A reasoning model based on the production of acceptable arguments

Leila AMGOUD (amgoud@info.univ-angers.fr)
LERIA, Universit d'Angers,
2, boulevard Lavoisier, 49045 Angers Cedex, FRANCE

Claudette CAYROL (ccayrol@irit.fr)
IRIT, Universit Paul Sabatier
118 route de Narbonne, 31062 Toulouse Cedex, FRANCE

Abstract. Argumentation is a reasoning model based on the construction of arguments and counter-arguments (or defeaters) followed by the selection of the most acceptable of them. In this paper, we refine the argumentation framework proposed by Dung by taking into account preference relations between arguments in order to integrate two complementary points of view on the concept of acceptability: acceptability based on the existence of direct counter-arguments and acceptability based on the existence of defenders. An argument is thus acceptable if it is preferred to its direct defeaters or if it is defended against its defeaters. This also refines previous works by Prakken and Sartor, by associating with each argument a notion of strength, while these authors embed preferences in the definition of the defeat relation. We propose a revised proof theory in terms of AND/OR trees, verifying if a given argument is acceptable, which better reflects the dialectical form of argumentation.

Keywords: Argumentation, Preference relations

1. Introduction

Argumentation is a promising model for reasoning with inconsistent knowledge, based on the construction and the comparison of arguments. It may also be considered as a different method for handling uncertainty. In particular, it should be possible to assess the reason why a fact holds, in the form of arguments, and combine these arguments to evaluate the certainty. Indeed, the process of combination may be viewed as a kind of reasoning about arguments themselves in order to determine the most acceptable of them. Note that the above definition encompasses two views of an argument:

– a local view, intended to give support in favour or against a conclusion; and

– a global view, intended to define the acceptable arguments.

Formal argumentation systems (Lin and Shoham, 1989; Vreeswijk, 1991; Vreeswijk, 1997; Pollock, 1992; Pinkas and Loui, 1992; Simari and Loui, 1992; Dung, 1995; Prakken and Sartor, 1996; Prakken and Sartor, 1997; Kohlas et al., 2000) are characterised by representing precisely some of these features of argumentation via formal languages, and by applying formal inference techniques.
The different approaches, which have been developed for reasoning within an argumentation system, use one of the two following kinds of acceptability:

Individual acceptability: an acceptability level is assigned to a given argument on the basis of the existence of direct defeaters. That leads to the concept of acceptability class introduced by (Elvang et al., 1993; Elvang and Hunter, 1995).

Joint acceptability: (Dung, 1993; Dung, 1995) the set of all the arguments that a rational agent accepts must defend itself against any defeater.

These notions of acceptability have been most often defined purely on the basis of defeaters. The resulting evaluation of arguments is only based on the interactions between (direct or indirect) defeaters. However, other criteria may be taken into account for comparing arguments such as for instance, specificity (Simari and Loui, 1992), or explicit priorities on the beliefs.

The notion of priority plays a crucial role in the study of knowledge-based systems. When priorities attached to pieces of knowledge are available, the task of coping with inconsistency is greatly simplified, since conflicts have a better chance to be resolved. More generaly, preference relations allow the comparison of arguments and in some cases the selection between conflicting arguments. Our aim is to take advantage of these priorities in argumentation frameworks. For that purpose, we introduce preference relations into argumentation systems.

In previous work (Amgoud et al., 1996), we studied different preference relations between arguments. In (Amgoud and Cayrol, 1997; Amgoud and Cayrol, 1998), we presented the principles of preference-based argumentation and how preference relations can be integrated into argumentation frameworks.
In this paper, we propose to refine the abstract argumentation framework proposed by Dung in (Dung, 1995) with explicit priorities. We propose a general preference-based argumentation framework where the definition of acceptability combines different independent evaluations: an evaluation based on direct or indirect defeaters and on defenders, and a preference-based comparison between conflicting arguments.

One basic idea is to accept an argument if it is not defeated, if it defends itself against its defeaters (because it is preferred or stronger than its defeaters), or if it is defended by other arguments. Finally, we show that a proof theory is available for testing whether an argument is acceptable. The concepts presented are illustrated in the particular framework of inconsistency handling in knowledge bases.
Note that Prakken and Sartor have also extended Dung's framework with priorities in (Prakken and Sartor, 1997). The route taken is quite different. They present a language with defeasible and strict rules, and strong negation (a sort of classical negation) and weak negation (a negation-by-failure), before considering the use of priorities and defeat. This language seems useful in modelling legal reasoning. However, that extra layer of logic programming notation and concepts masks the underlying simplicity of Dung's proposal. Our extension is more direct.

This paper is organized as follows: Section 2 introduces the basic argumentation framework of Dung (Dung, 1995). Section 3 is divided in two parts: the first one is devoted to the study of preference relations between arguments. The second part introduces a general and abstract preference-based argumentation framework. Section 4 introduces the proof theory. Finally, section 5 is devoted to some concluding remarks and perspectives. Proofs are given in the appendix.

2. The basic argumentation framework

2.1. Basic definitions

In Dung's work (Dung, 1993; Dung, 1995), an argumentation framework is defined as a pair consisting of a set of arguments and a binary relation representing the defeasibility relation between arguments. Here, an argument is an abstract entity whose role is only determined by its relation to other arguments. Then its structure and its origin are not known.

DEFINITION 2.1. *An* argumentation framework *is a pair* $<\mathcal{A}, \mathcal{R}>$ *where* $\mathcal{A}$ *is a set of arguments and* $\mathcal{R}$ *is a binary relation representing a defeasibility relationship between arguments, i.e.* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. *(A, B)* $\in \mathcal{R}$ *or equivalently "A* $\mathcal{R}$ *B" means that the argument A defeats* [1] *the argument B. We also say that A and B are in conflict.*

[1] Dung uses the word *Attack*.

Each defeasibility relation leads to an argumentation framework. Defeating arguments can in turn be defeated by other arguments so we need to define a notion of the *status* of arguments. This notion of *status* is the central element of any argumentation framework. Its definition takes as input the set of all possible arguments and their mutual relations of defeat, and produces as output a division of arguments into three classes of arguments:

- The class of *acceptable arguments*. They represent the "good" arguments. In the case of handling inconsistency in knowledge bases, for example, the formulas supported by such arguments will be inferred from the base.

- The class of *rejected arguments*. They are those arguments defeated by acceptable arguments. Such arguments would not be considered in the process of inference from a knowledge base, for example.

- The arguments which are neither acceptable nor rejected are gathered in the so-called class of *arguments in abeyance*.

Note that to define the rejected arguments and the arguments in abeyance of a given argumentation framework, we first need to determine the set of acceptable arguments of that framework.
The different approaches, which have been developed for reasoning within an argumentation framework, use either the individual acceptability proposed by Elvang and al. in (Elvang et al., 1993) or the joint acceptability proposed by Dung in ((Dung, 1993; Dung, 1995)).

DEFINITION 2.2. (Elvang et al., 1993) *Let* $<\mathcal{A}, \mathcal{R}>$ *be an argumentation framework. The class of* acceptable arguments, *denoted by* $\mathcal{C}_{\mathcal{R}}$, *is the set* $\{A \in \mathcal{A} \mid$ *there does not exist* $B \in \mathcal{A}$ *such that* $B \mathcal{R} A\}$. *The class of* rejected arguments, *denoted by* $Rej_{\mathcal{R}}$ *is the set* $\{A \in \mathcal{A} \mid \exists B \in \mathcal{C}_{\mathcal{R}}$ *such that* $B \mathcal{R} A\}$. *The set of* arguments in abeyance *is* $Ab_{\mathcal{R}} = \mathcal{A} \setminus (\mathcal{C}_{\mathcal{R}} \cup Rej_{\mathcal{R}})$.

EXAMPLE 2.1. *Let* $<\mathcal{A}, \mathcal{R}>$ *be the argumentation framework defined by* $\mathcal{A} = \{A, B, C\}$ *and* $\mathcal{R} = \{(A, B), (B, C)\}$. *The unique argument which is not defeated is A then* $\mathcal{C}_{\mathcal{R}} = \{A\}$. *The argument B is defeated by A which is acceptable so* $Rej_{\mathcal{R}} = \{B\}$. *The argument C is neither acceptable nor rejected so* $Ab_{\mathcal{R}} = \{C\}$.

The above definition of acceptability is very cautious. An argument is accepted if it is undefeated. With many defeasibility relations, when a given argument is defeated, the defeater itself is defeated and then unacceptable. So the acceptable arguments do not appear in any conflict. In

((Dung, 1993; Dung, 1995)), Dung extended the above definition using a notion of defence. The basic idea is that an argument is acceptable with respect to a set S of arguments if it is defended by that set S against all its defeaters.

DEFINITION 2.3. *(*Dung, 1995*) Let* $<\mathcal{A}, \mathcal{R}>$ *be an argumentation framework, and* $S \subseteq \mathcal{A}$. *An argument* A *is* defended *by* S *iff* $\forall\ B \in \mathcal{A}$, *if* $B\ \mathcal{R}\ A$ *then* $\exists\ C \in S$ *such that* $C\ \mathcal{R}\ B$.

Dung characterised the set of acceptable arguments by a monotonic function $\mathcal{F}$ that returns for each set of arguments the set of all arguments that are acceptable with respect to it.

DEFINITION 2.4. *(*Dung, 1995*)* $S \subseteq \mathcal{A}$. $\mathcal{F}(S) = \{A \in \mathcal{A}$ — A *is defended by* $S\}$.

Since the function $\mathcal{F}$ is monotonic, the set of acceptable arguments is defined as its least fixpoint. Moreover, Dung showed that if the argumentation framework $<\mathcal{A}, \mathcal{R}>$ is finitary (i.e. for each argument A there are finitely many arguments which defeat A), the function $\mathcal{F}$ is *continuous* and then its least fixpoint can be obtained by iterative application of $\mathcal{F}$ to the empty set.

DEFINITION 2.5. *(*Dung, 1995*) Let* $<\mathcal{A}, \mathcal{R}>$ *be an argumentation framework. The set of acceptable arguments, denoted by* $Acc_{\mathcal{R}}$, *is the least fixpoint of the function* $\mathcal{F}$.

EXAMPLE 2.2 (follows example 2.1). *Let* $<\mathcal{A}, \mathcal{R}>$ *be the argumentation framework defined by* $\mathcal{A} = \{A, B, C\}$ *and* $\mathcal{R} = \{(A, B), (B, C)\}$. *The set of acceptable arguments is* $\mathcal{C}_{\mathcal{R}} = \{A, C\}$. *In this case, the argument A defends C against its defeater B. The argument B is defeated by A which is acceptable so* $Rej_{\mathcal{R}} = \{B\}$. *The set of arguments in abeyance is empty,* $Ab_{\mathcal{R}} = \emptyset$.

2.2. Illustration

To illustrate the concepts of argument, defeat relation ($\mathcal{R}$) and acceptability, let's consider particular argumentation systems proposed for handling inconsistency in knowledge bases. The arguments are built from a propositional knowledge base Σ, which may be inconsistent. $\vdash$ stands for classical inference and $\equiv$ for standard logical equivalence.

DEFINITION 2.6. *An* argument *of* Σ *is a pair* (H, h) *where* $H \subseteq \Sigma$ *s.t:*

1. *H is consistent.*

2. $H \vdash h$.

3. *H is minimal (for set inclusion).*

H is called the support *and h the* conclusion *of the argument.* $\mathcal{A}(\Sigma)$ *denotes the set of all the arguments which are constructed from* Σ.

As examples of defeat relations, consider the *Rebut* and *Undercut* relations defined in (Elvang et al., 1993) as follows:

DEFINITION 2.7. *Let* (H_1, h_1), (H_2, h_2) *be two arguments of* $\mathcal{A}(\Sigma)$.

- (H_1, h_1) rebuts (H_2, h_2) *iff* $h_1 \equiv \neg h_2$.
- (H_1, h_1) undercuts (H_2, h_2) *iff* $\exists\ h \in H_2$ *such that* $h \equiv \neg h_1$.

Note that the definition of an argument and the definition of the defeasibility relation depend broadly on the considered application. See (Amgoud, 1999) for a thorough presentation of different definitions. In (Amgoud and Parsons, 2000) an argument may be seen as a plan for an agent to achieve one of its intentions. And in that particular case, and argument A defeats another argument B if the two arguments (the two plans) share the same resource.

In ((Prakken and Sartor, 1995; Prakken and Sartor, 1996)), an argument is a sequence of chained implicative rules. Each rule has a consequent part (consisting of one literal) and an antecedent part (consisting of a conjunction of literals). The consequent of each rule in a given argument is considered as a conclusion of that argument.

2.3. The limits of the basic framework

Dung's definition of acceptability disregards the quality of the arguments. However, the force of an argument can be often estimated by considering the beliefs used to build this argument. For example according to the preferences which can exist between the beliefs, an argument can be more or less strong than another argument. Let's consider the following example:

EXAMPLE 2.3. *Let* $<\mathcal{A}, \mathcal{R}>$ *be an argumentation framework with* $\mathcal{A} = \{A, B, C\}$ *and* $\mathcal{R} = \{(B, A), (C, B)\}$.
According to Dung, the set of acceptable arguments is $\{A, C\}$. *Yet, if we know that the argument B is preferred to A and C according to a preference relation Pref between arguments, then the argument C defeats B but B defends itself against C in some sense. Hence, the argument B might be considered as acceptable and A might be rejected.*

3. Preference-based argumentation framework

In order to refine the comparison between arguments and to enforce the concept of acceptability, we propose in this section to combine the preference relations between arguments with the defeasibility relations.

3.1. Preference relations

The notion of acceptability has been most often defined purely on the basis of other constructible arguments. However other criteria may be considered for comparing arguments. In the case of knowledge bases, for instance, specificity, or explicit priorities on the beliefs can be taken into account. More generally, preference relations can be used for comparing arguments.

We present below some of the preference relations proposed in the particular case of handling inconsistency in knowledge bases. The relations between the arguments are usually defined from priorities over the beliefs. Two kinds of priorities are most commonly encountered:

- Implicit priorities are extracted from the knowledge base. They are used in conditional approaches. Default rules can be (partially) ordered by exploiting specificity relations between the contexts. For example we know that all birds are animals; that generally animals do not fly and birds fly. For a given bird, the conflict is solved by privileging the rule about birds.

- Explicit priorities are specified outside the logical theory to which they apply. They may be given in the form of a partial order on the beliefs.

The preference relations between arguments are defined from a preordering over the supports of that arguments. The preordering over the supports is itself computed from the (explicit or implicit) priorities over the beliefs of the set Σ.

Reminder: A binary relation $\mathcal{P}$ defined on a set X is a preordering iff it is reflexive and transitive.
From a preordering $\mathcal{P}$, a relation of equivalence $\mathcal{EP}$ and a strict ordering relation $\mathcal{SP}$ can be defined as follows: x $\mathcal{EP}$ y iff x $\mathcal{P}$ y and y $\mathcal{P}$ x and x $\mathcal{SP}$ y iff x $\mathcal{P}$ y and not (y $\mathcal{P}$ x).
The relation of equivalence enables to partition X into several classes of equivalence. If the preordering is partial, which means that there exist elements of X not comparable, the different classes are not all

comparable too. However, if the preordering is total, a strict total ordering may be defined on the classes.
When a belief base Σ is equipped with a total preordering, it is equivalent to consider its partition into classes $\Sigma_1, \ldots, \Sigma_n$ such that Σ_1 contains the most preferred beliefs and Σ_n contains the less preferred ones. We say also that the belief base is stratified and we denote it by $\Sigma = \Sigma_1 \cup \ldots \cup \Sigma_n$.
Let's consider $\mathcal{P}$ a preordering over the elements of the belief base Σ. From $\mathcal{P}$ we define a preordering, denoted by Pref, between sets of beliefs. Pref will allow comparison of supports of arguments. Finally, the preference between arguments is defined as follows:

DEFINITION 3.1. *Let Pref be a (partial or total) preordering on subsets of Σ and let (H, h), (H', h') be two arguments of $\mathcal{A}(\Sigma)$. (H, h) is Pref-preferred to (H', h') iff H is preferred to H' w.r.t Pref.*

Notation: Let (H, h), (H', h') be two arguments of $\mathcal{A}(\Sigma)$. If Pref is a preordering then (H, h) Pref-preferred to (H', h') means that (H, h) is at least as "good" as (H', h'). $>>_{Pref}$ and $\equiv_{Pref}$ will denote respectively the strict ordering and the relation of equivalence associated with the preference between arguments. Hence, $(H, h) >>_{Pref} (H', h')$ means that (H, h) is strictly Pref-preferred to (H', h'). $(H, h) \equiv_{Pref} (H', h')$ means that (H, h) is Pref-preferred to (H', h') and (H', h') is Pref-preferred to (H, h).

An example of such a preference relation is the one based on the elitism principle (ELI-preference (Cayrol et al., 1993)). Let $\geq$ be a total preordering on Σ and $>$ be the associated strict ordering. In that case, the knowledge base Σ is assumed to be stratified into $(\Sigma_1, \ldots, \Sigma_n)$ such that Σ_1 is the set of $\geq$-maximal elements in Σ and Σ_{i+1} the set of $\geq$-maximal elements in $\Sigma \backslash (\Sigma_1 \cup \ldots \Sigma_n)$.
Let H and H' be two subbases of Σ. H is preferred to H' according to ELI-preference iff $\forall\ k \in H \setminus H'$, $\exists\ k' \in H' \setminus H$ such that $k > k'$.
Let (H, h), (H', h') be two arguments of $\mathcal{A}(\Sigma)$. $(H, h) >>_{ELI} (H', h')$ iff H is preferred to H' according to ELI-preference.

EXAMPLE 3.1. *$\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4$ is a knowledge base such that $\Sigma_1 = \{x, \neg r\}$, $\Sigma_2 = \{x \rightarrow t\}$, $\Sigma_3 = \{t \rightarrow r\}$ and $\Sigma_4 = \{\neg r \rightarrow p\}$. Let us consider the arguments A and B such that A : $(\{\neg r, \neg r \rightarrow p\}$, $p)$ and B : $(\{x, x \rightarrow t, t \rightarrow r\}$, $r)$. B undercuts A and B $>>_{ELI}$ A. Let us consider the arguments B and NR : $(\{\neg r\}$, $\neg r)$. NR rebuts B and NR $>>_{ELI}$ B.*

Other definitions of preference relations between arguments can be found in (Amgoud et al., 1996).

3.2. Argumentation framework

The result of combining preference relations with defeasibility relations leads to more complex argumentation frameworks defined as follows:

DEFINITION 3.2. *A* preference-based argumentation framework (PAF) *is a triplet* $<\mathcal{A}, \mathcal{R}, Pref>$ *where* $\mathcal{A}$ *is a set of arguments,* $\mathcal{R}$ *is a binary relation representing a defeat relationship between arguments,* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$, *and Pref is a (partial or complete) preordering on* $\mathcal{A} \times \mathcal{A}$.

Different definitions for the defeat relation ($\mathcal{R}$) and Pref lead to different preference-based argumentation systems.
Note that the defeat and preference relations are given independently. We propose to combine both relations in the following way:

DEFINITION 3.3. *Let A, B be two arguments of* $\mathcal{A}$. *B* attacks *A iff* $B\,\mathcal{R}\,A$ *and not*$(A >>_{Pref} B)$.

From a preference-based argumentation framework $<\mathcal{A}, \mathcal{R}, \text{Pref}>$, three categories of arguments will be defined:

- the set of acceptable arguments of the argumentation framework, denoted by $Acc_{\mathcal{R},Pref}$. It will be defined in Definition 3.6 below.
- $\{$A $\in \mathcal{A}$ — $\exists$ B $\in Acc_{\mathcal{R},Pref}$ such that B $\mathcal{R}$ A and not(A $>>_{Pref}$ B)$\}$ is the set of rejected arguments denoted by $Rej_{\mathcal{R},Pref}$.
  In other words, $Rej_{\mathcal{R},Pref}$ gathers the arguments which are attacked by acceptable arguments.
- $Ab_{\mathcal{R},Pref} = \mathcal{A} \setminus (Acc_{\mathcal{R},Pref} \cup Rej_{\mathcal{R},Pref})$ is the set of arguments which are in abeyance.

Note that the definition of the sets $Rej_{\mathcal{R},Pref}$ and $Ab_{\mathcal{R},Pref}$ (from the set of acceptable arguments) is the one present in most of the works about argumentation reasoning.

So next, we focus on the construction of $Acc_{\mathcal{R},Pref}$. Our purpose is to refine the abstract argumentation framework proposed by Dung in (Dung, 1995). Our contribution mainly concerns the concept of defence. We use two complementary notions of defence: a so-called *individual defence* (introduced via preference relations) and the notion of defence proposed by Dung, which may be called *joint defence.*

DEFINITION 3.4. *Let* $<\mathcal{A}, \mathcal{R}, Pref>$ *be a PAF. Let A, B be two arguments of* $\mathcal{A}$ *such that* $B\,\mathcal{R}\,A$. *A* defends itself *against B (wrt Pref) iff* $A >>_{Pref} B$. *An argument defends itself (wrt Pref) iff it is preferred w.r.t Pref to each of its defeaters.*
$\mathcal{C}_{\mathcal{R},Pref}$ *denotes the set of arguments defending themselves (wrt Pref) against their defeaters.*

EXAMPLE 3.2. *Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a PAF such that* $\mathcal{A} = \{A, B, C, D, E\}$, $\mathcal{R} = \{(C, D), (D, C), (A, E)\}$ *and* $C >>_{Pref} D$, *then* $\mathcal{C}_{\mathcal{R},Pref} = \{A, B, C\}$.

The set $\mathcal{C}_{\mathcal{R},Pref}$ contains also the arguments which are not defeated (in the sense of the relation $\mathcal{R}$). We have chosen to regard each argument in $\mathcal{C}_{\mathcal{R},Pref}$ as an acceptable argument. This corresponds to the individual point of view. It generalizes the concept of acceptability class proposed in (Elvang et al., 1993; Elvang and Hunter, 1995) in basic argumentation frameworks (i.e. without preference relations) since $\mathcal{C}_{\mathcal{R}}$ (as defined in Definition 2.2) is included in $\mathcal{C}_{\mathcal{R},Pref}$.

PROPOSITION 3.1. $\mathcal{C}_{\mathcal{R}} \subseteq \mathcal{C}_{\mathcal{R},Pref}$.

However, $\mathcal{C}_{\mathcal{R},Pref}$ is too restricted since it discards arguments which appear acceptable. Let us consider an argument A defeated by B such that B is preferred to A. It is clear that A does not belong to $\mathcal{C}_{\mathcal{R},Pref}$. Then assume that B itself is defeated by an argument C which is preferred to B. A might be regarded as an acceptable argument. This corresponds to the joint defence point of view.
When we instantiate the abstract schema of Dung by taking the defeasibility relation as the *Attack* relation given in Definition 3.3, the notion of joint defence becomes:

DEFINITION 3.5. *An argument A is* defended *by S (wrt Pref) iff* $\forall B \in \mathcal{A}$, *if* $B \; \mathcal{R} \; A$ *and* not($A >>_{Pref} B$) *then* $\exists \; C \in S$ *such that* $C \; \mathcal{R} \; B$ *and* not($B >>_{Pref} C$).

So, by applying the characteristic function $\mathcal{F}$ (as defined in Definition 2.4) to the empty set, and due to Definition 3.4, we obtain exactly the set of arguments defending themselves against their defeaters. More formally:

$$\mathcal{F}(\emptyset) = \mathcal{C}_{\mathcal{R},Pref}.$$

Now, we are ready to define our semantics of a general preference-based argumentation framework.

DEFINITION 3.6. *Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a finitary PAF in the sense of $\mathcal{R}$ (i.e. each argument is defeated by finitely many arguments). The set of acceptable arguments of the PAF <$\mathcal{A}$, $\mathcal{R}$, Pref> is defined as:*

$$Acc_{\mathcal{R},Pref} = \bigcup \mathcal{F}^{i>0}(\emptyset) = \mathcal{C}_{\mathcal{R},Pref} \cup [\bigcup \mathcal{F}^{i\geq 1}(\mathcal{C}_{\mathcal{R},Pref})]$$

The acceptable arguments are the ones which defend themselves against their defeaters ($\mathcal{C}_{\mathcal{R},Pref}$) and also the arguments which are defended (directly or indirectly) by the arguments of $\mathcal{C}_{\mathcal{R},Pref}$.

3.3. Application to inconsistency handling

Due to the use of propositional language and finite knowledge bases, in the particular case of handling inconsistency in knowledge bases, the two frameworks $<\mathcal{A}(\Sigma)$, Rebut, Pref$>$ and $<\mathcal{A}(\Sigma)$, Undercut, Pref$>$ are finitary. So, the associated sets of acceptable arguments are respectively:
$\mathcal{C}_{Rebut,Pref} \cup [\bigcup \mathcal{F}^{i\geq 1}(\mathcal{C}_{Rebut,Pref})]$,
$\mathcal{C}_{Undercut,Pref} \cup [\bigcup \mathcal{F}^{i\geq 1}(\mathcal{C}_{Undercut,Pref})]$.

EXAMPLE 3.3 (follows example 3.1). $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4$ *such that* $\Sigma_1 = \{x, \neg r\}$, $\Sigma_2 = \{x \rightarrow t\}$, $\Sigma_3 = \{t \rightarrow r\}$ *and* $\Sigma_4 = \{\neg r \rightarrow p\}$.
Let us consider the arguments A: ($\{\neg r, \neg r \rightarrow p\}$, p), B: ($\{x, x\rightarrow t, t\rightarrow r\}$, r) and C: ($\{x, x\rightarrow t, \neg r\}$, $\neg(t\rightarrow r)$).
B undercuts A and B $>>_{ELI}$ A, so A is not ELI-preferred to B and A does not defend itself against B: A $\notin \mathcal{C}_{Undercut,Pref}$.
C undercuts B and C $>>_{ELI}$ B, so C defends A against B. Moreover, C $\in \mathcal{C}_{Undercut,Pref}$. Then, A $\in \mathcal{F}(\mathcal{C}_{Undercut,Pref})$ and so A is an acceptable argument in the system $<\mathcal{A}(\Sigma)$, Undercut, Pref$>$.

4. Proof theory

So far, we have only provided a semantics for a preference-based argumentation framework, by defining the set of acceptable arguments. However, in practice we don't need to calculate the whole set $Acc_{\mathcal{R},Pref}$ in order to know the status of a given argument. In this section we propose a test for membership for an argument A, i.e. we propose a proof theory for our semantics.
For this purpose, we are motivated by the work of Prakken and Sartor in (Prakken and Sartor, 1997), developed in a logic programming like setting. We slightly improve their dialectical proof theory according two points: First, unlike Prakken and Sartor, we keep the original notion of defence (and so the original grounded semantics) proposed by Dung. Secondly, we allow a more general (i.e. less constrained) definition for dialogue trees.
Let us recall the main basic concepts of a dialectical proof theory, before presenting our improvement.

4.1. Definitions

In order to get a more efficient proof theory, Prakken and Sartor modify the characteristic function of an argumentation framework by using a notion of strict defence:

DEFINITION 4.1. *Let* $<\mathcal{A}, \mathcal{R}, Pref>$ *be a PAF, let* A *be an argument and* $S \subseteq \mathcal{A}$. A *is* strictly defended *by* S *iff* $\forall\, B \in \mathcal{A}$ *such that* B *attacks* A *then* $\exists\, C \in S$ *such that* C *strictly attacks* B *(i.e.* C *attacks* B *and* B *does not attack* C*).*

Indeed, the restriction to a strict defence is not necessary, since we have proved:

PROPOSITION 4.1. $\forall\, A \in Acc_{\mathcal{R},Pref}$, *A is strictly defended by* $Acc_{\mathcal{R},Pref}$. *In other words, the set of acceptable arguments strictly defends all its elements.*

This proposition is of great importance. It shows that to verify if an argument is acceptable (in the sense of Definition 3.6), we only have to take into account its strict defenders rather than all the defenders.

DEFINITION 4.2. *Let* $<\mathcal{A}, \mathcal{R}, Pref>$ *be a PAF.*
B indirectly attacks A *iff there exists a finite sequence of arguments* $A_0, \ldots, A_{2n+1}$ *such that:*

- $A = A_0$ *and* $B = A_{2n+1}$
- $\forall\, i$, $0 \leq i \leq 2n$, A_{i+1} *attacks* A_i.

B indirectly defends A *iff there exists a finite sequence of arguments* $A_0, \ldots, A_{2n}$ *such that:*

- $A = A_0$ *and* $B = A_{2n}$
- $\forall i$, $0 \leq i < 2n$, A_{i+1} *attacks* A_i.

The argument B *indirectly defends* A *against the argument* A_1.

The above definition was introduced by Dung using the defeasibility relation $\mathcal{R}$. The following results will enable us to prove the completeness of the proof theory.

PROPOSITION 4.2. *Let* $<\mathcal{A}, \mathcal{R}, Pref>$ *be a PAF.* $\forall\, x \in Acc_{\mathcal{R},Pref}$, x *is indirectly defended by arguments of* $\mathcal{C}_{\mathcal{R},Pref}$ *against all its defeaters.*

REMARK 4.1. *An argument indirectly defended against all its defeaters by arguments of $\mathcal{C}_{\mathcal{R},Pref}$ is not necessarily acceptable (i.e. does not necessarily belong to the set $Acc_{\mathcal{R},Pref}$). Consider the following example:*

EXAMPLE 4.1. *Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a PAF such that $\mathcal{A} = \{a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$, $\mathcal{R} = \{(a_1, a_0), (a_2, a_0), (a_4, a_2), (a_3, a_1), (a_5, a_3), (a_6, a_3), (a_7, a_6)\}$.*
Suppose that: $a_4 >>_{Pref} a_2 >>_{Pref} a_0$, $a_5 >>_{Pref} a_3 >>_{Pref} a_1 >>_{Pref} a_0$, $a_7 >>_{Pref} a_6 >>_{Pref} a_3$.
The argument a_0 is defeated by two arguments a_1 and a_2 and it does not defend itself. The argument a_0 is indirectly defended by a_7, which is in $\mathcal{C}_{\mathcal{R},Pref}$, against a_1. a_0 is also defended against a_2 by the argument a_4 which belongs to $\mathcal{C}_{\mathcal{R},Pref}$. However, a_0 is not in the set $Acc_{\mathcal{R},Pref}$ because it is indirectly attacked by the argument a_5 of $\mathcal{C}_{\mathcal{R},Pref}$.

PROPOSITION 4.3. *Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a PAF. If $x \in Rej_{\mathcal{R},Pref}$ then $\exists\ y \in \mathcal{C}_{\mathcal{R},Pref}$ such that y indirectly attacks x. In other words, if an argument is rejected then it is indirectly attacked by an argument of $\mathcal{C}_{\mathcal{R},Pref}$.*

4.2. The dialectical proof

According to Definition 3.6, an argument A is acceptable in a finite PAF if and only if it is in the result of finitely iterative applications of the function $\mathcal{F}$ to the set $\mathcal{C}_{\mathcal{R},Pref}$. Starting with A belonging to $\mathcal{F}^n$, for any argument B attacking A we find an argument C in $\mathcal{F}^{n-1}$ which defends A.
According to proposition 4.1, it is not necessary to find all the defenders, we just consider the strict defenders. The same process is repeated for each strict defender until there is no strict defender or defeater.

In a dialectical form, a proof that an argument A is acceptable will take the form of a dialogue tree, where each branch of the tree is a dialogue, and the root of the tree is the argument A. Each move in a dialogue consists of an argument of <$\mathcal{A}$, $\mathcal{R}$, $Pref$> which attacks the last move.

DEFINITION 4.3. *(*Prakken and Sartor, 1997*) A* dialogue *is a nonempty sequence of moves: $Move_i$ = ($Player_i$, Arg_i) ($i \geq 0$) such that:*

1. *$Player_i$ = PRO iff i is even, $Player_i$ = OPP iff i is odd.*

2. *$Player_0$ = PRO and Arg_0 = A.*

3. *If $Player_i$ = $Player_j$ = PRO and $i \neq j$ then $Arg_i \neq Arg_j$.*

4. *If $Player_i$ = PRO (i>1) then Arg_i strictly attacks Arg_{i-1}.*

5. *If $Player_i$ = OPP then Arg_i attacks Arg_{i-1}.*

Note that the player PRO begins the dialogue with the argument we are interested in. The players take turns, but have different roles. PRO must justify its initial argument A, while OPP wants to prevent A from being acceptable.

We define a *dialogue tree* as a finite tree where each branch is a dialogue. We adopt a very general definition for dialogue trees to allow for the dialectical form of argumentation: exchange of arguments. The main issue is then to define winning rules such that an argument A is acceptable iff the player PRO *wins* a suitable dialogue tree with root A. In the following, we propose a formalization which generalizes the one proposed by Prakken and Sartor in (Prakken and Sartor, 1997).

It is obvious that PRO wins a dialogue iff OPP cannot move (PRO ends the dialogue). However, defining the winning rule for a dialogue tree is less trivial. Let us consider the following example:

EXAMPLE 4.2. *Let $<\mathcal{A}, \mathcal{R}, Pref>$ be a PAF such that: $\mathcal{A} = \{a_0, a_{01}, a_{02}, a_{03}, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}\}$, $\mathcal{R} = \{(a_{10}, a_0), (a_{01}, a_{10}), (a_{12}, a_{02}), (a_{02}, a_{10}), (a_{03}, a_{11}), (a_{11}, a_0), (a_{13}, a_{14}), (a_{14}, a_{13})\}$.*
Suppose: $a_{03} >>_{Pref} a_{11} >>_{Pref} a_0$, $a_{01} >>_{Pref} a_{10} >>_{Pref} a_0$, $a_{12} >>_{Pref} a_{02} >>_{Pref} a_{10}$, $a_{13} >>_{Pref} a_{14}$. We are interested in the status of the argument a_0. The corresponding dialogue tree is presented in figure 1. PRO presents the arguments a_0, a_{01}, a_{02} and a_{03}; OPP presents the arguments a_{10}, a_{11} and a_{12}. a_{03} defends a_0 against a_{11} and a_{01} defends a_0 against a_{10}. Since a_{01} and a_{03} belong to $\mathcal{C}_{\mathcal{R},Pref}$, a_0 is acceptable.

In (Prakken and Sartor, 1997), the winning rule is defined by *a player wins a dialogue tree iff it wins all the branches of the tree.* That rule is very strict and in practice is only applied to particular dialogue trees, where each move labelled by OPP has only one child in the tree. For instance, the dialogue tree presented in Figure 1 is not won by PRO, in the sense of Prakken and Sartor. But a particular sub-tree will be won by PRO.

In the following, we propose a more general winning rule, which applies to arbitrary dialogue trees.

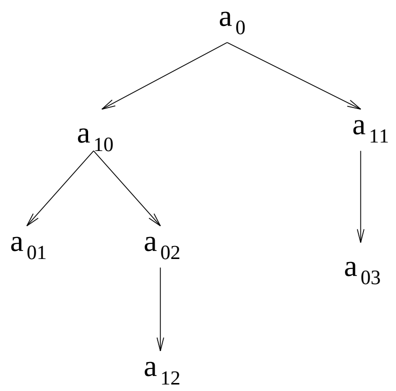

Figure 1. A dialogue tree

A dialogue tree can be considered as an AND/OR tree. A node corresponding to the player PRO is an AND node, and a node corresponding to the player OPP is an OR node. That distinction between nodes is due to the fact that an argument is acceptable if it is defended against all its defeaters. The children of a node containing an argument of PRO represent defeaters so they all must be defeated. In contrast, the children of a node containing an argument of OPP represent defenders of PRO so it is enough that one of them defeats the argument of OPP.

EXAMPLE 4.3 (follows example 4.2). *In the dialogue tree given in figure 1, a_0 is an AND node, while a_{10} is an OR node.*

DEFINITION 4.4. *A player* wins a dialogue *iff he ends the dialogue (he makes the last argument).*

A player who wins a dialogue does not necessarily win in all the subtrees of the dialogue tree. And if a player wins a dialogue, the last argument he makes is not necessarily acceptable. Consider the following example:

EXAMPLE 4.4. *Let $<\mathcal{A}, \mathcal{R}, Pref>$ be a PAF such that $\mathcal{A} = \{a_0, a_1, a_2, a_3, a_4\}$, $\mathcal{R} = \{(a_1, a_0), (a_2, a_0), (a_1, a_3), (a_3, a_1), (a_2, a_4), (a_4, a_2)\}$.*
Suppose that: $a_1 >>_{Pref} a_0$, $a_2 >>_{Pref} a_0$. Let's consider the argument a_0. The corresponding dialogue tree is presented in figure 2.
PRO presents the argument a_0 whereas OPP presents the two arguments a_1 and a_2. OPP wins the two dialogues and yet the arguments a_1 and a_2 are not acceptable.

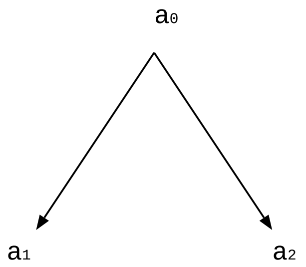

Figure 2. A dialogue tree

PROPOSITION 4.4. *If PRO wins a dialogue then his last move is an argument of* $\mathcal{C}_{\mathcal{R},Pref}$.

To formalize the winning rule of a dialogue tree, we define the concept of solution sub-tree.

DEFINITION 4.5. *A* candidate sub-tree *is a sub-tree of the dialogue tree containing all the edges of each AND node and exactly one edge of each OR node. A* solution sub-tree *is a candidate sub-tree whose branches are all won by PRO.*

EXAMPLE 4.5 (follows example 4.2). *The dialogue tree presented in example 4.2 has exactly two candidate sub-trees* S_1 *and* S_2 *(see Figure 3).*

EXAMPLE 4.6 (follows example 4.4). *The dialogue tree presented in example 4.4 has only one candidate sub-tree which is the dialogue tree itself.*

DEFINITION 4.6. *PRO* wins a dialogue tree *iff the dialogue tree has a solution sub-tree.*

EXAMPLE 4.7 (follows example 4.2). *PRO wins the dialogue tree presented in figure 1 because* S_2 *is a solution sub-tree.*

PROPOSITION 4.5. *If the player PRO wins the dialogue tree then:*

- *All the leaves of the solution sub-tree are arguments of* $\mathcal{C}_{\mathcal{R},Pref}$.
- *Each leaf of a solution sub-tree indirectly defends the arguments given by PRO in the dialogue leading to that leaf.*

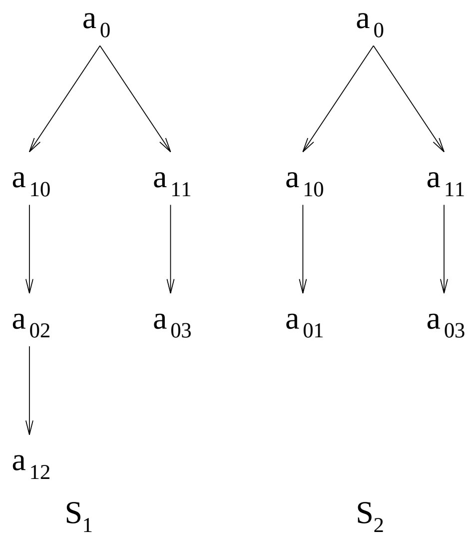

Figure 3. Candidate sub-trees

DEFINITION 4.7. *An argument A is* provably acceptable *iff there exists a complete dialogue tree whose root is A, won by the player PRO. A dialogue tree is* complete *iff for each move (PRO, Arg_i) ($i \leq 0$), the children are exactly all the defeaters of Arg_i.*

EXAMPLE 4.8 (follows example 4.1). *The argument a_0 is not provably acceptable because the dialogue tree whose root is a_0 is not won by PRO.*

EXAMPLE 4.9 (follows example 4.2). *The argument a_0 is provably acceptable because the player PRO won the dialogue tree.*

EXAMPLE 4.10 (follows example 4.4). *The argument a_0 is not provably acceptable because the player PRO didn't win the dialogue tree.*

Provably acceptable arguments exactly correspond to acceptable arguments:

PROPOSITION 4.6. *Let $<\mathcal{A}, \mathcal{R}, Pref>$ be a finite PAF.*

- $\forall\ x \in \mathcal{A}$, *if x is provably acceptable then each argument given by PRO belonging to the solution sub-tree is in $Acc_{\mathcal{R},Pref}$, in particular x.*

- $\forall\ x \in Acc_{\mathcal{R},Pref}$, *$x$ is provably acceptable.*

5. Conclusion

The work presented here concerns the acceptability of arguments in preference-based argumentation frameworks. Our first contribution is to identify two complementary notions of acceptability (individual acceptability and joint acceptability) and to present a unified general framework where both notions are used. Our second contribution is to take into account preference relations between arguments in order to select the most acceptable of them. The use of those preferences allows us to define a notion of individual defence and a notion of joint defence. We have proposed an argumentation framework in which an argument is acceptable if it is not defeated or if it defends itself against its defeaters or if it is defended by other arguments. We have also proposed a proof theory for this preference-based argumentation framework. The proof theory verifies whether a given argument A is acceptable or not. The proof theory is presented as a dialogue tree between two players PRO and OPP.

The idea of presenting the proof theory as a dialogue game between two players suggests the application of the argumentation framework to modelling dialogue and negotiation between agents.

In the past few years there have been a number of proposals for mechanisms for negotiation between agents that make use of argumentation. These proposals have largely been vague on the subject of how the generation and interpretation of arguments fits into the process of negotiation. In ((Amgoud et al., 2000a; Amgoud et al., 2000b)), this gap has been addressed by proposing a protocol based on our preference-based framework. One limit of our preference-based framework is that it is not able to take into account different preorderings on the set of arguments. These different preorderings can be considered to be contextual preferences, that is preferences which depend upon a particular context. In (Amgoud et al., 2000c) we have extended this argumentation framework by taking into account contextual preferences.

References

Amgoud, L.: July, 1999, 'Contribution a l'integration des préferences dans le raisonnement argumentatif'. In: *Thèse de doctorat de l'Université Paul Sabatier, Toulouse.*

Amgoud, L. and C. Cayrol: 1997, 'Integrating Preference Orderings into Argument-Based Reasoning'. In: *Proceedings of the 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR'97.* pp. 159–170.

Amgoud, L. and C. Cayrol: 1998, 'On the acceptability of arguments in preference-based argumentation framework'. In: *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*. pp. 1–7.

Amgoud, L., C. Cayrol, and D. LeBerre: 1996, 'Comparing arguments using preference orderings for argument-based reasoning'. In: *Proceedings of the 8th International Conference on Tools with Artificial Intelligence*. pp. 400–403.

Amgoud, L., N. Maudet, and S. Parsons: 2000a, 'Modelling dialogues using argumentation'. In: *Proceedings of the International Conference on Multi-Agent Systems, ICMAS'2000*. Boston, MA, pp. 31–38.

Amgoud, L. and S. Parsons: 2000, 'A dialogue framework based on argumentation'. *Submitted to Journal of Artificial Intelligence.*

Amgoud, L., S. Parsons, and N. Maudet: 2000b, 'Arguments, dialogue, and negotiation'. In: *Proceedings of the 14th European Conference on Artificial Intelligence*. pp. 338–342.

Amgoud, L., S. Parsons, and L. Perrussel: 2000c, 'An argumentation framework based on contextual preferences'. In: *Proceedings of the International Conference on Formal and Applied and Practical Reasoning*. pp. 59–67.

Cayrol, C., V. Royer, and C. Saurel: 1993, 'Management of preferences in Assumption-Based Reasoning'. *Lecture Notes in Computer Science (B. Bouchon-Meunier, L. Valverde, R.Y. Yager Eds.)* **682**, 13–22.

Dung, P. M.: 1993, 'On the acceptability of arguments and its fundamental role in non-monotonic reasoning and logic programming'. In: *Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI'93*. pp. 852–857.

Dung, P. M.: 1995, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games'. *Artificial Intelligence* **77**, 321–357.

Elvang, M., J. Fox, and P. Krause: 1993, 'Dialectic reasoning with inconsistent information'. In: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*. pp. 114–121.

Elvang, M. and A. Hunter: 1995, 'Argumentative logics: Reasoning with classically inconsistent information'. *Data and Knowledge Engineering* **12**, 125–145.

Kohlas, J., R. Haenni, and D. Berzati: 2000, 'Probabilistic argumentation systems and abduction'. In: *Proceedings of the 8th International Workshop on Non-Monotonic Reasoning, NMR'2000.*

Lin, F. and Y. Shoham: 1989, 'Argument systems - A uniform basis for non-monotonic reasoning'. In: *Proceedings of the first International Conference on Principles of Knowledge and Reasoning*. pp. 245–255.

Pinkas, G. and R. P. Loui: 1992, 'Reasoning from inconsistency: a taxonomy of principles for resolving conflicts'. In: *Proceedings of the 3rd International Conference on Principles of Knowledge representation and Reasoning*. pp. 709–719.

Pollock, J. L.: 1992, 'How to reason defeasibly'. *Journal of Artificial Intelligence* **57**, 1–42.

Prakken, H. and G. Sartor: 1995, 'On the relation between legal language and legal argument: assumptions, applicability and dynamic priorities'. In: *Proceedings of the 8th International Conference on Artificial Intelligence and Law*. pp. 1–10.

Prakken, H. and G. Sartor: 1996, 'A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning'. *Journal of Artificial Intelligence and Law* pp. 331–368.

Prakken, H. and G. Sartor: 1997, 'Argument-based extended logic programming with defeasible priorties'. *Journal of Applied Non-Classical Logics* **7**, 25–75.

Simari, G. R. and R. P. Loui: 1992, 'A mathematical treatment of defeasible reasoning and its implementation'. *Journal of Artificial Intelligence* **53**, 125–157.

Vreeswijk, G.: 1991, 'The feasibility of defeat in defeasible reasoning'. In: *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, KR'91.* pp. 526–534.

Vreeswijk, G.: 1997, 'Abstract argumentation systems'. *Journal of Artificial Intelligence* **90**, 225–279.

Appendix

Proposition 4.1. $\forall$ A $\in Acc_{\mathcal{R},Pref}$, A is strictly defended by $Acc_{\mathcal{R},Pref}$. In other words, the set of acceptable arguments strictly defends all its elements.

Proof

Let x $\in Acc_{\mathcal{R},Pref} = \bigcup \mathcal{F}^{i>0}(\emptyset) = \mathcal{C}_{\mathcal{R},Pref} \cup [\bigcup \mathcal{F}^{i\geq 1}(\mathcal{C}_{\mathcal{R},Pref})]$.
Let x' $\in \mathcal{A}$ such that x' attacks x. $Acc_{\mathcal{R},Pref}$ defends x then $Acc_{\mathcal{R},Pref}$ contains x" such that x" attacks x'.
Let i be the smallest index ≥ 0 such that $\mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref})$ contains x".
If x" does not strictly attack x' then x' attacks x".
x" $\in \mathcal{F}(\mathcal{F}^{i-1}(\mathcal{C}_{\mathcal{R},Pref}))$ then there exists y $\in \mathcal{F}^{i-1}(\mathcal{C}_{\mathcal{R},Pref})$ such that y attacks x'. Contradiction with the definition of i.

Proposition 4.2. Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a PAF. $\forall$ x $\in Acc_{\mathcal{R},Pref}$, x is indirectly defended by arguments of $\mathcal{C}_{\mathcal{R},Pref}$ against all its defeaters.

Proof

We show by induction on i that if x $\in \mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref})$ and there does not exist j < i such that x $\in \mathcal{F}^j(\mathcal{C}_{\mathcal{R},Pref})$ then $\exists$ y $\in \mathcal{C}_{\mathcal{R},Pref}$ such that y indirectly defends x.

- Let x $\in \mathcal{F}(\mathcal{C}_{\mathcal{R},Pref})$ such that x $\notin \mathcal{C}_{\mathcal{R},Pref}$ then $\exists$ x_1 s.t x_1 attacks x.
  $\exists$ $x_2 \in \mathcal{C}_{\mathcal{R},Pref}$ such that x_2 attacks x_1 then x_2 indirectly defends x.

- We assume that the property is true at an order i and we'll show that it is true at order i+1.
  x $\in \mathcal{F}^{i+1}(\mathcal{C}_{\mathcal{R},Pref})$ and $\forall$ j < i+1, x $\notin \mathcal{F}^j(\mathcal{C}_{\mathcal{R},Pref})$ (1)
  $\exists$ x_1 attacks x (due to (1)) and x $\in \mathcal{F}(\mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref}))$ then $\exists$ $x_2 \in \mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref})$ and x_2 attacks x_1.
  Let j be the smallest index $\leq$ i such that $x_2 \in \mathcal{F}^j(\mathcal{C}_{\mathcal{R},Pref})$, j ≥ 0.
  j = 0: $x_2 \in \mathcal{C}_{\mathcal{R},Pref}$ and indirectly defends x.
  j $\geq$ 1: according to induction hypothesis, $\exists$ y $\in \mathcal{C}_{\mathcal{R},Pref}$ such that y indirectly defends x_2 then y indirectly defends x.

Proposition 4.3. Let <$\mathcal{A}$, $\mathcal{R}$, Pref> be a PAF. If x $\in Rej_{\mathcal{R},Pref}$ then $\exists$ y $\in \mathcal{C}_{\mathcal{R},Pref}$ such that y indirectly attacks x. In other words, if an argument is rejected then it is indirectly attacked by an argument of $\mathcal{C}_{\mathcal{R},Pref}$.

Proof

$Rej_{\mathcal{R},Pref}$ = {x — $\exists$ y $\in Acc_{\mathcal{R},Pref}$ such that y attacks x}. Let x $\in Rej_{\mathcal{R},Pref}$, $\exists$y $\in Acc_{\mathcal{R},Pref}$ — y attacks x. According to proposition 4.2, $\exists$ z $\in \mathcal{C}_{\mathcal{R},Pref}$ such that z indirectly defends y. Due to Definition

4.2, since z indirectly defends y and y attacks x, we have z indirectly attacks x.

Proposition 4.4. If PRO wins a dialogue then his last move is an argument of $\mathcal{C}_{\mathcal{R},Pref}$.

Proof

Take A as the last argument provided by PRO. Since PRO wins the dialogue, then the player OPP has no argument which attacks A, so according to the definition of $\mathcal{C}_{\mathcal{R},Pref}$, A $\in \mathcal{C}_{\mathcal{R},Pref}$.

Proposition 4.5. If the player PRO wins the dialogue tree then:

- All the leaves of the solution sub-tree are arguments of $\mathcal{C}_{\mathcal{R},Pref}$.
- Each leaf of a solution sub-tree indirectly defends the arguments given by PRO in the dialogue leading to that leaf.

Proof

- Since PRO wins the dialogue tree then PRO wins all the dialogues of the solution sub-tree. According to proposition 4.4, the last moves of these dialogues are arguments of the class $\mathcal{C}_{\mathcal{R},Pref}$.
- Since PRO starts the game and gives the last argument in each dialogue of the solution sub-tree then there exists a sequence A_0 ... A_{2n} such that A_0 = A (the first argument given by PRO), A_{2n} = B $\in \mathcal{C}_{\mathcal{R},Pref}$ (B is a leaf of the solution sub-tree) and A_{i+1} attacks A_i. Then B indirectly defends A and indirectly defends each argument $A_{2(n-i)}$ with $1 \leq$ i $\leq$ n.

Proposition 4.6 Let $<\mathcal{A}$, $\mathcal{R}$, Pref$>$ be a finite PAF.

- $\forall$ x $\in \mathcal{A}$, if x is provably acceptable then each argument given by PRO belonging to the solution sub-tree is in $Acc_{\mathcal{R},Pref}$, in particular x.
- $\forall$ x $\in Acc_{\mathcal{R},Pref}$, x is provably acceptable.

Proof

1. Let x be a provably acceptable argument then there exists a dialogue tree whose root is x and won by PRO. So, there exists a solution sub-tree whose leaves are all arguments of the class $\mathcal{C}_{\mathcal{R},Pref}$ and given by PRO.
   Let 2i be the depth of the solution sub-tree (i.e. the maximum number of moves of all the dialogues in the solution sub-tree).

As usual, we define the height of a node N in a tree as the depth of the sub-tree of root N.
We show by induction on p that $\forall$ p such that $0 \leq$ p $\leq$ i, $\{y$ — y is an argument given by PRO in a node of height $\leq$ 2p belonging to the solution sub-tree$\}$ is included in $Acc_{\mathcal{R},Pref}$.

- Case p = 0: The leaves of the solution sub-tree are all elements of $\mathcal{C}_{\mathcal{R},Pref}$ and then of $Acc_{\mathcal{R},Pref}$.
- Assume that the property is true at an order p, and consider the order p+1.

It is sufficient to consider the arguments given by PRO in a node of height (2p+2) of the solution sub-tree. Let y be such an argument. Since y is given by PRO, all the arguments y'_k attacking y appear in the solution sub-tree as children of y, and each argument y'_k is itself strictly attacked by one argument z_k appearing in the solution sub-tree as a child of y'_k. So each z_k is given by PRO and appears in a node of height 2p of the solution sub-tree.
By induction hypothesis, each argument z_k belongs to $Acc_{\mathcal{R},Pref}$. Since all attackers of y have been considered, $Acc_{\mathcal{R},Pref}$ defends y. So, y belongs to $Acc_{\mathcal{R},Pref}$. And the property is shown at the order p+1.

2. $x \in Acc_{\mathcal{R},Pref}$. Construct a tree with root x. Let i be the smallest index ≥ 0 such that $x \in \mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref})$. We show by induction on i that there exists a solution sub-tree whose root is x and depth is $\leq$ 2i.
   Case i = 0: $x \in \mathcal{C}_{\mathcal{R},Pref}$, the depth of the tree is 0.
   Assume that the property is true at order i and consider the order i+1.
   $x \in \mathcal{F}^{i+1}(\mathcal{C}_{\mathcal{R},Pref})$ and $x \notin \mathcal{F}^j(\mathcal{C}_{\mathcal{R},Pref})$ with j < i+1.
   Let $x_1, \ldots, x_n$ be the arguments attacking x. x_j attacks x, and $x \in \mathcal{F}^{i+1}(\mathcal{C}_{\mathcal{R},Pref}) = \mathcal{F}(\mathcal{F}^i(\mathcal{C}_{\mathcal{R},Pref}))$.
   According to proposition 4.1, $\exists\, y_j \in Acc_{\mathcal{R},Pref}$ such that y_j strictly attacks x_j. Since y_j defends x (definition of $\mathcal{F}$), then $y_j \in F^i(\mathcal{C}_{\mathcal{R},Pref})$. By induction hypothesis applied to y_j, there exists a solution sub-tree whose root is y_j and the depth is $\leq$ 2i. The same construction is done for each x_j. So we obtain a solution sub-tree whose root is x and its depth is $\leq$ 2(i+1).