Book Title
Book Editors
IOS Press, 2003

Handling controversial arguments in bipolar argumentation systems

Claudette Cayrol [a], Caroline Devred [b] and Marie-Christine Lagasquie-Schiex [a]
[a] *IRIT, UPS, 31062 Toulouse, Cédex 9, {ccayrol, lagasq}@irit.fr*
[b] *CRIL, Université d'Artois, 62307 Lens, Cédex, devred@cril.univ-artois.fr*

Abstract. We consider bipolar argumentation frameworks, which extend Dung's argumentation frameworks by handling two independent kinds of interaction between arguments, attack and support. In this bipolar context, we propose new semantics for coping with the problem of controversial arguments (arguments which indirectly attack and indirectly defend a same argument).

Keywords. Argumentation Frameworks

1. Introduction

Argumentation has become an influential approach to model defeasible reasoning and dialogues between agents, based on the exchange of interacting arguments (see e.g. [17,19,1,16,20]). The following illustrative example presents the arguments exchanged during the meeting of the editorial board of a newspaper.

Ex. 1

Arg. a**:** If we have the agreement and without censoring, the important information I on the person X must be published.
Arg. b_1**:** I concerns the prime minister X who may use the right of censoring.
Arg. c_1**:** The prime minister has resigned. So, I no longer concerns a prime minister.
Arg. d**:** The resignation will be announced officially this evening on TV Channel 1.
Arg. b_2**:** I is a private information and X does not agree for publication.
Arg. c_2**:** Any information concerning the prime minister is a public information and not a private information.
repetition of Arg. c_1 **and** d**:** ...
Arg. c_3**:** But I is of national interest, so I cannot be considered as a private information.

In most existing systems, the interaction takes the form of a conflict, usually called attack. For example, an argument can be a pair (set of assumptions, conclusion), where the set of assumptions entails the conclusion according to some logical inference schema. Then, a conflict occurs between two arguments for instance if the conclusion of one of them contradicts an assumption of the other one. On Ex. 1, b_1 (resp. b_2) is in conflict with a. The main issue of any argumentation system is the selection of acceptable sets of arguments, based on the way arguments interact. Intuitively, an acceptable set of arguments must be in some sense coherent and strong enough (e.g. able to defend itself against all attacking arguments). The concept of acceptability has been explored through the use of argumentation frameworks, such as the fruitful Dung's argumentation framework [11], abstracting from the nature of the arguments. In such an abstract framework, from a set of arguments and a binary "attacks" relation, different semantics for acceptability are proposed, each one being characterized by several requirements that a *set of arguments* must satisfy in order to be selected. These selected sets of arguments are called extensions.

However, Dung's semantics do not always lead to expected conclusions, faced with the so-called controversial arguments. Roughly speaking, an argument c is *controversial w.r.t. an argument* a iff c indirectly defends a (e.g. c attacks an attacker of a) and also indirectly defends an attacker of a. Intuitively, even there is no direct conflict between c and a, it seems uncautious to accept together both arguments. On Ex. 1, c_1 defends a (against b_1) and also defends b_2 which is an attacker of a. So, c_1 is controversial w.r.t. a. In some sense, c_1 reinstates an attacker of a. That's why we find uncautious to accept both c_1 and a in the *same extension* because we are interested in the definition of "coherent" *sets of arguments*[1] (the simplest notion of "coherent" set proposed by Dung is the notion of conflict-free set). Moreover, since c_1 is the unique defender of a against b_1, it seems also uncautious to derive a from the discussion. However, whatever Dung's semantics, the unique extension contains a, c_1 and c_3. This problem has motivated the definition of new prudent semantics by [7,9][2], in which the notion of coherence is enforced: pairs of arguments which conflict indirectly cannot belong to a same extension.

Moreover, recent work on argumentation [5,16,20] has advocated for the representation of another kind of basic interaction between arguments. Indeed, it can be the case in a dialog that an agent brings to light some new piece of information and so advances an argument which justifies an assumption used by an argument provided by another agent (agents are assumed independent).This kind of interaction between arguments is not captured by the notion of defence. It is rather a kind of *support*. On Ex. 1, we consider that the argument d given by an agent Ag_1 supports the argument c_1 given by another agent Ag_2. It is not only a "dialog-like speech act": a new piece of information is really given. In [5], Dung's framework has been extended to cope with both kinds of interaction, into a so-called bipolar abstract argumentation framework. *Bipolarity* refers to the existence of two independent kinds of information which represent repellent forces. New semantics for acceptability have been defined, based on a more complex notion of attack, called the supported attack. On Ex. 1, the fact that d supports an attacker of b_1 may be considered as a supported attack on b_1 by d. These new semantics ensure that no supported attack can occur within an extension. However, a new kind of controversial arguments appears in a bipolar argumentation framework. On Ex. 1, d supports an indirect attacker of a and also supports a defender of a. Even if d is not directly controversial w.r.t. a, it seems uncautious to accept d and a in the same extension. So the purpose of this paper is to propose new semantics which handle this new kind of controversial arguments in a bipolar setting.

We first present the background concerning Dung's framework, its extension for handling controversial arguments and the bipolar extension (Section 2). Then, we present

[1] which is different of the definition of the derivability of an argument.
[2] There exist some other propositions for solving this problem, see Sect. 5.

our contribution (Section 3). The key issue is to define a new kind of controversial arguments, the b-controversial arguments, and a notion of conflict which encompasses both indirect attacks and supported attacks. Then, we propose an analog of the prudent semantics for a bipolar framework. Despite the apparent complexity of this new framework, most of the properties satisfied by the prudent semantics are preserved in the bipolar setting (Section 4).

2. Background

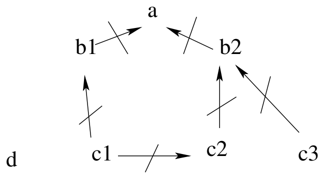
2.1. Dung's framework

Let us present some basic definitions at work in Dung's theory of argumentation [11].

Def. 1 *A* finite argumentation framework *is a pair* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ *where* $\mathcal{A}$ *is a finite set of so-called arguments and* $\mathcal{R}$ *is a binary relation over* $\mathcal{A}$ *(a subset of* $\mathcal{A} \times \mathcal{A}$*), the attacks relation.*

An argumentation framework can be represented by a directed graph in which each argument is a vertex and the edges are defined by the attacks relation: $\forall a, b \in \mathcal{A}$, $a\mathcal{R}b$ is represented by $a \not\to b$.

Ex. 1 (cont)
This example can be formalized by the framework AF_1 represented by the following interaction graph (note that the argument d remains isolated because it cannot be linked to the other arguments using the attacks relation):



In the following, we consider a finite argumentation framework $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ and its associated interaction graph $\mathcal{G}$. The first important notions are the notion of acceptability and the notion of conflict which are used for defining the extensions:

Def. 2 *Let* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ *be a finite argumentation framework.*

- *Let* $a \in \mathcal{A}$ *and* $S \subseteq \mathcal{A}$*.* a *is* acceptable *w.r.t.* S *iff* $\forall b \in \mathcal{A}$ *s.t.* $b\mathcal{R}a$*,* $\exists c \in S$ *s.t.* $c\mathcal{R}b$*. A set of arguments is acceptable w.r.t.* S *when each of its elements is acceptable w.r.t.* S*.*
- S *is* conflict-free *iff* $\nexists a, b \in S$*, s.t.* $a\mathcal{R}b$*.*
- *A subset* S *of* $\mathcal{A}$ *is* admissible *for* AF *iff* S *is conflict-free and acceptable w.r.t.* S*.*
- S *is* a preferred extension *of* AF *iff it is maximal w.r.t.* $\subseteq$ *among the admissible sets for* AF*.* S *is* a stable extension *of* AF *iff it is conflict-free for* AF *and* $\forall a \in \mathcal{A} \setminus S$*,* $\exists b \in S$ *s.t.* $b\mathcal{R}a$*.* S *is* the grounded extension *of* AF *iff it is the least element w.r.t.* $\subseteq$ *among the admissible sets s.t. each argument acceptable w.r.t.* S *belongs to* S*.*

In Ex. 1, $E_1 = \{c_1, c_3\}$ is an admissible set for AF_1 and $E_2 = \{a, c_1, c_3, d\}$ is the preferred extension, the stable extension and the grounded extension of AF_1.

Formally, the grounded extension of AF can be characterized as the least fixed point w.r.t. $\subseteq$ of the characteristic function $\mathcal{F}_{\mathsf{AF}}$.

Def. 3 *The* characteristic function*, denoted* $\mathcal{F}_{\mathsf{AF}}$*, of* AF *is defined by:* $\mathcal{F}_{\mathsf{AF}} : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ *and* $\mathcal{F}_{\mathsf{AF}}(S) = \{a \mid a \textit{ is acceptable w.r.t. } S\}$*.*

Dung has also identified problematical arguments, the controversial arguments.

Def. 4 *Let* $a, b \in \mathcal{A}$*.* a indirectly attacks b *iff* $\exists$ *an odd-length path from* a *to* b *in* $\mathcal{G}$*.* a indirectly defends b *iff* $\exists$ *an even-length path from* a *to* b *in* $\mathcal{G}$ *(length* ≥ 2*).*
a *is* controversial *w.r.t.* b *iff* a *indirectly attacks* b *and* a *indirectly defends* b*.*

In Ex. 1, c_1 indirectly attacks a, c_1 defends a, c_1 is controversial w.r.t. a.

2.2. Handling indirect conflict

Let $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ be a finite argumentation framework. In order to handle controversial arguments [7,9] define new semantics, the *p(rudent)* semantics. [7,9] refine Dung's notions of conflict-free and admissibility by exploiting the notion of indirect attack proposed by Dung.

Def. 5 *Let* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ *be a finite argumentation framework. Let* $S \subseteq \mathcal{A}$*.*

- S *is* p(rudent)-admissible *for* AF *iff* S *is acceptable w.r.t.* S *and* $\forall a, b \in S$*,* a *doesn't indirectly attack* b*.*
- S *is* a preferred p-extension *of* AF *iff it is maximal w.r.t.* $\subseteq$ *among the p-admissible sets for* AF*.* S *is* a stable p-extension *of* AF *iff* $\forall a, b \in S$*,* a *doesn't indirectly attack* b *and* $\forall a \in \mathcal{A} \setminus S$*,* $\exists b \in S$ *s.t.* $b\mathcal{R}a$*.*

In Ex. 1, $E_1 = \{c_1, c_3\}$ is a p-admissible set for AF_1. $E_2 = \{c_1, c_3, d\}$ is the preferred p-extension of AF_1. AF_1 has no stable p-extension.

A grounded p-extension is defined using the *p-characteristic function*.

Def. 6

- *The* p-characteristic function *of* AF *is defined as follows:* $\mathcal{F}^p_{\mathsf{AF}} : 2^{\mathcal{A}} \longrightarrow 2^{\mathcal{A}}$ *and* $\mathcal{F}^p_{\mathsf{AF}}(S) = \{a \mid a \textit{ is acceptable w.r.t. } S \textit{ and } \forall c, b \in S \cup \{a\}, c \textit{ doesn't indirectly attack } b.\}$*.*
- *Let* j *be the lowest integer s.t. the sequence* $(\mathcal{F}^{p,i}_{\mathsf{AF}}(\varnothing))_{i \in \mathbb{N}}$ *is stationary from rank* j*.* $\mathcal{F}^{p,j}_{\mathsf{AF}}(\varnothing)$ *is the* grounded p-extension *of* AF*.*

Contrariwise to $\mathcal{F}_{\mathsf{AF}}$, $\mathcal{F}^p_{\mathsf{AF}}$ is in general nonmonotonic w.r.t. $\subseteq$. This prevents from defining a notion of grounded p-extension as the least fixed point of $\mathcal{F}^p_{\mathsf{AF}}$. However, the sequence $(\mathcal{F}^{p,i}_{\mathsf{AF}}(\varnothing))_{i \in \mathbb{N}}$ is monotonic w.r.t. $\subseteq$. In Ex. 1, E_2 is the grounded p-extension of AF_1.

2.3. Bipolar argumentation frameworks

As already said, arguments may be conflicting. These conflicts are captured by the attacks relation in an argumentation framework, and may be considered as negative interactions. Then, the concept of defence has been introduced in order to reinstate some of the attacked arguments, namely those whose attackers are in turn attacked. So, most logical theories of argumentation assume that if an argument a_3 defends an argument a_1 against an argument a_2, then a_3 is a kind of support for a_1, so a positive interaction. It holds in the basic Dung's framework: only negative interaction is explicitly represented by the *attacks* relation, and positive interaction is implicitly represented through the notion of defence. In this case, support and attack are *dependent* notions. It is a parsimonious strategy, but it is not a correct description of the process of argumentation in realistic examples: in Ex. 1, the link between the argument d and the other arguments cannot be expressed with the attacks relation. So, we need a more complex argumentation framework, in order to formalize situations where two *independent* kinds of interaction are available: a positive and a negative one. Following [16,20], [5,4] propose a bipolar argumentation framework. This new framework[3] is an extension of the basic Dung's framework in which a new kind of interaction between arguments is represented by the *supports* relation[4]. This framework presents the following features:

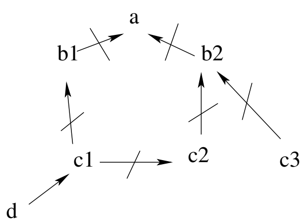
- an abstract point of view: arguments and interactions are considered as initial data;
- the existence of the support interaction is justified by the independence of the sources in a multi-agent system: different agents propose and exchange different arguments related to their own knowledge;
- the inference mechanism is based on the selection of acceptable sets of arguments and does not use a dialectical proof mechanism[5]: this selection is performed after the agents have exchanged their arguments.
- The supports relation is assumed to be totally independent of the attacks relation.

Def. 7 *A finite* bipolar argumentation framework $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ *consists of a finite set* $\mathcal{A}$ *of arguments, a binary relation* $\mathcal{R}_{att}$ *on* $\mathcal{A}$ *called the* attacks relation *and another binary relation* $\mathcal{R}_{sup}$ *on* $\mathcal{A}$ *called the* supports relation*.*

In the following, we consider a finite bipolar argumentation framework $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$. Note that BAF can still be represented by a directed graph $\mathcal{G}_b$ called the *bipolar interaction graph* with two kinds of edges, one for the attacks relation and another one for the supports relation. Consider $a, b \in \mathcal{A}$, $a\mathcal{R}_{att}b$ is represented by $a \not\to b$ and $a\mathcal{R}_{sup}b$ is represented by $a \to b$.

Ex. 1 (cont)

The whole discussion during the editorial board meeting can now be formalized by the bipolar framework BAF_1 represented by:



The fact that d supports an attacker of b_1 may be considered as a kind of negative interaction between d and b_1, which is however weaker than a direct attack. From a

[3]The bipolar argumentation framework presented here is a simplified version of what has been discussed in [5,4,18].

[4]If the support relation is removed, we retrieve Dung's framework.

[5]Even if there exist links between the selected acceptable sets and some particular dialectical proofs – see [3, 12].

cautious point of view, such arguments cannot appear together in a same extension. In order to address this problem, [5,4] introduce a new kind of attack which combines a sequence of supports with a direct attack.

Def. 8 *A* supported attack *for an argument b by an argument a is a sequence of supports followed by one attack: $a_1\mathcal{R}_1 \ldots \mathcal{R}_{n-1}a_n$, $n \geq 3$, with $a_1 = a$, $a_n = b$, s.t. $\forall i = 1 \ldots n-2$, $\mathcal{R}_i = \mathcal{R}_{sup}$ and $\mathcal{R}_{n-1} = \mathcal{R}_{att}$.*

In Ex. 1, there is a supported attack for b_1 by d.

3. Controversial arguments in a bipolar framework

In this paper, we are interested in handling the controversial arguments in a bipolar framework. Because we want to take into account the supports, some particular configurations may appear: in Ex. 1, d supports c_1 which is controversial w.r.t. a and it seems uncautious to accept d and a in the same extension, even if d is not directly controversial w.r.t. a. So, d and a illustrate a new kind of controversial arguments: the *b(ipolar)-controversial* arguments.

Def. 9 *Let $a, b \in \mathcal{A}$. a is* b-controversial *w.r.t. b iff a supports (by a sequence of supports) an argument x which indirectly attacks b and a supports (by a sequence of supports) an argument y which indirectly defends b.*

In this paper, we propose an analog of the p-semantics for a finite bipolar argumentation framework $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ in order to manage the b-controversial arguments: if a is b-controversial w.r.t. b, a and b cannot belong to the same extension.

So, the first step consists in enforcing the notion of coherence.

Def. 10 *Let $S \subseteq \mathcal{A}$. S is* bp-conflict-free[6] *iff $\nexists a, b \in S$ s.t. there exists a sequence $a_1\mathcal{R}_{sup} \ldots \mathcal{R}_{sup}a_n\mathcal{R}_{att} \ldots \mathcal{R}_{att}a_{n+m}$, $n \geq 1$, with $a_1 = a$, $a_{n+m} = b$, and m is an odd number.*

In Ex. 1, $\{a, d\}$ is not bp-conflict-free (via c_1, c_2, b_2).

When the sequence of supports is empty, this notion is exactly the notion defined by [7,9] applied to the partial framework $\langle \mathcal{A}, \mathcal{R}_{att} \rangle$. A bp-conflict-free set does not contain indirect attacks.

From this notion of bp-conflict-free, and keeping the notion of acceptability (Def. 2), different semantics for the acceptability can be proposed: the *b(ipolar)p(rudent)-semantics*.

Def. 11 *Let $S \subseteq \mathcal{A}$. S is* bp-admissible *for BAF iff S is bp-conflict-free and acceptable w.r.t. S. S is a* preferred bp-extension *of BAF iff S is maximal for $\subseteq$ among the bp-admissible sets for BAF. S is a* stable bp-extension *of BAF iff S is bp-conflict-free and $\forall a \notin S$, $\exists b \in S$ s.t. $b\mathcal{R}_{att}a$.*

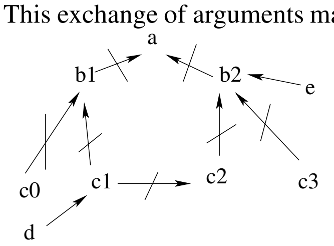
[6] bp means b(ipolar)p(rudent).

These new semantics are illustrated with the following example which is a complexification of Example 1.

Ex. 2 Some arguments are added to the dialog of Example 1 (arguments c_0 and e which are emphasized).

Arg. a**:** If we have the agreement and without censoring, the important information I on the person X must be published.
Arg. b_1**:** I concerns the prime minister X who may use the right of censoring.
Arg. c_0**:** *We are in democracy and even the prime minister cannot use the right of censoring.*
Arg. c_1**:** The prime minister has resigned. So, I no longer concerns a prime minister.
Arg. d**:** The resignation will be announced officially this evening on TV Channel 1.
Arg. b_2**:** I is a private information and X does not agree for publication.
Arg. e**:** *I concerns X's son which is suspected to be a spy of a foreign country.*
Arg. c_2**:** Any information concerning the prime minister is a public information and not a private information.
repetition of Arg. c_1 **and** d**:** ...
Arg. c_3**:** But I is of national interest, so I cannot be considered as a private information.

This exchange of arguments may be formalized by BAF_2 which is represented by:



$\{d, c_0, c_3, a\}$ does not contain indirect attacks but is not bp-conflict-free.
In this case, $\{a, c_0, c_3\}$ and $\{c_0, c_1, c_3, d, e\}$ are the two preferred bp-extensions of BAF_2 and BAF_2 has no stable bp-extension.

We define the bp-characteristic function as in Def. 6.

Def. 12 *The* bp-characteristic function *of* BAF *is defined as follows:* $\mathcal{F}^{bp}_{\mathsf{BAF}} : 2^{\mathcal{A}} \longrightarrow 2^{\mathcal{A}}$ *and* $\mathcal{F}^{bp}_{\mathsf{BAF}}(S) = \{a \mid a$ *is acceptable w.r.t.* S *and* $S \cup \{a\}$ *is bp-conflict-free*$\}$.

The bp-characteristic function is in general nonmonotonic. An important property of the p-characteristic function fails for the bp-characteristic function: the sequence $(\mathcal{F}^{bp,i}_{\mathsf{BAF}}(\varnothing))_{i\in\mathbb{N}}$ is nonmonotonic w.r.t. $\subseteq$. So, it is not possible to define a "grounded bp-extension"[7] as in Def. 6.

4. Some properties

Let $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup}\rangle$ be a finite bipolar argumentation framework. AF denotes the partial argumentation framework $\langle \mathcal{A}, \mathcal{R}_{att}\rangle$. First of all, note that, when $\mathcal{R}_{sup} = \varnothing$, the bp-semantics correspond exactly to p-semantics.

The following results establish links between bp-semantics and Dung's semantics or p-semantics.

[7]However, structural restrictions of BAF will allow for such a definition (it could be the subject of a future work). It is also possible to define a weaker notion of grounded extension inspired from the work of [8].

Prop. 1

(i) *Every bp-admissible set for* BAF *is also admissible and p-admissible for* AF*. The converse does not hold.*

(ii) *Every stable bp-extension of* BAF *is also a stable extension and a stable p-extension of* AF*. The converse does not hold.*

Proof:

(i) Obvious from the definition of bp-admissible. The negative result for the converse is given by Ex. 2: $\{a, c_0, c_3, d\}$ is an admissible set and a p-admissible set for AF_2 , but it is not a bp-admissible set for BAF_2.

(ii) Let S be a stable bp-extension of BAF. S contains no bp-conflict and attacks each argument outside of S. So S contains no indirect attack (and S is conflict-free) in AF and attacks each argument outside of S. Accordingly, S is a stable extension of AF and a stable p-extension of AF. The negative result for the converse is illustrated by $\mathsf{BAF} = \langle \mathcal{A} = \{a, b, c, d\}, \mathcal{R}_{att} = \{(d, b), (b, a)\}$ and $\mathcal{R}_{sup} = \{(c, b)\}\rangle$: $\{c, d, a\}$ is a stable extension and a stable p-extension of AF , but it is not a stable bp-extension of BAF.

■

Basic properties of Dung's framework are preserved. Since $\varnothing$ is a bp-admissible set, a bipolar argumentation framework has always at least one preferred bp-extension. Moreover, we have :

Prop. 2 *The set of all the bp-admissible sets for* BAF *forms a complete partial order w.r.t.* $\subseteq$*. And, for each bp-admissible set* S *for* BAF*, there exists a preferred bp-extension* E *of* BAF *s.t.* $S \subseteq E$*.*

Proof: The set of all the bp-admissible sets for BAF has a least element w.r.t. $\subseteq$ since $\varnothing$ is bp-admissible. Since $\mathcal{A}$ is a finite set, every chain of bp-admissible sets for BAF has a least upper bound w.r.t $\subseteq$ (namely the union of these sets). So, the set of all the bp-admissible sets for BAF is a complete partial order w.r.t. $\subseteq$. The second point follows immediately from the fact that $\mathcal{A}$ is finite. ■

Prop. 3 *Each stable bp-extension is a preferred bp-extension. The converse is false.*

Proof: Let S be a stable bp-extension of BAF. From Prop. 1, S is a stable extension of AF. So, S is acceptable w.r.t. S (see [11]). Moreover, S contains no bp-conflict. Accordingly, S is bp-admissible for BAF. Then, for each argument a not belonging to S, there exists a conflict in $S \cup \{a\}$, which is also a bp-conflict. So, S is maximal for $\subseteq$ among the bp-admissible sets for BAF. That is S is a preferred bp-extension of BAF.

The negative result for the converse is given by Ex. 2: $\{c_0, c_1, c_3, d, e\}$ is a preferred bp-extension of BAF_2, but it is not a stable bp-extension.

■

The bp-characteristic function characterizes the bp-admissible sets and the preferred bp-extensions.

Prop. 4 *Let $S \subseteq \mathcal{A}$.*

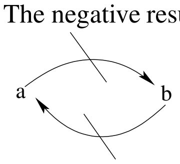
(i) *S is a bp-admissible set iff $S \subseteq \mathcal{F}^{bp}_{\mathsf{BAF}}(S)$.*

(ii) *If S is a preferred bp-extension, we have $S = \mathcal{F}^{bp}_{\mathsf{BAF}}(S)$. The converse does not hold.*

Proof: $S \subseteq \mathcal{F}^{bp}_{\mathsf{BAF}}(S)$ means that for each $a \in S$, a is acceptable w.r.t. S and $S \cup \{a\} = S$ contains no bp-conflict. It means exactly that S is bp-admissible for BAF.

Assume that S is a preferred bp-extension. It remains to prove that $\mathcal{F}^{bp}_{\mathsf{BAF}}(S) \subseteq S$. Let a be an element of $\mathcal{F}^{bp}_{\mathsf{BAF}}(S)$. a is acceptable w.r.t. S and $S \cup \{a\}$ contains no bp-conflict. So, $S \cup \{a\}$ is bp-admissible. If a does not belong to S, there is a contradiction with the maximality of S.

The negative result for the converse is given by the following example:



In this case, $\mathcal{R}_{sup} = \varnothing$. Assume that $S = \varnothing$, $\mathcal{F}^{bp}_{\mathsf{BAF}}(S) = \varnothing$. However, the two preferred bp-extensions are $\{a\}$ and $\{b\}$. ■

To sum up, every finite bipolar argumentation framework has at least one preferred bp-extension, and zero, one or many stable bp-extensions. A stable bp-extension is never empty (when $\mathcal{A} \neq \varnothing$).

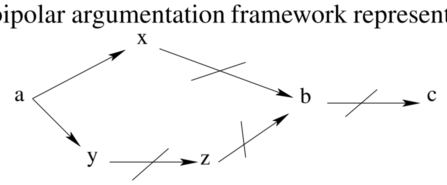
Our main purpose was to provide highly prudent semantics for which extensions cannot contain b-controversial arguments. This requirement is fulfilled, as shown by :

Prop. 5 *Let $a, b \in \mathcal{A}$. If a is b-controversial w.r.t. b then $\{a, b\}$ cannot be included in any bp-admissible set.*

Proof: If a is b-controversial w.r.t. b, there is a bp-conflict in $\{a, b\}$. ■

The above result does not prevent a or b from belonging to a bp-admissible set, but not to the same one.

Ex. 3 Let BAF_3 be a bipolar argumentation framework represented by:



The argument a is b-controversial w.r.t. c. The bp-extensions are $\{a, x, y\}$ and $\{x, c\}$. So, no bp-admissible set contains both a and c.

As a consequence of Prop. 5, we obtain that no argument belonging to an odd-length cycle in AF can belong to a bp-admissible set. It departs from [2] who handles in the same way odd-length and even-length cycles in an argumentation framework $\langle \mathcal{A}, \mathcal{R}_{att} \rangle$.

However, some other Dung's properties are not preserved.

The set of arguments which are not attacked is included in a preferred bp-extension, but not in all of them. In Ex. 3, $\{a, x, y\}$ is the set of arguments which are not attacked. It is a preferred bp-extension, but there is another one $\{x, c\}$.

Unlike preferred extensions, a bipolar argumentation framework can have more than one preferred bp-extension even if it is well-founded[8] (see BAF_2). Indeed, we have :

Prop. 6 *For every preferred bp-extension, E_{bp} of BAF, there exists at least one preferred extension E of AF such that $E_{bp} \subseteq E$.*

For every preferred bp-extension, E_{bp} of BAF, there exists at least one preferred p-extension E_p of AF such that $E_{bp} \subseteq E_p$.

Proof: A preferred bp-extension is bp-admissible, and from Prop. 1, it is also admissible and p-admissible. Dung has proved that every admissible set is included in a preferred extension. And [7] has proved that every p-admissible set is included in a preferred p-extension. ■

As a consequence of Prop. 6, when AF has a unique preferred extension E (for instance, when AF is well-founded, or without any even-length cycle), E includes every preferred bp-extension of BAF.

5. Related works

The work reported in this paper extends work by [7,9] on prudent semantics.

Some other works in the literature have extended Dung's semantics to cover the problem of controversial arguments.

In [13,15], the work is carried out in the specific context of argumentation in logic programming. The knowledge base is a logic program P. An argument A for a goal G is a set of negative hypotheses of the form ($\texttt{not}\ p$) such that G can be derived from P extended with A. A set of hypotheses A conflicts with another set A' when A is an argument for p with ($\texttt{not}\ p$) belonging to A'. The approach focusses on sets of hypotheses, not on sets of arguments. All the definitions for the different semantics are given for sets of hypotheses. So coherence is not defined for a set of arguments, as in our approach, but only for a set of hypotheses.

In [2], the issue is to handle in the same way odd-length and even-length cycles in an argumentation framework. It departs from our approach, since every argument belonging to an odd-length cycle is bp-controversial w.r.t. any argument of the cycle [10], no such argument can belong to a bp-extension.

[6] presents another semantics which handle the controversial arguments: the new careful semantics according to which two arguments cannot belong to the same extension whenever one indirectly attacks a third argument while the other one indirectly defends the third one (so, the controversial arguments cannot belong to the careful extensions).

As for handling support interaction, few works have been published [5,16,20]. Note that a notion of support has also appeared recently in work by [14]. This work is carried out in a specific logical framework of argumentation. The arguments are built using a knowledge base (containing a non defeasible part, called the background, and a defeasible part), classical inference and priorities between rules. Interaction between arguments is modelled by an attack relation which combines classical refutation and priorities. Then, an abductive component is added as follows. Given an available set of as-

[8] *i.e.* there does not exist an infinite sequence of attacks.

sumptions, the idea is to complete the background by some assumptions in order to be able to build new arguments. Namely, S is a supporting information for a goal G if a good argument for G can be provided from the defeasible knowledge, taking into account the background augmented with S. This kind of support is very different from the notion presented in this paper. Supporting information aims at providing new arguments, for instance for attacking old arguments. In our proposal, support occurs between arguments as a new kind of positive interaction which is essential in a multi-agent context.

6. Conclusion

Prudent semantics have been proposed in Dung's framework to cope with the problem of controversial arguments. The idea is to exclude indirect attacks in an extension. In a bipolar argumentation framework, where the presence of two kinds of interaction between arguments, conflict and support, enables to define more complex attacks, the supported attacks, the problem of controversial arguments becomes more complex because a new kind of controversial arguments appears, the b-controversial arguments. So, we have revisited the prudent semantics in a bipolar setting. We have shown that most of the properties satisfied by the preferred prudent semantics and the stable prudent semantics were preserved in the bipolar setting. However, it is no longer possible to define an equivalent for the grounded prudent semantics.

A direction for future work concerns computational issues. Algorithms described in [18] can serve as a basis for developing algorithms for computing prudent extensions in a bipolar framework. Moreover, it will be interesting to study the credulous decision problem associated with these prudent semantics: "given an argument (advanced for instance in a dialogue), is this argument in at least one prudent extension ?". For that purpose, an idea is to include heuristics into the dialectical proof theories described in [10] for the credulous decision problem.

A particular notion of attack can represent the notion of exception in nonmonotonic inheritance reasoning but it is not the only meaning[9]. Nevertheless, another direction for a future work could be the study of these particular cases.

Acknowledgements

We would like to thank the referees for their comments which helped us to improve this paper. The authors have been partly supported by the Région Nord/Pas-de-Calais, the IRCICA Consortium and the European Community FEDER Program.

References

[1] L. Amgoud, N. Maudet, and S. Parsons. Arguments, dialogue and negotiation. In *16th European Conference on Artificial Intelligence (ECAI 2004)*, pages 338–342, 2004.

[9] In our example, the agents propose arguments using their beliefs. These beliefs either are confirmed (supported), or invalidated (attacked) by other agents' arguments.

[2] P. Baroni and M. Giacomin. Solving semantic problems with odd-length cycles in argumentation. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty,7th European Conference, ECSQARU 2003*, volume LNAI 2711, pages 440–451, Aalborg, Denmark, July 2-5 2003.

[3] C. Cayrol, S. Doutre, and J. Mengin. Dialectical proof theories for the credulous preferred semantics of argumentation framework. In *he Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001)*, pages 668–679, 2001.

[4] C. Cayrol and M-C. Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In *Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 366–377, Barcelona - Spain, 6–8 july 2005.

[5] C. Cayrol and M-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 378–389, Barcelona - Spain, 6–8 july 2005.

[6] S. Coste-Marquis, C. Devred, and P. Marquis. Inference from controversial arguments. In *The 12th International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2005)*, 2–6 december 2005.

[7] S. Coste-Marquis, C. Devred, and P. Marquis. Prudent semantics for argumentation frameworks. In *The 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2005)*, 14–16 november 2005.

[8] S. Coste-Marquis, C. Devred, and P. Marquis. Contrained argumentation frameworks. In *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR' 06)*, pages 112–122, 2006.

[9] S. Coste-Marquis, C. Devred, and P. Marquis. Sémantiques prudentes pour les systèmes d'argumentation. In *15e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA 2006)*, 24–28 janvier 2006.

[10] S. Doutre. *Autour de la sémantique préférée des systèmes d'argumentation*. PhD thesis, Université Paul Sabatier de Toulouse, 2002.

[11] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[12] P.E. Dunne and T.J.M. Bench-Capon. Two party immediate response disputes: Properties and efficiency. *Artificial Intelligence*, 149(2):221–250, 2003.

[13] A.C. Kakas, P. Mancarella, and P.M. Dung. The acceptability semantics for logic programs. In *Proc. of International Conference on Logic Programming (ICLP)*, pages 504–519, 1994.

[14] A.C. Kakas and P. Moraïtis. Argumentation based decision making for autonomous agents. In *Proc. of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 883–890. ACM Press, 2003.

[15] A.C. Kakas and F. Toni. Computing Argumentation in Logic Programming. *Journal of Logic and Computation*, 9(4):515–562, August 1999.

[16] Nikos Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information systems*, 26(4):259–277, 2001.

[17] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty, 1995.

[18] M. Mardi, C. Cayrol, and M-C. Lagasquie-Schiex. Bipolarité en argumentation : acceptabilité et algorithmes. Rapport interne IRIT 2005-20-R, IRIT-UPS, octobre 2005.

[19] A. Prakken and G. Vreeswijk. Logics for defeaseable argumentation. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd edition*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht, 2002.

[20] Bart Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic in Computation*, 13:319–346, 2003.