicity and intuitiveness of AAFs, along with their ability to capture various types of non-monotonic reasoning, led to their wide adoption by the Knowledge Representation and Reasoning community. The only form of relation considered in AAFs is a binary attack relation on the set of arguments. From the definitions of acceptability semantics of AAFs, which provide a solid mechanism for selecting acceptable sets of arguments (called *extensions*), it is obvious that attacks actually have a dual role: on the one hand, they generate conflicts, i.e., they disallow two arguments that attack each other to be in the same extension; on the other hand, they can defend other arguments from attacks, thereby allowing arguments to be included in extensions, even if attacked, provided that all attacks are defended by other arguments in the extension.

Based on this observation, in previous work, we considered semantics where (some of the) attacks in the framework could be treated as having one of the two roles only. For example, consider the following exchange of arguments,  $a$ ,  $b$ , followed by one of  $c_1$ ,  $c_2$  and  $c_3$ .

$a$ : I will take antibiotic X for my dental infection. It was recommended by my dentist.

$b$ : X contains penicillin and you are allergic to it, so better take Y.

$c_1$ : I read in its label that X doesn't contain penicillin and X is more effective than Y.

$c_2$ : Y has serious and frequent side-effects, so I'd better avoid it.

$c_3$ : I was allergic to penicillin when I was a child, I think it has now faded away.

It is obvious that  $b$  attacks  $a$  and each of  $c_1$ ,  $c_2$  and  $c_3$  attacks  $b$ . However, the nature and effect of each of these attacks is different. If one accepts  $c_1$ , its attack to  $b$  should completely invalidate  $b$ , and, as a result,  $a$  should be accepted. This is captured by all acceptability semantics of AAFs that satisfy admissibility, i.e., they accept arguments that are not in conflict with any other accepted argument, and are defended by the accepted arguments against all attacks.

Now, let us consider  $c_2$ , which is also in conflict with  $b$  and, if accepted,  $b$  should be rejected. However,  $c_2$  does not defend  $a$  from  $b$ , in the sense that the information it conveys is not relevant to the attack from  $b$  to  $a$  (which relies on the claim that the person is allergic to penicillin and X contains penicillin). In other words, the attack from  $c_2$  to  $b$  has the single role of invalidating  $b$  (creating a conflict among  $b$  and  $c_2$ ).

Finally, the attack from  $c_3$  to  $b$  has a different effect. It does not invalidate  $b$ , but leaves some doubt about it. Given this attack, it would be reasonable to accept  $c_3$  and  $a$  and reject  $b$  (if one feels that the belief that the allergy has faded away

is strong enough to disregard the advice not to take the antibiotic), but it would also be reasonable to accept  $c_3$  and  $b$  and reject  $a$  (if the belief that the allergy has faded away is not strong enough to disregard the advice not to take the antibiotic).

The effect of the latter two types of attack (e.g., from  $c_2$  or  $c_3$  to  $b$ ) cannot be captured by any of the existing acceptability semantics of AAFs. To address this aim, we recently proposed a new abstract model of arguments, called multi-attack argumentation frameworks (MAAFs) [21], which allows each attack to be associated with a type, and we defined three types of semantics: *firm*, *restricted* and *loose*. In MAAF, a set of attack types is considered to have the “normal” behaviour, i.e., both conflict-generating and defending, whereas the rest are assumed to have only one (for firm or loose semantics), or neither of the two roles (for restricted semantics).

A shortcoming of MAAF is that conflict-generating and defending attacks cannot co-exist. In other words, a MAAF can only support two different “classes” of attacks: “normal” and “abnormal” ones, where the exact behaviour of “abnormal” attacks is determined by the type of semantics considered (firm, restricted or loose). In this paper, we generalize the idea of [21] by allowing the modeller to specify explicitly, and independently, what will be the role of each attack. For this purpose, we define a set of conflict-generating attacks  $\mathcal{R}_c$  and a set of defending attacks  $\mathcal{R}_d$ , as well as a family of semantics that takes into account the role of each attack while determining which arguments are attacked, which are defended, and which will be included in each extension.

This paper is based on and significantly extends our previous work presented in [21] by providing: (a) a more generalized framework that extends MAAF with the capability to explicitly specify the role of each attack; (b) a principle-based analysis of the semantics of this framework; (c) a formal study of the properties of the framework and its semantics, including its relation to AAFs; (d) a declarative implementation, in the context of a Web App<sup>1</sup>, where a user can develop their own instantiations of the argumentation framework and access the reasoning that our framework offers.

In the remainder of the paper, we define our new proposed framework, called AAFs with Attack Classification, along with some properties of the framework and its semantics (Section 3). We study the behaviour of the semantics with respect to a standard set of principles proposed for AAFs (Section 4). To better illustrate our approach, we present a use case on argumentation schemes (Section 5). Next, we describe an implementation of the framework based on the language of ASP (Section 6). Finally, we discuss the related work (Section 7) and conclude (Section 8). A detailed introduction to MAAF (as defined in [21]), proofs of formal properties of the new framework, as well as ASP encodings for the Web App are provided in the Appendix.

---

<sup>1</sup><http://139.91.183.45:8070/>

## 2 Abstract Argumentation Frameworks with Attack Classification

In this section, we define the notions of conflict-generating, normal, defending and irrelevant attacks, as well as the notion of classification of attacks for an AAF, and we define the standard extension classes for the new argumentation framework.

### 2.1 Framework

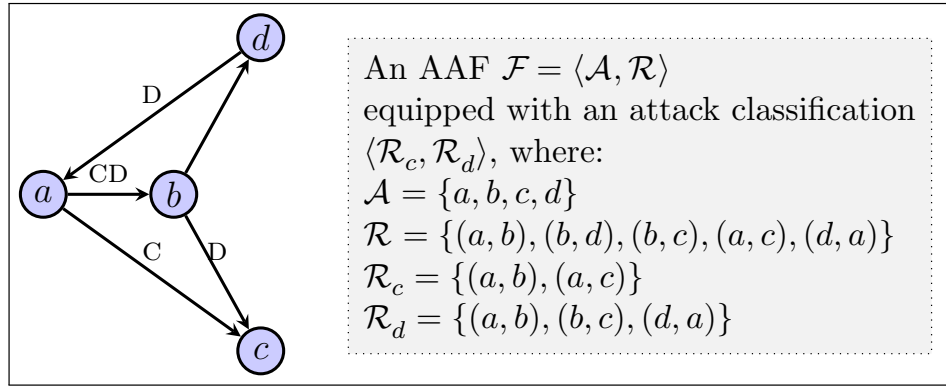


Figure 1: Running example

We recall that an argumentation framework is defined as a pair  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , where  $\mathcal{A}$  is the set of arguments and  $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$  is the set of attacks.

An *attack classification* over an AAF is a structure that determines which attacks should be treated as conflict-generating and which attacks should be treated as defending:

**Definition 1** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ . An attack classification over  $\mathcal{F}$  is a pair  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , such that  $\mathcal{R}_c \subseteq \mathcal{R}$ ,  $\mathcal{R}_d \subseteq \mathcal{R}$ .

An example of an AAF equipped with an attack classification is shown in Figure 1. Note that an attack classification essentially breaks down the attacks into four disjoint classes, or *types*, as follows:

- *Normal attacks*, i.e., attacks that behave in the classical manner. Such attacks are both conflict-generating and defending, and are the ones that belong in  $\mathcal{R}_c \cap \mathcal{R}_d$ . In our example,  $(a, b)$  is the only such attack.
- *Conflict-generating attacks*, i.e., attacks that can generate conflicts, but cannot defend against one. Such attacks are the ones that belong in  $\mathcal{R}_c \setminus \mathcal{R}_d$ . In our example,  $(a, c)$  is conflict-generating.

- *Defending attacks*, i.e., attacks that can defend, but do not generate conflicts. Such attacks are the ones that belong in  $\mathcal{R}_d \setminus \mathcal{R}_c$ . In our example,  $(b, c)$  and  $(d, a)$  are defending.
- *Irrelevant attacks*, i.e., attacks that are neither conflict-generating, nor defending. These attacks play no role in our semantics, and are essentially ignored. These are the attacks that belong in  $\mathcal{R} \setminus (\mathcal{R}_c \cup \mathcal{R}_d)$ . In our example,  $(b, d)$  is an irrelevant attack.

Note that symmetric attacks between two arguments  $a$  and  $b$  need not be of the same type, e.g., it is possible that  $(a, b) \in \mathcal{R}_c$  and  $(b, a) \in \mathcal{R}_d$ . For a given AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , equipped with an attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , we write  $a \rightarrow b$  whenever  $(a, b) \in \mathcal{R}$ ,  $a \rightarrow_c b$  whenever  $(a, b) \in \mathcal{R}_c$ , and  $a \rightarrow_d b$  whenever  $(a, b) \in \mathcal{R}_d$ .

We extend notation to sets of arguments, and, for  $B, C \subseteq \mathcal{A}$ , we write  $B \rightarrow C$  if and only if  $\exists b \in B, c \in C$  such that  $b \rightarrow c$  (analogously for  $B \rightarrow_c C$ ,  $B \rightarrow_d C$ ). For singleton sets, we often write  $b \rightarrow C$  and  $B \rightarrow c$  instead of  $\{b\} \rightarrow C$  and  $B \rightarrow \{c\}$ , respectively (analogously for  $\rightarrow_c, \rightarrow_d$ ).

## 2.2 Semantics

We will now recast the definitions associated with the standard extension classes (already known from the work of Dung [11]) for our setting. In the following, we use shorthands to refer to the various types of semantics. In particular, we use **cf** for conflict-free, **ad** for admissible, **co** for complete, **pr** for preferred, **gr** for grounded, and **st** for stable. We also use  $\sigma$  as a catch-all symbol to indicate any of these extension types.

To define our semantics, following the tradition of Dung [11], we first refine the notion of defense, in a way that takes into account the role of attacks in the attack classification:

**Definition 2** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . Given an argument  $a \in \mathcal{A}$  and a set  $\mathcal{E} \subseteq \mathcal{A}$ , we say that  $\mathcal{E}$  defends  $a$  w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or simply  $\mathcal{E}$  defends  $a$ , when  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  is obvious from the context), if and only if  $\mathcal{E} \rightarrow_d b$  whenever  $b \rightarrow_c a$ .

Returning to our running example, and the chosen attack classification, we observe that  $\{b, d\}$  defends  $c$ , because there is only one conflict-generating attack targeting  $c$  (namely  $(a, c)$ , i.e.,  $a \rightarrow_c c$ ), and there is a respective defending attack  $(d, a)$  that originates from  $\{b, d\}$  and targets  $a$  (i.e.,  $\{b, d\} \rightarrow_d a$ ).

Now, we can recast the standard definitions for the different types of semantics given in [11], using the above ideas:

**Definition 3** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is conflict-free w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or simply **cf**, when  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  is obvious from the context) if and only if it is not the case that  $\mathcal{E} \rightarrow_c \mathcal{E}$ .

In our running example,  $\{b, d\}$  is a **cf**-extension; indeed, note that, although  $b \rightarrow d$ ,  $(b, d)$  is not a conflict-generating attack.

The same ideas are applied to admissible and complete extensions, whose definition essentially mimics the one typically used in AAFs, but considers the attack classification through the alternative notion of defense (Definition 2):

**Definition 4** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is an admissible extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or **ad**-extension for short, omitting  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  when obvious from the context) if and only if:

- $\mathcal{E}$  is **cf**
- If  $a \in \mathcal{E}$ , then  $\mathcal{E}$  defends  $a$

**Definition 5** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a complete extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or **co**-extension for short, omitting  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  when obvious from the context) if and only if:

- $\mathcal{E}$  is an **ad**-extension
- If  $\mathcal{E}$  defends  $a$  and  $\mathcal{E} \cup \{a\}$  is **cf**, then  $a \in \mathcal{E}$

Note that, in the above definition (Definition 5), instead of only requiring that  $a \in \mathcal{E}$  whenever  $\mathcal{E}$  defends  $a$ , we have included the additional requirement that  $\mathcal{E} \cup \{a\}$  is **cf**, thereby deviating somewhat from the definition pattern used in AAFs for **co**-semantics [11]. This additional requirement is redundant in the AAF setting, because it results as a corollary of the weaker definition. However, it is necessary here, for reasons that will be clarified below through an example.

In our running example,  $\{b, d\}$  is an **ad**-extension, because: although  $a \rightarrow_c b$ , it is also the case that  $d \rightarrow_d a$ , so  $b$  is defended by  $\{b, d\}$ ; and  $d$  is not attacked by a conflict-generating attack. However,  $\{b, d\}$  is not a **co**-extension because  $c$  is defended by  $\{b, d\}$  but not included.

Further, it can be shown that  $\{b, c, d\}$  is a **co**-extension. Indeed,  $\{b, c, d\}$  is **cf**, as there is no conflict-generating attack among its members. Also, it is an **ad**-extension, as it defends its members, as explained above. Moreover, although  $a$  is also defended by  $\{b, c, d\}$  (in a trivial manner, as  $a$  is not attacked by a conflict-generating attack), it is also the case that  $\{a, b, c, d\}$  is not **cf**; thus, by definition,  $\{b, c, d\}$  is a **co**-extension. This example also shows why the extra requirement in Definition 5 was necessary (as was also in MAAF's [21]); without the extra requirement, neither  $\{b, c, d\}$  nor  $\{a, b, c, d\}$  would be **co**-extensions, which would be absurd. Moreover, without this extra requirement, no **co**-extension would exist in the above example, because any such extension should include  $a, d$ , and, since  $d$  is included,  $b, c$  should be included too, leading to a non-**cf** set.

Grounded and preferred semantics are defined analogously:

**Definition 6** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a grounded extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or **gr**-extension for short, omitting  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  when obvious from the context) if and only if  $\mathcal{E}$  is a minimal with respect to set inclusion **co**-extension.

**Definition 7** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a preferred extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or **pr**-extension for short, omitting  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  when obvious from the context) if and only if  $\mathcal{E}$  is a maximal with respect to set inclusion **ad**-extension.

In our running example, as already shown above,  $\{b, c, d\}$  is a maximal **ad**-extension (because  $\{a, b, c, d\}$  is not **cf**-extension), thus it is also a **pr**-extension. Interestingly,  $\{b, c, d\}$  is also a **gr**-extension, because, we can easily verify that none of its subsets are **co**-extensions ( $d$  is trivially defended and  $d$  defends both  $b$  and  $c$ ). Moreover,  $\{a, d\}$  is also a **gr**-extension. This shows that the **gr**-extension is not necessarily unique in our setting (as also was the case in MAAF's [21]), in contrast to the standard Dung semantics [11].

Stable semantics also follow a similar pattern:

**Definition 8** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a stable extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  (or **st**-extension for short, omitting  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  when obvious from the context) if and only if:

- $\mathcal{E}$  is maximally **cf** (i.e., a **cf** set that is maximal w.r.t. the subset relation among all other **cf** sets)
- $\mathcal{E} \rightarrow_d a$  whenever  $a \notin \mathcal{E}$

Note that Definition 8 also deviates somewhat from the definition pattern of **st** semantics in standard AAFs. In particular, instead of requiring that  $\mathcal{E}$  is **cf**, we have required that it is maximally **cf**. Moreover, we have required that a defending attack against all arguments not in  $\mathcal{E}$  exists.

This alternative definition is necessary to capture the underlying intuition behind the respective definition in [11], namely that a **st**-extension attacks all arguments not in the extension, therefore (a) the addition of any further argument will render it conflicting, thus, it is maximally **cf**; (b) it defends itself against any attack from such arguments *even if* such an attack does not really exist.

This will become clearer if we analyse the examples illustrated in Figure 2. In the left example  $\{a\}$  is a stable extension in the respective AAF, whereas in the right example it is not, which shows that a stable extension needs to defend itself from all external arguments, even if they do not attack the extension itself. In the setting with the attack classification, this is captured by the requirement that there should exist a defending attack  $(a, b)$  to ensure that  $\{a\}$  is stable. In our running example (Figure 1), note that  $\{b, c, d\}$  is a **st**-extension, but  $\{a, d\}$  is not.

### 3 Formal Properties

In this section, we study various properties of AAFs with an Attack Classification. Most of the results below are reformulations of the respective propositions that apply for AAFs, although there are exceptions. The proofs are given in Section B of the Appendix.

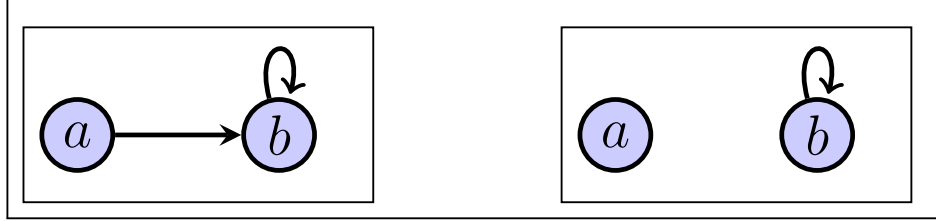


Figure 2: Two AAFs motivating our definition of stable extensions

### 3.1 Initial results and special cases

We start by showing the analogous of Dung’s fundamental lemma (Lemma 10 in [11]):

**Proposition 1** *Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and some  $\mathcal{E} \subseteq \mathcal{A}$ . Then:*

1. *If  $\mathcal{E}$  is an **ad**-extension,  $\mathcal{E}$  defends  $a$  and  $\mathcal{E} \cup \{a\}$  is **cf**, then  $\mathcal{E} \cup \{a\}$  is an **ad**-extension.*
2. *If  $\mathcal{R}_d \subseteq \mathcal{R}_c$ ,  $\mathcal{E}$  is an **ad**-extension and  $\mathcal{E}$  defends  $a$ , then  $\mathcal{E} \cup \{a\}$  is an **ad**-extension.*

Notice that Dung’s fundamental lemma is not generally true for AAFs with Attack Classification, except from the special case where there are no attacks which are defending but not conflict-generating (case #2 of Proposition 1).

Using the second bullet of Proposition 1, we can show that Definitions 5 and 8 can be formulated more simply (i.e., using the pattern of [11]), when  $\mathcal{R}_d \subseteq \mathcal{R}_c$ :

**Proposition 2** *Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , such that  $\mathcal{R}_d \subseteq \mathcal{R}_c$  and some  $\mathcal{E} \subseteq \mathcal{A}$ . Then:*

1.  *$\mathcal{E}$  is a **co**-extension if and only if  $\mathcal{E}$  is an **ad**-extension and  $a \in \mathcal{E}$  whenever  $\mathcal{E}$  defends  $a$ .*
2.  *$\mathcal{E}$  is a **st**-extension if and only if  $\mathcal{E}$  is **cf** and  $\mathcal{E} \rightarrow_d a$  whenever  $a \notin \mathcal{E}$ .*

### 3.2 Reductions of AAFs with an attack classification

The following two propositions show how AAFs with Attack Classification reduce to standard AAFs [11] and MAAF [21] for special cases. We omit further details on MAAF here, but the interested reader can see [21], or subsection A of the Appendix in this paper.



**Proposition 3** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  such that  $\mathcal{R}_c = \mathcal{R}_d = \mathcal{R}$ . Then, for any  $\sigma$ ,  $\mathcal{E}$  is a  $\sigma$ -extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  if and only if  $\mathcal{E}$  is a  $\sigma$ -extension in the AAF  $\mathcal{F}$ .

**Proposition 4** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . Consider also a MAAF  $\mathcal{F}_M = \langle \mathcal{A}_M, \mathcal{T}_M, \mathcal{R}_M \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}_M$ . Then, for any  $\sigma$ -extension, the following hold:

1. If  $\mathcal{R}_c = \mathcal{R}$ ,  $\mathcal{R}_d = \{(a, b) \mid (a, b, \tau) \in \mathcal{R}_M, \tau \in \mathcal{T}_0\}$ , then  $\mathcal{E}$  is a  $\sigma$ -extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  if and only if  $\mathcal{E}$  is a **fr**- $\sigma$ -extension w.r.t.  $\mathcal{T}_0$  in the MAAF  $\mathcal{F}_M$ .
2. If  $\mathcal{R}_c = \mathcal{R}_d = \{(a, b) \mid (a, b, \tau) \in \mathcal{R}_M, \tau \in \mathcal{T}_0\}$ , then  $\mathcal{E}$  is a  $\sigma$ -extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  if and only if  $\mathcal{E}$  is a **re**- $\sigma$ -extension w.r.t.  $\mathcal{T}_0$  in the MAAF  $\mathcal{F}_M$ .
3. If  $\mathcal{R}_d = \mathcal{R}$ ,  $\mathcal{R}_c = \{(a, b) \mid (a, b, \tau) \in \mathcal{R}_M, \tau \in \mathcal{T}_0\}$ , then  $\mathcal{E}$  is a  $\sigma$ -extension w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  if and only if  $\mathcal{E}$  is a **lo**- $\sigma$ -extension w.r.t.  $\mathcal{T}_0$  in the MAAF  $\mathcal{F}_M$ .

### 3.3 Hierarchy and Existence in Extensions

Initially, we show that a hierarchy of extensions, similar to the one shown in [11], holds for AAFs with Attack Classification:

**Proposition 5** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and some  $\mathcal{E} \subseteq \mathcal{A}$ . Then:

1. If  $\mathcal{E}$  is an **ad**-extension, then  $\mathcal{E}$  is a **cf**-extension
2. If  $\mathcal{E}$  is a **co**-extension, then  $\mathcal{E}$  is an **ad**-extension
3. If  $\mathcal{E}$  is a **gr**-extension, then  $\mathcal{E}$  is a **co**-extension
4. If  $\mathcal{E}$  is a **pr**-extension, then  $\mathcal{E}$  is a **co**-extension
5. If  $\mathcal{E}$  is a **st**-extension, then  $\mathcal{E}$  is a **pr**-extension

Existence results are also analogous to [11]; in particular, for any AAF equipped with an attack classification, all types of extensions exist, except maybe **st**-extensions. To show this, we will first need an intermediate result, showing that we can “incrementally” construct minimally-complete extensions starting from an **ad** one. The proof uses an iterative function, similar to the function  $F_{AF}$  used by Dung in [11]. However, for AAFs with an Attack Classification, there are two subtleties.

First,  $F_{AF}$  (as defined in [11]) adds all acceptable arguments in each iteration. In our case, this could lead to a set that is not **cf**. For example, consider Figure 1, where we observe that each of the arguments  $a, b, c$  is accepted by  $\{d\}$ , and adding any one of them in  $\{d\}$  would result in an **ad**-extension; however adding them all at the same time would result in a set that is not **cf** (and thus not **ad**). Therefore, Dung’s construction is inappropriate for our purposes, and a more elaborate one is needed.

Second, for infinite frameworks (i.e., argumentation frameworks with an infinite set of arguments), the existence of a minimal fixpoint for  $F_{AF}$  (in [11]) is guaranteed by the implicit use of the Knaster-Tarski theorem [18], which requires an order preserving function. Although  $F_{AF}$  is order-preserving, our alternative is not.

To overcome these problems, our proof uses a more complex iterative function, employing ordinals. Our proof is analogous to the one employed in [21] (Proposition 7); as a matter of fact, the proof of Proposition 7 in [21] is a special case of our proof here (see also Proposition 4). Importantly, our construction applies also to standard AAFs, so it can be viewed also as an alternative proof for a well-known property of AAFs. Note also that the proof employs the Axiom of Choice.

**Proposition 6** *Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and some  $\mathcal{E}_* \subseteq \mathcal{A}$  such that  $\mathcal{E}_*$  is an **ad**-extension. Then, there exists some  $\mathcal{E}$  such that  $\mathcal{E} \supseteq \mathcal{E}_*$ , and the following hold:*

1.  $\mathcal{E}$  is a **co**-extension.
2. For any  $\mathcal{E}'$  such that  $\mathcal{E}_* \subseteq \mathcal{E}' \subset \mathcal{E}$ , there exists  $a \in \mathcal{E} \setminus \mathcal{E}'$  which is defended by  $\mathcal{E}'$  and  $\mathcal{E}' \cup \{a\}$  is **cf**.
3. For any  $\mathcal{E}'$  such that  $\mathcal{E}_* \subseteq \mathcal{E}' \subset \mathcal{E}$ ,  $\mathcal{E}'$  is not a **co**-extension.

Now we are ready to show our existence result, namely that all extensions (except **st**-extensions) exist in any AAF equipped with any attack classification:

**Proposition 7** *Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . Then, for any  $\sigma \in \{\mathbf{cf}, \mathbf{ad}, \mathbf{co}, \mathbf{gr}, \mathbf{pr}\}$ , there exists a  $\sigma$ -extension.*

The fact that **st**-extensions are not guaranteed to exist is an obvious corollary of the fact that AAFs are special cases of AAFs equipped with an attack classification (see Proposition 3). Propositions 5 and 7 are summarised in Figure 3.

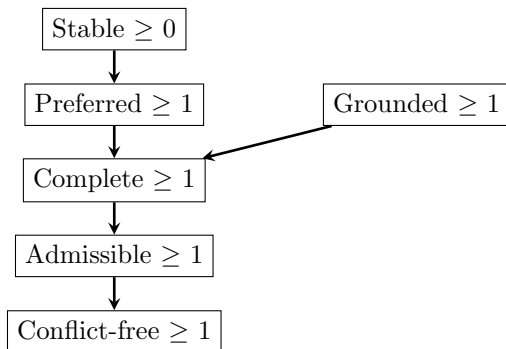


Figure 3: Properties of extensions for AAFs with an attack classification

In AAFs, a **gr**-extension is unique. The counter-example of Figure 1 shows that this is not the case here. However, in the special case where  $\mathcal{R}_d \subseteq \mathcal{R}_c$ , the uniqueness of **gr**-extensions is guaranteed:

**Proposition 8** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , such that  $\mathcal{R}_d \subseteq \mathcal{R}_c$ . Then, there exists a unique **gr**-extension.

### 3.4 Studying the effect of attack classifications

We will now show some results considering the effect of the attack classification on the various extensions. Our first result shows that, as expected, irrelevant attacks (i.e., those that are neither conflict-generating nor defending) have no effect:

**Proposition 9** Consider two AAFs  $\mathcal{F}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$ ,  $\mathcal{F}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$  and some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , such that  $\mathcal{A}_1 = \mathcal{A}_2$  and  $\mathcal{R}_c \cup \mathcal{R}_d \subseteq \mathcal{R}_1 \subseteq \mathcal{R}_2$ . Then, for any  $\sigma \in \{\mathbf{cf}, \mathbf{ad}, \mathbf{co}, \mathbf{pr}, \mathbf{gr}, \mathbf{st}\}$  and any  $\mathcal{E} \subseteq \mathcal{A}_1 = \mathcal{A}_2$ , it holds that  $\mathcal{E}$  is a  $\sigma$ -extension in  $\mathcal{F}_1$  w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  if and only if  $\mathcal{E}$  is a  $\sigma$ -extension in  $\mathcal{F}_2$  w.r.t.  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ .

Another relevant question is what happens if we change the status of some attacks, e.g., if a “normal” attack becomes conflict-generating or if a defending attack becomes “normal”. Towards this, a series of results can be shown, summarised in the following proposition:

**Proposition 10** Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and two attack classifications  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ ,  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , such that  $\mathcal{R}_c^2 \subseteq \mathcal{R}_c^1 \subseteq \mathcal{R}$  and  $\mathcal{R}_d^1 \subseteq \mathcal{R}_d^2 \subseteq \mathcal{R}$ . Then, for any  $\mathcal{E} \subseteq \mathcal{A}$ , the following hold:

1. If  $\mathcal{E}$  is **cf** w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ , then it is **cf** w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .
2. If  $a \in \mathcal{A}$  and  $\mathcal{E}$  defends  $a$  w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ , then  $\mathcal{E}$  defends  $a$  w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .
3. If  $\mathcal{E}$  is an **ad**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ , then it is an **ad**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .
4. If  $\mathcal{E}$  is a **co**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  and an **ad**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  then it is a **co**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ .
5. If  $\mathcal{E}$  is a **pr**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  and an **ad**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  then it is a **pr**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ .
6. If  $\mathcal{E}$  is a **st**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and maximally **cf** w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  then it is a **st**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .

Some comments regarding Proposition 10 are in line here. First, it is interesting to note how the direction of the implication changes depending on the semantics considered. Moreover, in most cases, additional assumptions are needed, i.e., the fact that  $\mathcal{E}$  is a **co**-extension in one of the attack classifications does not necessarily imply that it is also a **co**-extension w.r.t. the other. Also, notice that  $\mathcal{R}_c^2 \subseteq \mathcal{R}_c^1 \subseteq \mathcal{R}$  and  $\mathcal{R}_d^1 \subseteq \mathcal{R}_d^2 \subseteq \mathcal{R}$ .

More importantly, **gr**-extensions are missing from Proposition 10. As a matter of fact, no similar condition can be devised for **gr**-extensions. In particular, if a

set is a **gr**-extension under one attack classification and a **co**-extension under the other, this is not enough to guarantee that it is also a **gr**-extension under both. More formally, we will consider the following two conditions (both of which turn out to be false):

1. If  $\mathcal{E}$  is a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and a **co**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  then it is a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .
2. If  $\mathcal{E}$  is a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  and a **co**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  then it is a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ .

The first condition is shown to be false via Figure 4. In that figure, we observe the same AAF, with two different attack classifications ( $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle, \langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , shown in the left and right part of the figure respectively), where:

- $\mathcal{R}_c^1 = \{(a', b'), (b, b'), (b', b), (c', b), (c, c'), (c', c)\}$
- $\mathcal{R}_d^1 = \{(a, a'), (b, b'), (b', b), (c, c'), (c', c)\}$
- $\mathcal{R}_c^2 = \{(a', b'), (b, b'), (b', b), (c, c'), (c', c)\}$
- $\mathcal{R}_d^2 = \{(a, a'), (b, b'), (b', b), (c, c'), (c', c)\}$

We observe that  $\mathcal{R}_c^2 \subseteq \mathcal{R}_c^1$  and  $\mathcal{R}_d^1 = \mathcal{R}_d^2$ .

Let us now consider the AAF in Figure 4 from the perspective of  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ . We observe that  $\{a, b, c\}$  is a **gr**-extension. Indeed, it is a **co**-extension, because it defends all its members, and the addition of any extra argument would render the set non-**cf**. Now, let us consider its subclasses:  $\emptyset, \{b\}, \{c\}$  and  $\{b, c\}$  are not **co**-extensions (because they defend  $a$  and if we add  $a$  in the respective set we get a **cf** one);  $\{a\}$  and  $\{a, c\}$  are not **co**-extensions (because they defend  $a'$  and if we add  $a'$  in the respective set we get a **cf** one); finally,  $\{a, b\}$  is not an **ad**-extension (due to the attack  $c' \rightarrow_c b$ ). Thus,  $\{a, b, c\}$  is indeed a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ .

Now, let us consider the same AAF, but from the perspective of  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ . Here, we observe that  $\{a, b, c\}$  is again a **co**-extension (as can be easily verified), but  $\{a, b\}$  is also a **co**-extension, since the conflict-generating attack  $(c', c)$  is missing in  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ . Thus,  $\{a, b, c\}$  is not a **gr**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .

The second condition is shown to be false via the counter example of Figure 5. There, we depict the same AAF under two different attack classifications ( $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle, \langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , shown in the left and right part of the figure respectively), where:

- $\mathcal{R}_c^1 = \{(a, b), (b, a)\}$
- $\mathcal{R}_d^1 = \{(a, b), (b, a)\}$
- $\mathcal{R}_c^2 = \{(b, a)\}$
- $\mathcal{R}_d^2 = \{(a, b), (b, a)\}$

We observe again that  $\mathcal{R}_c^2 \subseteq \mathcal{R}_c^1$  and  $\mathcal{R}_d^1 = \mathcal{R}_d^2$ .

We now note that  $\{b\}$  is a **co**-extension from the perspective of  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ , but not a **gr**-extension (because  $\emptyset$  is also a **co**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ ). On the other hand,  $\{b\}$  is a **gr**-extension from the perspective of  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , as can be easily verified.

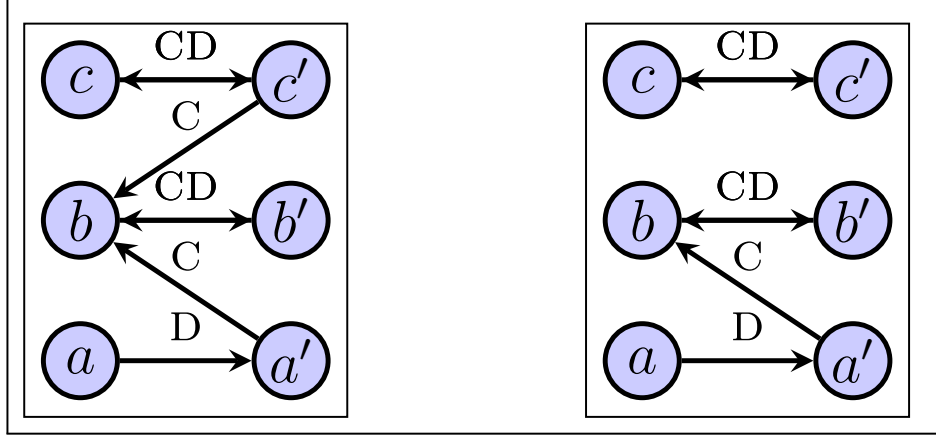


Figure 4: An AAF with two attack classifications (counter-example #1)

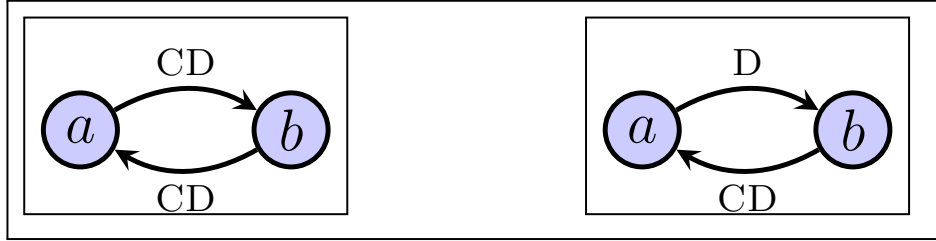


Figure 5: An AAF with two attack classifications (counter-example #2)

## 4 A principle-based analysis

In order to better understand the behaviour of the different acceptability semantics of AAFs, recent studies proposed a set of principles and examined which semantics satisfy each of these principles. In this section, we present a similar principle-based analysis for AAFs with Attack Classification and their semantics. The definitions of the principles we consider are from [20]. They were originally defined for AAFs, so, where necessary, we adjust them to fit the definitions of AAFs with attack classification. We focus our attention to complete, preferred, grounded and stable semantics. At the end of the section, Table 1 provides an overview of the results and highlights the differences with AAFs.

We adopt the standard definition of *isomorphic* argumentation frameworks.

**Definition 9 (Isomorphic)** *Two argumentation frameworks  $\mathcal{F}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$  and  $\mathcal{F}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$  are isomorphic if and only if there exists a bijective function  $m$ :*

$\mathcal{A}_1 \rightarrow \mathcal{A}_2$ , such that  $(a, b) \in \mathcal{R}_1$  if and only if  $(m(a), m(b)) \in \mathcal{R}_2$ . This is denoted by  $\mathcal{F}_1 \doteq_m \mathcal{F}_2$ .

We introduce the notion of isomorphic attack classifications to describe isomorphic argumentation frameworks whose attacks are classified in the same way.

**Definition 10 (Isomorphic with Classification)** *Two argumentation frameworks  $\mathcal{F}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$  and  $\mathcal{F}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$  such that  $\mathcal{F}_1 \doteq_m \mathcal{F}_2$  are equipped with isomorphic attack classifications  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$  if and only if for every  $(a, b) \in \mathcal{R}_1$ ,  $(m(a), m(b)) \in \mathcal{R}_2$  it holds that  $(a, b) \in \mathcal{R}_c^1$  iff  $(m(a), m(b)) \in \mathcal{R}_c^2$  and  $(a, b) \in \mathcal{R}_d^1$  iff  $(m(a), m(b)) \in \mathcal{R}_d^2$ .*

The *language independence* principle holds for the semantics that only take into account the topology of the argumentation graph and the classification of the attacks, and not the names of the arguments. We have extended its original definition [3], to take into account the classification of the attacks.

**Definition 11 (Language Independence)** *A semantics  $\sigma$  satisfies the language independence principle if and only if for every two argumentation frameworks  $\mathcal{F}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$ ,  $\mathcal{F}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$ , such that  $\mathcal{F}_1 \doteq_m \mathcal{F}_2$ , with isomorphic attack classifications  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , respectively, it holds that  $\sigma(\mathcal{F}_2) = \{m(\mathcal{E}) \mid \mathcal{E} \in \sigma(\mathcal{F}_1)\}$ , where  $\sigma(\mathcal{F}_1)$  is the set of  $\sigma$ -extensions of  $\mathcal{F}_1$  with respect to  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and  $\sigma(\mathcal{F}_2)$  is the set of  $\sigma$ -extensions of  $\mathcal{F}_2$  with respect to  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .*

In the following and all other propositions that appear in this section, the mentioned semantics refer to the semantics of AAFs with Attack Classification, as defined in Section 2.2. The proofs of all propositions are presented in the Appendix.

**Proposition 11** *Language independence is satisfied by **co**, **gr**, **pr** and **st** semantics.*

Dung [11], introduced the notion of conflict-freeness, which was later stated as a principle by [3]. Here, we extend the original definition to take into account the classification of attacks.

**Definition 12 (Conflict Freeness)** *A semantics  $\sigma$  satisfies the conflict-freeness principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ ,  $\mathcal{E}$  is conflict-free with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ .*

**Proposition 12** *Conflict-freeness is satisfied by **co**, **gr**, **pr** and **st** semantics.*

Defense and admissibility were also proposed as principles by Baroni and Giacomin [3]. Here, we adjust their definitions to take into account the classification of attacks.

**Definition 13 (defense)** A semantics  $\sigma$  satisfies the defense principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every argument  $a \in \mathcal{E}$ ,  $\mathcal{E}$  defends  $a$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ .

**Definition 14 (Admissibility)** A semantics  $\sigma$  satisfies the admissibility principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  is admissible with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ .

**Proposition 13** Defense and admissibility are satisfied by **co**, **gr**, **pr** and **st** semantics.

Baroni and Giacomin [3] introduced the notions of strong defense and strong admissibility. In our framework, these are defined as follows.

**Definition 15 (Strong defense)** Let  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ ,  $S \subseteq \mathcal{A}$  and an attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . An argument  $a \in \mathcal{A}$  is strongly defended by  $S$  if and only if for every  $b \in \mathcal{A}$  such that  $b \rightarrow_c a$  there exists  $c \in S \setminus \{a\}$  such that  $c \rightarrow_d b$  and  $c$  is strongly defended by  $S \setminus \{a\}$ .

**Definition 16 (Strong Admissibility)** A semantics  $\sigma$  satisfies the strong admissibility principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ ,  $\mathcal{E}$  strongly defends all arguments it contains.

Strong admissibility is satisfied by the grounded semantics of AAFs, but this does not hold for AAFs with Attack Classification. This is the first difference we observe between the two frameworks.

**Proposition 14** Strong admissibility is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.

Van der Torre and Vesic [20] proposed the principle of *naivety*, which we redefine as follows:

**Definition 17 (Naivety)** A semantics  $\sigma$  satisfies the naivety principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ ,  $\mathcal{E}$  is maximal for set inclusion conflict-free set in  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ .

**Proposition 15** Naivety is satisfied by **st** semantics, but not by **co**, **gr** or **pr**.

Coste-Marquis et al. [10] introduced the notion of *indirect conflicts*, which they defined as follows:

**Definition 18 (Indirect Conflict)** Let  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ ,  $S \subseteq \mathcal{A}$  and  $a, b \in \mathcal{A}$ . Then,  $a$  indirectly attacks  $b$  if and only if there is an odd-length path from  $a$  to  $b$  with respect to the attack relation.  $S$  is without indirect conflicts if and only if there exist no  $x, y \in S$  such that  $x$  indirectly attacks  $y$ .

The indirect conflict-freeness principle is then defined as follows.

**Definition 19 (Indirect Conflict Freeness)** A semantics  $\sigma$  satisfies the indirect conflict-freeness principle if and only if for every argumentation framework  $\mathcal{F}$  and for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  is without indirect conflicts.

**Proposition 16** Indirect conflict-freeness is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.

Another principle proposed by Baroni and Giacomin in [3] was *reinstatement*, according to which an extension must contain all the arguments it defends.

**Definition 20 (Reinstatement)** A semantics  $\sigma$  satisfies the reinstatement principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every  $a \in \mathcal{A}$  it holds that if  $\mathcal{E}$  defends  $a$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , then  $a \in \mathcal{E}$ .

Interestingly, in contrast with AAFs, reinstatement is not satisfied by any of the semantics of AAFs with Attack Classification.

**Proposition 17** Reinstatement is not satisfied by any of **co**, **pr**, **gr** and **st** semantics.

Baroni and Giacomin studied in [3] another similar principle called *weak reinstatement*, which has the following definition.

**Definition 21 (Weak Reinstatement)** A semantics  $\sigma$  satisfies the weak reinstatement principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every  $a \in \mathcal{A}$  it holds that if  $\mathcal{E}$  strongly defends  $a$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , then  $a \in \mathcal{E}$ .

As shown in the following proposition, weak reinstatement is also not satisfied by any of the semantics of AAFs with Attack Classification, signifying another important difference with AAFs.

**Proposition 18** Weak reinstatement is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.



Another similar principle introduced by Baroni and Giacomin in [3] is *CF-reinstatement*. Here, this principle can be defined as follows.

**Definition 22 (CF-Reinstatement)** *A semantics  $\sigma$  satisfies the CF-reinstatement principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $\sigma$ -extension  $\mathcal{E}$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every  $a \in \mathcal{A}$  it holds that if  $\mathcal{E}$  defends  $a$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and  $\mathcal{E} \cup \{a\}$  is conflict-free, then  $a \in \mathcal{E}$ .*

While AAFs and AAFs with Attack Classification behave totally differently with respect to the two other forms of reinstatement, their behaviour is exactly the same with respect to *CF-reinstatement*.

**Proposition 19** *CF-reinstatement is satisfied by **co**, **gr**, **pr** and **st** semantics.*

The principle of *I-maximality*, originally proposed in [3] states that an extension cannot contain another extension.

**Definition 23 (I-maximality)** *A semantics  $\sigma$  satisfies the I-maximality principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and for every  $\sigma$ -extensions  $\mathcal{E}_1, \mathcal{E}_2$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , if  $\mathcal{E}_1 \subseteq \mathcal{E}_2$  then  $\mathcal{E}_1 = \mathcal{E}_2$ .*

**Proposition 20** *I-maximality is satisfied by **gr**, **pr** and **st** semantics, but not by **co**.*

We next consider the *allowing abstention* principle, introduced by [1] and defined as follows.

**Definition 24 (Allowing Abstention)** *A semantics  $\sigma$  satisfies the allowing abstention principle if and only if for every argumentation framework  $\mathcal{F}$ , for every classification of its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , for every  $a \in \mathcal{A}$ , if there exist two  $\sigma$ -extensions  $\mathcal{E}_1, \mathcal{E}_2$  of  $\mathcal{F}$  with respect to  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , such that  $a \in \mathcal{E}_1$  and  $\mathcal{E}_2 \rightarrow_c a$  then there exists an extension  $\mathcal{E}_3$  such that  $a \notin \mathcal{E}_3$  and  $\mathcal{E}_3$  does not attack (with a conflict-generating attack)  $a$ .*

AAFs and AAFs with Attack Classification behave differently with respect to this principle for the complete and grounded semantics:

**Proposition 21** *Allowing abstention is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.*

In order to define crash resistance [8] in our framework, we first need to introduce the following two definitions.

**Definition 25 (Disjoint Argumentation Frameworks)** Let  $\mathcal{F}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$  and  $\mathcal{F}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$  be two argumentation frameworks. The frameworks are called disjoint if and only if  $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$ .

An argumentation framework  $\mathcal{F}^*$  with classification of attacks  $\langle \mathcal{R}_c^*, \mathcal{R}_d^* \rangle$ , is contaminating if joining  $\mathcal{F}^*$  with an arbitrary disjoint framework  $\mathcal{F}$  with classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , results in a framework  $\mathcal{F}^* \cup \mathcal{F}$  having the same extensions as  $\mathcal{F}^*$ .

**Definition 26 (Contamination)** An argumentation framework  $\mathcal{F}^* = \langle \mathcal{A}^*, \mathcal{R}^* \rangle$  with a classification of attacks  $\langle \mathcal{R}_c^*, \mathcal{R}_d^* \rangle$  is contaminating for a semantics  $\sigma$  if and only if, for any argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  disjoint from  $\mathcal{F}^*$  with any classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  it holds that  $\mathcal{F}^* \cup \mathcal{F}$  and  $\mathcal{F}^*$  have the same  $\sigma$ -extensions. We shall denote this by  $\sigma(\mathcal{F}^* \cup \mathcal{F}) = \sigma(\mathcal{F}^*)$ .

**Definition 27 (Crash Resistance)** A semantics  $\sigma$  satisfies the crash resistance principle if and only if there are no contaminating argumentation frameworks for  $\sigma$ .

**Proposition 22** Crash resistance is satisfied by **co**, **gr** and **pr** semantics but not by **st**.

Crash resistance forbids only the most extreme form of interferences between disjoint subgraphs. A stronger property, non-interference, was defined by Caminada et al. [8]. We first need to define a notion of isolated set, i.e., a set that neither attacks outside arguments nor is attacked by them.

**Definition 28 (Isolated)** Let  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  be an argumentation framework, with any type of classification for its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ . A set  $S \subseteq \mathcal{A}$  of arguments is isolated in  $\mathcal{F}$  if and only if the following holds.

$$((S \times (\mathcal{A} \setminus S)) \cup ((\mathcal{A} \setminus S) \times S)) \cap \mathcal{R} = \emptyset$$

A semantics satisfies non-interference if for every isolated set  $S$ , the intersections of the extensions with the set  $S$  coincide with the extensions of the restriction of the frameworks on  $S$ .

**Definition 29 (Non-Interference)** A semantics  $\sigma$  satisfies the non-interference principle if and only if for every argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , with any classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for every  $S \subseteq \mathcal{A}$  set of arguments isolated in  $\mathcal{F}$ , it holds  $\sigma(\mathcal{F}_{\downarrow S}) = \{\mathcal{E} \cap S \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ , where  $\mathcal{F}_{\downarrow S} = (S, \mathcal{R} \cap (S \times S))$ .

**Proposition 23** Non-interference is satisfied by **co**, **pr** and **gr** semantics but not by **st**.

The previous principle can be made even stronger by considering the case when the set  $S$  is not attacked by the rest of the framework but can attack the rest of the framework with any attack. Such sets are called *unattacked* and are defined as follows.

**Definition 30 (*Unattacked*)** *Given an argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  with a classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , a set of arguments  $U \subseteq \mathcal{A}$  is unattacked if and only if  $\nexists a \in \mathcal{A} \setminus U$  such that  $a$  attacks  $U$ . The set of unattacked sets in  $\mathcal{F}$  is denoted by  $\mathcal{US}(\mathcal{F})$ .*

Using the notion of unattacked sets, the principle of directionality, introduced by Baroni and Giacomin in [3], is defined as follows.

**Definition 31 (*Directionality*)** *A semantics  $\sigma$  satisfies the directionality principle if and only if for every argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , with any classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for every  $U \in \mathcal{US}(\mathcal{F})$  it holds that  $\sigma(\mathcal{F}_{\downarrow U}) = \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ .*

While AAFs and AAFs with Attack Classification behave in the same way with respect to crash resistance and non-interference, their behaviour is different for the complete, preferred and grounded semantics with respect to directionality.

**Proposition 24** *Directionality is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.*

We now consider two variants of directionality, called weak directionality and semi-directionality, originally defined in [20].

**Definition 32 (*Weak Directionality*)** *A semantics  $\sigma$  satisfies the weak-directionality principle if and only if for every argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , with any type of classification for its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for every  $U \in \mathcal{US}(\mathcal{F})$  it holds that  $\sigma(\mathcal{F}_{\downarrow U}) \supseteq \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ .*

**Definition 33 (*Semi Directionality*)** *A semantics  $\sigma$  satisfies the semi-directionality principle if and only if for every argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , with any type of classification for its attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for every  $U \in \mathcal{US}(\mathcal{F})$  it holds that  $\sigma(\mathcal{F}_{\downarrow U}) \subseteq \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ .*

Similar to directionality, AAFs with Attack Classification demonstrate a different behaviour from AAFs with respect to weak directionality and semi-directionality for the complete, grounded and preferred semantics.

**Proposition 25** *Weak directionality is satisfied by **st** semantics but not by **co**, **gr** or **pr**.*

**Proposition 26** *Semi-directionality is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.*

Table 1 provides an overview of the principles that are satisfied or violated by each of the semantics of AAFs with Attack Classification, highlighting the differences with AAFs. As expected, AAFs with Attack Classification satisfy a subset of the principles satisfied by AAFs. This is because AAFs are a special case of AAFs with Attack Classification, as shown in Proposition 3. Some notable findings are: (i) the two frameworks behave in the same way with respect to most of the principles; (ii) the grounded semantics does not satisfy strong admissibility; (iii) while none of the semantics satisfy reinstatement or weak reinstatement, they all satisfy  $\mathcal{CF}$ -reinstatement; (iv) allowing abstention is not satisfied by any of the semantics; and (v) with respect to all directionality principles, AAFs with Attack Classification behave in the same way with AAFs for the stable semantics, but, in contrast with AAFs, they violate these principles for the complete, grounded and preferred semantics.

Table 1: Principles and Semantics for Argumentation Framework with Attack Classification. We write highlight to denote a difference with AAFs.

<b>Principle</b>	<b>co</b>	<b>gr</b>	<b>pr</b>	<b>st</b>
Language Independence (Def. 11)	Yes	Yes	Yes	Yes
Conflict Freeness (Def. 12)	Yes	Yes	Yes	Yes
Defense (Def. 13)	Yes	Yes	Yes	Yes
Admissibility (Def. 14)	Yes	Yes	Yes	Yes
Strong Admissibility (Def. 16)	No	No	No	No
Naivety (Def. 17)	No	No	No	Yes
Indirect Conflict Freeness (Def. 19)	No	No	No	No
Reinstatement (Def. 20)	No	No	No	No
Weak Reinstatement (Def. 21)	No	No	No	No
$\mathcal{CF}$ -Reinstatement (Def. 22)	Yes	Yes	Yes	Yes
I-maximality (Def. 23)	No	Yes	Yes	Yes
Allowing Abstention (Def. 24)	No	No	No	No
Crash Resistance (Def. 27)	Yes	Yes	Yes	No
Non-Interference (Def. 29)	Yes	Yes	Yes	No
Directionality (Def. 31)	No	No	No	No
Weak Directionality (Def. 32)	No	No	No	Yes
Semi Directionality (Def. 33)	No	No	No	No

## 5 Use case: Argumentation Schemes

Argumentation schemes are patterns of arguments used in everyday conversational argumentation [23]. Each scheme is informally described in terms of a set of defeasible premises and a conclusion, and is associated with a set of critical questions, which are possible ways to undermine an argument matching the scheme. For example, the scheme for arguments from position to know is described as follows:

*Major Premise:* Source  $a$  is in a position to know about things in a certain subject domain  $S$  containing proposition  $A$ .

*Minor Premise:*  $a$  asserts that  $A$  (in domain  $S$ ) is true (false).

*Conclusion:*  $A$  is true (false).

The critical questions associated with this scheme are:

CQ<sub>1</sub>: Is  $a$  in a position to know whether  $A$  is true (false)?

CQ<sub>2</sub>: Is  $a$  an honest (trustworthy, reliable) source?

CQ<sub>3</sub>: Did  $a$  assert that  $A$  is true (false)?

A formal model of argumentation schemes, which is still missing, would enable evaluating arguments matching the schemes using computational methods. Here, we show how they can be formalised as AAFs equipped with an attack classification. Our approach consists of the following steps (for an example, see Figure 6):

1. Model the argument matching a scheme, each critical question associated to the scheme, and each response to a critical question as arguments in an AAF.
2. Add a normal attack from each critical question to the main argument.
3. Add an attack from each response to the corresponding critical question. If the response adequately addresses the critical question, model the attack as normal. If the response leaves a doubt, model it as a defending attack.

By modelling the attack from the response to the critical question as a normal attack, we make sure that the attack from the critical question to the argument is ineffective. If all critical questions are attacked with a normal attack, and if there are no other counter-arguments, the argument becomes sceptically accepted. On the other hand, by modelling the attack to the critical question as defending, we leave some doubt regarding the acceptability of the argument. If there are no other counter-arguments, the argument is not rejected but is included in some of the extensions of the framework.

To illustrate this behaviour, we use the classic example of a dialogue in which a tourist, wandering around in a foreign city, asks a stranger where the Central Station can be found. The stranger says that the station is behind building X, so the tourist believes that the station is there ( $a$ ). The three questions that could then be asked are: Is the stranger in position to know where the station is ( $cq_1$ )? Is the stranger a reliable source ( $cq_2$ )? Did the stranger indeed say that the Central

Station is located behind building X ( $cq_3$ )? Consider the following responses to these questions: The stranger said she works at the station, she is therefore in position to know ( $r_1$ ). She was wearing a uniform, so she is a reliable source ( $r_2$ ). Her English was great, so I am sure she said that the station is located behind building X ( $r_3$ ).

We can model this dialogue as an AAF as described above and as shown in Figure 6. Considering the responses to the critical questions as adequate, we model the attacks to the arguments representing the critical questions as normal. This framework has only one (complete, grounded, preferred, stable) extension, which contains the argument concluding that the Central Station is behind building X.

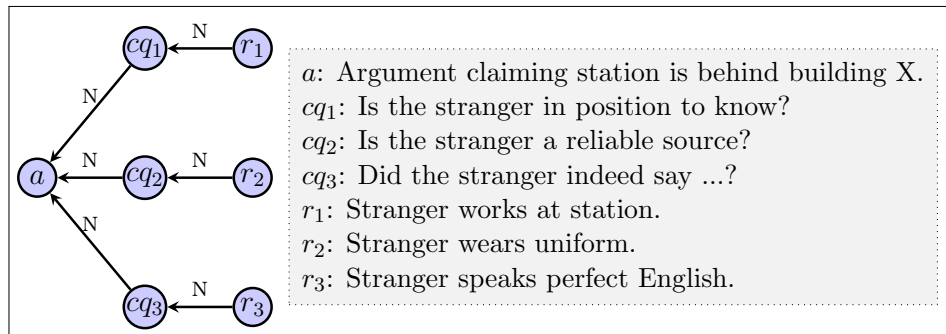


Figure 6: Model of a dialogue with an argument from position to know

Consider now the case that the response to  $cq_3$  is the following: “There was too much noise and I couldn’t hear what the stranger was saying, but I could read her lips” ( $r'_3$ ). Obviously, such a response would leave a doubt about whether the station is indeed located behind building X. To capture this case, we model the attack from  $r'_3$  to  $cq_3$  as a defending attack. The framework then has two complete extensions, one in which  $a$  is accepted ( $\{r_1, r_2, r'_3, a\}$ ) and another in which  $a$  is not accepted ( $\{r_1, r_2, r'_3, cq_3\}$ ). While there might be other ways to end up with the same extensions by using standard AAFs and by adding additional arguments or attacks, our approach offers a standard and intuitive way of modelling argumentation schemes, with critical questions attacking the main argument and responses attacking the critical questions. Following this approach, the acceptability of the main argument depends only on characterising the attack to the critical question.

## 6 A Declarative Implementation

The proposed framework has been implemented as a Web App<sup>2</sup>, through which users can write their own AAFs with Attack Classification and request the compu-

<sup>2</sup><http://139.91.183.45:8070/>

tation of all/one of the  $\sigma$ -extensions, where  $\sigma \in \{\mathbf{cf}, \mathbf{ad}, \mathbf{co}, \mathbf{st}, \mathbf{pr}, \mathbf{gr}\}$ . For each request, the App returns the number of  $\sigma$ -extensions (if the user requests all  $\sigma$ -extensions), the  $\sigma$ -extension(s) and the time that the solver needed to compute the extensions.

The various semantics have been encoded in the language of Answer Set Programming (ASP), a declarative, non-monotonic formalism suitable for solving combinatorial and optimization problems. The Web App uses the Clingo<sup>3</sup> and Asprin ASP reasoners<sup>4</sup> to execute the logic programs. Asprin is applied when optimizations, such as subset/superset relations, are needed.

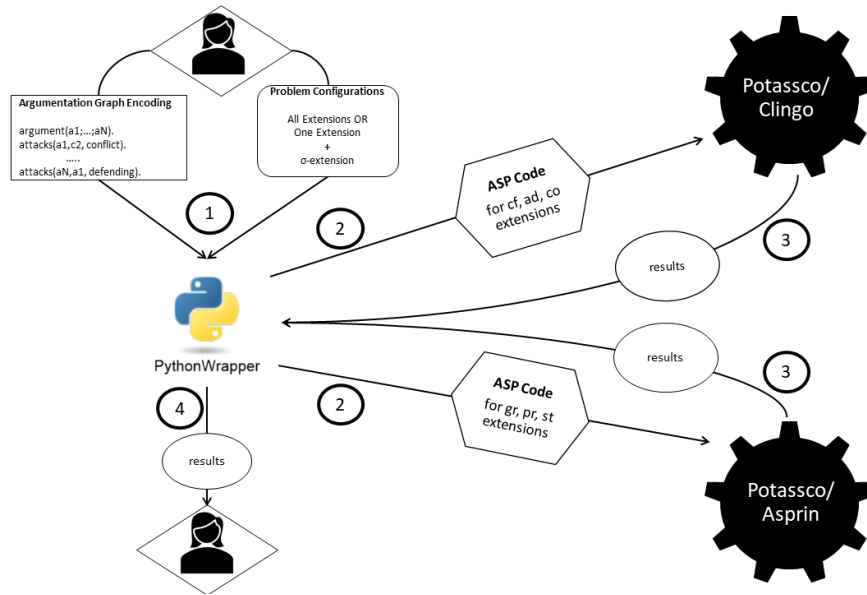


Figure 7: Workflow of the Application

The main components and the workflow of our implementation are shown in Figure 7. The wrapper and the ASP code are available in our GitHub repository<sup>5</sup>. The ASP encoding of the different semantics is also shown in subsection C of the Appendix.

In brief, the system receives an ASP program representing an AAF with Attack Classifications, along with the user's choice of semantics to compute. Through a GUI, the user can specify the type of extensions she is interested in and also if she wishes one or all extensions (step 1 in Figure 7). Then, the system computes the corresponding extensions by applying the associated (to the selected semantics) ASP code. It uses Clingo to compute the **cf**, **ad** and **co** extensions, and Asprin

<sup>3</sup><https://potassco.org/clingo/>

<sup>4</sup><https://potassco.org/asprin/>

<sup>5</sup><https://github.com/valexande/Argumentation-with-Classification>

to compute the **gr**, **pr** and **st** extensions, taking advantage of its capability to compute maximal or minimal answer sets (step 3). The wrapper deals with all data communication and transformation among components (steps 2 and 4). Finally, the results, along with some analytics, are presented on the GUI (step 4).

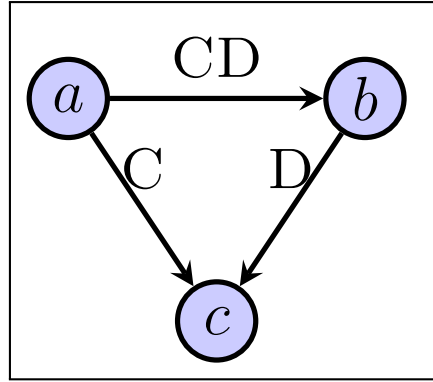


Figure 8: Running Example

Next, we discuss the rationale of our ASP encoding with the help of an example. Consider the framework shown in Figure 8, containing the arguments  $\{a, b, c\}$ , a conflict-generating attack from argument  $a$  to  $c$ , a defending attack from argument  $b$  to  $c$ , and a normal attack from  $a$  to  $b$ . The encoding of this input in ASP, according to our implementation, is as follows:

```
argument(a;b;c).
attacks(a, b, conflict).
attacks(a, c, conflict).
attacks(a, b, defending).
attacks(b, c, defending).
```

The first rule defines the available arguments of the framework, and then all attack relations are specified by means of the *attacks/3* predicate.

The ASP program for **cf** begins by generating combinations of arguments that may comprise a possible answer:

```
{cf_extension(A) : argument(A)}.
```

The above so-called choice rule generates every possible subset of the set of arguments, and for each subset, it instantiates the *cf\_extension/1* predicate with every argument it contains. In our example, a program having this rule alone would generate 8 answer sets, each containing instances of cf-extension with arguments from each of the following sets:  $\{\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$ .



Apparently, not all of these answer sets are valid *cf*-extensions for the given framework. In order to compute the valid conflict-free sets, an appropriate constraint needs to be applied:

```
:- cf_extension(A1), cf_extension(A2), attacks(A1, A2, conflict).
```

The above constraint eliminates any answer where an internal conflict-generating attack among arguments exists. In our running example, when this constraint is applied, together with the choice rule and the input, it will generate 5 answer sets, one for each of the following sets:  $\{\}$ ,  $\{a\}$ ,  $\{b\}$ ,  $\{c\}$ ,  $\{b, c\}$ .

The same rationale is applied to the other extensions. For instance, for computing the **ad** extensions, we initially consider all **cf** extensions, and we start pruning any answer set that does not satisfy the following constraint:

```
:- ad_extension(A), not defenededByExtension(A).
```

The following rules specify the cases that the *defenededByExtension/1* predicate is true:

```
defenededByExtension(A) :-
    argument(A),
    not argumentGetsConfAttack(A).
defenededByExtension(A) :-
    argument(A),
    attacks(A1, A, conflict),
    attacks(A2, A1, defending),
    cf_extension(A2).
argumentGetsConfAttack(A) :- argument(A),
    attacks(_, A, conflict).
```

The first rule represents the case that an argument does not receive a conflict-generating attack. The second rule describes the case that the argument is defeneded against conflict-generating attacks by a conflict-free set. The third rule is used to describe arguments receiving a conflict-generating attack. These rules implement the conditions of admissibility as described in Definition 4. In our running example, these rules will generate two **ad** answer sets:  $\{\}$ ,  $\{a\}$ .

The computation of **pr** and **gr** extensions, is trivial with Asprin. We first compute all the **ad** and **co** answer sets, respectively for **pr** and **gr**, and then we find the maximal (for **pr**) or minimal (for **gr**) sets among them. For the **pr** answer sets, the rules that compute the maximal w.r.t. subset relation **ad** answer sets are:

```
#preference(p1, superset){ad_extension(A)}.
#optimize(p1).
```

The first line is a *preference* predicate with name *p1* and type *superset* for the atom *ad\_extension/1*. In the second line, we ask to optimize *p1*. Eventually, Asprin will

compute the maximal subsets for the atom *ad\_extension/1*. If more than one maximal answer sets exist, all of them will be returned.

Similar is the case for **gr**. With the following rules, we find the minimal w.r.t. subset relation **co** answer sets.

```
#preference(p1,subset){co_extension(A)}.
#optimize(p1).
```

In our running example, the **gr**, **pr**, and **co** answer sets are the same:  $\{a\}$ .

For the **st** semantics, we use a similar set of rules to compute the maximal conflict-free sets. We then use the following two rules to find the arguments attacked (with a defending attack) by these sets, and to prune the answers that contain arguments that are neither contained nor attacked by these sets:

```
other(A) :- argument(A), not cf_extension(A),
           cf_extension(B), attacks(B,A, defending).

:- argument(A), not other(A), not cf_extension(A).
```

In our running example, there is no **st** extension.

## 7 Related Work

The need to further refine the notion of attack in argumentation frameworks has led to several different extensions of AAFs. For example, AAFs with Recursive Attacks (AFRA) [2] and Extended Argumentation Frameworks (EAF) [17] extended the definition of attack, allowing attacks to be directed not only to arguments, but also to other attacks. The difference between the two is that, while in EAFs only attacks whose target is an argument can be attacked, in AFRA any attack can be attacked. This idea is orthogonal to our approach that considers a classification of attacks, which are, however, all directed to arguments, and studying the combination of these two approaches, e.g., by allowing different types of attack that can be directed to either arguments or attacks is an interesting research direction.

Commonsense Argumentation Frameworks [22], on the other hand, included two types of attacks, which differ in the type of arguments they are directed to, i.e., deductive arguments or commonsense arguments. In our framework, all arguments are of the same type and attacks are not characterized by the arguments they are directed to, but by the role that the modeler would like them to have in the process that selects an acceptable set of arguments.

Bipolar Argumentation Frameworks (BAF) [9] introduced support as a new kind of interaction among arguments in abstract argumentation. The definition and role of *direct defeats* is the same as that of attacks in AAFs, while the role of support is to help arguments establish their rationality. A combination of supports and defeats in a chain of arguments can lead to different types of defeat among arguments such as *indirect* and *supported defeats*. The main difference with our

approach is that such types of interaction are not primitive but result from the combination of the pairwise interactions in a chain of arguments. Another important difference is that all types of defeat in BAFs have the same dual role with the attacks in AAFs, i.e. to generate conflicts and to defend other arguments.

Some other studies have introduced weights or preferences on attacks following quantitative or qualitative approaches. For example, Weighted Argumentation Systems [13] assign weights to attacks as a way to describe their strength, and use the idea of an inconsistency budget as a way to disregard attacks up to a certain weight. The idea of weighted attacks is also used in [14], where the acceptability of arguments is not defined in terms of the standard Dung-style extensions, but in terms of numerical values derived from a set of equations describing the arguments and the attacks. Weighted Argumentation Systems are also presented in [6]. The main difference with our approach is that our aim is not to capture the strength of attacks, but the different roles that attacks may have in an argumentation framework. These two ideas are orthogonal, and combining them to develop a weighted argumentation framework with attack classification would be an interesting future research direction. While social networks is indeed a domain where numerical weights can be derived from the reactions of the users, in many other domains such types of data may not be available.

A qualitative approach to represent preferences among attacks was proposed in [16]. Similarly to our approach, the authors defined a framework with (an arbitrary number of) types of attack. These are partially ordered and each attack is assigned one of these types. This allows for a finer grained definition of defense (compared to AAFs), which can roughly be described as follows: an argument is defended against an attack from a counter-argument, if the latter receives a stronger attack from another argument. It also allows for a finer definition of acceptability semantics, which take into account the relative difference of strength between defensive and offensive attacks.

All such preference-based approaches and many others, such as [15, 7], which use either numerical values or priorities to represent the (relative) strength of attacks, have a common aim: to capture the (absolute or relative) strength of arguments and to resolve conflicts by comparing the conflicting arguments according to their strength. Our aim, on the other hand, is to capture the different roles that attacks may have in an argumentation framework. These two ideas are orthogonal, and combining them to develop a weighted or preference-based argumentation framework with attack classification would be an interesting future research direction.

Similarly to our approach, [19] considers different types of attack (or attack relations) among arguments. Their approach is based on the intuition that each attack relation can represent a different criterion according to which the arguments can be evaluated one against another. The evaluation of arguments is based on the aggregation of the different attack relations using methods from social choice theory, such as majority voting, and the use of the standard acceptability semantics of AAFs in the aggregate argumentation framework. They do not, therefore, provide ways to treat certain criteria differently than others, which is one of the main characteristics of the AAFs with Attack Classification where conflict-generating

and defending attacks have different effects on the acceptability of the arguments they attack or defend.

Deductive argumentation [5] also supports different types of attack, which depend on the underlying logic. For example, choosing classical logic as the base logic provides seven different types of attack. The different types of attack in such frameworks are associated with the internal structure of arguments and cannot therefore be directly compared with our framework where arguments are abstract. Extending those frameworks with the ability to explicitly represent the role of attacks, as in our framework, would enable alternative ways to reason with structured arguments, which might be useful in some domains.

Abstract Argumentation Frameworks with Attack Classification generalize AAFs in a way that cannot be captured by any of the other extensions of AAFs discussed in this section. As shown in Proposition 3, the framework we propose preserves compatibility with AAFs retaining many of its nice properties (see Section 4). This also makes it possible to extend AAFs with Attack Classification with additional features of other frameworks (e.g., preferences, weights on arguments or attacks, etc.), which are also compatible with AAFs, resulting in even more expressive argumentation models. The study of such potential extensions is among our plans for future work.

## 8 Conclusion

Motivated by the observation we made in [21] on the dual role of attacks in abstract argumentation frameworks, in this paper we introduce a new abstract model of arguments called Abstract Argumentation Frameworks with Attack Classification. Its main characteristic is that it allows specifying the role of each attack in a given argumentation graph. Specifically, it allows classifying attacks into four disjoint sets: *normal attacks*, which are equivalent to the standard attacks of AAFs; *conflict-generating attacks*, which generate conflicts, but cannot defend against one; *defending attacks*, which can defend other arguments, but do not generate conflicts; and *irrelevant attacks*, which are neither conflict-generating nor defending and are, therefore, ignored. This classification allows our framework to more accurately model real-world dialogues where different responses to the same argument may have a different effect on its acceptability. We recast the definitions of conflict-free, admissible, complete, preferred, grounded and stable extension-based semantics to account for the attack classification that we introduced to AAFs. To examine the behaviour of the new framework and its semantics, we studied their properties and their relation with the semantics of AAFs, and we assessed the different semantics against a set of standard principles. Finally, to enable testing our framework, we developed a Web App that supports the creation of argumentation frameworks with attack classification and the computation of their extensions.

As we explain in Section 7, the attack classification that we extend AAFs with, is orthogonal to other extensions that previous studies have proposed. The combination of different extensions can, therefore, lead to more expressive frameworks

that introduce preferences or weights on different types of attack, or apply the proposed classification to other types of attack such as the second-order attacks of EAFs [17], the high-order attacks of AFRA [2] or the structure-aware attacks of structured argumentation frameworks.

Our plans for future work include extending both the theoretical and the practical results we present in this paper. In terms of theory, we plan to define alternative semantics for our framework (e.g., semi-stable, eager, ideal, etc.) based on their corresponding definitions for AAFs. We also want to extend our principle-based analysis with other principles that are related to the notion of skepticism, such as *skepticism adequacy* and *resolution adequacy* [3] or to expressiveness, such as *tightness*, *conflict-sensitiveness* and *com-closure* [12]. Such an analysis will help us better understand the advantages and limitations of our approach and its differences with AAFs. We also plan to define new principles, which will be specific to AAFs with Attack Classification. This will help further analyse the behaviour of this framework, and may also lead to the definition of new semantics, tailored to its distinctive features.

Another possible theoretical extension of our work would be to make the “defending” property attack-specific. In particular, one could argue that a defending attack from argument  $a$  to argument  $b$  may be defending some (but not all) of the conflict-generating attacks originating from  $b$ . To support this, we could consider a formalism, where the “characterization” of the attack as defending would also include the (conflict-generating) attacks that it can be used to defend other arguments from. This would, of course, result in a more complex formalism, but also a more expressive one.

In terms of practical work, we want to evaluate and refine the Web App in terms of performance and usability. We also want to explore potential applications of our framework in domains where the proposed classification of attacks fits well with the types of information they use, such as persuasion dialogues or debates in Web forums and social media.

## Acknowledgment

This work has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 188. The second author was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “1st Call for H.F.R.I. Research Projects to support Faculty Members and Researchers and the procurement of high-cost research equipment” grant (Project Number: 4195).

## References

- [1] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *The knowledge engineering review*, 26(4):365–410, 2011.

- [2] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. AFRA: Argumentation framework with recursive attacks. *IJAR*, 52(1):19–37, 2011.
- [3] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.
- [4] Ringo Baumann and Christof Spanring. Infinite argumentation frameworks. In Thomas Eiter, Hannes Strass, Mirosław Truszczyński, and Stefan Woltran, editors, *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation: Essays Dedicated to Gerhard Brewka on the Occasion of His 60<sup>th</sup> Birthday*, pages 281–295. Springer International Publishing, 2015.
- [5] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001.
- [6] Stefano Bistarelli and Francesco Santini. Weighted argumentation. *FLAP*, 8(6):1589–1622, 2021.
- [7] Jean-Rémi Bourguet, Leila Amgoud, and Rallou Thomopoulos. Towards a unified model of preference-based argumentation. In *International Symposium on Foundations of Information and Knowledge Systems*, pages 326–344. Springer, 2010.
- [8] Martin Caminada. Semi-stable semantics. *COMMA*, 144:121–130, 2006.
- [9] Claudette Cayrol and Marie-Christine Lagasque-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In Lluís Godo, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005, Barcelona, Spain, July 6-8, 2005, Proceedings*, volume 3571 of *Lecture Notes in Computer Science*, pages 378–389. Springer, 2005.
- [10] Sylvie Coste-Marquis, Caroline Devred, and Pierre Marquis. Prudent semantics for argumentation frameworks. In *17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’05)*, pages 5–pp. IEEE, 2005.
- [11] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [12] Paul E. Dunne, Wolfgang Dvorák, Thomas Linsbichler, and Stefan Woltran. Characteristics of multiple viewpoints in abstract argumentation. *Artif. Intell.*, 228:153–178, 2015.
- [13] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, February 2011.

- [14] Sinan Egilmez, João Martins, and João Leite. Extending social abstract argumentation with votes on attacks. In *Theory and Applications of Formal Argumentation - Second International Workshop, TAFA*, volume 8306, pages 16–31. Springer, 2013.
- [15] Souhila Kaci and Leendert van der Torre. Preference-based argumentation: Arguments supporting multiple values. *International Journal of Approximate Reasoning*, 48(3):730–751, 2008.
- [16] Diego C Martinez, Alejandro Javier Garcia, and Guillermo R Simari. An abstract argumentation framework with varied-strength attacks. In *KR-08*, pages 135–144, 2008.
- [17] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artificial intelligence*, 173(9-10):901–934, 2009.
- [18] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 1955.
- [19] Fernando Tohmé, G. Bodanza, and Guillermo R. Simari. Aggregation of attack relations: a social-choice theoretical analysis of defeasibility criteria. In *FoIKS*, pages 8–23, 2008.
- [20] Leendert Van der Torre and Srdjan Vesic. *The Principle-Based Approach to Abstract Argumentation Semantics*, volume 1, page 797–838. College Publications, 2018.
- [21] Alexandros Vassiliades, Giorgos Flouris, Theodore Patkos, Antonis Bikakis, Nick Bassiliades, and Dimitris Plexousakis. A multi attack argumentation framework. In *Proceedings of the 4<sup>th</sup> International Conference on Logic and Argumentation (CLAR-21)*, 2021.
- [22] Alexandros Vassiliades, Theodore Patkos, Antonis Bikakis, Giorgos Flouris, Nick Bassiliades, and Dimitris Plexousakis. Preliminary notions of arguments from commonsense knowledge. In *11th Hellenic Conference on Artificial Intelligence*, pages 211–214, 2020.
- [23] Douglas Walton, Christopher Reed, and Fabrizio Macagno. *Argumentation schemes*. Cambridge University Press, 2008.

# Appendix

## A Multi Attack Argumentation Frameworks

In this section, we provide the background material which refers to MAAF's and their semantics, originally defined in [21]. A multi-attack argumentation framework is an argumentation framework where attacks are of multiple types.

**Definition 34** A multi-attack argumentation framework (*MAAF for short*) is a tuple  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ , such that:

- $\mathcal{A}$  is a set of arguments
- $\mathcal{T}$  is a set of attack types
- $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A} \times \mathcal{T}$  is a set of type-annotated attacks among arguments

Note that  $\mathcal{A}$  and/or  $\mathcal{T}$  can be infinite, so  $\mathcal{R}$  can be infinite too. Intuitively an attack  $(a, b, \tau) \in \mathcal{R}$  represents that  $a$  attacks  $b$ , and that the attack is of type  $\tau$ . Note that the same two arguments may be related with attacks of different types, in which case each attack type is represented as a different triple in  $\mathcal{R}$ .

For any given set of types  $\mathcal{T}_0 \subseteq \mathcal{T}$ , we say that  $a$  attacks  $b$  w.r.t.  $\mathcal{T}_0$  (denoted by  $a \rightarrow_{\mathcal{T}_0} b$ ) if there exists  $\tau \in \mathcal{T}_0$ , such that  $(a, b, \tau) \in \mathcal{R}$ . For simplicity, we often write  $\rightarrow_{\tau}$  to denote  $\rightarrow_{\{\tau\}}$ , and  $\rightarrow$  to denote  $\rightarrow_{\mathcal{T}}$ . We extend notation to sets of arguments, and, for  $B, C \subseteq \mathcal{A}$ , we write  $B \rightarrow_{\mathcal{T}_0} C$  if and only if  $\exists b \in B, c \in C$  such that  $b \rightarrow_{\mathcal{T}_0} c$ . For singleton sets, we often write  $b \rightarrow_{\mathcal{T}_0} C$  and  $B \rightarrow_{\mathcal{T}_0} c$  instead of  $\{b\} \rightarrow_{\mathcal{T}_0} C$  and  $B \rightarrow_{\mathcal{T}_0} \{c\}$ , respectively.

The *restriction* of an MAAF to a specific set of types  $\mathcal{T}_0$  is the AAF that is generated from the MAAF by considering only the attacks in  $\mathcal{T}_0$ . Formally, given an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ , the *restriction* of  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  to  $\mathcal{T}_0$  is an AAF  $\langle \mathcal{A}', \mathcal{R}' \rangle$ , where  $\mathcal{A}' = \mathcal{A}$  and  $\mathcal{R}' = \{(a, b) \mid (a, b, \tau) \in \mathcal{R} \text{ for some } \tau \in \mathcal{T}_0\}$ .

To define MAAF extensions, we introduce three new classes of semantics: *firm*, *restricted* and *loose*. For each type of semantics defined in [11] (e.g., admissible, complete, etc), we define its counterpart for each class (e.g., firmly admissible, restrictively stable, loosely complete, etc.). The three classes differ in how certain types of attack are considered. As already mentioned, the idea behind our semantics is the treatment of certain types of attacks as being conflict-generators only or attackers only. To do this, we consider a certain set of types, say  $\mathcal{T}_0$ , which are treated in the “normal” manner. Different types of semantics can now result depending on the exact behaviour of the attacks in  $\mathcal{T} \setminus \mathcal{T}_0$ . In particular:

1. *Firm semantics* (e.g., admissible, complete etc) w.r.t. a certain set of attack types (say  $\mathcal{T}_0$ ) requires a candidate extension to be defended against all types of attacks, and an attack can be defended only by attacks from  $\mathcal{T}_0$ . In other words, attacks in  $\mathcal{T}_0$  have the standard behaviour, but attacks in  $\mathcal{T} \setminus \mathcal{T}_0$  act as conflict-generators only, not as defenders. We call them firm because, while they allow any type of argument to unleash offensive attacks, they only



allow certain types of attack (those in  $\mathcal{T}_0$ ) to defend an argument, making its defense more difficult.

2. *Restricted semantics* (e.g., admissible, complete etc) w.r.t. a certain set of attack types (say  $\mathcal{T}_0$ ) require a candidate extension to be defended against attacks from  $\mathcal{T}_0$  only, and an attack can be defended only by attacks from  $\mathcal{T}_0$ . Thus, restricted semantics essentially consider only the attacks in  $\mathcal{T}_0$ , both for the attacks and for defending against them, i.e., attacks in  $\mathcal{T} \setminus \mathcal{T}_0$  are totally ignored.
3. *Loose semantics* (e.g., admissible, complete etc) w.r.t. a certain set of attack types (say  $\mathcal{T}_0$ ) are the most “relaxed” ones, as they require a candidate extension to be defended only against attacks from  $\mathcal{T}_0$ , while defense can happen by any type of attack. In other words, in loose semantics, attacks in  $\mathcal{T} \setminus \mathcal{T}_0$  are treated as defenders only, and cannot generate attacks. Loose semantics allows attacks to be ignored, so they may result to extensions that are not defended against all attacks, specifically against attacks that are of types not in  $\mathcal{T}_0$ .

In the following, we use shorthands to refer to the various types and classes of semantics. In particular, for the three classes of semantics, we use **fr** for firm, **re** for restricted, and **lo** for loose semantics. We also use  $\theta$  as a catch-all variable that refers to any of these classes. Similarly, for types of extensions, we use **cf** for conflict-free, **ad** for admissible, **co** for complete, **pr** for preferred, **gr** for grounded, and **st** for stable. We also use  $\sigma$  as a catch-all variable to indicate any of these extension types. For example, we write **fr-co**-extension to refer to a firmly complete extension, and  $\theta$ - $\sigma$ -extension to refer to an extension of class  $\theta$  and the type denoted by  $\sigma$ .

The definitions for the above semantics in MAAF's follow:

**Definition 35** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ , some  $\mathcal{T}_0 \subseteq \mathcal{T}$ , some  $a \in \mathcal{A}$  and some set  $\mathcal{E} \subseteq \mathcal{A}$ . We define the notion of defense for the different classes of semantics as follows:

- $\mathcal{E}$  firmly defends  $a$  (or **fr**-defends  $a$ ) w.r.t.  $\mathcal{T}_0$  if and only if  $\mathcal{E} \rightarrow_{\mathcal{T}_0} b$  whenever  $b \rightarrow a$
- $\mathcal{E}$  restrictively defends  $a$  (or **re**-defends  $a$ ) w.r.t.  $\mathcal{T}_0$  if and only if  $\mathcal{E} \rightarrow_{\mathcal{T}_0} b$  whenever  $b \rightarrow_{\mathcal{T}_0} a$
- $\mathcal{E}$  loosely defends  $a$  (or **lo**-defends  $a$ ) w.r.t.  $\mathcal{T}_0$  if and only if  $\mathcal{E} \rightarrow b$  whenever  $b \rightarrow_{\mathcal{T}_0} a$

**Definition 36** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}$ . For  $\theta \in \{\mathbf{fr}, \mathbf{re}, \mathbf{lo}\}$ , a set  $\mathcal{E} \subseteq \mathcal{A}$  is a  $\theta$ -**ad** extension w.r.t.  $\mathcal{T}_0$  (in words: firmly/restrictively/loosely admissible) if and only if:

- $\mathcal{E}$  is  $\theta$ -**cf**
- If  $a \in \mathcal{E}$ , then  $\mathcal{E}$   $\theta$ -defends  $a$  w.r.t.  $\mathcal{T}_0$

**Definition 37** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}$ . For  $\theta \in \{\mathbf{fr}, \mathbf{re}, \mathbf{lo}\}$ , a set  $\mathcal{E} \subseteq \mathcal{A}$  is a  $\theta$ -**co** extension w.r.t.  $\mathcal{T}_0$  (in words: firmly/restrictedly/loosely complete) if and only if:

- $\mathcal{E}$  is  $\theta$ -**ad**
- If  $\mathcal{E}$   $\theta$ -defends  $a$  w.r.t.  $\mathcal{T}_0$ , and  $\mathcal{E} \cup \{a\}$  is  $\theta$ -**cf** w.r.t.  $\mathcal{T}_0$ , then  $a \in \mathcal{E}$

**Definition 38** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a  $\theta$ -**gr** extension w.r.t.  $\mathcal{T}_0$  (in words: firmly/restrictedly/loosely grounded) if and only if  $\mathcal{E}$  is a minimal with respect to set inclusion  $\theta$ -**co** extension w.r.t.  $\mathcal{T}_0$ .

**Definition 39** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is a  $\theta$ -**pr** extension w.r.t.  $\mathcal{T}_0$  (in words: firmly/restrictedly/loosely preferred) if and only if  $\mathcal{E}$  is a maximal with respect to set inclusion  $\theta$ -**ad** extension w.r.t.  $\mathcal{T}_0$ .

**Definition 40** Consider an MAAF  $\langle \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$  and some  $\mathcal{T}_0 \subseteq \mathcal{T}$ . A set  $\mathcal{E} \subseteq \mathcal{A}$  is:

- A firmly stable extension (**fr-st**) w.r.t.  $\mathcal{T}_0$  if and only if:
  - $\mathcal{E}$  is maximally **fr-cf** w.r.t.  $\mathcal{T}_0$
  - $\mathcal{E} \rightarrow_{\mathcal{T}_0} a$  whenever  $a \notin \mathcal{E}$
- A restrictedly stable extension (**re-st**) w.r.t.  $\mathcal{T}_0$  if and only if:
  - $\mathcal{E}$  is maximally **re-cf** w.r.t.  $\mathcal{T}_0$
  - $\mathcal{E} \rightarrow_{\mathcal{T}_0} a$  whenever  $a \notin \mathcal{E}$
- A loosely stable extension (**lo-st**) w.r.t.  $\mathcal{T}_0$  if and only if:
  - $\mathcal{E}$  is maximally **lo-cf** w.r.t.  $\mathcal{T}_0$
  - $\mathcal{E} \rightarrow a$  whenever  $a \notin \mathcal{E}$

## B Proofs for Formal Properties

### Proof of Proposition 1.

Case #1 of the proposition is obvious by the fact that  $\mathcal{E}$  defends all its elements, as well as  $a$  and thus it defends all elements of  $\mathcal{E} \cup \{a\}$ . Also,  $\mathcal{E} \cup \{a\}$  is **cf**, by the hypothesis.

For case #2, again, we observe that  $\mathcal{E} \cup \{a\}$  defends all its elements, so it suffices to show that  $\mathcal{E} \cup \{a\}$  is **cf**. We note 4 different cases:

- If  $\mathcal{E} \rightarrow_c \mathcal{E}$ , then we get a contradiction, because  $\mathcal{E}$  is **cf**.
- If  $\mathcal{E} \rightarrow_c a$ , then, since  $\mathcal{E}$  defends  $a$ , it follows that  $\mathcal{E} \rightarrow_d \mathcal{E}$ , thus  $\mathcal{E} \rightarrow_c \mathcal{E}$ , a contradiction since  $\mathcal{E}$  is **cf**.
- If  $a \rightarrow_c \mathcal{E}$ , then, since  $\mathcal{E}$  is an **ad**-extension, it follows that  $\mathcal{E} \rightarrow_d a$ , thus  $\mathcal{E} \rightarrow_c a$ , a contradiction by the second bullet.

- If  $a \rightarrow_c a$ , then, since  $\mathcal{E}$  defends  $a$  it follows that  $\mathcal{E} \rightarrow_d a$ , i.e.,  $\mathcal{E} \rightarrow_c a$ , a contradiction again.

It follows that  $\mathcal{E} \cup \{a\}$  is **cf**, and thus an **ad**-extension.  $\square$

**Proof of Proposition 2.**

The first result is direct by combining Definition 5, Proposition 1 (second bullet), and the fact that if  $\mathcal{E}$  is **ad**, then it is **cf** by definition.

For the second result, note that if  $\mathcal{E}$  is **cf** and  $\mathcal{E} \rightarrow_d a$  whenever  $a \notin \mathcal{E}$ , then  $\mathcal{E} \rightarrow_c a$  whenever  $a \notin \mathcal{E}$ , thus  $\mathcal{E}$  is maximally **cf**. The result is now direct from Definition 8.  $\square$

**Proof of Proposition 3.**

We observe that, for any  $a, b \in \mathcal{A}$ :  $a \rightarrow_c b$  if and only if  $a \rightarrow_d b$  if and only if  $(a, b) \in \mathcal{R}$ . We also observe that Proposition 2 applies, since  $\mathcal{R}_d \subseteq \mathcal{R}_c$ . Combining the above facts with the definitions of the various semantics in Subsection 2.2 and in [11], the result follows trivially.  $\square$

**Proof of Proposition 4.**

For the first case, we observe that, for all  $a, b \in \mathcal{A}$ :

- $a \rightarrow_c b$  in the AAF if and only if  $a \rightarrow b$  in the MAAF
- $a \rightarrow_d b$  in the AAF if and only if  $a \rightarrow_{\mathcal{T}_0} b$  in the MAAF

Combining these two facts with the respective definitions in Subsection 2.2 and [21], the result follows trivially.

Analogously, for the second case and for all  $a, b \in \mathcal{A}$ :

- $a \rightarrow_c b$  in the AAF if and only if  $a \rightarrow_{\mathcal{T}_0} b$  in the MAAF
- $a \rightarrow_d b$  in the AAF if and only if  $a \rightarrow_{\mathcal{T}_0} b$  in the MAAF

Again, the result follows trivially from the respective definitions.

Finally, for the third case and for all  $a, b \in \mathcal{A}$ :

- $a \rightarrow_c b$  in the AAF if and only if  $a \rightarrow_{\mathcal{T}_0} b$  in the MAAF
- $a \rightarrow_d b$  in the AAF if and only if  $a \rightarrow b$  in the MAAF

Again, the result follows trivially from the respective definitions.  $\square$

**Proof of Proposition 5.**

#1, #2 and #3 are obvious by the respective definitions.

For #4, it suffices to show that if  $\sigma$  defends  $a$  and  $\mathcal{E} \cup \{a\}$  is **cf**, then  $a \in \sigma$ . Indeed, we observe that, under the above assumptions  $\mathcal{E} \cup \{a\}$  is an **ad**-extension, so if  $a \notin \sigma$ , then  $\mathcal{E} \cup \{a\} \supset \mathcal{E}$ , a contradiction by the fact that  $\mathcal{E}$  is a maximal **ad**-extension.

For #5, note that if  $\mathcal{E}$  is a **st**-extension, then it is **cf**, and also defends itself against all external attacks, thus it is **ad**. Furthermore, it is maximally **ad**, since it is maximally **cf**. Therefore,  $\mathcal{E}$  is a **pr**-extension.  $\square$

**Proof of Proposition 6.**

We will prove the result constructively. First, we will describe a construction over  $\mathcal{F}$ , and then we will show that this construction generates some  $\mathcal{E}$  with the properties required. The proof is broken down in steps, represented as claims proved individually below; the last claim (Claim 5) shows the result.

**Construction.** We assume a well-order  $\prec$  over  $\mathcal{A}$  (its existence is guaranteed by the Axiom of Choice). For a given set  $E \subseteq \mathcal{A}$ , we denote by  $\min_{\prec} E$  the minimal element of  $E$  according to  $\prec$ .

Moreover, for  $E \subseteq \mathcal{A}$ , set  $E^{\mathbf{v}} = \{a \in \mathcal{A} \setminus E \mid E \text{ defends } a, E \cup \{a\} \text{ is } \mathbf{cf}\}$ , i.e., the arguments that are defended by  $E$ , and do not conflict with  $E$ .

We define the function:  $\phi : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$  as follows:

$$\phi(E) = \begin{cases} E & , \text{ when } E^{\mathbf{v}} = \emptyset \\ E \cup \{\min_{\prec}(E^{\mathbf{v}})\} & , \text{ when } E^{\mathbf{v}} \neq \emptyset \end{cases}$$

Finally, we define a function  $\mathcal{G}$  recursively on the ordinals as follows:

$$\begin{aligned} \mathcal{G}(\beta) &= \mathcal{E}_* & , \text{ when } \beta = 0 \\ \mathcal{G}(\beta + 1) &= \phi(\mathcal{G}(\beta)) & , \text{ when } \beta \text{ is a successor ordinal} \\ \mathcal{G}(\beta) &= \bigcup \{\mathcal{G}(\gamma) \mid \gamma < \beta\} & , \text{ when } \beta \text{ is a limit ordinal} \end{aligned}$$

**Claim 1.** For two ordinals  $\beta, \gamma$ , if  $\beta < \gamma$ , then  $\mathcal{G}(\beta) \subseteq \mathcal{G}(\gamma)$ .

*Proof of Claim 1.* We will use transfinite induction on  $\gamma$ .

If  $\gamma = 0$ , then the result holds trivially as there is no  $\beta$  for which  $\beta < \gamma$ . Suppose that the result holds for all  $\gamma < \delta$ ; we will show that it holds for  $\gamma = \delta$ .

If  $\delta$  is a successor ordinal, then there exists some  $\delta^-$  such that  $\delta = \delta^- + 1$ . Clearly, by the definition of  $\mathcal{G}$  and  $\phi$ ,  $\mathcal{G}(\delta) \supseteq \mathcal{G}(\delta^-)$ . Furthermore, by the inductive hypothesis,  $\mathcal{G}(\delta^-) \supseteq \mathcal{G}(\beta)$ , which shows the result.

If  $\delta$  is a limit ordinal, then the result follows directly by the definition of  $\mathcal{G}$ .  $\circ$

**Claim 2.** For any ordinals  $\beta$ ,  $\mathcal{G}(\beta) \supseteq \mathcal{E}_*$ .

*Proof of Claim 2.* If  $\beta = 0$  the result follows by the definition of  $\mathcal{G}$ . If  $\beta > 0$ , the result follows by Claim 1.  $\circ$

**Claim 3.** For any ordinal  $\beta$ ,  $\mathcal{G}(\beta)$  is an **ad**-extension.

*Proof of Claim 3.* We will use transfinite induction over  $\beta$ . For  $\beta = 0$ , the result follows by our assumption on  $\mathcal{E}_*$ . Now suppose that it holds for all  $\beta < \gamma$ . We will show that it holds for  $\beta = \gamma$ .

If  $\gamma$  is a successor ordinal, then take  $\gamma^-$  such that  $\gamma = \gamma^- + 1$ . Then, by definition,  $\mathcal{G}(\gamma) = \phi(\mathcal{G}(\gamma^-))$ . By the inductive hypothesis  $\mathcal{G}(\gamma^-)$  is an **ad**-extension. Moreover, by the definition of  $\phi$ ,  $\phi(E)$  is an **ad**-extension whenever  $E$  is an **ad**-extension, so  $\mathcal{G}(\gamma)$  is an **ad**-extension.

If  $\gamma$  is a limit ordinal, then suppose that  $\mathcal{G}(\gamma)$  is not **cf**. Then, there exist  $a_1, a_2 \in \mathcal{G}(\gamma)$  such that  $\{a_1, a_2\}$  is not **cf**, and, thus, there exist ordinals  $\delta_1, \delta_2$  such that  $\delta_1 < \gamma$ ,  $\delta_2 < \gamma$ ,  $a_1 \in \mathcal{G}(\delta_1)$ ,  $a_2 \in \mathcal{G}(\delta_2)$ . If  $\delta_1 = \delta_2$  then  $\mathcal{G}(\delta_1)$  is not **cf**, a contradiction by the inductive hypothesis. If  $\delta_1 < \delta_2$  then  $\mathcal{G}(\delta_2) \supseteq \mathcal{G}(\delta_1)$  (by Claim 1), so  $a_1, a_2 \in \mathcal{G}(\delta_2)$ , a contradiction by the inductive hypothesis. The case of

$\delta_2 < \delta_1$  is analogous. Thus,  $\mathcal{G}(\gamma)$  is **cf**.

Now consider some  $a \in \mathcal{G}(\gamma)$ . Then, by the definition of  $\mathcal{G}$ , there exists some  $\delta < \gamma$  such that  $a \in \mathcal{G}(\delta)$ . Since  $\mathcal{G}(\delta)$  is an **ad**-extension by the inductive hypothesis, it follows that  $\mathcal{G}(\delta)$  defends  $a$ , so, given that  $\mathcal{G}(\gamma) \supseteq \mathcal{G}(\delta)$  (Claim 1), we conclude that  $\mathcal{G}(\gamma)$  defends  $a$ . Thus,  $\mathcal{G}(\gamma)$  is an **ad**-extension.  $\circ$

**Claim 4.** There exists ordinal  $\beta$  such that  $\mathcal{G}(\beta) = \mathcal{G}(\beta + 1)$ .

*Proof of Claim 4.* By Claim 1, we conclude that  $\mathcal{G}$  is an increasing function from the ordinals into  $2^A$ . It cannot be strictly increasing, as if it were we would have an injective function from the ordinals into a set, violating Hartogs' lemma. Therefore the function must be eventually constant, so for some  $\beta$ ,  $\mathcal{G}(\beta) = \mathcal{G}(\beta + 1)$ .  $\circ$

**Claim 5.** There exists some  $\mathcal{E}$  such that  $\mathcal{E} \supseteq \mathcal{E}_*$ , and the following hold:

1.  $\mathcal{E}$  is a **co**-extension.
2. For any  $\mathcal{E}'$  such that  $\mathcal{E}_* \subseteq \mathcal{E}' \subset \mathcal{E}$ , there exists  $a \in \mathcal{E} \setminus \mathcal{E}'$  which is defended by  $\mathcal{E}'$  and  $\mathcal{E}' \cup \{a\}$  is **cf**.
3. For any  $\mathcal{E}'$  such that  $\mathcal{E}_* \subseteq \mathcal{E}' \subset \mathcal{E}$ ,  $\mathcal{E}'$  is not a **co**-extension.

*Proof of Claim 5.* By Claim 4, there exists ordinal  $\beta$  such that  $\mathcal{G}(\beta) = \mathcal{G}(\beta + 1)$ . Set  $\mathcal{E} = \mathcal{G}(\beta)$ . By Claim 2,  $\mathcal{E} \supseteq \mathcal{E}_*$ , so it is an adequate choice. We will show that  $\mathcal{E}$  satisfies the required properties.

For the first result, note that by Claim 3,  $\mathcal{E}$  is an **ad**-extension. Moreover,  $\mathcal{E} = \mathcal{G}(\beta) = \mathcal{G}(\beta + 1) = \phi(\mathcal{G}(\beta)) = \phi(\mathcal{E})$ , which implies that  $\mathcal{E}^\blacktriangleright = \emptyset$ , which, in tandem with the fact that  $\mathcal{E}$  is an **ad**-extension leads to the conclusion that  $\mathcal{E}$  is a **co**-extension.

For the second result, take some  $\mathcal{E}'$  such that  $\mathcal{E}_* \subseteq \mathcal{E}' \subset \mathcal{E}$ .

Set  $S = \{\gamma \mid \mathcal{G}(\gamma) \not\subseteq \mathcal{E}'\}$ . We observe that  $\beta \in S$ , so  $S \neq \emptyset$ . Set  $\delta = \min_{<} S$ . Obviously,  $\delta = \beta$  or  $\delta < \beta$ .

If  $\delta = 0$ , then  $\mathcal{G}(\delta) = \mathcal{E}_* \subseteq \mathcal{E}'$ , a contradiction.

If  $\delta$  is a successor ordinal, then take  $\delta^-$  such that  $\delta = \delta^- + 1$ . Thus,  $\mathcal{G}(\delta) = \phi(\mathcal{G}(\delta^-))$ . By construction,  $\mathcal{G}(\delta^-) \subseteq \mathcal{E}'$  and  $\mathcal{G}(\delta) \not\subseteq \mathcal{E}'$ , therefore  $\mathcal{G}(\delta) = \mathcal{G}(\delta^-) \cup \{a\}$ , for some  $a$  for which  $\mathcal{G}(\delta^-)$  defends  $a$  and  $\mathcal{G}(\delta^-) \cup \{a\}$  is **cf**. If  $a \in \mathcal{E}'$ , then  $\mathcal{G}(\delta) \subseteq \mathcal{E}'$ , a contradiction by the choice of  $\delta$ , so  $a \notin \mathcal{E}'$ . Moreover,  $a \in \mathcal{G}(\delta)$ . If  $\delta = \beta$  then  $\mathcal{G}(\delta) = \mathcal{E}$ , so  $a \in \mathcal{E}$ . If  $\delta < \beta$  then  $a \in \mathcal{G}(\delta) \subseteq \mathcal{G}(\beta)$  (by Claim 1), so  $a \in \mathcal{E}$ . We conclude that  $a \in \mathcal{E} \setminus \mathcal{E}'$ . Thus, we have found some  $a$  with the required properties. If  $\delta$  is a limit ordinal, then, by the definition of  $\delta$ ,  $\mathcal{G}(\delta') \subseteq \mathcal{E}'$  for all  $\delta' < \delta$ . Therefore,  $\mathcal{G}(\delta) = \bigcup_{\delta' < \delta} \mathcal{G}(\delta') \subseteq \mathcal{E}'$ , a contradiction by the choice of  $\delta$ .

The third result follows from the second: indeed, as there exists  $a \in \mathcal{E} \setminus \mathcal{E}'$  which is defended by  $\mathcal{E}'$  and  $\mathcal{E}' \cup \{a\}$  is **cf**, it cannot be the case that  $\mathcal{E}'$  is a **co**-extension.  $\circ$

□

### **Proof of Proposition 7.**

We first observe that  $\emptyset$  is **cf** and an **ad**-extension, so the claim is true for  $\sigma \in \{\mathbf{cf}, \mathbf{ad}\}$ .

Let us now turn our attention to the case where  $\sigma = \mathbf{pr}$ . Our proof follows the

lines of the respective proof in [4]. Set  $\mathcal{AD} = \{\mathcal{E} \mid \mathcal{E} \text{ is an ad-extension}\}$  ( $\mathcal{AD} \neq \emptyset$ , as shown above). We will show that any  $\subseteq$ -chain  $(\mathcal{E}_i)_{i \in I}$  (for some appropriate set of indexes  $I$ ) in  $\mathcal{AD}$  possesses an upper bound. Indeed, set  $\mathcal{E} = \bigcup \mathcal{E}_i$ . Obviously  $\mathcal{E} \supseteq \mathcal{E}_i$ , so it is an upper bound; it remains to show that  $\mathcal{E} \in \mathcal{AD}$ , i.e., that  $\mathcal{E}$  is an **ad**-extension.

Suppose that  $\mathcal{E}$  is not **cf**, i.e., there exist  $a_1, a_2 \in \mathcal{E}$  such that  $a \rightarrow_c b$ . By the definition of  $\mathcal{E}$ , there exist  $\mathcal{E}_i, \mathcal{E}_j$  such that  $a_1 \in \mathcal{E}_i, a_2 \in \mathcal{E}_j$  for some  $i, j \in I$ . It is the case that  $\mathcal{E}_i \subseteq \mathcal{E}_j$  or  $\mathcal{E}_j \subseteq \mathcal{E}_i$ , so suppose, without loss of generality, that  $\mathcal{E}_i \subseteq \mathcal{E}_j$ . Then  $a_1, a_2 \in \mathcal{E}_j$ , a contradiction, since  $\mathcal{E}_j$  is an **ad**-extension (thus **cf**). Thus,  $\mathcal{E}$  is **cf**. It remains to show that  $\mathcal{E}$  defends all  $a \in \mathcal{E}$ . Indeed, take some  $a \in \mathcal{E}$ . Then,  $a \in \mathcal{E}_i$  for some  $i \in I$ , and, thus  $\mathcal{E}_i$  defends  $a$ , which implies that  $\mathcal{E}$  defends  $a$ , since  $\mathcal{E} \supseteq \mathcal{E}_i$ . Thus, any  $\subseteq$ -chain  $(\mathcal{E}_i)_{i \in I}$  in  $\mathcal{AD}$  possesses an upper bound, which, by Zorn's Lemma, implies that  $\mathcal{AD}$  has a maximal element, i.e., that there exists a **pr**-extension.

By proposition 5, the existence of a **pr**-extension implies that there exists a **co**-extension as well.

For **gr**-extensions, note that  $\emptyset$  is an **ad**-extension, so applying Proposition 6 for  $\mathcal{E}_* = \emptyset$  we ensure the existence of some  $\mathcal{E}$  which is minimally **co**-extension, i.e.,  $\mathcal{E}$  is a **gr**-extension.  $\square$

### **Proof of Proposition 8.**

Given that  $\emptyset$  is an **ad**-extension, we can apply Proposition 6 for  $\mathcal{E}_* = \emptyset$  to get some  $\mathcal{E}$  which is a minimal **co**-extension, i.e.,  $\mathcal{E}$  is a **gr**-extension. Now suppose that there is a second **gr**-extension, say  $\mathcal{E}'$  ( $\mathcal{E}' \neq \mathcal{E}$ ). Obviously,  $\mathcal{E} \not\subseteq \mathcal{E}'$  and  $\mathcal{E}' \not\subseteq \mathcal{E}$ . Set  $\mathcal{E}_0 = \mathcal{E} \cap \mathcal{E}'$ . It follows that  $\emptyset \subseteq \mathcal{E}_0 \subset \mathcal{E}$ , so by Proposition 6 again there exists some  $a \in \mathcal{E} \setminus \mathcal{E}_0$  which is defended by  $\mathcal{E}_0$  and  $\mathcal{E}_0 \cup \{a\}$  is **cf**. Moreover,  $\mathcal{E}_0 \subset \mathcal{E}'$ , so  $a$  is defended by  $\mathcal{E}'$ . But  $\mathcal{E}'$  is a **gr**-extension, thus a **co**-extension and also  $\mathcal{E}'$  defends  $a$ . Thus, given that  $\mathcal{R}_d \subseteq \mathcal{R}_c$ , we can apply Proposition 2 (bullet #1) to conclude that  $a \in \mathcal{E}'$ , a contradiction by the choice of  $a$ .  $\square$

### **Proof of Proposition 9.**

The result is direct by the fact that the relations  $\rightarrow_c, \rightarrow_d$  are identical in the two frameworks.  $\square$

### **Proof of Proposition 10.**

For #1: suppose that there exist  $a, b \in \mathcal{E}$  such that  $(a, b) \in \mathcal{R}_c^2$ , then  $(a, b) \in \mathcal{R}_c^1$  so  $\mathcal{E}$  is not **cf** w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ , a contradiction.

For #2: consider some  $b \in \mathcal{A}$  such that  $(b, a) \in \mathcal{R}_c^2$ . Then,  $(b, a) \in \mathcal{R}_c^1$ , so, by the hypothesis, there exists  $c \in \mathcal{E}$  such that  $(c, b) \in \mathcal{R}_d^1 \subseteq \mathcal{R}_d^2$ , which proves the result.

For #3: direct from the first two bullets.

For #4: it suffices to show that if  $\mathcal{E}$  defends  $a$  w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  and  $\mathcal{E} \cup \{a\}$  is **cf** w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$  then  $a \in \mathcal{E}$ . Indeed, by the second result, we conclude that  $\mathcal{E}$  defends  $a$  w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , whereas by the first result we conclude that  $\mathcal{E} \cup \{a\}$  is **cf** w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ . Given that  $\mathcal{E}$  is a **co**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , it follows that  $a \in \mathcal{E}$ , which proves the result.

For #5: suppose, for the sake of contradiction, that  $\mathcal{E}$  is not a **pr**-extension w.r.t.  $\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ . Then, there exists some  $\mathcal{E}' \supset \mathcal{E}$ , such that  $\mathcal{E}'$  is an **ad**-extension w.r.t.

$\langle \mathcal{R}_c^1, \mathcal{R}_d^1 \rangle$ . By the second result,  $\mathcal{E}'$  is also an **ad**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ , a contradiction by the fact that  $\mathcal{E}$  is a **pr**-extension w.r.t.  $\langle \mathcal{R}_c^2, \mathcal{R}_d^2 \rangle$ .  
 For #6: direct by the fact that  $\mathcal{R}_d^1 \subseteq \mathcal{R}_d^2$  and the hypotheses.  $\square$

**Proof of Proposition 11.**

It follows directly from the definitions of **co**, **gr**, **pr** and **st** semantics, according to which, whether a set of arguments  $\mathcal{E}$  is an extension of an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  or not depends only on its attack relation,  $\mathcal{R}$ , and the classification of the attacks,  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , and not on the names or labels of the arguments in  $\mathcal{R}$ .  $\square$

**Proof of Proposition 12.**

It follows directly from the definitions of **co**, **gr**, **pr** and **st** semantics, according to which every **co**, **gr**, **pr** or **st** extension must be **cf**.  $\square$

**Proof of Proposition 13.**

Admissibility follows directly from the definitions of **co**, **gr** and **pr** semantics, according to which every **co**, **gr** or **pr** extension must be **ad**. According to the definition of **st** semantics, every **st**-extension is **cf** and defends itself against all arguments it doesn't contain; it is, therefore, **ad**. Since admissibility is a special case of defense, defense is also satisfied by **co**, **gr**, **pr** and **st** semantics.  $\square$

We introduce the following lemma to simplify the proofs of some of the remaining propositions.

**Lemma 1** *Any principle that is not satisfied by the  $\sigma$ -semantics of AAF is not also satisfied by the  $\sigma$ -semantics of AAFs with Attack Classification.*

**Proof of Lemma 1.**

Consider a semantics of AAFs  $\sigma$  and a principle  $p$  not satisfied by  $\sigma$ . This means that there is an AAF,  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  and a  $\sigma$ -extension of it  $\mathcal{E}$  not satisfying the requirements of  $p$ . From Proposition 3,  $\mathcal{E}$  is also a  $\sigma$ -extension of  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  with attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , where  $\mathcal{R}_c = \mathcal{R}_d = \mathcal{R}$ . Since  $\mathcal{E}$  does not satisfy the requirements of  $p$ , the  $\sigma$  semantics of AAFs with Attack Classification violate  $p$ .  $\square$

**Proof of Proposition 14.**

As shown in [3], strong admissibility is not satisfied by the **co**, **pr** and **st** semantics of AAFs. By Lemma 1, it is also not satisfied by the **co**, **pr** and **st** semantics of AAFs with Attack Classification. For the **gr** semantics, consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  with Attack Classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , where:  $\mathcal{A} = \{a, b, c\}$ ,  $\mathcal{R}_c = \{(a, b), (a, c)\}$  and  $\mathcal{R}_d = \{(c, a)\}$ . Its grounded extensions are  $\mathcal{E}_1 = \{a\}$  and  $\mathcal{E}_2 = \{b, c\}$  and  $\mathcal{E}_2$  does not strongly defend  $c$ . Therefore, strong admissibility is not satisfied by **gr** semantics either.  $\square$

**Proof of Proposition 15.**

As shown in [20], naivety is not satisfied by the **co**, **gr** and **pr** semantics of AAFs. By Lemma 1, it is also not satisfied by the **co**, **gr** and **pr** semantics of AAFs with Attack Classification. For the **st** semantics, naivety follows directly from its definition.  $\square$

**Proof of Proposition 16.**

As shown in [10], indirect conflict-freeness is not satisfied by any of **co**, **gr**, **pr** and **st** semantics of AAFs. By Lemma 1, it is also not satisfied by any of **co**, **gr**, **pr** and **st** semantics of AAFs with Attack Classification.  $\square$

**Proof of Proposition 17.**

Consider the AAF shown in Figure 1.  $\{b, c, d\}$  is a complete, grounded, preferred and stable extension, which defends  $a$  but does not contain it. Therefore, reinstatement is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.  $\square$

**Proof of Proposition 18.**

In the AAF shown in Figure 1,  $\{b, c, d\}$  strongly defends  $a$  but does not contain it. Therefore, weak reinstatement is not satisfied by any of **co**, **gr**, **pr** and **st** semantics.  $\square$

**Proof of Proposition 19.**

$\mathcal{CF}$ -reinstatement follows directly from the definitions of **co**, **gr** and **pr** semantics. For the **st** semantics: Suppose there is a **st** extension  $\mathcal{E}$  and an argument  $a$  such that  $\mathcal{E} \cup \{a\}$  is conflict-free and  $a$  does not belong to  $\mathcal{E}$ . This cannot be the case because  $\mathcal{E}$  would not be maximally **cf**. Therefore,  $\mathcal{CF}$ -reinstatement is also satisfied by the **st** semantics.  $\square$

**Proof of Proposition 20.**

As shown in [3], I-maximality is not satisfied by the **co** semantics of AAFs. By Lemma 1, it is also not satisfied by the **co** semantics of AAFs with Attack Classification. I-maximality follows directly from the definitions of **gr**, **pr** and **st** semantics.  $\square$

**Proof of Proposition 21.**

Consider the AAF shown in Figure 1.  $\{b, c, d\}$  and  $\{a, d\}$  are **co**, **gr** and **pr** extensions. Argument  $b$  is contained in  $\{b, c, d\}$  and attacked by  $\{a, d\}$  and there is no other (**co**, **gr** or **pr**) extension. The non-satisfiability of allowing abstention by the **st** semantics follows directly from its definition, which requires that every **st** extension must be maximally **cf**.  $\square$

The following lemma has been proved for AAFs in [1]. We will prove that this result holds also for AAFs with Attack Classification. This lemma will be used for proving the remaining propositions.

**Lemma 2** *For all semantics of AAFs with Attack Classification: (i) directionality implies non-interference and (ii) non-interference implies crash resistance.*

**Proof of Lemma 2.**

(i) follows directly from Definitions 28 and 30, according to which every isolated set is also an unattacked set.

For (ii), suppose that for a semantics  $\sigma$  that satisfies non-interference, there is a contaminating argumentation framework  $\mathcal{F}^* = (\mathcal{A}^*, \mathcal{R}^*)$ . Then for any framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  disjoint from  $\mathcal{F}^*$  it holds that  $\sigma(\mathcal{F}^*) = \sigma(\mathcal{F}^* \cup \mathcal{F})$ . However,



$\sigma$  satisfies non-interference and  $\mathcal{A}$  is isolated in  $\mathcal{F}^* \cup \mathcal{F}$ , thus:  $\sigma((\mathcal{F}^* \cup \mathcal{F})_{\downarrow \mathcal{A}}) = \{\mathcal{E} \cap \mathcal{A} \mid \mathcal{E} \in \sigma(\mathcal{F}^* \cup \mathcal{F})\}$ .  $\mathcal{F}^* = (\mathcal{A}^*, \mathcal{R}^*)$  and  $\mathcal{F} = (\mathcal{A}, \mathcal{R})$  are disjoint, therefore,  $\sigma((\mathcal{F}^* \cup \mathcal{F})_{\downarrow \mathcal{A}}) = \sigma(\mathcal{F})$  and by the initial supposition,  $\mathcal{F}^*$  and  $\mathcal{F}$  have the same set of  $\sigma$ -extensions. Therefore, the previous equation implies:  $\sigma(\mathcal{F}) = \{\mathcal{E} \cap \mathcal{A} \mid \mathcal{E} \in \sigma(\mathcal{F}^*)\}$ , and, since  $\mathcal{F}^*$  and  $\mathcal{F}$  are disjoint,  $\sigma(\mathcal{F}) = \sigma(\mathcal{F}^*)$ . This is obviously a contradiction, unless  $\sigma$  returns only the empty set as an extension for any framework, which is not the case for any of the semantics that we study.  $\square$

**Proof of Proposition 22.**

By Proposition 23 and Lemma 2, **co**, **gr** and **pr** semantics satisfy crash resistance. As shown in [8], crash resistance is not satisfied by the **st** semantics of AAFs. By Lemma 1, it is also not satisfied by the **st** semantics of AAFs with Attack Classification.  $\square$

**Proof of Proposition 23.**

We show that for  $\sigma = \mathbf{co}$ , for any argumentation framework  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  with any classification of attacks  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and for any isolated set  $S \subseteq \mathcal{A}$ :  $\sigma(\mathcal{F}_{\downarrow S}) = \{\mathcal{E} \cap S \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ . Let  $S_1$  an isolated set of  $\mathcal{F}$  and  $\mathcal{E}_1$  a **co**-extension of  $\mathcal{F}_{\downarrow S_1}$ . Then  $S_2 = \mathcal{A} \setminus S_1$  should also be isolated, since it doesn't attack or receive any attacks from  $S_1$ . Let  $\mathcal{E}_2$  be a **co**-extension of  $\mathcal{F}_{\downarrow S_2}$ . For both  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , it holds that they are conflict-free and they defend all arguments they contain from attacks from  $S_1$  (resp.  $S_2$ ). Since there are no attacks between  $S_1$  and  $S_2$ , both  $\mathcal{E}_1$  and  $\mathcal{E}_2$  defend all arguments they contain from attacks from  $\mathcal{A}$  and  $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$  is conflict-free.  $\mathcal{E}$  is, therefore, admissible in  $\mathcal{F}$ . Furthermore,  $\mathcal{E}_1$  (resp.  $\mathcal{E}_2$ ) cannot defend any argument in  $S_2$  (resp.  $S_1$ ). Therefore,  $\mathcal{E}$  contains all arguments in  $\mathcal{A}$  that it defends from attacks from  $\mathcal{A}$ . It is, therefore, a **co**-extension of  $\mathcal{F}$  and  $\mathcal{E}_1 = \mathcal{E} \cap S_1$ . Since this holds for any extension of  $S_1$ ,  $\mathbf{co}(\mathcal{F}_{\downarrow S_1}) \subseteq \{\mathcal{E} \cap S_1 \mid \mathcal{E} \in \mathbf{co}(\mathcal{F})\}$ .

Consider now any **co**-extension of  $\mathcal{F}$ ,  $\mathcal{E}'$ , such that  $\mathcal{E}'_1 = \mathcal{E}' \cap S_1$ .  $\mathcal{E}'$  is conflict-free and defends all arguments it contains from attacks from  $\mathcal{A}$ , therefore, given that there are no attacks between  $S_1$  and  $S_2 = \mathcal{A} \setminus S_1$ ,  $\mathcal{E}'_1$  is conflict-free and defends all arguments it contains from attacks from  $S_1$ . This means that  $\mathcal{E}'_1$  is admissible in  $\mathcal{F}_{\downarrow S_1}$ . Since  $\mathcal{E}'$  is complete, there is no argument  $a_1$  in  $\mathcal{A} \setminus \mathcal{E}'$  and therefore also in  $S_1 \setminus \mathcal{E}'_1$ , such that  $\mathcal{E}'$  defends  $a_1$  and  $\mathcal{E}' \cup \{a_1\}$  is conflict-free. Given that there are no attacks between  $S_1$  and  $S_2$ , this means that there is no argument  $a_1 \in S_1 \setminus \mathcal{E}'_1$  such that  $\mathcal{E}'_1$  defends  $a_1$  and  $\mathcal{E}'_1 \cup \{a_1\}$  is conflict-free.  $\mathcal{E}'_1$  is, therefore, a **co**-extension of  $\mathcal{F}_{\downarrow S_1}$ . Since this holds for any **co**-extension of  $\mathcal{F}$ ,  $\mathbf{co}(\mathcal{F}_{\downarrow S_1}) \supseteq \{\mathcal{E} \cap S_1 \mid \mathcal{E} \in \mathbf{co}(\mathcal{F})\}$ . The **co** semantics, therefore, satisfies non-interference.

Using a similar proof process, it can be easily verified that **pr** and **gr** semantics also satisfy non-interference.

As shown in [3], non-interference is not satisfied by the **st** semantics of AAFs. By Lemma 1, it is also not satisfied by the **st** semantics of AAFs with Attack Classification.  $\square$

**Proof of Proposition 24.**

From Definitions 31, 32 and 33, it follows that a semantics  $\sigma$  satisfies directionality if and only if  $\sigma$  satisfies both weak- and semi-directionality. Since, by Propositions

25 and 26, none of the **co**, **gr**, **pr** and **st** semantics satisfy both weak- and semi-directionality, they do not satisfy directionality either.  $\square$

**Proof of Proposition 25.**

For **co**, **gr** and **pr** semantics: Consider an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$  with attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$ , where:  $\mathcal{A} = \{a, b, c\}$ ,  $\mathcal{R}_c = \{(b, a)\}$  and  $\mathcal{R}_d = \{(c, b)\}$ . Consider the unattacked set  $U = \{b, c\}$ . The only **co**-, **pr**- and **gr**-extension of  $\mathcal{F}_{\downarrow U}$  is  $\mathcal{E} = \{b, c\}$ .  $\mathcal{F}$  has two **co**-, **pr**- and **gr**-extensions:  $\mathcal{E}_1 = \{a, b\}$  and  $\mathcal{E}_2 = \{a, c\}$ . Neither  $\mathcal{E}_1 \cap U$  nor  $\mathcal{E}_2 \cap U$  is equal to  $\mathcal{E}$ , the only extension of  $\mathcal{F}_{\downarrow U}$ . Therefore, **co**, **gr** and **pr** semantics do not satisfy weak directionality.

For **st** semantics: Consider any **st**-extension,  $\mathcal{E}$ , of an AAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ , equipped with some attack classification  $\langle \mathcal{R}_c, \mathcal{R}_d \rangle$  and an unattacked set of arguments  $U \in \mathcal{US}(\mathcal{F})$ . By definition,  $\mathcal{E}$  is maximal among the conflict-free subsets of  $\mathcal{A}$ .  $\mathcal{E} \cap U$  is, therefore, maximal among the conflict-free subsets of  $U$ .  $\mathcal{E}$  attacks (with a defending attack) all arguments that are not in  $\mathcal{E}$ , therefore  $\mathcal{E} \cap U$  attacks (with a defending attack) all arguments in  $U \setminus \{\mathcal{E} \cap U\}$ .  $\mathcal{E} \cap U$ , therefore, fulfils all the conditions for being a **st**-extension of  $\mathcal{F}_{\downarrow U}$ , which means that for  $\sigma = \mathbf{st}$ , for every  $U \in \mathcal{US}(\mathcal{F})$  it holds that  $\sigma(\mathcal{F}_{\downarrow U}) \supseteq \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}$ . Weak directionality is, therefore, satisfied by the **st** semantics.  $\square$

**Proof of Proposition 26.**

The example used in the Proof of Proposition 25 is a proof that **co**, **gr** and **pr** semantics do not satisfy semi-directionality.

By Propositions 24 and 25, **st** semantics does not satisfy directionality but satisfies weak directionality. Given that a semantics  $\sigma$  satisfies directionality if and only if  $\sigma$  satisfies both weak- and semi-directionality, **st** semantics does not satisfy semi-directionality.  $\square$

## C ASP Code for Web App

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Conflict free Extension (Def 3)

% Generate all possible extensions
{cf_extension(A) : argument(A)}.

% Generate the Answer Sets without
% internal conflict-generating attacks
:- cf_extension(A1), cf_extension(A2),
attacks(A1, A2, conflict).

#show cf_extension/1.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

% Admissible Extension (Def 4)
% Import code for computation of cf_extensions

% Consider all cf extensions
ad_extension(A) :-
    cf_extension(A).

% If an argument (a) of a cf set (S)
% receives an attack
argumentGetsConfAttack(A) :-
    argument(A),
    attacks(_, A, conflict).

% The argument (a) must be defended
% by the set (S)
defendedByExtension(A) :-
    argument(A),
    not argumentGetsConfAttack(A).
defendedByExtension(A) :-
    argument(A),
    attacks(A1, A, conflict),
    attacks(A2, A1, defending), ad_extension(A2).

% Remove the cf sets that contain an argument (a)
% not defended by the set (S)
:- ad_extension(A), not defendedByExtension(A).

#show ad_extension/1.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Complete Extension (Def 5)
% Import code for computation of ad_extensions

% Gather all admissible extensions
co_extension(A) :-
    ad_extension(A).

% Given an admissible set (S) and an argument (a)
% then  $S \cup \{a\}$  must be cf (*)
attacksExtension(A) :-
    attacks(A, A1, conflict),
    co_extension(A1).
getsAttackedByExtension(A) :-
    attacks(A1, A, conflict),
    co_extension(A1).

```

```

% If S defends a. Then, a belongs in S (**)
compliantArg(A) :-
    defenedByExtension(A),
    not getsAttackedByExtension(A),
    not attacksExtension(A).

% Remove the sets admissible sets which
% do not obey (*) and (**)
:- compliantArg(A), not co_extension(A).
#show co_extension/1.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Grounded Extension (Def 6)
% Import code for computation of co_extensions

% Find the smallest (w.r.t subset relation) co extension(s)
#preference(p1,subset){co_extension(A)}.
#optimize(p1).

gr_extension(A) :- co_extension(A).
#show gr_extension/1.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Preferred Extension (Def 7)
% Import code for computation of ad_extensions

% Find the biggest (w.r.t subset relation) ad extension(s)
#preference(p1,superset){ad_extension(A)}.
#optimize(p1).

pr_extension(A) :- ad_extension(A).
#show pr_extension/1.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Stable Extension (Def 8)
% Import code for computation of cf_extensions

% Find the biggest (w.r.t subset relation) cf extension(s)
#preference(p1,superset){cf_extension(A)}.
#optimize(p1).

% The argument(s) (A) not contained in

```

```
% the biggest cf extension(s) must receive
% a defending attacks

other(A) :- argument(A), not cf_extension(A),
             cf_extension(B), attacks(B,A, defending).
:- argument(A), not other(A), not cf_extension(A).

st_extension(A) :- cf_extension(A).
#show st_extension/1.
```