# On the Extension of Argumentation Logic

Antonis Kakas[1] and Paolo M. Mancarella[2]

[1] Department of Computer Science, University of Cyprus, Cyprus
antonis@ucy.ac.cy
[2] Department of Computer Science, University of Pisa, Italy
paolo.mancarella@unipi.it

**Abstract**

This paper shows how Argumentation Logic can be further extended to cover more fully paraconsistent forms of reasoning. The extension is based on the notion of non-acceptable self-defeating arguments as a generalization of the Reductio ad Absurdum principle.

## 1 Motivation and Background

Argumentative inference relies on the central normative condition of the acceptability of a (set of) argument(s). Informally, this condition states that "a (set of) argument(s) is acceptable iff it defends against all its counter-arguments". An acceptable argument thus forms a "case" that supports *satisfactorily* its claim and hence the claim is a possible or credulous conclusion under the argumentative reasoning.

When this is applied to formal logical reasoning where arguments are sets of logical formulae, e.g., propositional formulae, a case corresponds to a set of formulae which can be enveloped in a model of the theory and thus a credulous conclusion corresponds to a *satisfiable* formula. We can then show that such a form of Argumentation Logic (AL) is logically equivalent to classical Propositional Logic (PL). This equivalence though holds only when reasoning under a set of given premises that are classically consistent. When the premises are inconsistent, AL does not trivialize like PL but smoothly extends PL into a paraconsistent logic.

Technically, AL does this by encompassing the proof rule of Reduction ad absurdum through the notion of *non-acceptability* of arguments, namely the contrary notion of acceptability of arguments. Non-acceptable arguments are "self-defeating" arguments. Informally, such an argument is one that either forms a counter-argument to itself or that it is a counter-argument to an argument that it necessarily needs in order to defend against some counter-argument to it. In other words, a non-acceptable or self-defeating argument invalidates its possible case of support by rendering the set of arguments in the case incompatible with each other.

In this paper, we will explore further the notion of non-acceptable arguments and study how this can give in the AL reformulation of PL new acceptable sets of arguments (under inconsistent premises) that were not recognized as such before. Whereas in the previous work on AL in [3] this was carried out only for the limiting case of non-acceptability of a self-attacking counter-argument, in this paper we will show how more complex forms of self-defeating non-acceptable arguments can be identified and used to "neutralize" the effect of such arguments when they appear as counter-arguments to other arguments.

Section 2, reviews the acceptability semantics for general abstract argumentation frameworks under which the classical Propositional Logic is reformulated as an Argumentation logic. Section 3 discusses the non-acceptability of arguments. Section 4 defines the proposed extension of the acceptability semantics and applies this to the specific case of AL as a reformulation of PL. Section 5 concludes with a brief discussion of future work.

## 2 Acceptability semantics of Argumentation

Let us briefly review the area of Abstract Argumentation and its semantics [1, 2] as developed and used in the area of Artificial Intelligence. In abstract argumentation we are not interested in the internal structure of arguments but only in their relative properties.

**Definition 1.** *[Abstract Argumentation Framework]*
*An abstract argumentation framework is a triple, $\langle Arg, Att, \mathcal{D}ef \rangle$, where*

- *$Arg$ is a set (of arguments)*
- *$Att$ is a binary (partial) relation on $Arg$ (attack relation)*
- *$\mathcal{D}ef$ is a binary (partial) relation on $Arg$ (defence relation)*

Given $A, \Delta, D \subseteq Arg$, we say that $A$ **attacks** $\Delta$ (written $A \leadsto \Delta$) iff there exists $a \in A$ and $b \in \Delta$ such that $(a, b) \in Att$ and that $D$ **defends against** $A$ (written $D \twoheadrightarrow A$) iff $(d, c) \in \mathcal{D}ef$ for some $d \in D$ and $c \in A$.

A typical realization of a triple argumentation framework in some language, $\mathcal{L}$, for constructing and comparing arguments is given by: (1) $a$ is in conflict in $\mathcal{L}$ with $b$ for $(a, b) \in Att$ to hold and (2) $a$ is at least as strong in $\mathcal{L}$ as $b$ for $(a, b) \in \mathcal{D}ef$ to hold. In such realizations, the attack relation is symmetric and the defence relation is a subset of the attack relation.

The semantics of an abstract argumentation framework is defined via subsets of arguments that satisfy an acceptability property, $\mathcal{A}cc(\Delta, \Delta_0)$, whose informal meaning is that the set of arguments $\Delta$ is acceptable in the context of a given set of arguments $\Delta_0$, only when $\Delta$ can defend against all its counter-arguments.

**Definition 2** (Acceptability property)**.**
*Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be an abstract argumentation framework and $\Delta, \Delta_0 \subseteq Arg$. Then:*

- *$\mathcal{A}cc(\Delta, \Delta_0)$ iff*
  - *$\Delta \subseteq \Delta_0$, or*
  - *for any $A \subseteq Arg$ such that $A \leadsto \Delta$:*
    - *$A \not\subseteq \Delta \cup \Delta_0$, and*
    - *there exists $\Delta' \subseteq Arg$ such that $\Delta' \twoheadrightarrow A$ and $\mathcal{A}cc(\Delta', \Delta \cup \Delta_0)$*

In other words, counter-arguments must be defended against by arguments that are themselves acceptable in the extended context of $\Delta \cup \Delta_0$, and hence $\Delta$ can contribute to its own defense. Formally, the acceptability property is defined through the least fixed point of an associated monotonic operator on the binary Cartesian product of sets of arguments $\mathcal{R} = 2^{Arg} \times 2^{Arg}$.

**Definition 3** (Acceptability operator)**.** *Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be an abstract argumentation framework. The acceptability operator $\mathcal{A} : \mathcal{R} \rightarrow \mathcal{R}$ is defined as follows. Given $r \in \mathcal{R}$ and $\Delta, \Delta_0 \subseteq Arg$, $(\Delta, \Delta_0) \in \mathcal{A}(r)$ iff:*

- *$\Delta \subseteq \Delta_0$, or*
- *for any $A \subseteq Arg$ such that $A \leadsto \Delta$:*
  - *$A \not\subseteq \Delta \cup \Delta_0$, and*
  - *there exists $\Delta' \subseteq Arg$ such that $\Delta' \twoheadrightarrow A$ and $(\Delta', \Delta \cup \Delta_0) \in r$*

We denote by $\mathcal{A}^{fix}$ the least fixed point of this operator. Then the semantics of an argumentation framework is given through the subsets of arguments $\Delta$ that are acceptable with respect to the empty set of arguments, i.e. such that $(\Delta, \{\}) \in \mathcal{A}^{fix}$ holds. We say that such sets of arguments are *acceptable*.

**Example 1.** *Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be the abstract argumentation framework where*

- $Arg = \{a, b\}$
- $Att = \{(a, b), (b, a)\}$
- $\mathcal{D}ef = \{(b, a)\}$

*In this framework its two arguments attack each other but only $b$ is able to defend against its counter-argument of $a$, e.g., because $b$ is stronger than $a$. We can then see that the set $\{b\}$ is acceptable whereas the set $\{a\}$ is not acceptable as it cannot defend against its counter-argument $A = \{b\}$. Instead, if the defense relation contained also $(a, b)$, e.g., when the two arguments are of equal strength, then both $\{a\}$ and $\{b\}$ would be acceptable sets of arguments.*

## 2.1 Propositional Logic as Argumentation Logic

We will review the reformulation [3] of classical Propositional Logic and its paraconsistent extension of Argumentation Logic as a realization of the abstract argumentation framework and its acceptability semantics.

**Definition 4** (Argumentation Logic Framework)**.** *We denote by $\vdash_{MRA}$ the Natural Deduction direct derivation relation of propositional logic modulo Reduction ad Absudrum (MRA), i.e. without the proof rule of Reduction ad Absudrum.*

*Let $T$ be a propositional theory. The argumentation logic framework corresponding to $T$ is the triple $AF^T = \langle Args, Att, Def \rangle$ with:*

- $Arg = \{\Sigma \mid \Sigma$ *is a finite set of propositional sentences*$\}$
- *given* $\Delta, \Gamma \in Arg$*, with* $\Delta \neq \{\}$*,* $(\Gamma, \Delta) \in Att$ *iff* $T \cup \Gamma \cup \Delta \vdash_{MRA} \bot$
- *given* $\Delta \in Arg$*,* $(\{\overline{\phi}\}, \Delta) \in Def$*, where* $\overline{\phi}$ *is the complement of some sentence* $\phi \in \Delta$ *and* $(\{\}, \Delta) \in Def$ *whenever* $T \cup \Delta \vdash_{MRA} \bot$*.*

We see that the attack relation is symmetric, i.e. arguments are always counter-arguments of each other when together they are directly inconsistent in the context of the given premises $T$. The defense relation essentially expresses the fact that any argument can be defended against by *undermining* one of its premises. In logical terms the defense relation expresses the property that for any formula $\phi$ we are free to choose this or its complement. The second part of the defense relation expresses the fact that if an argument is self-inconsistent with respect to the given premises then this can be trivially defended against by the "safe" empty argument (which in turn can not be attacked). We will see below that when we extend the acceptability semantics, this second part of the defense relation will not need to be stated explicitly at this level but will be captured at the extended acceptability semantic level.

We will denote by $\mathcal{AL}^{fix}$ (or simply by $\mathcal{AL}$) the least fix point of the corresponding operator $\mathcal{A}$ in the general abstract argumentation frameworks as above in definition 3. We then have [3] a logical correspondence between propositional logic (for classically consistent premises $T$) and the argumentation acceptability semantics. For any formula $\phi$: $\phi$ is acceptable, i.e.,

$(\{\phi\}, \{\}) \in \mathcal{AL}^{fix}$ if and only if there is a model of $T$ in which $\phi$ is true. Furthermore, for classically inconsistent premises which are directly consistent, i.e. consistent under the restricted derivation of $\vdash_{MRA}$, the argumentation semantics does not trivialize but smoothly extends the propositional deductive semantics into such cases of inconsistent premises. The full technical details of these results can be found in [3]. For the purposes of this paper, it is important to point out that the results rest on the correspondence between proofs via Reductio ad Absurdum and the non-acceptability of formulae, i.e. formulae $\phi$ such that $(\{\phi\}, \{\}) \notin \mathcal{AL}^{fix}$ holds.

# 3 Non-acceptable Arguments

In this section, we will examine further the nature of non-acceptable arguments and the relative defeatedness of such arguments in the context of a given set of arguments.

**Example 2** (Motivating Example 1). *Let $AF = \langle Arg, Att, \mathcal{D}ef\rangle$ be the abstract argumentation framework where*

- $Arg = \{a, b\}$
- $Att = \{(a, b)\}$
- $\mathcal{D}ef = \{\}$

*The argument set $\{b\}$ is attacked by the argument set $\{a\}$. Trivially then, $(\{b\}, \{a\}) \notin \mathcal{A}^{fix}$ i.e., $\{b\}$ is non-acceptable in the context of $\{a\}$, as $\{b\}$ is attacked by an argument that belongs to the context . We will also say that $\{b\}$ is **defeated in the context of** $\{a\}$.*

**Example 3** (Motivating Example 2). *Let $AF = \langle Arg, Att, \mathcal{D}ef\rangle$ be the abstract argumentation framework where*

- $Arg = \{a, b\}$
- $Att = \{(a, a), (a, b)\}$
- $\mathcal{D}ef = \{\}$

*The argument set $\{a\}$ is self-attacking and hence it is non-acceptable or defeated in its own context. We consider this argument as a **self-defeating argument** exactly because it contains one of its attacks. This property of $\{a\}$ being self-defeating is not affected by the argument $\{b\}$.*

The above example shows a simple (and limiting) case of a non-acceptable self-defeating argument. More complex forms of such arguments exist as it is illustrated in the next example.

**Example 4** (Motivating Example 3). *Let $AF = \langle Arg, Att, \mathcal{D}ef\rangle$ be the abstract argumentation framework where*

- $Arg = \{a, b, a_1, d_1\}$
- $Att = \{(a, b), (a_1, a), (a, d_1), (d_1, a_1)\}$
- $\mathcal{D}ef = \{(d_1, a_1)\}$

*Argument $a$ is attacked by $a_1$ which can only be defended against by argument $d_1$. But $a$ attacks this defense of $d_1$, i.e. $d_1$ is defeated in the context of $a$. Hence, as in the example above, $a$ is non-acceptable and we can consider it as self-defeating, but now in an indirect way, because $a$ renders its necessary defending argument(s) non-acceptable or defeated in its own context.*

These more complex forms of self-defeated arguments arise from the recursive nature of non-acceptability given by negating the recursive definition of acceptability.

**Proposition 1** (Non-acceptability)**.**
*Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be an abstract argumentation framework and $\Delta, \Delta_0 \subseteq Arg$. Let $non\_Acc(\Delta, \Delta_0)$ denote the statement $(\Delta, \Delta_0) \notin \mathcal{A}^{fix}$. Then the following holds*[1]*:*

$non\_Acc(\Delta, \Delta_0)$ *iff* $\Delta \not\subseteq \Delta_0$ *and*

$\exists A \subseteq Arg$ *such that* $A \rightsquigarrow \Delta$ *and*

- $A \subseteq \Delta \cup \Delta_0$*, or*
- $\forall \Delta' \subseteq Arg$ *s.t.* $\Delta' \twoheadrightarrow A$*:* $non\_Acc(\Delta', \Delta \cup \Delta_0)$.

A non-acceptable argument $A$ such that $(A, \{\}) \notin \mathcal{A}^{fix}$ holds, is one where when we collect recursively the defenses against one of its counter-arguments and recursively the defenses against attacks of the earlier defenses we end up with a collection of defenses that is self-attacking.

## 4 Extended Acceptability semantics

The extension of the notion of acceptability of arguments follows the simple idea that counter-arguments that are non-acceptable or self-defeating can be dealt with without the need to explicitly defend against them. It is sufficient to recognize that such attacks are self-defeating.

**Definition 5** (Extended Acceptability)**.**
*Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be an abstract argumentation framework and $\Delta, \Delta_0 \subseteq Arg$. Then a set of arguments $\Delta$* **is acceptable in the context of** *$\Delta_0$, denoted by $\mathcal{A}cc^+(\Delta, \Delta_0)$, when the following holds:*

$\mathcal{A}cc^+(\Delta, \Delta_0)$ *iff*

- $\Delta \subseteq \Delta_0$*, or*
- *for any* $A \subseteq Arg$ *such that* $A \rightsquigarrow \Delta$*:*
  - $A \not\subseteq \Delta \cup \Delta_0$*, and*
  - $(A, \{\}) \notin \mathcal{A}^{fix}$*, or* $\exists\, \Delta' \subseteq Arg$ *such that* $\Delta' \twoheadrightarrow A$ *and* $(\Delta', \Delta \cup \Delta_0) \in \mathcal{A}^{fix}$

**Proposition 2.** *Let $AF = \langle Arg, Att, \mathcal{D}ef \rangle$ be an abstract argumentation framework and $\Delta, \Delta_0 \subseteq Arg$. Then $(\Delta, \Delta_0) \in \mathcal{A}^{fix} \implies \mathcal{A}cc^+(\Delta, \Delta_0)$.*

**Example 5** (Examples 1 and 4 cnt.)**.** *In both of these examples $(\{b\}, \{\})$ does not belong to $\mathcal{A}^{fix}$, i.e. the argument set $\{b\}$ is not acceptable. However, $\mathcal{A}cc^+(\{b\}, \{\})$ holds because the only (minimal) attack against $\{b\}$, namely the set $\{a\}$, is self-defeating. Hence the argument set $\{b\}$ is acceptable in the extended semantics.*

[1] Proofs in this paper are omitted due to lack of space.

### 4.1 AL$^+$: Extended Argumentation Logic

We will now apply the extended acceptability semantics to obtain an extended form of Argumentation Logic (AL). We simply apply definition 5 to the case of AL. In effect, this will give us a generalized form of proof by contradiction under inconsistent premises.

**Definition 6** (Extended Argumentation Logic)**.** *Let $AF^T = \langle Args, Att, Def\rangle$ be the argumentation logic framework corresponding to a (directly consistent) propositional theory $T$. Then the extended argumentation logic, $\mathcal{AL}^+$, is given by:*

*$\mathcal{AL}^+(\Delta, \{\})$ holds iff for any $A \subseteq Arg$ such that $A \rightsquigarrow \Delta$:*

- *$A \not\subseteq \Delta$, and*
- *$(A, \{\}) \notin \mathcal{AL}$, or there exists $\Delta' \subseteq Arg$ such that $\Delta' \twoheadrightarrow A$ and $(\Delta', \Delta) \in \mathcal{AL}$*

Hence a set of formulae is acceptable in $\mathcal{AL}^+$ either because its attacks could be defended acceptably, as before in the basic logic of $\mathcal{AL}$, or because its attacks are non-acceptable in $\mathcal{AL}$.

The following result shows that the extended argumentation logic, $\mathcal{AL}^+$, is a "proper" extension of $\mathcal{AL}$ when the given premises $T$ are classically consistent.

**Theorem 1.** *Let $T$ be a classically consistent theory and $AF^T = \langle Arg, Att, \mathcal{D}ef\rangle$ its corresponding argumentation logic framework. Let also $\phi$ be a propositional formula such that $(\{\phi\}, \{\}) \notin \mathcal{AL}$ holds. Then $\mathcal{AL}^+(\{\phi\}, \{\})$ does not hold.*

Thus the extension of the logic does not trivialize the original logic and specifically classical Propositional Logic for consistent premises. We also know from proposition 2 that $\mathcal{AL}^+$ contains the original logic of $\mathcal{AL}$. The following example, taken from [3], clarifies the link between the extended AL and the original AL and how the former gives genuinely new cases of acceptable formulae.

**Example 6.** *Consider the following two theories of propositional logic:*

- $T_1 = \{\neg(\beta \wedge \alpha), \neg\alpha\}$ $\qquad$ $T_2 = \{\neg(\beta \wedge \alpha), \neg(\alpha \wedge \gamma), \neg(\alpha \wedge \neg\gamma)\}$

*It is easy to see that the argument $\{\beta\}$ is acceptable in $\mathcal{AL}$ relative to theory $T_1$. Its minimal attack $\{\alpha\}$ is directly self-inconsistent and hence self-attacking (i.e. $T_1 \cup \{\alpha\} \vdash_{MRA} \bot$) and so it can be defended by $\{\}$. The argument $\{\beta\}$ is also acceptable in $\mathcal{AL}$ relative to theory $T_2$, even though its attack $\alpha$ is not directly inconsistent. The defense against the attack of $\{\alpha\}$, namely $\{\neg\alpha\}$, is such that $(\{\neg\alpha\}, \{\beta\}) \in \mathcal{AL}$. Notice, however, that this attack of $\{\alpha\}$ is itself a non-acceptable self-defeating argument, as it cannot defend acceptably against its attack by $\{\gamma\}$: the only possible defense of $\{\neg\gamma\}$ in non-acceptable in the context of $\{\alpha\}$, because $\{\alpha\}$ attacks $\{\neg\gamma\}$. Therefore, recognizing the non-acceptability of the attack $\{\alpha\}$ is an alternative way to enforce the acceptability of $\{\beta\}$. The extended acceptability semantics of $\mathcal{AL}^+$ uses this alternative way. Importantly, it does so in the same way for both theories $T_1$ and $T_2$.*

The extended acceptability semantics becomes relevant when the theory of premises is inconsistent, and attacks like $\{\alpha\}$ above cannot be defended acceptably by $\{\neg\alpha\}$.

**Example 7** (Example 6 cnt.)**.** *Consider the following theory, obtained from $T_2$ by making also $\neg\alpha$ non acceptable: $T_3 = T_2 \cup \{\neg(\neg\alpha \wedge \delta), \neg(\neg\alpha \wedge \neg\delta)\}$. The attack $\{\alpha\}$ cannot be acceptably defended because the possible defense of $\{\neg\alpha\}$ is non-acceptable in a way similar to the non-acceptability of $\{\alpha\}$ shown above (replacing $\{\gamma\}$ with $\{\delta\}$). Nevertheless, as $\{\alpha\}$ is by itself non-acceptable, it is reasonable to accept $\{\beta\}$ as acceptable, as $\mathcal{AL}^+$ does.*

# 5 Conclusions

We have shown how to extend Argumentation Logic to capture the intuitive idea that for attacks which are by themselves self-defeating it is not necessary to defend against. This extension is based on definition 5. We can then consider applying this definition iteratively to give possible further extensions of acceptability and study the properties of such extensions.

# References

[1] 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.

[2] A.C. Kakas and P. Mancarella. On the semantics of abstract argumentation. *Journal of Logic and Computation [electronic only]*, 23(5):991–1015, 2013.

[3] Antonis Kakas, Paolo Mancarella, and Francesca Toni. On Argumentation Logic and Propositional Logic. *Studia Logica*, 106(2):237–279, 2018.