An Argumentation Framework with Backing and Undercutting

Andrea Cohen, Alejandro J. García, and Guillermo R. Simari

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Artificial Intelligence Research and Development Laboratory (LIDIA)
Department of Computer Science and Engineering (DCIC)
Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina
e-mail: {ac,ajg,grs}@cs.uns.edu.ar

Abstract. In this work we will combine two important notions for the argumentation community into Abstract Argumentation Frameworks (AFs). These notions correspond to Toulmin's backings and Pollock's undercutting defeaters. We will define Backing-Undercutting Argumentation Frameworks (BUAFs), an extension of AFs that includes a specialized support relation, a distinction between different attack types, and a preference relation among arguments. Thus, BUAFs will provide a more concrete approach to represent argumentative or non-monotonic scenarios where information can be attacked and supported.

1 Introduction

The study of argumentation within the field of Artificial Intelligence has grown lately [3]. Briefly, argumentation is a form of reasoning where a claim is accepted or rejected according to the analysis of the arguments for and against it. Then, argumentation provides a reasoning mechanism where contradictory, incomplete and uncertain information may appear. In the last decade several approaches were proposed to model argumentation on an abstract basis [7], using classical logics [4], or using logic programming [8].

Argumentation models usually consider an argument as a piece of reasoning that provides a connection between some premises and a conclusion. Notwithstanding, in [13] Toulmin argued that arguments had to be analyzed using a richer format than the traditional one of formal logic. Whereas a formal logic analysis uses the dichotomy of premises and conclusion, Toulmin proposed a model for the layout of arguments that in addition to data and claim distinguishes four elements: warrant, backing, rebuttal and qualifier. However, Toulmin did not elaborate much on the nature of rebuttals, but simply stated that they provide conditions of exception for the argument. That is, without loss of generality, the notion of rebuttal can be paired to the notion of defeater for an argument, as proposed in the literature [12].

An important contribution to the field of argumentation which regards the nature of defeaters was proposed by Pollock. In [10] Pollock stated that defeasible reasons (which can be assembled to comprise arguments) have defeaters and that there are two kinds of defeaters: rebutting defeaters and undercutting defeaters. The former attack the conclusion of an inference by supporting the opposite one (*i. e.* they are reasons for denying the conclusion), while the latter attack the connection between the premises and conclusion without attacking the conclusion directly.

In this work, we will combine the notions presented by Toulmin and Pollock into an abstract argumentation framework. We will incorporate Pollock's categorization of defeaters and the modeling of Toulmin's scheme elements, in particular, focusing in undercutting defeaters and backings. We will follow the approach of [6] in which Pollock's undercutting defeaters can be regarded as attacking Toulmin's warrants. Thus, Toulmin's backings can be regarded as aiming to defend their associated warrants against undercutting attacks, by providing support for them. In that way, we will be able to capture both attack and support for an inference, that is, for Toulmin's warrants.

We will extend Abstract Argumentation Frameworks (AFs) [7] to incorporate a specialized type of support and preference relation among arguments, as well as distinguishing between different types of attacks. In particular, the support relation will correspond to the support that Toulmin's backings provide for their associated warrants. On the other hand, we will distinguish three different types of attack within Dung's original attack relation, more specifically, rebutting attacks, undermining attacks and undercutting attacks; the former and the latter being related to rebutting and undercutting defeaters, as proposed by Pollock. The remaining type of attack we will consider corresponds to undermining defeaters, which are widely considered in the literature (see *e. g.* [11]) and originate from attacks to an argument's premise.

The rest of this work is organized as follows. Section 2 briefly reviews Dung's Abstract Argumentation Frameworks (AFs). In Section 3 we present Backing-Undercutting Argumentation Frameworks (BUAFs), an extension of AFs that incorporates attack and support for inferences, as well as a preference relation to decide between conflicting arguments. In Section 4 we introduce the different types of defeat that can be obtained from a BUAF by applying preferences to the conflicting arguments as indicated by the attack relation. Later we define the requirements for conflict-free sets of arguments in a BUAF. Section 5 introduces some semantic notions, followed by the formal definitions of the acceptability semantics for BUAFs. Finally, in Section 6 some conclusions and related work are discussed.

2 Dung's Abstract Argumentation Frameworks

In this section we will briefly review Dung's Abstract Argumentation Frameworks, as defined in [7].

Definition 1. *An* Abstract Argumentation Framework (AF) *is a pair* $\langle Args, \mathbb{R} \rangle$*, where* $Args$ *is a set of arguments and* $\mathbb{R} \subseteq Args \times Args$ *is an attack relation.*

Here, arguments are abstract entities that will be denoted using calligraphic uppercase letters. No reference to the underlying logic is needed since we are abstracting from argument's structure. The attack relation between two arguments $\mathcal{A}$ and $\mathcal{B}$ denotes the fact that these arguments cannot be accepted simultaneously since they contradict each other. We say that an argument $\mathcal{A}$ *attacks* an argument $\mathcal{B}$ iff $(\mathcal{A}, \mathcal{B}) \in \mathbb{R}$, and it is noted as $\mathcal{A} \rightarrow \mathcal{B}$. For instance, in the AF of Figure 1 $\mathcal{A}$ and $\mathcal{B}$ attack each other, $\mathcal{B}$ attacks $\mathcal{C}$, and so on.

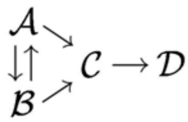

Fig. 1. A Dung's Abstract Argumentation Framework.

Dung then defines the acceptability of arguments and the admissible sets of the framework.

Definition 2. *Let AF= $\langle Args, \mathbb{R} \rangle$ and $S \subseteq Args$ a set of arguments. Then:*

1. *S is* conflict-free *iff $\nexists \mathcal{A}, \mathcal{B} \in S$ s.t. $(\mathcal{A}, \mathcal{B}) \in \mathbb{R}$.*
2. *$\mathcal{A}$ is* acceptable *w.r.t. S iff $\forall \mathcal{B} \in Args$: if $(\mathcal{B}, \mathcal{A}) \in \mathbb{R}$ then $\exists \mathcal{C} \in S$ s.t. $(\mathcal{C}, \mathcal{B}) \in \mathbb{R}$.*
3. *If S is conflict-free, then S is an* admissible set *of AF iff each argument in S is acceptable w.r.t. S.*

Intuitively, an argument $\mathcal{A}$ is acceptable w.r.t. S if for any argument $\mathcal{B}$ that attacks $\mathcal{A}$ there is an argument $\mathcal{C}$ in S that attacks $\mathcal{B}$, in which case $\mathcal{C}$ is said to defend $\mathcal{A}$. An admissible set S can then be interpreted as a coherent defendable position. For instance, in the AF of Figure 1, argument $\mathcal{D}$ is acceptable w.r.t. the sets $\{\mathcal{A}\}$, $\{\mathcal{B}\}$ and $\{\mathcal{A}, \mathcal{B}\}$; however, only the first two of these sets are admissible.

Taking into account the notion of admissibility Dung then defines the acceptability semantics of the framework.

Definition 3. *Let $AF = \langle Args, \mathbb{R} \rangle$ be an argumentation framework and $S \subseteq Args$ a conflict-free set of arguments. Then:*

- *S is a* complete extension *of AF iff all arguments acceptable w.r.t. S belong to S.*
- *S is a* preferred extension *of AF iff it is a maximal (w.r.t. set-inclusion) admissible set (i. e., a maximal complete extension).*
- *S is a* stable extension *of AF iff it is a preferred extension that attacks all arguments in $Args \backslash S$.*
- *S is the* grounded extension *of AF iff it is the smallest (w.r.t. set-inclusion) complete extension.*

The complete extensions of the framework in Figure 1 are $\emptyset$, $\{\mathcal{A}, \mathcal{D}\}$ and $\{\mathcal{B}, \mathcal{D}\}$; the preferred and stable extensions are $\{\mathcal{A}, \mathcal{D}\}$ and $\{\mathcal{B}, \mathcal{D}\}$; and the grounded extension is $\emptyset$.

3 Backing-Undercutting Argumentation Frameworks

A classical abstract argumentation framework is characterized by a set of arguments and an attack relation among them. In this section, we will introduce an extension of Dung's argumentation frameworks called Backing-Undercutting Argumentation Frameworks (BUAF). In the extended framework we will: distinguish between different types of attack, incorporate a special kind of support relation, and include a preference relation among arguments. Thus, the BUAF will provide the means for representing both attack and support for an argument's inference, allowing to express Pollock's undercutting defeaters and Toulmin's backings.

Definition 4 (Backing-Undercutting Argumentation Framework). *A Backing-Undercutting Argumentation Framework (BUAF) is a tuple* $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ *where:*

- $\mathbb{A}$ *is a set of arguments,*
- $\mathbb{D} \subseteq \mathbb{A} \times \mathbb{A}$ *is a set of attacks,*
- $\mathbb{B}_k \subseteq \mathbb{A} \times \mathbb{A}$ *is a backing relation, and*
- $\preceq \subseteq \mathbb{A} \times \mathbb{A}$ *is a preference relation.*

We will distinguish the three different types of attack in $\mathbb{D}$, where the set of rebutting attacks is denoted as $\mathbb{R}_b$, the set of undercutting attacks is denoted as $\mathbb{U}_c$, and the set of undermining attacks is denoted as $\mathbb{U}_m$ ($\mathbb{D} = \mathbb{R}_b \cup \mathbb{U}_c \cup \mathbb{U}_m$). In addition, when two arguments $\mathcal{A}$ and $\mathcal{B}$ are related by the preference relation (*i. e.* $(\mathcal{A}, \mathcal{B}) \in \preceq$) it means that argument $\mathcal{B}$ is at least as preferred as argument $\mathcal{A}$, denoting it as $\mathcal{A} \preceq \mathcal{B}$. Furthermore, following the usual convention, $\mathcal{A} \prec \mathcal{B}$ means $\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{B} \not\preceq \mathcal{A}$.

From hereon, we may use the following notation:

- $\mathcal{A} \dashrightarrow \mathcal{B}$ denotes $(\mathcal{A}, \mathcal{B}) \in \mathbb{D}$.
- $\mathcal{A} \Longrightarrow \mathcal{B}$ denotes $(\mathcal{A}, \mathcal{B}) \in \mathbb{B}_k$.

In order to illustrate, let us consider one of Toulmin's famous examples which discusses whether Harry is a British subject or not [13], as shown in Figure 2.

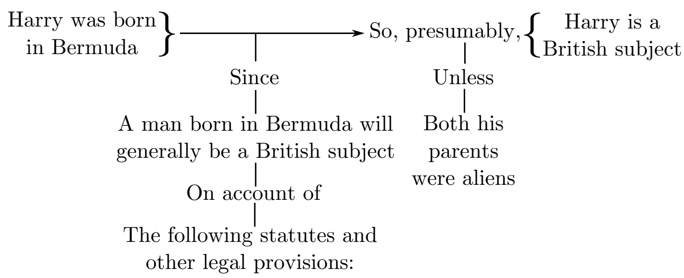

Fig. 2. Toulmin's example about Harry.

The following arguments can represent situation depicted in Toulmin's example:

$\mathcal{H}$: *"Harry was born in Bermuda. A man born in Bermuda will generally be a British subject. So, Harry is a British subject"*
$\mathcal{B}$: *"On account of the following statutes and other legal provisions..."*
$\mathcal{U}$: *"Both Harry's parents are aliens"*

Example 1 *A possible representation for Toulmin's example about Harry is given by the BUAF* $\Delta_1 = \langle \mathbb{A}_1, \mathbb{D}_1, \mathbb{B}_{k1}, \preceq_1 \rangle$, *where*

$$\mathbb{A}_1 = \{\mathcal{H},\ \mathcal{B},\ \mathcal{U}\} \qquad \mathbb{B}_{k1} = \{(\mathcal{B},\mathcal{H})\}$$
$$\mathbb{U}_{c1} = \{(\mathcal{U},\mathcal{H})\} \qquad \preceq_1 = \{(\mathcal{B},\mathcal{U})\}$$

Here, that the statutes and other legal provisions provide support for the warrant is expressed by the pair $(\mathcal{B},\mathcal{H})$ *in the backing relation. In addition, the fact that Harry's parents were aliens is considered as an undercut for the inference, as expressed by the pair* $(\mathcal{U},\mathcal{H})$ *in the attack relation.*

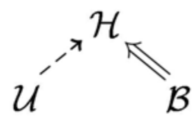

Fig. 3. The BUAF of Example 1.

4 Defeat and Conflict-Freenes

Before defining any semantics-related notion, we must first consider the concept of defeat. Intuitively, given that in a BUAF there is a preference relation among arguments, an argument $\mathcal{A}$ would defeat an argument $\mathcal{B}$ iff $\mathcal{A}$ attacks $\mathcal{B}$ and $\mathcal{A}$ is not less preferred than $\mathcal{B}$. Following this intuition, in this section we will define the notion of defeat in the context of a BUAF, where we will distinguish between two types of defeat. Then, we will define the basic restriction that an acceptable set of arguments in a BUAF must satisfy, that is, the notion of conflict-freenes for a set of arguments.

The first type of defeat we will distinguish is called *primary defeat* and is obtained directly by resolving the attacks given on the attack relation through the use of preferences. It is important to note that, in the case of undercutting attacks, the attacks will always succeed as defeats, like in [11]. On the other hand, for rebutting and undermining attacks we must compare the attacking and the attacked arguments in order to determine the existence of a defeat.

Definition 5 (Primary Defeat). *Let* $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ *be a BUAF and* $\mathcal{A}, \mathcal{B} \in \mathbb{A}$. *We will say that* $\mathcal{A}$ primary defeats $\mathcal{B}$ *iff one of the following conditions hold:*

- $(\mathcal{A},\mathcal{B}) \in (\mathbb{R}_b \cup \mathbb{U}_m)$ *and* $\mathcal{A} \not\prec \mathcal{B}$, *or*
- $(\mathcal{A},\mathcal{B}) \in \mathbb{U}_c$.

Observe that in the above definition rebutting and undermining attacks are grouped together. This is because, given the abstract nature of arguments, we can not distinguish an attack an argument's premise from an attack to its conclusion. Thus, the only way to determine the existence of a defeat in the presence of an undermining attack or a rebutting attack is to compare the attacking and attacked arguments. In contrast, for instance, if we had considered a notion of sub-argument the analysis for rebutting and undermining attacks would be different.

Example 2 *In the AF of Example 1, argument* $\mathcal{U}$ *primary defeats argument* $\mathcal{H}$*.*

As stated before, likewise [11], an undercutting attack will always result in defeat; however, in that approach the existence of arguments supporting an inference is not considered. Hence, following [6]'s approach, we will consider that backings are intended to defend their associated warrants against undercutting attacks. Therefore, it will be necessary to establish the relation between backing and undercutting arguments.

It is clear that backing and undercutting arguments are conflicting: while the latter attacks the connection between premises and conclusion of an argument, the former provides support for it. Thus, they should not be jointly accepted. Moreover, given that the conflict between backing and undercutting arguments may not always be explicit in the attack relation of a BUAF, it is necessary to ensure this acceptability restriction. To achieve this, we will define a new type of defeat called *implicit defeat*.

Definition 6 (Implicit Defeat). *Let* $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ *be a BUAF and* $\mathcal{A}, \mathcal{B} \in \mathbb{A}$*. We will say that* $\mathcal{A}$ implicitly defeats $\mathcal{B}$ *iff one of the following conditions hold:*

- $(\mathcal{A}, \mathcal{C}) \in \mathbb{U}_c$ *and* $(\mathcal{B}, \mathcal{C}) \in \mathbb{B}_k$*, and* $\mathcal{A} \not\prec \mathcal{B}$*, or*
- $(\mathcal{A}, \mathcal{C}) \in \mathbb{B}_k$ *and* $(\mathcal{B}, \mathcal{C}) \in \mathbb{U}_c$*, and* $\mathcal{A} \not\prec \mathcal{B}$*.*

Example 3 *Given the AF of Example 1, argument* $\mathcal{U}$ *implicitly defeats argument* $\mathcal{B}$*.*

Then, an argument will be defeated in a BUAF if it is primary or implicitly defeated.

Definition 7 (Defeat). *Let* $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ *be a BUAF and* $\mathcal{A}, \mathcal{B} \in \mathbb{A}$*. Then* $\mathcal{A}$ defeats $\mathcal{B}$*, noted as* $\mathcal{A} \rightsquigarrow \mathcal{B}$*, iff* $\mathcal{A}$ *primary defeats or implicitly defeats* $\mathcal{B}$*.*

From a BUAF Δ we can construct a directed graph called the *defeat graph*. The nodes in the graph are the arguments in Δ and the edges correspond the defeat relation obtained by Definiton 7.

Example 4 *Consider the BUAF* $\Delta_2 = \langle \mathbb{A}_2, \mathbb{D}_2, \mathbb{B}_{k2}, \preceq_2 \rangle$*, where*

$$\begin{array}{ll} \mathbb{A}_2 = \{\mathcal{E}, \mathcal{F}, \mathcal{G}, \mathcal{H}, \mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{L}\} & \mathbb{U}_{m2} = \{(\mathcal{I}, \mathcal{H})\} \\ \mathbb{R}_{b2} = \{(\mathcal{F}, \mathcal{E}), (\mathcal{J}, \mathcal{G})\} & \mathbb{B}_{k2} = \{(\mathcal{G}, \mathcal{E}), (\mathcal{L}, \mathcal{J})\} \\ \mathbb{U}_{c2} = \{(\mathcal{H}, \mathcal{E}), (\mathcal{K}, \mathcal{J})\} & \preceq_2 = \{(\mathcal{F}, \mathcal{E}), (\mathcal{H}, \mathcal{G}), (\mathcal{G}, \mathcal{J}), (\mathcal{J}, \mathcal{K})\} \end{array}$$

A graphical representation of Δ_2 is shown below on the left and its corresponding defeat graph is shown on the right:

The primary defeats obtained from Δ_2 are $\mathcal{I} \rightsquigarrow \mathcal{H}$, $\mathcal{H} \rightsquigarrow \mathcal{E}$, $\mathcal{J} \rightsquigarrow \mathcal{G}$ and $\mathcal{K} \rightsquigarrow \mathcal{J}$; and the implicit defeats are $\mathcal{G} \rightsquigarrow \mathcal{H}$, $\mathcal{L} \rightsquigarrow \mathcal{K}$ and $\mathcal{K} \rightsquigarrow \mathcal{L}$.

Note that in Example 4 argument $\mathcal{G}$ is a backing for argument $\mathcal{E}$, thus defending it against the undercut of $\mathcal{H}$. In addition, argument $\mathcal{I}$ defeats argument $\mathcal{H}$, becoming a defender for $\mathcal{E}$. Notwithstanding, the nature of the defenses provided by $\mathcal{G}$ and $\mathcal{I}$ is different. The former is a backing for argument $\mathcal{E}$, having the support between these two arguments explicitly determined by the backing relation; on the other hand, the latter merely defeats one of $\mathcal{E}$'s defeaters, in particular, the undercutting defeater $\mathcal{H}$.

Next, conflict-free sets of arguments are characterized directly, by requiring the absence of defeats.

Definition 8 (Conflict-free Set). *Let $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ be a BUAF. A set $S \subseteq \mathbb{A}$ is* conflict-free *iff $\nexists \mathcal{A}, \mathcal{B} \in S$ s.t. $\mathcal{A} \rightsquigarrow \mathcal{B}$.*

Example 5 *Given the BUAF of Example 4, some conflict-free sets of arguments are $\emptyset$, $\{\mathcal{E}\}$ and $\{\mathcal{F}, \mathcal{I}, \mathcal{G}, \mathcal{L}\}$.*

5 Acceptability Semantics

Since arguments in a BUAF can defeat each other, conflicting arguments should not be accepted simultaneously. Therefore, arguments in a BUAF will be subject to a status evaluation in which an argument will be accepted if it somehow "survives" the defeats it receives, or rejected otherwise. This evaluation process will be determined by the acceptability semantics.

In this section, we will define the basic semantic notions required for obtaining the set of acceptable arguments. Then, we will formally define the acceptability semantics for BUAFs. Finally, a characterization of BUAFs as Dung's AFs is presented, establishing the relation between these two frameworks.

Definition 9 (Acceptability). *Let $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ be a BUAF. An argument $\mathcal{A} \in \mathbb{A}$ is* acceptable *w.r.t. $S \subseteq \mathbb{A}$ iff $\forall \mathcal{B} \in \mathbb{A}$ s.t. $\mathcal{B} \rightsquigarrow \mathcal{A}$, $\exists \mathcal{C} \in S$ s.t. $\mathcal{C} \rightsquigarrow \mathcal{B}$.*

Intuitively, an argument $\mathcal{A}$ will be acceptable with respect to a set of arguments S iff S defends $\mathcal{A}$ against all its defeaters.

Example 6 *In the BUAF of Example 4, the argument $\mathcal{E}$ is acceptable w.r.t. the sets $\{\mathcal{I}\}$, $\{\mathcal{F}, \mathcal{G}\}$, $\{\mathcal{I}, \mathcal{J}, \mathcal{K}\}$, and $\{\mathcal{F}, \mathcal{I}, \mathcal{G}, \mathcal{K}\}$ among others.*

In the literature, a usual requirement when defining the set of acceptable arguments of an AF is the conflict-freenes of the set (see *e. g.*, [7, 2]). This implies that the set of collectively acceptable arguments must be internally coherent, in the sense that no pair of arguments in the set defeat each other. Thus, it is reasonable to accept only those arguments that are acceptable. We will follow this approach and therefore, the set of accepted arguments in a BUAF will be the set of arguments that defends itself against all defeats on it, leading to a classical definition of admissibility for BUAFs.

Definition 10 (Admissibility). *Let $\langle \mathbb{A}, \mathbb{D}, \mathbb{B}_k, \preceq \rangle$ be a BUAF. A set $S \subseteq \mathbb{A}$ is* admissible *iff it is conflict-free and all elements of S are acceptable w.r.t. S.*

Example 7 *From the sets of arguments listed in Example 6, only the sets $\{\mathcal{I}\}$ and $\{\mathcal{F}, \mathcal{I}, \mathcal{G}, \mathcal{K}\}$ are admissible.*

Recall that acceptability semantics identify a set of extensions of an argumentation framework, namely sets of arguments which are collectively acceptable. The complete, preferred, stable and grounded extensions of a BUAF are now defined in the same way as for Dung's frameworks.

Definition 11 (Extensions). *Let $\Delta = \langle \mathbb{A}, \mathbb{D}, \mathbb{B}_{k,} \preceq \rangle$ be a BUAF and $S \subseteq \mathbb{A}$ a conflict-free set of arguments. Then:*

- *S is a* complete extension *of Δ iff all arguments acceptable w.r.t. S belong to S.*
- *S is a* preferred extension *of Δ iff it is a maximal (w.r.t. set-inclusion) admissible set of Δ (i. e., a maximal complete extension).*
- *S is a* stable extension *of Δ iff it is a preferred extension that defeats all arguments in $\mathbb{A} \backslash S$.*
- *S is the* grounded extension *of Δ iff it is the smallest (w.r.t. set-inclusion) complete extension.*

Given a BUAF and a semantics s, an argument $\mathcal{A}$ is *skeptically accepted* if it belongs to all s-extensions; $\mathcal{A}$ is *credulously accepted* if it belong to some (not all) s-extensions; and $\mathcal{A}$ is *rejected* if does not belong to any s-extension.

Example 8 *From the BUAF of Example 4 we can obtain the following sets of extensions:*

- *the complete extensions $\{\mathcal{F}, \mathcal{I}, \mathcal{E}\}$, $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{G}\}$, $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{J}\}$, $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{G}, \mathcal{K}\}$, and $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{J}, \mathcal{L}\}$;*
- *the preferred and stable extensions $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{G}, \mathcal{K}\}$ and $\{\mathcal{F}, \mathcal{I}, \mathcal{E}, \mathcal{J}, \mathcal{L}\}$; and*
- *the grounded extension $\{\mathcal{F}, \mathcal{I}, \mathcal{E}\}$.*

Definitions 9, 10 and 11 correspond to those presented for Dung's argumentation frameworks. Recall that classical a argumentation framework is characterized by a set of arguments and an attack relation among them. Thus, using the defeat relation from Definition 7 and the set of arguments of a BUAF we can characterize an abstract argumentation framework which accepts exactly the same arguments as the BUAF under a given semantics.

Proposition 1. *Let* $\Delta = \langle \mathbb{A}, \mathbb{D}, \mathbb{B}_{k,} \preceq \rangle$ *be a BUAF. There exists an abstract argumentation framework* $AF = \langle \mathbb{A}, \leadsto \rangle$ *such that the sets of extensions of* Δ *and* AF *under a given semantics are equal.*

Proof. Straightforward from definitions 2, 3, 9, 10 and 11. □

Therefore, by Proposition 1, BUAFs will inherit all properties from abstract argumentation frameworks (refer to [7] for details). Moreover, it will be possible to determine the acceptability of arguments in a BUAF using its associated Dung's AF. We first obtain the associated AF and then, acceptability semantics are applied to this AF.

6 Conclusions and Related Work

In this work, an extension of Abstract Argumentation Frameworks called Backing-Undercutting Argumentation Frameworks (BUAFs) was proposed, inspired by the work of Pollock [10] and Toulmin [13]. This extension allows to express attack and support for an inference by distinguishing different types of attacks and incorporating a specialized support relation among arguments. In that way, the extended framework enables the representation of Toulmin's backings and Pollock's undercutting defeaters, two important notions in the argumentation community. Several approaches address these two notions separately, yet they were not widely considered together in the formalizations provided so far. For instance, in [11] an extension of AFs is presented, where arguments are partly provided of an internal structure and a categorization of defeaters is also given; however, in that work there is no consideration for support among arguments.

Likewise [1], our approach incorporates a preference relation among arguments in order to determine the success of attacks. Other works that consider preferences among arguments include [9] and [2], but the difference between those approaches and ours is that they express preferences in the object level, by incorporating attacks to attacks.

Among other approaches that address support between arguments, in addition to the attack relation, are the Bipolar Argumentation Frameworks (BAFs) [5]. A Bipolar Argumentation Framework extends Dung's framework to incorporate a support relation between arguments. The main difference between BAFs and BUAFs is that the support relation in a BAF is general, while the backing relation proposed in this work corresponds to the specific support relation between Toulmin's backings and warrants. Therefore, the implicit defeats as presented in Definition 6 could not be modeled in BAFs. On the other hand, some additional requirements for an admissible set of arguments are considered in [5], such as external coherence or consistency. Although for BUAFs we have only considered the conflict-freenes (internal consistency) of the set, those requirements are also satisfied by the notion of admissibility given in Definition 10; however, a detailed explanation is left for future work.

Finally, it was shown that BUAFs can be mapped to AFs by considering the set of arguments and the corresponding defeat relation. Thus, it is clear that the examples and applications shown for BUAFs can also be modeled with Dung's abstract frameworks. Notwithstanding, besides formalizing the backing relation and different types of attack, BUAFs will provide a more concrete approach to represent argumentative or non-monotonic scenarios where inferences can be attacked and supported.

References

1. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Ann. Math. Artif. Intell. 34(1-3), 197–215 (2002)
2. Baroni, P., Cerutti, F., Giacomin, M., Guida, G.: Afra: Argumentation framework with recursive attacks. International Journal of Approximate Reasoning 52(1), 19–37 (2011)
3. Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in artificial intelligence. Artif. Intell. 171(10-15), 619–641 (2007)
4. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artif. Intell. 128(1-2), 203–235 (2001)
5. Cayrol, C., Lagasquie-Schiex, M.C.: Bipolar abstract argumentation systems. In: Simari, G.R., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 65–84. Springer US (2009)
6. Cohen, A., García, A.J., Simari, G.R.: Backing and undercutting in defeasible logic programming. In: ECSQARU. pp. 50–61 (2011)
7. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–358 (1995)
8. García, A.J., Simari, G.R.: Defeasible logic programming: An argumentative approach. Theory and Practice of Logic Programming 4(1-2), 95–138 (2004)
9. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artificial Intelligence 173(9-10), 901–934 (2009)
10. Pollock, J.L.: Defeasible reasoning. Cognitive Science 11(4), 481–518 (1987)
11. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation 1, 93–124 (2009)
12. Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. 4, pp. 218–319. Kluwer Academic Pub. (2002)
13. Toulmin, S.E.: The Uses of Argument. Cambridge University Press (1958)