Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex

IRIT – Institut de Recherche en Informatique de Toulouse, UMR 5505

Mémoire présenté en vue de l'obtention de

l'Habilitation à diriger les recherches

spécialité : informatique

par

MARIE-CHRISTINE LAGASQUIE-SCHIEX

Inconsistency handling in nonmonotonic reasoning: from consistency restoration to argumentation

Soutenue le 10 juillet 2009 devant :

Mme	C. Cayrol	Professeur	Univ. de Toulouse III	Examinateur
Mr	A. Hunter	Senior scientist	Univ. College of London	Rapporteur
Mr	J. Lang	DR CNRS	Univ. Paris Dauphine	Examinateur
Mr	P. Marquis	Professeur	Univ. d'Artois	Examinateur
Mme	O. Papini	Professeur	Univ. de Marseille	Rapporteur

Abstract

Nonmonotonic reasoning enables the simulation of a very human property: the ability to make a correct reasoning with inconsistent, uncertain, incomplete information.

This domain has been largely studied in Artificial Intelligence for many years. Many systems, often ad-hoc, have been defined. The most reliable systems, from the reasoning quality point of view, are very expensive in time and also in space. My first works in this domain were dedicated to this point: the subject of my PHD Thesis has been a comparative study of some inference relations using temporal complexity and some other criteria. Precisely, the studied problem was the deduction in a *consistency restoration* framework: as starting point, an *inconsistent* knowledge base is given, representing an agent's knowledge, then we select some *preferred* and *consistent* subbases of the initial base, and we define nonmonotonic inference relations using the results obtained with the classical deduction process on the preferred subbases.

The mechanisms used in this context are often simple: a conclusion is kept if "all the subbases agree with this conclusion", or if "at least one subbase agrees with this conclusion", or if "at least one subbase agrees with this conclusion and none of the subbases agrees with the negation of this conclusion". Therefore, in a first step, I have proposed a generalization of this notion of nonmonotonic inference relation using a mechanism of vote in order to obtain more intuitive conclusions but also sound conclusions (I did not want to infer at the same time a conclusion and its negation); among the new relations obtained in this new framework, one can find a relation which infers a conclusion if "the majority of the subbases agrees".

An important part of nonmonotonic reasoning can be easily translated in a problem of conflict resolution. So the next part of my work has been the study of conflict resolution using a specific point of view: what are the basic principles used in conflict resolution? They can be very different: optimality of the solutions with regard to set-inclusion, respect of preferences over knowledge, ...

Generally, solutions of a conflict resolution problem are sets of knowledge on which classical deduction can be applied and these solutions are mutually in conflict. Very naturally, this property leads us to the notion of arguments and attack between arguments, and so to a specific type of reasoning, "argumentation", based on exchange and valuation of interacting arguments. This process can be decomposed into several steps: creation of the arguments, identification of interactions between these arguments, valuation of these arguments, selection of arguments considered as the most acceptable and then choice of conclusions of the argumentation system (for instance, which can correspond to a nonmonotonic inference).
Argumentation allows the modelization of interactions between rational agents, so it is a tool for making nonmonotonic reasoning over a group of agents, and not only over one agent.
In this framework, my work is focused on a generalization of a particular argumentation system: the abstract system proposed by Dung. The following aspects have been pointed out: "valuation" of arguments, their "acceptability", use of "bipolar" interactions, "merging" or "revision" in argumentation and also use of argumentation in order to make *practical reasoning*.

This part of my work is very important in terms of devoted time, so it also will be important in term of the number of pages in this document.

Eventually, the final point presented in this document will be about a specific part of Logic and Game Theory: Boolean games. In these games, I have reused the notions of logic and preferences and I have highlighted some new links between argumentation and games.

Contents

Synthesis of my work

Context

I do my research in the “Plausible reasonings, Decision and Proof Methods” team (RPDMP) at IRIT in Paul Sabatier University (UPS) in Toulouse. This team is specialised in the formalisation of mechanisms for plausible reasoning and decision making (for instance, maintaining or restoring consistency, reasoning under uncertainty, ...). My works belong to this framework and mainly concern the handling of inconsistency. Let me introduce this concept in a few words.

One of the aims of Artificial Intelligence is the study of the methods enabling the manipulation of knowledge in order to perform reasoning (deduction of information, explanation of facts, decision ...). Classical logic is a very usual and suitable tool in this domain as long as the knowledge is consistent. But, it becomes insufficient when the knowledge becomes inconsistent or incomplete, uncertain (so, one talks about belief); for example, using inconsistent information, with classical logic one can deduce anything! Nevertheless, the manipulation of this kind of information is essential in order to formalize and to automate “common sense” reasoning. In this case, it may be necessary to call old conclusions into question because of the presence of new facts, and also to make a choice between several opposed propositions; this is the nonmonotonic aspect of the reasoning. In this context there exist at least three main points which can be taken into account in order to classify my works:

How to handle an inconsistent base? Does one have additional elements in order to solve conflicts?

Generally, when one must handle an inconsistency, three different methods can be considered:

- one *tries to avoid it*, with mechanisms which really remove the inconsistencies from the base; this may be done gradually, as information arrives (this is belief revision); this may be also done if information arrives from distinct sources (this is knowledge merging);
- one *restores consistency*, *i.e*, one identifies some *information subsets* which are consistent (here, the main idea is the fact that these subsets represent possible states of the world) and one applies classical deduction on them; in this framework, there exist several methods which depend essentially on the way in which one identifies the *consistent subsets*; in this case, the initial information set[1] remains unchanged but one reasons using its consistent subsets;
- one *accepts inconsistency*; in this case, one needs new kinds of languages (nonmonotonic logics, paraconsistent logics, ...); here too, the initial information set remains unchanged and it is completely used for reasoning.

Of course, the distinction between these three methods is not so clear. For instance, some works have historically been defined in the context of inconsistency acceptation but they also can be defined as a method of consistency restoration in the largest sense of this concept.

[1] I use the term “information set” in a very general sense. In a logical context, it can be viewed as a synonym of the notion of (knowledge or belief) base. In a more informal context, it can be used for pointing out a set of different elements which are used for reasoning (informal sentences in natural language, ...).

This is the case of argumentation. Initially, argumentation in AI has been defined as a way of accepting inconsistency without identifying all the consistent subsets representing the states of the world; it was a tool for reasoning towards a specific conclusion; it identified arguments "for this conclusion" and arguments "against this conclusion" then these arguments and their relations were used for reasoning. However, if we consider an enlarged definition for consistency restoration, argumentation can also be viewed as an example of this approach: each argument is a consistent subset on which we can apply classical deduction (even if this subset represents only a very partial state of the world).

All my works have been realized in the framework of **inconsistency handling** using as starting point the **consistency restoration** approach and as a final point a particular type of **inconsistency acceptation**. Indeed, my main goal is *to exploit the decomposition of the initial inconsistent information set into* ***consistent subsets*** *which are related* (generally, by a conflict relation) using possibly some additional elements in order to solve conflicts. Among these elements, one can find some **preferences between beliefs** which can be classified according to the type of ordering relations used on beliefs:

- those which are defined *a priori*,
- those which are deduced from explicit relations given on beliefs (attack relations, support relations, ...),
- those which are deduced from implicit relations between beliefs (logical dependency, ...),
- those which must satisfy some constraints,
- ....

In my works, I have only used preferences defined either *a priori*, or deduced from explicit relations given on beliefs.

What is the goal of the reasoning? (to deduce new information? To explain some information? ...)

Initially, my work concerned deduction (with the study of nonmonotonic inference), then it has evolved to abduction with the study of argumentation. However, it is interesting to note that argumentation is a sufficiently powerful tool to be used for deduction, decision making, handling negotiation dialogues, ...

How many agents are concerned by the reasoning: One agent? Several agents?

If there is only one agent, then this agent reasons with one inconsistent information set; if there are several agents then each agent may have one information set, consistent or not, and it will be necessary to study the way to build a reasoning between all these agents.

On this point, my work has considered both cases but not for all the studied topics; for instance, the study of nonmonotonic inferences concerns only one agent, whereas the argumentation study relates to successively one agent then several agents.

Thematic evolution of my work

In this section, I give a more precise description of the studied topics, transitions from a topic to another one, collaborations, supervision works and associated publications.

During my PHD Thesis, and in the context described in the previous section, I made a classification of several nonmonotonic entailment relations using different formal points of view:

- the problem of computational complexity (worst case time complexity);
- the cautiousness, *i.e.* the ability of the entailment relation to produce the "correct" conclusions;
- the properties of the studied relations, *i.e.* "the adequacy of the relation to the model we wish to formalize".

At the end of my PhD Thesis, I also studied several algorithms for nonmonotonic entailment relations. This has allowed me to compare several algorithms on numerous benchmarks for one nonmonotonic entailment relation. The most efficient algorithm is a direct consequence of the study of the computational complexity and another one is very interesting in the case of a "precompilation" of the knowledge.

The aim of all these studies was to synthesize the numerous works on this subject in order to define an efficient system for nonmonotonic reasoning.

From a "timing" point of view, this first contribution corresponds to my PHD thesis (1992–1995) supervised by Claudette and Michel CAYROL and concerns only the topic "nonmonotonic inference"[2]. All these works have been presented in an invited paper at the PRC-GDR Artificial Intelligence conference [LS95c] and have been published in my PhD Thesis, and in several articles and reports [CLS93, LS94, CLS94b, CLS94a, CLS95, LS95a, LS95b, CLSS96a]. The part dedicated to the algorithmic study has been published in three publications (one in an international journal) [CLSS96a, CLSS96b, CLSS98].

After this preliminary step, all my works are oriented towards:

- handling inconsistency by selecting consistent information subsets,
- generalizing existing methods and enriching them,
- exhibiting links between these methods.

Therefore, since the end of my PHD Thesis, I made research contribution in the four following topics: nonmonotonic inference, conflict resolution, argumentation and games.

Topic 1: Nonmonotonic inference These works directly followed the works initiated during my PhD Thesis. I have defined a generic framework for describing deduction in nonmonotonic reasoning. With this framework, one can retrieve the nonmonotonic inference relations I have used in my PhD Thesis. Moreover one defines some new nonmonotonic inference relations whose results intuitively seem better.

The idea lies in the selection, in the knowledge base, of some subsets representing "opinions" (one can use preferences between beliefs in order to assign a strength to the different "opinions"). Then, these "opinions" "vote" using classical logic and the deduction of a formula is determined by the result of this "vote".

The aim of this general framework is to define a single framework capturing existing nonmonotonic inference relations and some new nonmonotonic inference relations which are able to conclude (on a formula) even when there exist some "opinions against" (this formula).

These works have been done from 1997 to 2000 in collaboration with Claudette CAYROL. They were presented in a technical report [CLS97] and in a national publication [CLS00].

Transition from nonmonotonic inference to conflict resolution one question appeared during the study of nonmonotonic inferences using the strict framework of consistency restoration: "why should one choose some selection mechanisms rather than other ones?" The part concerning the "theoretical comparison of nonmonotonic inference relations" of my PHD Thesis gave some lines of thinking on this subject but which involved too much things at the same time (the selection mechanism but also the entailment principle). I wanted to answer this question in an independent way; therefore I studied the principles implemented in conflict resolution.

[2]These works have been introduced by the subject of my research master: the study of a formal unified framework and its tools allowing to use several approaches defined in Artificial Intelligence (classical logic, default logic, modal logics ...) see [LS92].

Topic 2: Conflict resolution is a different approach allowing the handling of inconsistency again in the strict framework of consistency restoration. One identifies conflicts between beliefs. Then, using partial preferences (an ordering relation given for each conflict), one tries to restore the consistency of the beliefs, *i.e.* to choose the beliefs to remove in order to obtain a new belief base without inconsistency (on which one can use classical inference if one wants to make deduction, and thus to join Topic 1 "nonmonotonic inference").

My approach is original by its methodological aspect: I have tried to identify principles that must be satisfied. There exist different kinds of solutions and I worked to classify them and to compare them with the solutions obtained using nonmonotonic entailment relations.

These works have been done from 2000 to 2003, with Claudette CAYROL and Jérôme MENGIN, respectively professor and assistant professor at UPS, and have been presented in [CLSM00]. They have also led to a co-supervision of a student of a professional master degree[3] of UPS, Laurent VIDAL, in charge of the implementation of some solutions proposed in our study (see [Vid99]).

Then, they led to a collaboration with Hélène FARGIER, CNRS senior researcher at IRIT, aimed at studying links between conflict resolution and decision in the context of real problems in project management (in this collaboration, I co-supervised two students of research master degree: Loïc DUSSEUX in 2002 and Pascal LAFOURCADE in 2003 – see [Laf03]).

> ***Transition from nonmonotonic inference to argumentation*** A study by Claudette CAYROL (see [Cay95]) highlighted an equivalence between some argumentation processes and nonmonotonic inference relations. This work thus quite naturally directed me towards the field of argumentation; moreover this field is sufficiently powerful to simulate many kinds of nonmonotonic reasoning. The main idea was to establish a footbridge towards other fields that I knew less, like the study of interactions between rational agents (for instance, under the form of – argumentation – dialogues), but also with the same finality: handling inconsistency by the selection of consistency information subsets.

Topic 3: Argumentation is a tool for reasoning from an inconsistent information set using a enlarged consistency restoration: there exists a notion of "argument" linked to a given conclusion (for example a subset of the initial base which induces the conclusion). We can also use these arguments with different explicit relations between arguments (for example, an attack relation: an argument "defeats" another argument, or a support relation: an argument "helps" another argument). With these arguments and these relations, one can build an axiomatic system that can take into account the classical notion of "arguments for" and "arguments against", but also the notions of "valuations of arguments", "strengths of arguments", "acceptable arguments", ... Then these acceptable arguments are taken into account in order to define the conclusion of the argumentation system. Of course, these conclusions depend on the goal of the system (deduction, decision taking, negotiation, ...).

Note that argumentation is closely related to the previous domains since argumentation offers a general framework in which one can define nonmonotonic inference relations and handle conflict resolution.

It is also important to note that argumentation, approached here by the way of an enlarged concept of consistency restoration, is a much more general process, initially introduced for accepting inconsistency (and not for consistency restoration in the strict sense of this concept) and which also appears in many other domains than Artificial Intelligence (for instance, in Philosophy or in Cognitive Psychology).

These works have been done from 2001 to 2009 and have been presented in several reports [CLS01, CLS02b, CDLSM02a, CLS03a, CLS04, MCLS05, CDLS06b, ALS07], and in national and international publications [CDLSM02b, CLS02a, CLS03e, CLS03b, CLS03c, CLS05d, ACLS04b, ACLSL08a, CMDK$^+$05], [CLS05c, CLS05e, CDLS06a, CLS07a, CDLS07, CMDK$^+$07, ADLS08, ADLS09, CLS09] (three are book parts on the domain, see [CLS03c, ADLS09, CLS09], and several others are publications in international journals, see [CLS05d, CMDK$^+$07, ACLSL08a]).

This topic being among the main topics developed in my group, I have taken part to the supervision of one PHD Thesis on this subject (Sylvie DOUTRE, see [Dou02]) and in the supervision of different students in research master degree (Dominique GAY in 2004, Mathieu MARDI in 2005 and Aurore MIQUEL in 2007).

[3] DESS IRR: DESS en Intelligence artificielle, Reconnaissance des formes et Robotique.

These works have been realized in collaboration with other members of RPDMP team (Claudette CAYROL, Leila AMGOUD, Florence DUPIN DE SAINT-CYR and Jérôme MENGIN), and also with members of other french universities (Sylvie and Pierre MARQUIS and Sébastien KONIECZNY at CRIL laboratory in Lens and Caroline DEVRED at LERIA laboratory in Angers).

Transition from argumentation to games Argumentation, as Game Theory, aims at representing and reasoning with interactions between rational agents. Moreover, many links between these two topics have been highlighted.

Topic 4: Games This part of my work follows my works on argumentation. Indeed, argumentation can be viewed as an interaction process between rational agents; moreover, games which can model agents' interactions, are already used for simulating some argumentation processes (see [Dun95]). So it appears natural to explore the links between argumentation and game theory. After a first study on this topic, I have become mostly interested in a specific class of games, called Boolean games, which present three points of interest:

- they use classical logic for modelling agents' preferences,
- therefore they can easily be improved by some preference representation languages already used in non-monotonic inference,
- they propose solution concepts which correspond to consistent subsets.

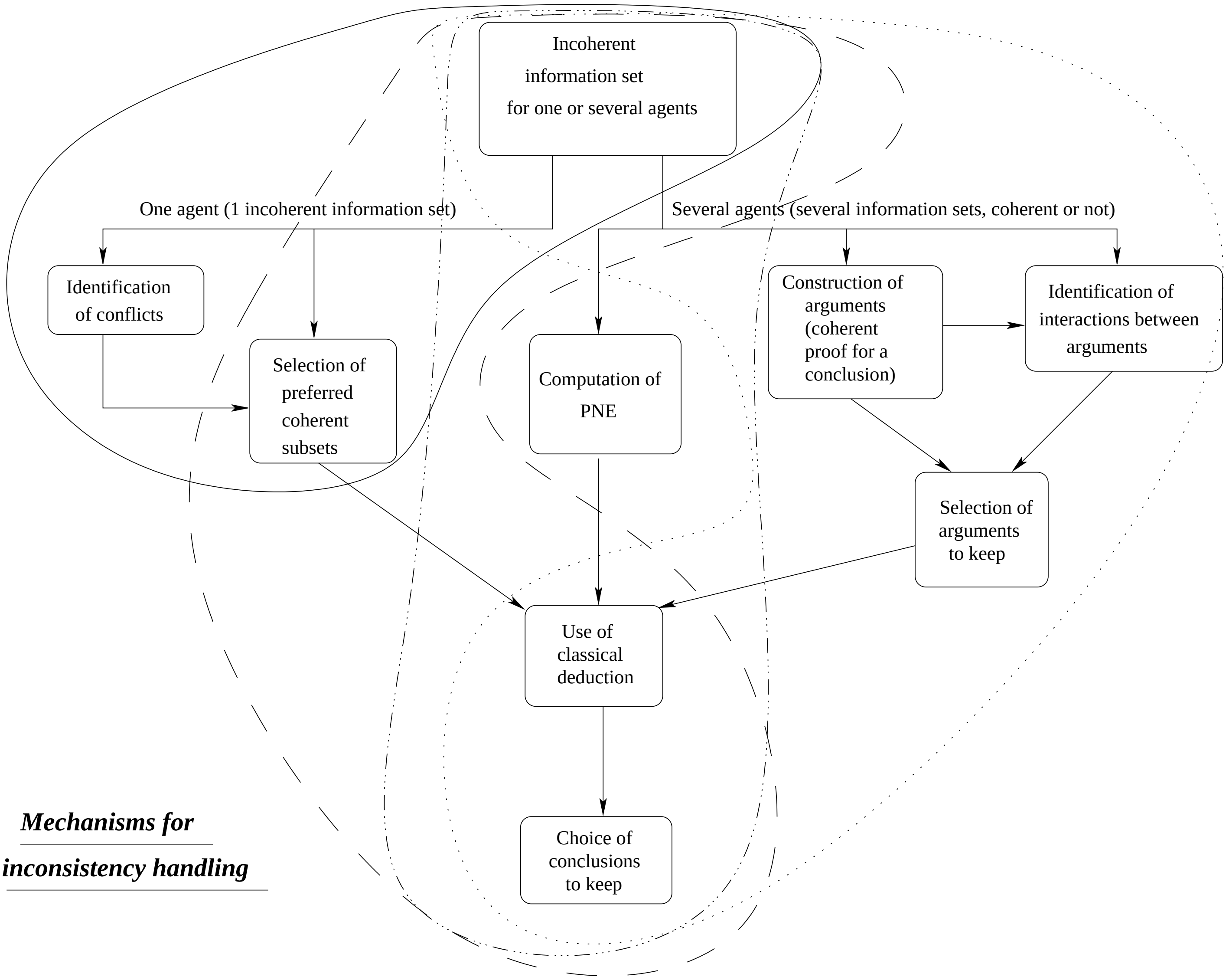
At this time, my study of the link between argumentation and Boolean games is preliminary but looks promising.

On this subject, from 2004 to 2009, I worked with Jérôme LANG, CNRS senior researcher at Paris Dauphine and I co-supervised with him Elise BONZON's PHD thesis (see [Bon07]) and a student of research master degree (Denis SIREYJOL in 2004). Moreover, during Elise's PHD Thesis, we have also collaborated with Bruno ZANUTTINI (GREYC Laboratory at Caens).

All these works have been published in technical reports [BLSL06b, BLSL07c], and in national and international publications [BLSL05, BLSL06a, BLSLZ06, BLSL07b, BLSL07d, BLSL07a, BLSL08b, BLSL08a, BLSLZ09] (one is an article in a book, see [BLSL08b] and another one is a publication in an international journal, see [BLSLZ09]).

All these topics are represented on the following figure.

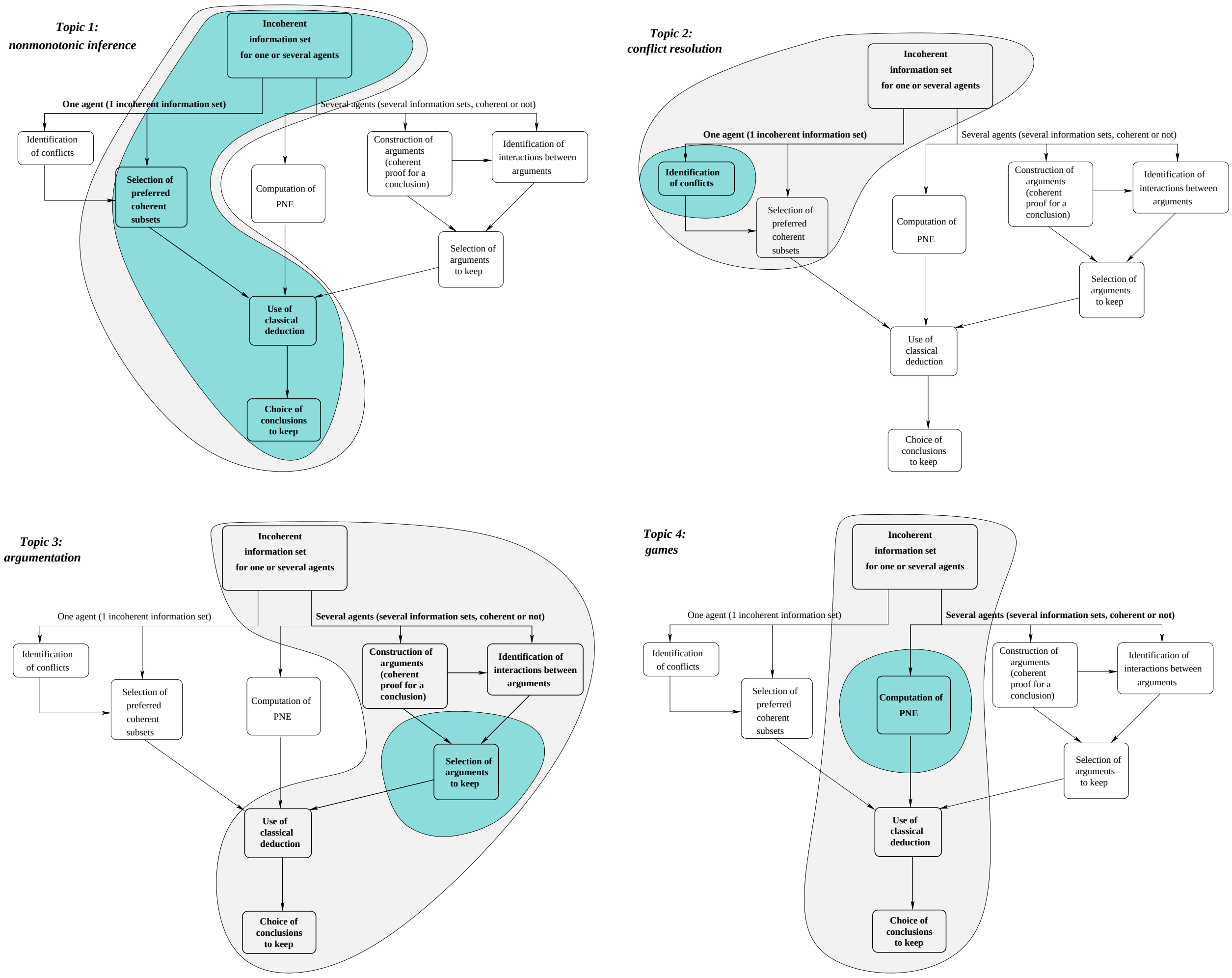
- Surrounded by a dashed line, Topic 1 "nonmonotonic inference" contains the initial information set, the selection of preferred subbases, the use of classical deduction then the choice of the final conclusions.
- Surrounded by a plain line, Topic 2 "conflict resolution" identifies conflicts in the initial information set before selection of preferred subbases (junction with Topic 1 "nonmonotonic inference").
- Surrounded by a dotted line, and using the initial information set, Topic 3 "argumentation" produces arguments and their interactions, then selects acceptable arguments before classical deduction (junction with Topics 1 and 2).
- And ultimately, surrounded by a line with dashes and dots, Topic 4 "games" computes different solution concepts (Pure Nash Equilibrium – PNE –) before doing classical deduction (junction with all other topics).



In this figure, the “entry point” is clearly the inconsistent information set which can be built either for one agent, or for several agents (in this case, it is the **union** of each agent’s information set). The word **union** is essential: in all my works, I have studied neither the merging nor the revision of the initial information *in order to obtain an initial consistent information set*. I have always used the method consisting *in extracting consistent information subsets from the initial inconsistent information set* and *to reason with these consistent subsets*.

The “exit point” is also very clear: this is the classically deduced information by the system independently of the intermediate method used and of the type of this deduced information (new formula, decision, . . .).

On the following figures, each topic is represented by a clear surrounded part and the subparts of these topics corresponding to my works are represented by a dark surrounded part. For Topic 1, I have worked on each subpart; for Topic 2, I have worked only on the subpart “identification of conflicts”; for Topic 3, my main work has concerned the selection of acceptable arguments; and for Topic 4, I have only studied the computation of solution concepts:



Organisation of the document

Chapter 1 is dedicated to nonmonotonic inference relations (Topic 1), then methods for conflict resolution (Topic 2) are presented in Chapter 2. Chapter 3 which is the most important with regard to size (because of the many years I have worked on this subject) focuses on argumentation (Topic 3) and Chapter 4 is dedicated to games (Topic 4).

At the beginning of each chapter, the reader will find an introduction describing the topic and its links with the other topics.

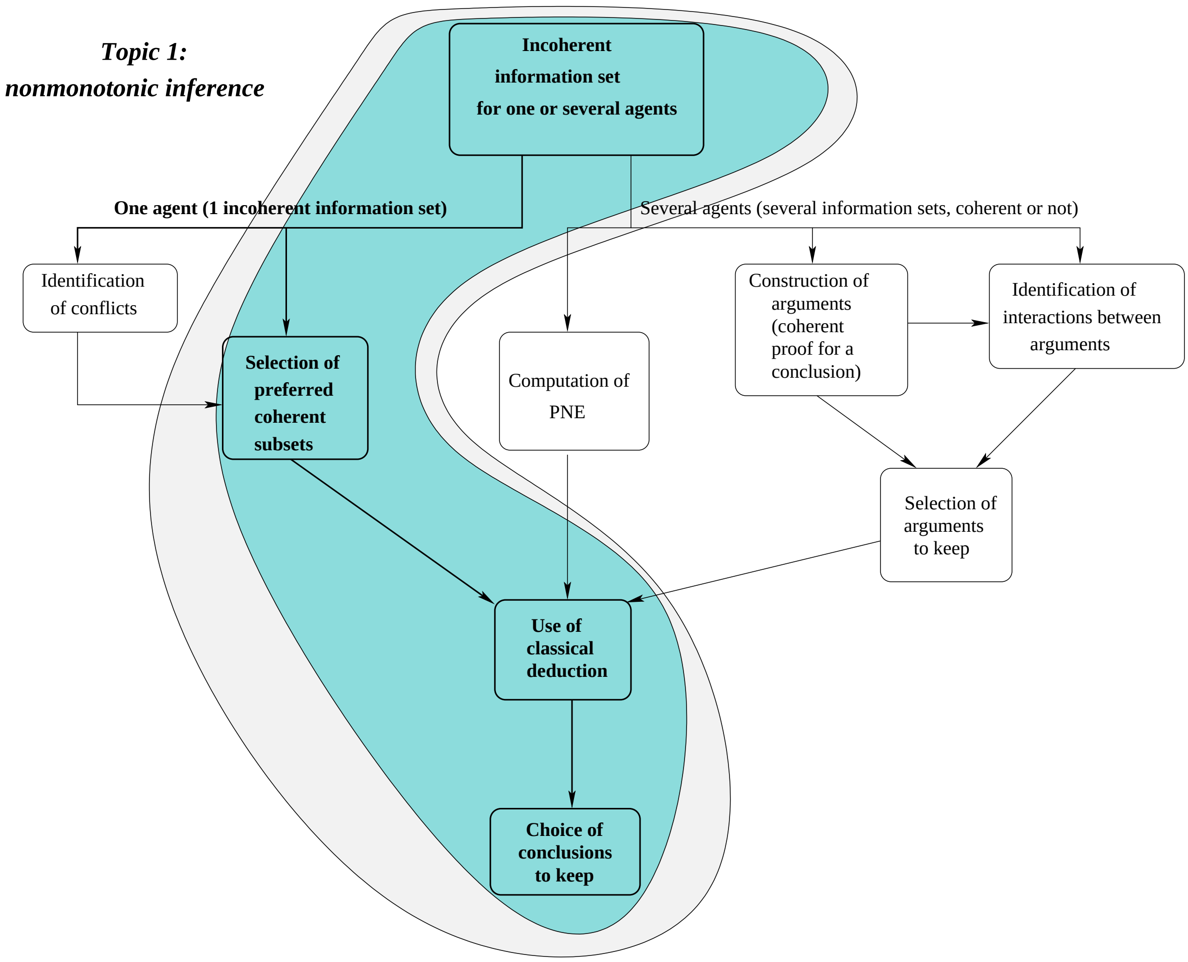
Chapter 5 describes my research project.

Finally, in Appendix A, some overall information is given about my research activity.

Chapter 1

Topic 1: definition and study of nonmonotonic inference relations

The topic presented in this chapter is represented on the following figure by the clear surrounded part and my works on this topic correspond to the dark surrounded part:



The main idea is: from an *inconsistent* information set one extracts some *consistent* information subsets representing possible states of the world and on which the classical deduction applies; nonmonotonic inference relations are defined using the results of classical deduction on the extracted subsets. During my PHD Thesis, I studied the theoretical and algorithmic properties of these relations. Then, after my PHD Thesis, I continued this study in order to generalize these relations.

1.1 Main results of the PHD Thesis

In this section, the reader can find the main definitions and notations about nonmonotonic inference. See the English papers [CLS94b, CLS95, CLSS96a, CLSS98] for more details.

1.1.1 Formal definition of a nonmonotonic inference

The process of consistency restoration can be represented as follows:

$(E, \leq)$	Inconsistent stratified belief base
$\downarrow$	*Step 1: selection mechanism m*
$\{Y_i\}$	Set of m-preferred subbases
$\downarrow$	*Step 2: entailment principle p*
$(E, \leq) \mid\!\sim^{p,m} \Phi$	Conclusion about $(E, \leq)$ and Φ

with:

Selection mechanisms.

The most famous methods use the $\subseteq$-maximal consistent subbases of $(E, \leq)$. The differences between all these methods are given by the use of the relation $\leq$ on E (see [CLS94b] for a synthesis of these works). In this document, I just present three methods: "Best-Out" preference induced by possibilistic logic, a preference relation combining $\leq$ and the $\subseteq$-maximality, and a preference relation combining $\leq$ and the maximality for cardinality.

Definition 1 *Let $X = (X_1 \cup \ldots \cup X_q)$ and $Y = (Y_1 \cup \ldots \cup Y_q)$ be two consistent subbases of $(E, \leq)$ (with $X_i = (X \cap E_i)$ and $Y_i = (Y \cap E_i)$).*

Best-Out preference *(see [BCD+93]): for a consistent subbase X of E, $a(X) = \min\{i \mid \exists \Phi \in E_i \setminus X\}$. Best-Out preference is the complete preordering defined by: $X \ll^{bo} Y$ iff $a(X) \leq a(Y)$.*

Set-inclusion-based preference *(see [CRS93] and [Gef92]): $X \ll^{incl} Y$ (denoting Y is incl-preferred to X) iff $\exists i$ such that $X_i \subset Y_i$ and $\forall j \mid 1 \leq j < i$, $X_j = Y_j$.*

Cardinality-based preference *(see [Leh95, BCD+93]) : $X \ll^{card} Y$ (denoting Y is card-preferred to X) iff $\exists i$ such that $|X_i| < |Y_i|$ and $\forall j \mid 1 \leq j < i$, $|X_j| = |Y_j|$ ($|Y|$ = cardinality of Y).*

Cardinality-based preference refines set-inclusion-based preference. The converse is false.

T (resp. INCL, CARD, BO) denotes the selection mechanism producing the set of maximal consistent subbases (resp. incl-preferred, card-preferred, bo-preferred subbases) of $(E, \leq)$.

Entailment principles.

The most known entailment principles are the skeptical one and the credulous one (see [PL92] for a more complete taxonomy). In this document, I will use three instances of these principles:

Definition 2 *Let $m(E,\leq)$ be a set of preferred consistent subbases of $(E,\leq)$ (for instance, $m(E,\leq)$ can be obtained with one of the mechanisms* T, INCL, CARD *or* BO*). Let Φ be a propositional formula.*

UNI principle*: Φ is inferred from $m(E,\leq)$ using the skeptical (or universal) principle iff Φ is classically inferred by each element of $m(E,\leq)$.*

EXI principle*: Φ is inferred from $m(E,\leq)$ using the credulous (or existential) principle iff Φ is classically inferred by at least one element of $m(E,\leq)$.*

ARG principle*: Φ is inferred from $m(E,\leq)$ using the argumentative principle iff Φ is classically inferred by at least one element of $m(E,\leq)$ and no element of $m(E,\leq)$ classically infers $\neg\Phi$.*

Nonmonotonic inference relation.

A nonmonotonic inference relation is defined using a selection mechanism m and an entailment principle p:

Definition 3 *Let $(E,\leq)$ be a stratified belief base and Φ be a propositional formula. $(E,\leq) \mid\!\sim^{p,m} \Phi$ iff Φ is inferred from $m(E,\leq)$ using the entailment principle p, $m(E,\leq)$ denoting the set of the preferred consistent subbases of $(E,\leq)$ obtained by the mechanism m.*

For instance, $m \in$ {T, INCL, CARD, BO} and $p \in$ {UNI, EXI, ARG}. So a nonmonotonic inference relation will be denoted with p-m.

An example

Consider the following belief base $(E,\leq)$ with 3 strata and representing the famous penguins problem (p means "penguin", o means "bird", v means "fly", a means "have wings") :

$$\begin{array}{c} p \\ p \rightarrow o \\ \hline p \rightarrow \neg v \\ o \rightarrow v \\ \hline o \rightarrow a \\ a \rightarrow v \end{array}$$

(formulae p and $p \rightarrow o$ have a more important priority than formulae $p \rightarrow \neg v$ and $o \rightarrow v$ which also have a more important priority than the formulae $o \rightarrow a$ and $a \rightarrow v$).

Among all the consistent subbases (58), five ones are interesting:

- $Y_1 = \{p, p \rightarrow o, p \rightarrow \neg v, o \rightarrow a\}$ which entails p, o, $\neg v$, a,
- $Y_2 = \{p, p \rightarrow o, p \rightarrow \neg v, a \rightarrow v\}$ which entails p, o, $\neg v$, $\neg a$,
- $Y_3 = \{p, p \rightarrow o, o \rightarrow v, o \rightarrow a, a \rightarrow v\}$ which entails p, o, v, a,
- $Y_4 = \{p, p \rightarrow \neg v, o \rightarrow v, o \rightarrow a, a \rightarrow v\}$ which entails p, $\neg o$, $\neg v$, $\neg a$,
- $Y_5 = \{p \rightarrow o, p \rightarrow \neg v, o \rightarrow v, o \rightarrow a, a \rightarrow v\}$ which entails $\neg p$.

(They are the T-preferred subbases, the 3 first ones being also the incl-preferred subbases and the third one being card-preferred).

Note that the bo-preferred subbases are not given here, but all of them contain the two formulae p and $p \rightarrow o$ and entail p and o. The nonmonotonic inference relations concerning literals are the following ones:

- with UNI-T, nothing is inferred,

- with UNI-BO and UNI-INCL, p and o are inferred,
- with EXI-T, all the literals are inferred,
- with EXI-BO and EXI-INCL, v, $\neg v$, a, $\neg a$, p and o are inferred,
- with ARG-T, nothing is inferred,
- with ARG-BO and ARG-INCL, p and o are inferred,
- with EXI-CARD, ARG-CARD and UNI-CARD, p, o, v and a are inferred.

1.1.2 Comparison of the nonmonotonic inference relations p-m

This analysis uses three theoretical comparison criteria and one practical comparison criterion:

Temporal complexity This is a very important criterion from a computational point of view; what is the computational cost of answering to the following question "does the base E infer the formula Φ?" (see [GJ79, Neb91, Got92]);

Cautiousness This is a subjective criterion depending on the context; nevertheless it is important because it shows the realism of a relation from the point of view of the number of inferred conclusions (see [PL92]);

Deductive properties This is also an important criterion which corresponds to an axiomatization of nonmonotonic inference (see [KLM90, GM94]);

Algorithms From the practical point of view, it is an essential criterion.

Knowing that this part of my work is not the subject of this document, I do not give here all obtained results; the reader interested can be found them in my publications:

- Complexity results are given in [CLS94b, CLSS96a, CLSS98].
- Results about cautiousness and satisfied postulates of deduction are given in [CLS95].
- Conclusions about the algorithmic study are given in [CLSS96a, CLSS98].

1.2 An additional study: generalization of nonmonotonic inference

This study has been realized after my PHD Thesis and describes a more general framework for defining nonmonotonic inferences in which one can retrieve the already existing nonmonotonic inference relations used in my PHD Thesis. See [CLS97, CLS00] for more details (only in french, sorry!).

1.2.1 Introduction

The example of the penguin given in Section 1.1.1 on page 10 highlights some problems related to the classical definition of a nonmonotonic inference relation:

- first, *how can we define a "good" selection mechanism, neither too selective, nor not selective enough*:
  - if too many subbases are selected, then a computational problem appears;
  - if not enough subbases are selected, the risk is to "forget" important information (in the example of the penguin, if Y_3 is the only selected subbase, we forget that penguin cannot fly).

- moreover, *how can we define a "good" entailment principle neither too cautious, nor too permissive*:
  - with the universal principle, very few conclusions are obtained (in the example of the penguin, with the selected subbases Y_1, Y_2 and Y_3, one cannot conclude about v and a; it is the same case with the argumentative principle);
  - with the credulous principle, one obtains too many conclusions and in particular some formulae and their negation (in the example of the penguin, with the selected subbases Y_1, Y_2 and Y_3, one concludes v and $\neg v$, a and $\neg a$) which is unsound.

In order to solve these problems, we have proposed a new approach with the following ideas:

- the inference of a conclusion Φ is viewed as a decision problem ("is Φ inferred by the system?") and we want the used process to be sound (if Φ is inferred then $\neg\Phi$ is not inferred);
- generally, selected subbases can be viewed as conflicting "opinions";
- so the selection of the "preferred subbases" is the definition of a set of "opinions";
- these "opinions" can have a weight; so we need a *weighting mechanism* which takes into account the fact that some "opinions" are more significant than other ones;
- the entailment principle corresponds to a *vote mechanism* defined using the meaning of the "opinions"; thus, we can have a compensation process between "opinions for" and "opinions against" which takes into account the weight of each opinion, and we are able to conclude about a formula Φ even there exist "opinions against" Φ (so "for" $\neg\Phi$);
- and finally, the conclusion is deduced from vote counts (*counting mechanism*).

1.2.2 A general process for handling inconsistency

With these ideas, a general process for handling inconsistency can be defined as follows:

$(E, \leq)$	Inconsistent stratified belief base
$\downarrow$	*Step 1: selection mechanism*
$\{Y_i\}$	Set of "opinions" (consistent subbases)
$\downarrow$	*Step 2: weighting mechanism*
$\{(Y_i, p(Y_i))\}$	Set of weighted "opinions"
$\downarrow$	*Step 3: voting mechanism*
$n(\Phi)$	Vote of "opinions" for a formula Φ
$n(\neg\Phi)$	Vote of "opinions" against a formula Φ
$n(\text{neither } \Phi, \text{ nor} \neg\Phi)$	Vote of indifferent "opinions" to Φ
$\downarrow$	*Step 4: counting mechanism*
$(E, \leq) \mid\!\sim \Phi$	Conclusion about $(E, \leq)$ and Φ

Note that, even if this general process was original (in 2000), different subparts of this process had already been studied (each of these works has been done with a specific point of view[1]):

- several selection mechanisms have been proposed in [Bre89, CRS93, DLP91, BCD$^+$93];

[1] [PL92]: Definition of a taxonomy for nonmonotonic inference relations; [Sme93]: belief revision; [Lan94]: diagnosis; [Bre89, CRS93, DLP91, BCD$^+$93]: nonmonotonic reasoning and/or default logic.

- Step 2 has been studied in possibilistic and probabilistic contexts (see [DLP91, Lan94]);
- Step 3 has been defined in [Lan94] with belief functions (see [Sme93]) ;
- Step 4 has been described in [PL92, Lan94].

1.2.3 Application to nonmonotonic reasoning

1.2.3.1 To retrieve existing nonmonotonic inference relations

The p-m nonmonotonic inference relations presented in Section 1.1 on page 10 can be retrieved in our general process:

- Step 1: we choose the selection mechanism m (with $m \in$ {T, INCL, CARD, BO});
- Step 2: we consider that each subbase has the same weight equal to 1;
- Step 3:
  - $n(\Phi) =$ sum of the weights of preferred subbases which classically entail the formula Φ;
  - $n(\neg\Phi) =$ sum of the weights of preferred subbases which classically entail the formula $\neg\Phi$;
  - $n(\text{neither } \Phi,\ \text{nor}\neg\Phi) =$ sum of the weights of preferred subbases which classically entail neither the formula Φ, nor the formula $\neg\Phi$;
- Step 4:
  - entailment principle UNI can be defined with: $n(\Phi) > 0$, $n(\neg\Phi) = 0$ and $n(\text{neither } \Phi,\ \text{nor}\neg\Phi) = 0$;
  - entailment principle EXI can be defined with: $n(\Phi) > 0$;
  - entailment principle ARG can be defined with: $n(\Phi) > 0$ and $n(\neg\Phi) = 0$.

1.2.3.2 Some "new" nonmonotonic inference relations

Among all the possible relations which can be defined with our general process, I just present two examples in this document:

- The first example is a variant of the p-INCL relations (with $p \in$ {UNI, EXI, ARG}) corresponding to a change in the counting mechanism; the GRAD-EQUADD-INCL relation is defined by:
  - Step 1: selection mechanism INCL;
  - Steps 2 and 3: they are the same steps as those presented in Section 1.2.3.1;
  - Step 4: $n(\Phi) > n(\neg\Phi)$ and $n(\Phi) \geq n(\text{neither } \Phi,\ \text{nor } \neg\Phi)$.

  So this relation uses a democratic vote.
- The second example is more complex and uses probabilistic reasoning; the GRAD-POSMAX-T relation is defined by:
  - Step 1: selection mechanism T;
  - for Step 2, we use the following definition:

    Definition 4 *(see [DLP91]) Let Y_i be a subbase of $(E, \leq)$, the* possibilistic weight *of Y_i denoted by $p_{po}(Y_i)$ is defined by: if $Y_i \neq E$ then $p_{po}(Y_i) = x$ (x = number of the stratum having the greatest priority and in which a formula has been removed for restoring consistency); otherwise $p_{po}(E) = q + 1$ (q = number max of strata in $(E, \leq)$).*

– Step 3:
  - $n(\Phi) =$ max of the weights of preferred subbases which classically entail Φ ;
  - $n(\neg\Phi) =$ max of the weights of preferred subbases which classically entail $\neg\Phi$;
  - $n(\text{neither } \Phi, \text{ nor} \neg\Phi) =$ max of the weights of preferred subbases which classically entail neither Φ, nor $\neg\Phi$;
– Step 4: $n(\Phi) > n(\neg\Phi)$ and $n(\Phi) \geq n(\text{neither } \Phi, \text{ nor}\neg\Phi)$.

1.2.3.3 An example

Consider the following belief base partitioned in 4 strata and representing a new version of the penguin problem (p means "penguin", o means "bird", v means "fly", a means "have wings", pl means "have feathers") :

$$\begin{array}{c} p \\ p \to o \\ \hline p \to \neg v \\ \hline o \to a \\ a \to v \\ \hline a \to pl \end{array}$$

Note that the validity of the results depends on the intuition we have. For instance, in this example, it seems intuitively correct to conclude $\neg v$ and pl.

The different nonmonotonic inference relations give the following results concerning these two literals:

- neither $\neg v$, nor pl with UNI-T,
- only $\neg v$ with UNI-BO, UNI-INCL and UNI-CARD,
- v, $\neg v$, pl with EXI-T,
- only pl with ARG-T,
- $\neg v$, pl with p-m ($p \in$ {EXI, ARG} and $m \in$ {BO, INCL, CARD}), and with GRAD-EQUADD-INCL and GRAD-POSMAX-T.

In [CLS97], many other examples are given which show that the "new" relations GRAD-EQUADD-INCL and GRAD-POSMAX-T generally give good intuitive results.

1.2.3.4 Theoretical study

In this section, some results concerning the cautiousness[2] and the satisfaction of rationality postulates are given for these new relations (all the proofs are given in [CLS97]).

So, the cautiousness of these new relations is the same as the cautiousness of the relations ARG-m (with $m \in$ {T, INCL}).

About rationality postulates[3], the obtained results are:

[2]Consider 2 inference relations R_1 and R_2, R_1 is *more cautious than* R_2 iff each conclusion obtained with R_1 is also obtained with R_2.

[3]The postulates used in this section are:
reflexivity: $\alpha \mid\!\sim \alpha$;
left logical equivalence: $\dfrac{\models \alpha \leftrightarrow \beta \,;\, \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma}$;

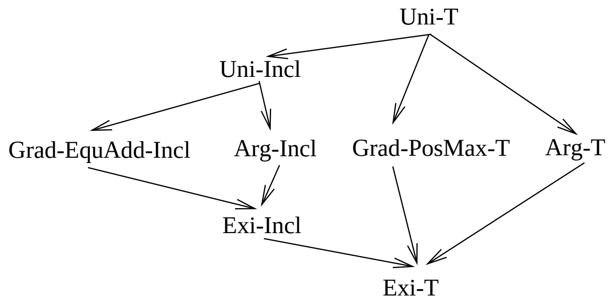


Figure 1.1: Comparison from the cautiousness point of view (the arrow means “more cautious than”)

- Grad-EquAdd-Incl satisfies only the properties of reflexivity, left logical equivalence, supra-classicity and consistency preservation;
- Grad-PosMax-T satisfies only the properties of reflexivity, left logical equivalence, supra-classicity, consistency preservation, right weakening and weak conditionalisation.

These results confirm the fact that the new relations have the same behaviour as the relations Arg-m.

1.2.4 Conclusion about the additional study

In order to handle inconsistency in nonmonotonic reasoning, the general process described in Section 1.2.2 on page 13 allows us:

- to capture the existing nonmonotonic inference relations described in Section 1.1 on page 10 and in [CLS95] ;
- to define “new” nonmonotonic inference relations (see for instance, the two relations described in Section 1.2.3.2 on page 14) which seem very promising from the point of view of the intuitiveness;
- to compare all these relations (old and new) in a same framework (see for instance, a comparison of cautiousness and of the respect of rationality postulates in Section 1.2.3.4 on the preceding page).

It is not the only application domain; for instance, the process could be applied to the merging of belief bases: each belief base can be considered as an “opinion” and the weights of these “opinions” could represent a hierarchy between sources.

1.3 Conclusion on Topic 1

In this chapter, I very briefly present some works I realized on Topic 1 “nonmonotonic inference” in the strict framework of consistency restoration during 8 years including my PHD Thesis.

The obtained results can be synthesised as follows:

right weakening: $\dfrac{\models \alpha \rightarrow \beta\,;\, \gamma \mathrel{|\!\sim} \alpha}{\gamma \mathrel{|\!\sim} \beta}$;

supra-classicity: $\dfrac{\alpha \vdash \beta}{\alpha \mathrel{|\!\sim} \beta}$;

weak conditionalisation: $\dfrac{\alpha \mathrel{|\!\sim} \beta}{\mathrel{|\!\sim} \alpha \rightarrow \beta}$;

consistency preservation: $\dfrac{\alpha \mathrel{|\!\sim} \bot}{\alpha \vdash \bot}$ (with $\bot$ representing the contradiction).

- during my PHD Thesis: study of theoretical (computational complexity, cautiousness, satisfaction of rationality postulates) and algorithmic properties of some nonmonotonic inference relations built using the consistency restoration method,
- after my PHD Thesis: definition of a general framework enabling the description of nonmonotonic inference relations (old ones and some new ones) and study of the new ones.

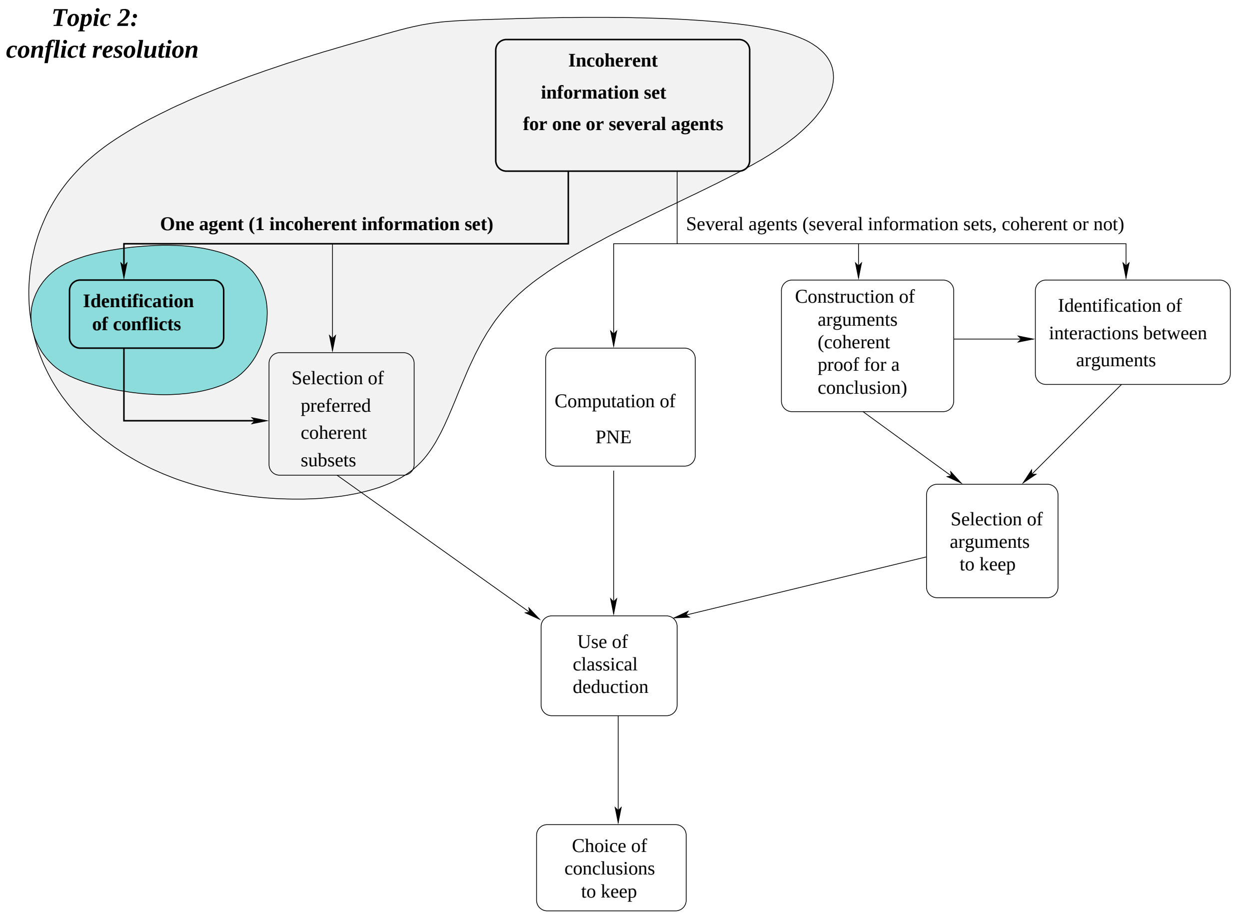
In this context, we have exhibited a relation which seems very interesting from the point of view of the properties it satisfies; we have also identified some other relations, obtained by the general process, which give very intuitive conclusions, even if these relations are less interesting from the point of view of properties they satisfy.

Another conclusion of this work is the fact that the mechanisms used for comparing all these relations are very different from each other and it is difficult to use them together because their conclusions are very often conflicting. This remark will be also an important point of my research project (see Chapter 5 on page 151): how to define "good" comparison criteria? It is also the source of some questions among which the following one: "how can we choose one selection mechanism rather than another one?" The following chapter and topic try to answer to this question.

Chapter 2

Topic 2: conflict resolution

On the following figure, Topic 2 is represented by the “clear surrounded” part and my works on this subject are represented by the “dark surrounded” part:



In this topic, I focus on the study of the selection mechanisms used for selecting consistent information subsets. There exist two dual approaches:

- Either we directly select some consistent subbases of the inconsistent initial base[1]. This is the approach used in

[1]It is possible without computing all the conflicts.

Chapter 1 on page 9.

- Or we identify all the conflicts of the initial base (which corresponds to an "ATMS" approach with no-goods computation – see [Kle86]) then we choose the beliefs we must remove in order to solve these conflicts and to get some consistent subbases.

When there is no preference between beliefs, these two approaches give the same results. Otherwise:

- with the first approach, we give a *global* semantics to belief preferences (how is a formula located with regard to the other formulae *of the base*?).
- whereas, with the second approach, preferences are used *locally* (how is a formula located with regard to the other formulae *of a given conflict*?).

The local approach is essential when preferences are partial (in this case, incompatibility can appear: for instance, belief a is preferred to belief b in the conflict c_1 whereas in the conflict c_2 this is the opposite case). So the local approach is more general than the global one. Note also that the local approach is more adapted to context-dependent preferences or conditional preferences (see [Bre94, PS97, BG98, Gar98]).

This chapter concerns the local approach but with a specific point of view: I study this problem in order to try to identify some principles and to characterize solutions which satisfy these principles. This work has been done in collaboration with Claudette CAYROL and Jérôme MENGIN. All the proofs and many additional details can be found in [CLSM00] (only in french, sorry!).

2.1 Methodology

2.1.1 Basic definitions and principles

Let L be a logical language. Let E be a set of formulae of L representing beliefs (E is the belief base), and K be a consistent set of formulae of L representing certain knowledge (the union $E \cup K$ may be inconsistent).

The consistency restoration problem can be described by:

> *Restoring the K-consistency* consists in finding a K-consistent subset of E (generally, we use $\subseteq$-maximal solutions – see [Bre89, CRS93, DLP91, BCD$^+$93, LS95a] and Chapter 1 on page 9).

The conflict resolution problem can be described by:

> E is assumed to be K-inconsistent and it is given as a set of *K-conflicts* (*i.e.* subsets of E $\subseteq$-minimal among those which are K-inconsistent). This set of K-conflicts is denoted by $\mathcal{C}(E) = \{c_i\}$ (each c_i being a K-conflict); $\mathcal{N}(E) = \bigcup_i c_i$ denotes the core of E and $\mathcal{L}(E)$ denotes the set of formulae of E which do not take part in any conflict (*free* formulae of E); we have $E = \mathcal{N}(E) \cup \mathcal{L}(E)$[2].
>
> *Solving the K-conflicts* consists in finding a subset of E which does not contain any of the K-conflicts. So for each conflict we must find at least one formula to remove and it is sufficient to work on $\mathcal{N}$ since the formulae of $\mathcal{L}$ are not concerned by K-inconsistency.

A solution of a conflict resolution problem is defined by:

Definition 5 *Let $\mathcal{C} = \{c_i\}$ be a conflict resolution problem. S is a* solution[3] *of $\mathcal{C}$ iff S is a subset of $\mathcal{N}$ such that $c_i \cap S \neq \varnothing, \forall i$.*

[2]To simplify, one will use the notation $\mathcal{C}$, $\mathcal{N}$ and $\mathcal{L}$ in place of $\mathcal{C}(E)$, $\mathcal{N}(E)$ and $\mathcal{L}(E)$.

[3]This corresponds to the notion of *hitting-set*" (see [Rei87]).

Example 1 *Let $\mathcal{C}$ be the following set of conflicts: $\mathcal{C} = \{\{a, b\}, \{b, c\}\}$ (a, b, c being 3 propositional formulae).*
On this example, 5 solutions exist: $S_1 = \{a, b, c\}$, $S_2 = \{a, b\}$, $S_3 = \{b, c\}$, $S_4 = \{a, c\}$, $S_5 = \{b\}$.
Note that S_4 and S_5 seem more interesting than the other ones because only one formula is removed for each conflict.

The *minimality principle* corresponds to the removal of as few formulae as possible (from the inclusion point of view):

Definition 6 *Let $\mathcal{C} = \{c_i\}$ be a conflict resolution problem. S is a* minimal solution *of $\mathcal{C}$ iff S is a subset of $\mathcal{N}$ such that:*

- $c_i \cap S \neq \varnothing, \forall i$,
- $\forall x \in S$, $\exists i$ *such that* $c_i \cap S = \{x\}$.

Example 1 (cont'd) *On this example, only S_4 and S_5 are minimal solutions.*

Definition 6 gives a more interesting result than the result given by Definition 5 on the facing page. However, one does not take into account the fact that, in a conflict, one can prefer to remove one formula rather than another one.

2.1.2 Use of preferences

We consider that each K-conflict of E (denoted by c_i) is equipped with a complete preordering and $\forall i$, $f(c_i)$ denotes the set of the minimal elements of c_i with regard to this preordering. Notation: $\mathcal{N}_{min} = \bigcup_i f(c_i)$.
We can now explicit some principles in order to obtain "good" solutions:

Minimality principle: solutions must be $\subseteq$-minimal (as in Definition 6);

Parsimony principle: to remove the smallest number of beliefs (minimality in the sense of cardinality) ;

Preference principle: removed belief are chosen as much as possible in the $f(c_i)$ (the idea is that a formula is removed for a "good" reason) ;

Efficiency principle: to select the smallest number of solutions (in order to facilitate the future use of these solutions – if there are too many solutions, it will be difficult to exploit them).

Of course, these principles are not always compatible.

Example 1 (cont'd) *The minimal elements of each conflict are underlined: $\mathcal{C} = \{\{\underline{a}, b\}, \{b, \underline{c}\}\}$. S_4 and S_5 satisfy the minimality principle, S_4 satisfies the preference principle, S_5 satisfies the parsimony principle. And there is no solution which satisfies all the principles at the same time.*

Using minimality principle and preference principle leads to the following definition:

Definition 7 *Let $\mathcal{C} = \{c_i\}$ be a conflict resolution problem equipped with a complete preordering for each c_i. S is a* preferred minimal solution *iff S is a subset of $\mathcal{N}$ such that:*

- $c_i \cap S \neq \varnothing, \forall i$,
- $\forall x \in S$, $\exists i$ *such that* $c_i \cap S = \{x\}$,
- $\forall x \in S$, $\exists j$ *such that* $x \in f(c_j)$.

Proposed solutions	Satisfied principles: Minimality	Parsimony	Preference
Basic Sol. (Def. 5 on page 20)			
Minimal Sol. (Def. 6 on the previous page)	X		
Preferred Minimal Sol. (Def. 7 on the preceding page)	X		X
Parsimonious Sol. (Def. 8)	X	X	
Card-min Preferred Sol. (Def. 9)	X		X

Table 2.1: Principles and solutions for conflict resolution (X means that the proposed solution satisfies the given principle)

Example 1 on the previous page (cont'd) *S_4 is the only preferred minimal solution.*

Example 2 *Consider the following conflict resolution problem:*

$\mathcal{C} = \{\{a_6, \underline{a_1, a_2}\}, \{a_7, \underline{a_2, a_3}\}, \{a_3, \underline{a_4, a_8}\}, \{a_4, \underline{a_5}\}\}$.

The 5 preferred minimal solutions are:

$S_1 = \{a_2, a_4\}$, $S_2 = \{a_2, a_8, a_5\}$, $S_3 = \{a_1, a_3, a_5\}$, $S_4 = \{a_2, a_3, a_5\}$, $S_5 = \{a_1, a_3, a_4\}$.

Using the parsimony principle leads to the following definition:

Definition 8 *Let $\mathcal{C} = \{c_i\}$ be a conflict resolution problem. S is a* parsimonious solution *of $\mathcal{C}$ iff S is a subset of $\mathcal{N}$ minimal for cardinality which intersects each conflict.*

Example 1 on the preceding page (cont'd) *S_5 is the only parsimonious solution.*

Note that the parsimony principle refines the minimality principle, and so the parsimony principle is a way to satisfy the efficiency principle.

However, parsimony principle and preference principle are not compatible. But we can apply the cardinality-minimality to the efficiency principle:

Definition 9 *Let $\mathcal{C} = \{c_i\}$ be a conflict resolution problem equipped with a complete preordering for each c_i. S is a* card-min preferred solution *iff S is a subset of $\mathcal{N}_{min}$ minimal for the cardinality which intersects each conflict.*

Example 1 on the previous page (cont'd) *S_4 is the only card-min preferred solution.*

Example 2 (cont'd) *S_1 is the only card-min preferred solution.*

There also exist some other possibilities in order to respect the efficiency principle (see the "élitiste" preference of [CRS93]), but I do not present them in this document (see [CLSM00] for more details).

All the given definitions are presented in Table 2.1 with the corresponding principles.

2.2 Computation of preferred minimal solutions

The computation of preferred minimal solutions corresponds exactly to a famous problem in AI: the computation of intersecting sets for a set of sets[4].

There already exist many algorithms for this computation (for diagnosis [Rei87], for propositional logical deduction [Cas96, Cas97], for nonmonotonic logics [JK90, Lev92, Men95, Nie95], …).

See in [CLSM00] more details about this computation.

Note also that, in the framework of belief revision, an approach respecting the parsimony principle has been developed in [Pap92] in order to find the removed sets of beliefs. Then this approach has been applied in the applicative framework of Geographic Information Systems (GIS), see [WPJ00, BBPW04] using algorithms also based on the notion of hitting sets.

2.3 Link with direct selection of K-consistent subsets

This is the link between Topic 1 described in Chapter 1 and the current topic.

2.3.1 Duality between consistency problem and conflict problem

Let E be a K-inconsistent belief base.

- For a consistency problem given by (K, E), solutions are K-consistent subsets of E.
- For a conflict problem given by $(\mathcal{L}, \mathcal{C})$ (where $\mathcal{L}$ is the set of the free formulae of E and $\mathcal{C} = \{c_i\}$ is the set of the K-conflicts built from E), solutions are the subsets of E which do not contain any conflict.

The duality between the consistency problem and the conflict problem is given by the following proposition:

Proposition 1 *Let E be a belief base and K be a certain knowledge base.*

- *The subset F of E is a solution of the consistency problem iff $(\mathcal{N} \setminus F)$ is a solution (in the sense of Def. 5 on page 20) of the conflict problem.*
- *The subset S of $\mathcal{N}$ is a minimal solution (in the sense of Def. 6 on page 21) of the conflict problem iff $(\mathcal{N} \setminus S)$ is a $\subseteq$-maximal subset among the K-consistent subsets of $\mathcal{N}$.*
- *The maximal solutions of the consistency problem are under the form of $\mathcal{L} \cup (\mathcal{N} \setminus S)$ with S a minimal solution (in the sense of Def. 6 on page 21) of the conflict problem.*

2.3.2 With preferences

Links between the consistency problem and the conflict problem still exist when if there are preferences. In this case, we must use a special preference relation between subbases, the democratic preference:

Notation 1 *Let $\leq$ be a partial preordering on E. $<$ denotes the strict ordering associated with $\leq$ and defined by $x < y$ iff $(x \leq y) \wedge \neg(y \leq x)$. In this case, the consistency problem is denoted by $(K, E, <)$.*

[4] An *intersecting set* (also called *hitting set*) for a set of sets $\mathcal{C}$ is a part of the union of the elements of $\mathcal{C}$ whose intersection with each element of $\mathcal{C}$ is non empty.

Definition 10 *Let F_1 and F_2 be two K-consistent subsets of E. F_2 is demo-preferred to F_1 iff for all $x \in F_1 \setminus F_2$, there exists $y \in F_2 \setminus F_1$ such that $x < y$.*

The solutions of the consistency problem $(K, E, <)$ which are maximal for the demo-preference relation are said demo-preferred.

Note that if $\leq$ is complete then the demo-preferred solutions of the consistency problem are exactly the subbases incl-preferred evoked in Chapter 1 on page 9 (see [Bre89, CRS93]).

If $\mathcal{N}_{min}$ is strictly included in $\mathcal{N}$, we can consider the stratification of E in (E_1, E_2, E_3) with $E_1 = \mathcal{L}$, $E_2 = \mathcal{N} \setminus \mathcal{N}_{min}$ and $E_3 = \mathcal{N}_{min}$. And the following property holds:

Proposition 2 *The subset S of $\mathcal{N}$ is a preferred minimal solution of the conflict problem iff $E \setminus S$ is a demo-preferred solution of the consistency problem (K, E_1, E_2, E_3).*

Example 2 on page 22 (cont'd) *Let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$ be a conflict problem with:*

$c_1 = \{a_6, \underline{a_1, a_2}\}$, $c_2 = \{a_7, \underline{a_2, a_3}\}$, $c_3 = \{a_3, \underline{a_4, a_8}\}$, $c_4 = \{a_4, \underline{a_5}\}$ *(for each c_i, underlined elements are minimal).*

So we have:

$E = \mathcal{N}$,

$\mathcal{N} \setminus \mathcal{N}_{min} = \{a_6, a_7\}$

and $\mathcal{N}_{min} = \{a_1, a_2, a_3, a_4, a_5, a_8\}$.

The demo-preferred solutions (which are also the incl-preferred subbases) of the consistency problem defined by $(K, \{a_6, a_7\}, \{a_1, a_2, a_3, a_4, a_5, a_8\})$ are:

$F_1 = \{a_6, a_7, a_1, a_3, a_5, a_8\}$,

$F_2 = \{a_6, a_7, a_3, a_1, a_4\}$,

$F_3 = \{a_6, a_7, a_2, a_4, a_8\}$,

$F_4 = \{a_6, a_7, a_1, a_4, a_8\}$

and $F_5 = \{a_6, a_7, a_2, a_5, a_8\}$.

These solutions correspond respectively to the preferred minimal solutions of the conflict problem:

$S_1 = \{a_2, a_4\}$,

$S_2 = \{a_2, a_8, a_5\}$,

$S_3 = \{a_1, a_3, a_5\}$,

$S_4 = \{a_2, a_3, a_5\}$

and $S_5 = \{a_1, a_3, a_4\}$.

In [CLSM00], some other results are given, in particular a link between the democratic preference and the "élitiste" preference.

2.4 Conclusion on Topic 2

The definition of solutions for a conflict resolution problem presents the same difficulty as the definition of solutions for the consistency restoration problem: taking into account two constraints whose satisfaction is intuitively necessary. The first constraint implies that the solutions are as small (resp. large) as possible. The other says that one can satisfy preferences on beliefs.

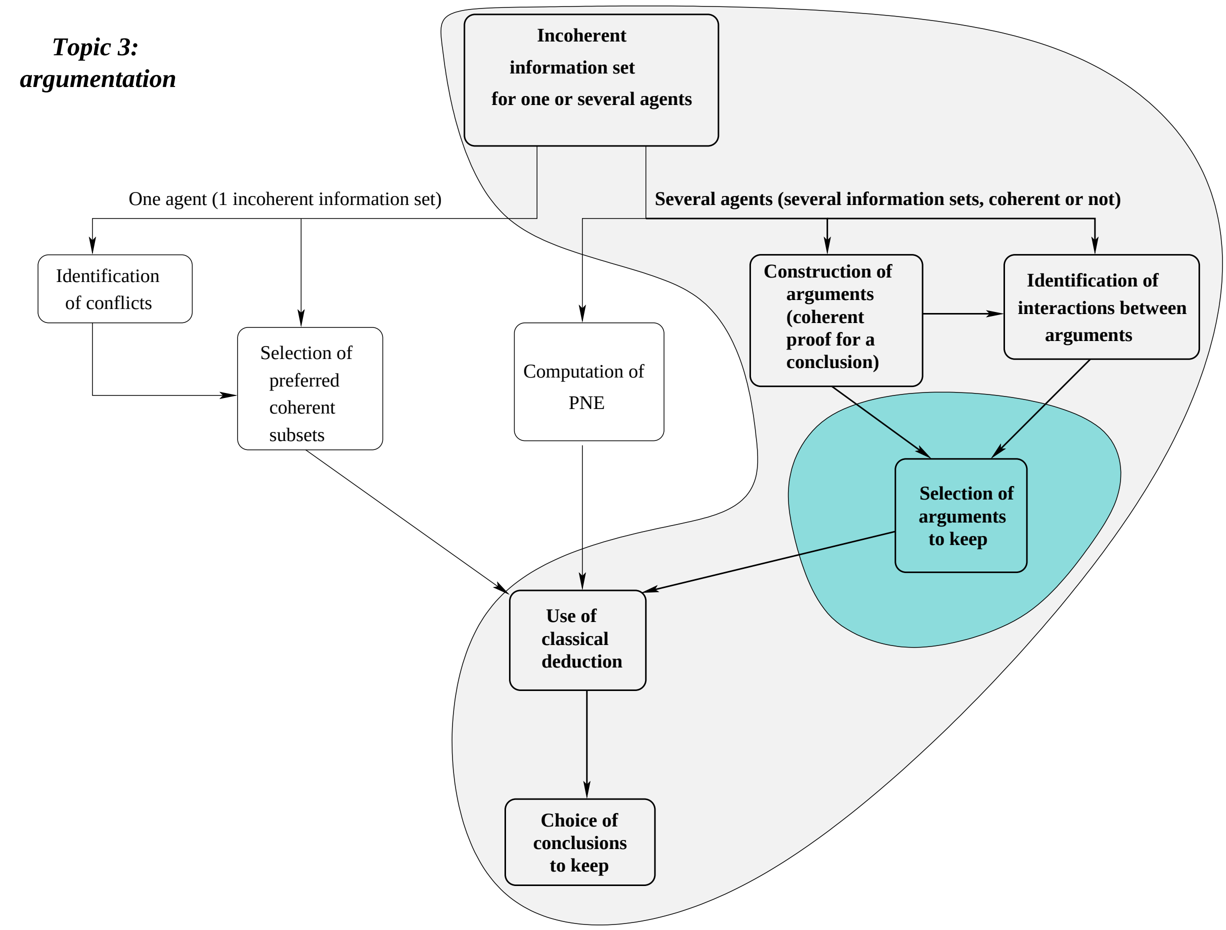
In order to study how these two constraints interact during the conflict resolution process, we have proposed several principles. These principles, alone or together, lead to different types of solutions for this problem and allow a comparison between solutions of this problem and solutions of consistency restoration problem corresponding to the direct selection of consistent subbases.

One definition seems interesting from the computational point of view: *preferred minimal solutions* which can be computed as the minimal intersecting sets of the set of minimal parts of conflicts (in the sense of preferences). See the algorithmic study in [CLSM00].

Chapter 3

Topic 3: argumentation

In the following figure, Topic 3 is represented by the “clear surrounded” part and my work on this subject is represented by the “dark surrounded” part:



This topic relates to both previous topics because this is again a method for handling inconsistency by the selection of consistent subsets (even if, in argumentation, these subsets do not represent possible states of the world).

Argumentation is not only a kind of enlarged restoration of consistency, because it is an important framework, as nonmonotonic reasoning, which covers many different domains: AI, philosophy and cognitive psychology. My works in argumentation only concerns AI domain.

In this chapter, I present argumentation as a complex mechanism in two main stages (these stages will be more precisely defined and partitioned in the following section):

- using as starting point an inconsistent information set, the *construction* of consistent subsets for or against some given conclusions

- and then the *analysis* of these subsets (the arguments) and their interactions in order to conclude.

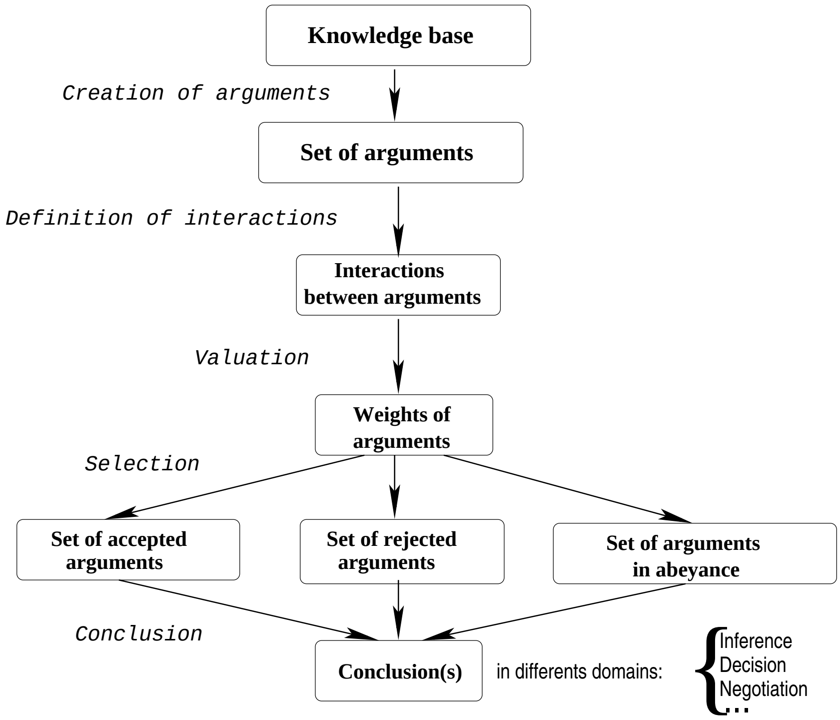
With this aim, I forget about the logical aspect of argumentation and instead take a particular assumption: arguments and interactions are given, so the starting element of my works in argumentation is a directed graph in which vertices are arguments and edges are binary interactions between arguments. This assumption is usual in this domain when one wants to work only on the *analysis* part of the argumentation process.

In Section 3.1, the general argumentation system is described. Then two different instances of this system are presented. The first one corresponds to "unipolar" argumentation (see Section 3.2 on page 30) and second one is about "bipolar" argumentation (see Section 3.3 on page 109); in the first case, interactions only represent conflicts between arguments whereas, in the second case, one takes into account conflicts and also another kind of interaction which represents a more "positive" relation between arguments: support.

For each argumentation systems instance, I have studied the notions of valuation of arguments (how to define a "weight" for an argument – see Sections 3.2.2 on page 35 and 3.3.4 on page 119) and the acceptability of arguments (how to select an argument – see Sections 3.2.3 on page 54 and 3.3.2 on page 113). In the case of an unipolar argumentation system, I have also worked on the notions of merging and revision of argumentation systems (see Sections 3.2.4 on page 62 and 3.2.5 on page 80), and I have used an unipolar argumentation system in order to simulate practical reasoning (see Section 3.2.6 on page 91).

3.1 General scheme

Argumentation can be viewed as a process composed of several steps. The following picture describes this process which is a refinement of the previous figure:



1. Definition of *arguments*: the notion of argument generally relates to the notions of explanation, proof, justification. The aim of an argument is to "support" beliefs, or to criticize another agent in order to obtain a particular behaviour. Arguments can have many different forms: informal text in natural language as a piece of dialogue, formal proof in a formalized logic language.

2. Definition of the different *interactions between arguments*: arguments built from a knowledge base cannot been considered as independent elements. Many kinds of interactions between these arguments can appear. Generally, one considers two main types of interactions: arguments are either in conflict (interaction representing a conflict) or they help other arguments (interaction representing a support). If there only exist conflicts then one talks about *unipolar argumentation systems*, and if there are conflicts and supports, one talks about *bipolar argumentation systems*.

3. *Valuation of arguments*: this gives a *weight* to each argument. These weights can be used for comparing arguments. The definition of these weights can use different criteria. For instance, in [AC02a], implicit and explicit priorities are given over the knowledge or over the arguments, whereas, in [CLS03c, ACLS04b], weights are defined using interactions between arguments.

4. *Selection of the most acceptable arguments*: this selection can be achieved using interactions between arguments and/or valuation of these arguments. The set of the acceptable arguments can be considered as an "output" of the argumentation process. Informally, acceptable arguments are those which would be able "to win" an argumentation dialogue between agents. Note that the valuation step could be sufficient for selecting arguments on the base of their weights. However one distinguishes the valuation step from the selection step because the selection mechanism can be more complex and can use notions completely different from the notion of weight.

5. *Conclusion of argumentation*: this last step uses the status of the arguments (acceptable or not) in order to define the status of the "conclusions" of the system; it depends on the application domain of the system (nonmonotonic inference, decision making, negotiation, . . .).

3.2 Unipolar argumentation

The content of this section is the following: basic concepts of unipolar argumentation are presented in Section 3.2.1, Section 3.2.2 on page 35 concerns gradual valuation of arguments, Section 3.2.3 on page 54 is for gradual acceptability of arguments, in Section 3.2.4 on page 62 I present my work on the merging of unipolar argumentation systems, Section 3.2.5 on page 80 describes the notion of revision of unipolar argumentation systems and Section 3.2.6 on page 91 shows the use of an unipolar argumentation system for simulating practical reasoning.

All my works presented in this section on valuation and acceptability have been done with Claudette CAYROL and published in several articles which have been synthesised in [CLS05d] (all the proofs of the propositions given can be found in [CLS05d]). For works on merging, revision and practical reasoning, the reader will find references in the corresponding sections.

3.2.1 Background on unipolar argumentation

In 1995, in a seminal article for argumentation domain in AI ([Dun95]), Dung proposed an abstract argumentation system in which he defined several methods for selecting arguments. He considered that arguments and interactions (only conflicts) were given, he thus ignored the steps corresponding to the creation of arguments and to the definition of the interactions. Moreover, he was not interested by the valuation step.

The main definitions given by Dung are:

Definition 11 *An abstract argumentation system is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$ with the set of arguments $\mathcal{A}$ and a binary relation $\mathcal{R}$ on $\mathcal{A}$ called* attack relation[1].

$A_i \mathcal{R} A_j$ means that A_i attacks *A_j. An argumentation system can be represented by a directed graph in which vertices are arguments and edges represent the attack relation.*

Let $A \in \mathcal{A}$, $\mathcal{R}^-(A)$ denotes the set of arguments which attack A and $\mathcal{R}^+(A)$ denotes the set of arguments attacked by A.

Dung also proposed a notion allowing the "reinstatement" of an argument following an attack. This is the *defence* defined from the attack as follows:

Definition 12 *An argument A_i defends an argument A_j against an argument B iff $B \mathcal{R} A_j$ and $A_i \mathcal{R} B$.*

In the abstract unipolar argumentation system proposed by Dung, the selection mechanism of the most acceptable arguments uses a notion of "collective acceptability", *i.e.* the membership to some particular sets satisfying special properties:

Definition 13

- Conflict-free*: $S \subseteq \mathcal{A}$ is conflict-free iff there are no arguments A_i and A_j in S such that $A_i \mathcal{R} A_j$.*
- Collective defence*: $S \subseteq \mathcal{A}$ collectively defends an argument A_i iff for each argument B such that $B \mathcal{R} A_i$ there exists $C \in S$ such that $C \mathcal{R} B$.*

Using these properties, Dung proposed in [Dun95] different semantics for acceptability:

[1] For us, the "attack" is a *generic notion* corresponding to several cases; for instance, an argument can attacked another one because:

- their conclusions are in conflict,
- the conclusion of one argument is in contradiction with a premise of the other one,
- one argument attacks and is preferred to the other one,
- …

Definition 14 *Let $S \subseteq \mathcal{A}$ be a set of arguments.*

- Admissible semantics*: S is an admissible set iff S is conflict-free and S collectively defends all its elements.*
- Preferred semantics*: S is a preferred extension iff S is $\subseteq$-maximal among the admissible subsets of $\mathcal{A}$.*
- Stable semantics*: S is a stable extension iff S is conflict-free and S attacks each argument which does not belong to S.*
- Grounded semantics*: S is the grounded extension iff S is the smallest fixpoint of the characteristic function $\mathcal{F}$ of $\langle \mathcal{A}, \mathcal{R} \rangle$ ($\mathcal{F}$: $2^{\langle \mathcal{A}, \mathcal{R} \rangle} \rightarrow 2^{\langle \mathcal{A}, \mathcal{R} \rangle}$ with $\mathcal{F}(S) = \{A$ such that S collectively defends $A\}$).*

Note that, in Dung's work, each "attacker" is always considered individually (the "direct attacks" considered together as a unique notion is not used by Dung).

Some interesting properties are given in [Dun95]:

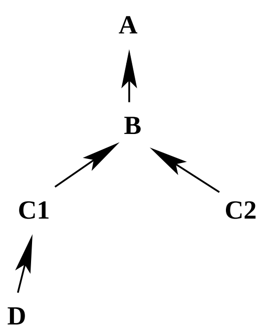
Proposition 3

- *An admissible set is always included in a preferred extension.*
- *There always exists at least one preferred extension.*
- *There always exists only one grounded extension.*
- *If the argumentation system is well-founded[2] then there exists only one preferred extension which is also the only stable extension and the grounded extension.*
- *Each stable extension is also a preferred extension (the converse is false).*
- *There is not always a stable extension.*
- *The grounded extension is included in each preferred extension.*
- *If $\mathcal{R}$ is finite, the grounded extension can be computed in applying iteratively the function $\mathcal{F}$ from the empty set.*

Proposition 4 *The set of unattacked arguments (i.e. $\{A|\mathcal{R}^{-}(A) = \varnothing\}$) is included in the grounded extension, in every preferred extension and in every stable extension.*

All these notions are illustrated on the following argumentation system:

Example 3



In this system, A is defended by C_1 and C_2, and $\{D, C_2, A\}$ is the only preferred extension of the system and it is also stable and grounded.

[2] An argumentation system is well-founded iff there is no infinite sequence of arguments A_0, A_1, $\ldots$, A_n, $\ldots$ such that $\forall i, A_i \in \mathcal{A}$ and $A_{i+1} \mathcal{R} A_i$.

We also need some basic notions related to the graphical representation of an argumentation system:

Definition 15 *Let $\mathcal{G}$ be the attack graph representing the argumentation system $\langle \mathcal{A}, \mathcal{R} \rangle$.*

Leaf of the attack graph *A* leaf of *$\mathcal{G}$ is an unattacked argument $A \in \mathcal{A}$*[3].

Path in the attack graph *A* path from *A* to *B is a sequence of arguments $\mathcal{C} = A_1 - \ldots - A_n$ such that:*

- $A = A_1$,
- $A_1 \mathcal{R} A_2$,
- $\ldots$,
- $A_{n-1} \mathcal{R} A_n$,
- $A_n = B$.

The length of the path *is equal to $n - 1$ (i.e. the number of edges used in the path) and is denoted by $l_{\mathcal{C}}$.*

A particular case is the path[4] *from A to A whose length is 0.*

The set of paths from A to B is denoted by $\mathcal{C}(A, B)$.

Dependency, independence, root-dependency of a path

Let $\mathcal{C}_A \in \mathcal{C}(A_1, A_n)$ and $\mathcal{C}_B \in \mathcal{C}(B_1, B_m)$ be two paths.

These paths are dependent *iff $\exists A_i \in \mathcal{C}_A$, $\exists B_j \in \mathcal{C}_B$ such that $A_i = B_j$. Otherwise they are* independent.

These paths are root-dependent in *A_n iff $A_n = B_m$ and $\forall A_i \neq A_n \in \mathcal{C}_A$, $\nexists B_j \in \mathcal{C}_B$ such that $A_i = B_j$.*

Cycles in the attack graph *A* cycle[5] *is a path $\mathcal{C} = A_1 - \ldots - A_n - A_1$ such that $\forall i, j \in [1, n]$, if $i \neq j$, then $A_i \neq A_j$.*

A cycle $\mathcal{C}$ is isolated *iff $\forall A \in \mathcal{C}$, $\nexists B \in \mathcal{A}$ such that $B \mathcal{R} A$ and $B \notin \mathcal{C}$.*

Two cycles $\mathcal{C}_A = A_1 - \ldots - A_n - A_1$ and $\mathcal{C}_B = B_1 - \ldots - B_m - B_1$ are interconnected *iff $\exists i \in [1, n], \exists j \in [1, m]$ such that $A_i = B_j$.*

We also use the notions of indirect attack or indirect defence. These notions are inspired by notions proposed by Dung in [Dun95] but they are not strictly equivalent[6].

Definition 16 (Direct/Indirect Attacker/Defender of an argument) *Let $A \in \mathcal{A}$:*

- *The* direct attackers *of A are the elements of $\mathcal{R}^-(A)$.*
- *The* direct defenders *of A are the direct attackers of the elements of $\mathcal{R}^-(A)$.*
- *The* indirect attackers *of A are the elements A_i defined by:*
  $$\exists \mathcal{C} \in \mathcal{C}(A_i, A) \text{ such that } l_{\mathcal{C}} = 2k + 1, \text{ with } k \geq 1.$$
- *The* indirect defenders *of A are the elements A_i defined by:*
  $$\exists \mathcal{C} \in \mathcal{C}(A_i, A) \text{ such that } l_{\mathcal{C}} = 2k, \text{ with } k \geq 2.$$

Note that an attacker may also be a defender. This leads us to the notion of controversial arguments (see [Dun95]).

Definition 17 (Controversial arguments) *Let A and $B \in \mathcal{A}$ be two arguments. B is* controversial *with regard to A iff B is at the same time a direct or indirect defender and a direct or indirect attacker of A.*

The next definition gives the link between the graphical representation and the notions of attack and defence:

[3] A is a leaf iff $\mathcal{R}^-(A) = \varnothing$.

[4] We assume that there is an infinity of such paths.

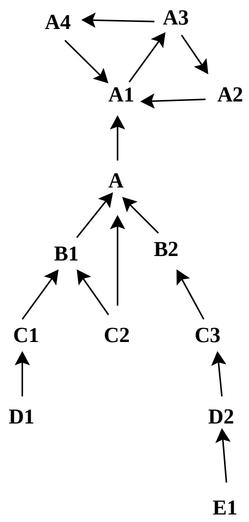
[5] This definition corresponds to the definition of an elementary cycle in Graph Theory (an elementary cycle does not contain two edges having the same origin or the same end).

[6] In [Dun95], direct attackers (resp. defenders) are also indirect attackers (resp. defenders); this is not the case with our definitions.

Definition 18 (Attack or defence branches of an argument) *Let $A \in \mathcal{A}$, an* attack branch *(resp.* defence branch*) for A in a path in $\mathcal{G}$ from a leaf to A whose length is odd (resp. even). A is considered as the* root of an attack branch *(resp.* defence branch*).*

All these notions are illustrated with the following example:

Example 4



On this graph $\mathcal{G}$, one can found:

- *a path from C_2 to A whose length is 2 ($C_2 - B_1 - A$);*
- *2 cycles $A_1 - A_3 - A_2 - A_1$ and $A_1 - A_3 - A_4 - A_1$ whose length is 3; these cycles are not isolated (note that $A_1 - A_3 - A_2 - A_1 - A_3 - A_4 - A_1$ is not a cycle with our definition);*
- *the two previous cycles are interconnected (in A_1 and in A_3);*
- *the paths $D_1 - C_1 - B_1$ and $C_3 - B_2 - A$ are independent, the paths $D_1 - C_1 - B_1 - A$ and $C_3 - B_2 - A$ are root-dependent and the paths $D_1 - C_1 - B_1 - A$ and $C_2 - B_1 - A$ are dependent;*
- *D_1, C_2, E_1 are the leaves of $\mathcal{G}$;*
- *$D_1 - C_1 - B_1 - A$ is an attack branch for A whose length is 3, $C_2 - B_1 - A$ is a defence branch for A whose length is 2;*
- *C_2, B_1 and B_2 are the direct attackers of A;*
- *C_1, C_2 (which is already a direct attacker of A) and C_3 are the direct defenders of A;*
- *D_1 and D_2 are the two indirect attackers of A;*
- *E_1 is the only one indirect defender of A;*
- *C_2 is controversial with regard to A, and to each A_i; moreover A_1, A_2, A_3, and A_4 are controversial with regard to themselves.*

There exist many generalizations of Dung's system, one of them corresponds to constrained argumentation systems proposed by [CMDM06] whose aim is to use *constraints* between arguments that the arguments must satisfy in order to belong to extensions. For instance, one could want that two arguments A and B belong to the same extension. These constraints are generally expressed with a propositional logical formula built using $\mathcal{A}$ as vocabulary.

Definition 19 ([CMDM06] – Constraint, Completion) *Let $\mathcal{A}$ be a set of arguments and $\mathcal{L}_\mathcal{A}$ be the propositional language defined using $\mathcal{A}$ as the set of propositional variables.*

- *C is a* constraint *over arguments of $\mathcal{A}$ iff C is a formula of $\mathcal{L}_\mathcal{A}$.*
- *The* completion *of a set $E \subseteq \mathcal{A}$ is: $\widehat{E} = \{A \mid A \in E\} \cup \{\neg A \mid A \in \mathcal{A} \setminus E\}$.*
- *A set $E \subseteq \mathcal{A}$* satisfies *C iff $\widehat{E}$ is a model of C ($\widehat{E} \vdash C$).*

A constrained argumentation system is defined by:

Definition 20 ([CMDM06] – Constrained argumentation system) *A* constrained argumentation system *is a triple $\mathsf{CoAF} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ with C being a constraint over arguments on the set $\mathcal{A}$.*

The notion of Dung's extension is extended in order to take into account constraints:

Definition 21 ([CMDM06] – C-admissible set) *Let $E \subseteq \mathcal{A}$. E is C-*admissible *iff*

1. *E is admissible,*

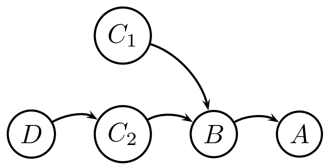
2. *E satisfies the constraint C.*

Note that the empty set is always admissible, but it is not always C-admissible since $\widehat{\varnothing}$ does not always entail C.

Definition 22 ([CMDM06] – C-preferred, C-stable extensions) *Let $E \subseteq \mathcal{A}$.*

- *E is a C-*preferred extension *iff E is $\subseteq$-maximal among the C-admissible sets.*
- *E is a C-*stable extension *iff E is a C-preferred extension which attacks all the arguments of $\mathcal{A} \setminus E$.*

Example 5 *Let* AF *be the argumentation system defined by $\mathcal{A} = \{A, B, C_1, C_2, D\}$ and $\mathcal{R} = \{(D, C_2), (C_1, B), (C_2, B), (B, A)\}$. The system* AF *is graphically represented by:*



There are 4 admissible sets in this system: $E_1 = \varnothing$, $E_2 = \{D, C_1\}$, $E_3 = \{A, C_1\}$ and $E_4 = \{D, C_1, A\}$. Only E_4 is a preferred extension, and in this example, E_4 is also a stable extension.

We transform AF *into a constrained system,* CoAF*, adding for instance a constraint which prevents from having arguments A and D in the same extension. This constraint is defined by: $C = D \rightarrow \neg A$. In this case, $E_4 = \{D, C_1, A\}$ is not a C-admissible set, since the set $\widehat{E_4} = \{D, C_1, A, \neg B, \neg C_2\}$ does not entail the formula $D \rightarrow \neg A$. On the other hand, the admissible sets $E_2 = \{D, C_1\}$ and $E_3 = \{A, C_1\}$ are both C-admissible and are also C-preferred extensions. Note that, on this example, there is no C-stable extension.*

The last important basic notion is the status of an argument. This notion relates to the notion of acceptability, independently of the type of argumentation system used:

Definition 23 (Status of an argument) *Consider an argumentation system (in Dung's sense, or generalized in the sense proposed in [CMDM06]). Let $E_1, \ldots, E_x$ be the extensions of this system for a given semantics. Let $A \in \mathcal{A}$ be an argument.*

1. *A is* accepted *iff $A \in E_i$, $\forall E_i$ with $i = 1, \ldots, x$.*
2. *A is* rejected *iff $\nexists E_i$ such that $A \in E_i$.*
3. *A is* undefined *(or* in abeyance*) iff A is neither accepted, nor rejected. That means that A belongs to some extensions and does not belong to some other ones.*

It is easy to see that a rejected argument in an argumentation system AF in Dung's sense will be also rejected in all the constrained systems CoAF built from AF.

Proposition 5 *Let $A \in \mathcal{A}$. Under stable or preferred semantics, if A is rejected in* AF $= \langle \mathcal{A}, \mathcal{R} \rangle$*, then A is also rejected in* CoAF $= \langle \mathcal{A}, \mathcal{R}, C \rangle$*, $\forall C$.*

Example 5 (cont'd) In AF, D, C_1 and A are accepted; C_2 and B are rejected.
In CoAF, C_1 is accepted; C_2 and B are always rejected; D and A are now in abeyance.

3.2.2 Valuation of arguments

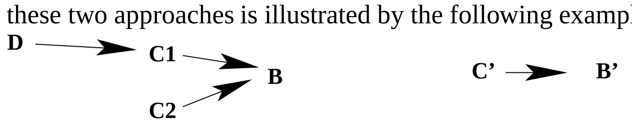
In this section, we assume that the argumentation system is unipolar, abstract and represented by an attack graph.

There already exist some works on this subject: some of them propose a valuation using preferences on belief which are taken into account for creating arguments; other ones exploit the link between premises and conclusion during the building of the argument. My work follows another approach consisting in using the interactions between arguments.

We consider two different valuation methods for taking into account the quality of attackers and defenders of an argument in order to define the value of an argument using only the interaction between arguments:

- In the first approach, the value of an argument only depends on the values of the direct attackers of this argument. Therefore, defenders are taken into account through the attackers. This approach is called *local*.
- In the second approach, the value of an argument represents the set of all the attack and the defence branches for this argument. This approach is called *global*.

The main difference between these two approaches is illustrated by the following example:



In the local approach, B has two direct attackers (C_2 and C_1) whereas B' has only one (C'). Thus B' is better than B (since B' suffers one attack whereas B suffers two attacks).

In the global approach, two branches (one of attack and one of defence) lead to B whereas only one branch of attack leads to B'. Thus B is better than B' (since it has at least one defence whereas B' has none). In this case, C_1 loses its negative status of attacker, since it is in fact "carrying a defence" for B.

3.2.2.1 Local approach (generic valuation)

Some existing proposals can already be considered as examples of *local valuations*.

In [JV99b] approach, a labelling of a set of arguments assigns a status (accepted, rejected, undecided) to each argument using labels from the set $\{+, -, ?\}$. $+$ (resp. $-$, $?$) represents the "accepted" (resp. "rejected", "undecided") status. Intuitively, an argument labelled with $?$ is both supported and weakened.

Definition 24 (Jakobovits and Vermeir's labellings, 1999) *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system. A complete labelling of $\langle \mathcal{A}, \mathcal{R} \rangle$ is a function $Lab : \mathcal{A} \to \{+, ?, -\}$ such that:*

1. *If $Lab(A) \in \{?, -\}$ then $\exists B \in \mathcal{R}^-(A)$ such that $Lab(B) \in \{+, ?\}$*
2. *If $Lab(A) \in \{+, ?\}$ then $\forall B \in \mathcal{R}^-(A) \cup \mathcal{R}^+(A)$, $Lab(B) \in \{?, -\}$*

The underlying intuition is that an argument can only be weakened (label $-$ or $?$) if one of its direct attackers is supported (Condition 1); an argument can get a support only if all its direct attackers are weakened and an argument which is supported (label $+$ or $?$) weakens the arguments it attacks (Condition 2). So:

- If A has no attacker $Lab(A) = +$.
- If $Lab(A) =?$ then $\exists B \in \mathcal{R}^-(A)$ such that $Lab(B) =?$.
- If ($\forall B \in \mathcal{R}^-(A)$, $Lab(B) = -$) then $Lab(A) = +$.
- If $Lab(A) = +$ then $\forall B \in \mathcal{R}^-(A) \cup \mathcal{R}^+(A)$, $Lab(B) = -$.

Every argumentation system can be completely labelled. The associated semantics is that S is an acceptable set of arguments iff there exists a complete labelling Lab of $\langle \mathcal{A}, \mathcal{R} \rangle$ such that $S = \{A | Lab(A) = +\}$.

Other types of labellings are introduced by [JV99b] among which the so-called "rooted labelling" which induces a corresponding "rooted" semantics. The idea is to reject only the arguments attacked by accepted arguments: an attack by an "undecided" argument is not rooted since an "undecided" attacker may become rejected.

Definition 25 (Jakobovits and Vermeir's labellings, 1999 – cont'd)
The complete labelling Lab is rooted iff $\forall A \in \mathcal{A}$, if $Lab(A) = -$ then $\exists B \in \mathcal{R}^-(A)$ such that $Lab(B) = +$.

The rooted semantics enables to clarify the links between all the other semantics introduced by [JV99b] and some semantics introduced by [Dun95].

Example 6 *On the following example:*

$$\mathbf{An} \longrightarrow \mathbf{An{-}1} \text{ - - - - - } \mathbf{A2} \longrightarrow \mathbf{A1}$$

For n even, we obtain $Lab(A_n) = Lab(A_{n-2}) = \ldots = Lab(A_2) = +$ and $Lab(A_{n-1}) = Lab(A_{n-3}) = \ldots = Lab(A_1) = -$.

For n odd, we obtain $Lab(A_n) = Lab(A_{n-2}) = \ldots = Lab(A_1) = +$ and $Lab(A_{n-1}) = Lab(A_{n-3}) = \ldots = Lab(A_2) = -$

Another type of *local valuation* has been introduced by [BH01] for "deductive" arguments. The approach can be characterised as follows. An argument is structured as a pair $\langle$*support*, *conclusion*$\rangle$, where *support* is a consistent set of formulae that enables to prove the formula *conclusion*. The attack relation considered here is strict and cycles are not allowed. The notion of a "tree of arguments" allows a concise and exhaustive representation of attackers and defenders of a given argument, root of the tree. A function, called a "categoriser", assigns a value to a tree of arguments. This value represents the relative strength of an argument (root of the tree) given all its attackers and defenders. Another function, called an "accumulator", synthesises the values assigned to all the argument trees whose root is an argument for (resp. against) a given conclusion. The phase of categorisation therefore corresponds to an interaction-based valuation. [BH01] introduce the following function Cat:

- if $\mathcal{R}^-(A) = \varnothing$, then $Cat(A) = 1$
- if $\mathcal{R}^-(A) \neq \varnothing$ with $\mathcal{R}^-(A) = \{A_1, \ldots, A_n\}$, $Cat(A) = \frac{1}{1 + Cat(A_1) + \ldots + Cat(A_n)}$

Intuitively, the larger the number of direct attackers of an argument, the lower its value. The larger the number of defenders of an argument, the larger its value.

Example 6 (cont'd) *We obtain:*

$Cat(A_n) = 1$, $Cat(A_{n-1}) = 0.5$, $Cat(A_{n-2}) = 0.66$, $Cat(A_{n-3}) = 0.6, \ldots$, *and* $Cat(A_1) = (\sqrt{5} - 1)/2$ *when* $n \to \infty$ *(this value is the inverse of the golden ratio[7]).*

So, we have:

If n is even $Cat(A_{n-1}) \leq \ldots \leq Cat(A_3) \leq Cat(A_1) \leq Cat(A_2) \leq \ldots \leq Cat(A_n) = 1$

If n is odd $Cat(A_{n-1}) \leq \ldots \leq Cat(A_2) \leq Cat(A_1) \leq Cat(A_3) \leq \ldots \leq Cat(A_n) = 1$

Our approach for *local valuations* is a generalization of these two previous proposals in the sense that [BH01] Cat function and [JV99b] labellings are instances of our approach.

The main idea is that the value of an argument is obtained with the composition of two functions:

- one for aggregating the values of all the direct attackers of the argument; so, this function computes the value of the "direct attack";

[7] The golden ratio is a famous number since the antiquity which has several interesting properties in several domains (architecture, for example).

- the other for computing the effect of the “direct attack” on the value of the argument: if the value of the “direct attack” increases then the value of this argument decreases, if the value of the “direct attack” decreases then the value of this argument increases.

Let $(W, \geq)$ be a totally ordered set with a minimum element (V_{Min}) and a subset V of W, that contains V_{Min} and with a maximum element V_{Max}.

Definition 26 (Generic gradual valuation) *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system. A valuation is a function v : $\mathcal{A} \to V$ such that:*

1. $\forall A \in \mathcal{A}$, $v(A) \geq V_{Min}$
2. $\forall A \in \mathcal{A}$, *if* $\mathcal{R}^-(A) = \varnothing$, *then* $v(A) = V_{Max}$
3. $\forall A \in \mathcal{A}$, *if* $\mathcal{R}^-(A) = \{A_1, \ldots, A_n\} \neq \varnothing$, *then* $v(A) = g(h(v(A_1), \ldots, v(A_n)))$

with $h : V^* \to W$ such that (V^* denotes the set of all finite sequences of elements of V)

- $h(x) = x$
- $h() = V_{Min}$
- *For any permutation* $(x_{i1}, \ldots, x_{in})$ *of* $(x_1, \ldots, x_n)$, $h(x_{i1}, \ldots, x_{in}) = h(x_1, \ldots, x_n)$
- $h(x_1, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_n)$
- *if* $x_i \geq x'_i$ *then* $h(x_1, \ldots, x_i, \ldots, x_n) \geq h(x_1, \ldots, x'_i, \ldots, x_n)$

and $g : W \to V$ such that

- $g(V_{Min}) = V_{Max}$
- $g(V_{Max}) < V_{Max}$
- *g is non-increasing (if $x \leq y$ then $g(x) \geq g(y)$)*

Note that $h(x_1, \ldots, x_n) \geq \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)$ is a logical consequence of the properties of the function h.

A first property on the function g explains the behaviour of the local valuation in the case of an argument which is the root of only one branch (like in Example 6 on the facing page):

Proposition 6 *The function g satisfies for all $n \geq 1$:*

$$g(V_{Max}) \leq g^3(V_{Max}) \leq \ldots \leq g^{2n+1}(V_{Max}) \leq g^{2n}(V_{Max}) \leq \ldots \leq g^2(V_{Max}) \leq V_{Max}$$

Moreover, if g is strictly non-increasing and $g(V_{Max}) > V_{Min}$, the previous inequalities become strict.

A second property shows that the local valuation induces an ordering relation on arguments:

Proposition 7 (Complete preordering) *Let v be a valuation in the sense of Definition 26. v induces a complete*[8] *preordering $\succeq$ on the set of arguments $\mathcal{A}$ defined by: $A \succeq B$ iff $v(A) \geq v(B)$.*

A third property handles the cycles:

[8]A complete preordering on $\mathcal{A}$ means that any two elements of $\mathcal{A}$ are comparable.

Proposition 8 (Value in a cycle) *Let $\mathcal{C}$ be an isolated cycle of the attack graph, whose length is n. If n is odd, all the arguments of the cycle have the same value and this value is a fixpoint of the function g. If n is even, the value of each argument of the cycle is a fixpoint of the function g^n.*

The following property shows the underlying principles satisfied by all the local valuations defined according to our schema:

Proposition 9 (Underlying principles) *The gradual valuation given by Definition 26 on the preceding page respects the following principles:*

P1 *The valuation is maximal for an argument without attackers and non maximal for an attacked and undefended argument.*

P2 *The valuation of an argument is a function of the valuation of its direct attackers (the "direct attack").*

P3 *The valuation of an argument is a non-increasing function of the valuation of the "direct attack".*

P4 *Each attacker of an argument contributes to the increase of the valuation of the "direct attack" for this argument.*

The last properties explain why [JV99b] and [BH01] propose instances of the local valuation described in Definition 26 on the previous page:

Proposition 10 (Link with [JV99b])
Every rooted labelling of $\langle \mathcal{A}, \mathcal{R} \rangle$ in the sense of [JV99b] can be defined as an instance of the generic valuation such that:

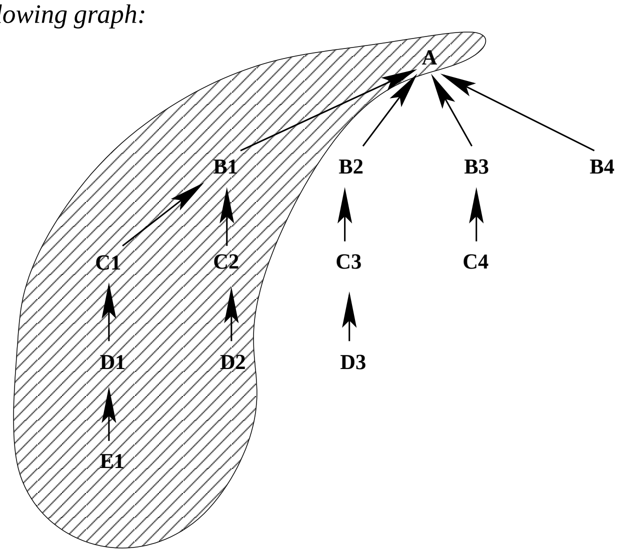
- $V = W = \{-, ?, +\}$ *with* $- < ? < +$,
- $V_{Min} = -$,
- $V_{Max} = +$,
- g *defined by* $g(-) = +$, $g(+) = -$, $g(?) = ?$
- *and* h *is the function* $\mathcal{M}\mathtt{ax}$.

Proposition 11 (Link with [BH01]) *The gradual valuation of [BH01] can be defined as an instance of the generic valuation such that:*

- $V = [0, 1]$,
- $W = [0, \infty[$,
- $V_{Min} = 0$,
- $V_{Max} = 1$,
- $g : W \to V$ *defined by* $g(x) = \frac{1}{1+x}$
- *and* h *defined by* $h(x_1, \ldots, x_n) = x_1 + \ldots + x_n$.

Note that, in the work of [BH01], the valued graphs are acyclic. However, it is easy to show that the valuation proposed by [BH01] can be generalized to graphs with cycles (in this case, we must solve second degree equations – see Example 8 on page 40).

Example 7 *Consider the following graph:*



In this example, with the generic valuation, we obtain:

- $v(E_1) = v(D_2) = v(D_3) = v(C_4) = v(B_4) = V_{Max}$
- $v(D_1) = v(C_2) = v(C_3) = v(B_3) = g(V_{Max})$
- $v(C_1) = v(B_2) = g^2(V_{Max})$
- $v(B_1) = g(h(g^2(V_{Max}), g(V_{Max})))$
- $v(A) = g(h(g(h(g^2(V_{Max}), g(V_{Max}))), g^2(V_{Max}), g(V_{Max}), V_{Max}))$

So, we have:

$$\begin{array}{c} E_1, D_2, D_3, C_4, B_4 \\ \succeq \\ C_1, B_2 \\ \succeq \\ D_1, C_2, C_3, B_3 \end{array}$$

However, the constraints on $v(A)$ and $v(B_1)$ are insufficient to compare A and B_1 with the other arguments.

The same problem exists if we reduce the example to the hatched part of the graph in the previous figure; we obtain $E_1, D_2 \succeq C_1 \succeq D_1, C_2$, but A and B_1 cannot be compared with the other arguments[9].

Now, we use the instance of the generic valuation proposed by [BH01]:

- $v(E_1) = v(D_2) = v(D_3) = v(C_4) = v(B_4) = 1$,
- $v(D_1) = v(C_2) = v(C_3) = v(B_3) = \frac{1}{2}$,
- $v(C_1) = v(B_2) = \frac{2}{3}$,
- $v(B_1) = \frac{6}{13}$,
- $v(A) = \frac{78}{283}$.

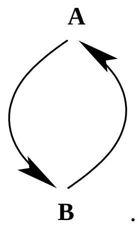
[9] $v(A) = g^2(h(g^2(V_{\text{Max}}, g(V_{\text{Max}})))$ and $v(B_1) = g(h(g^2(V_{\text{Max}}), g(V_{\text{Max}})))$.

So, we have:

$$\begin{array}{c} E_1, D_2, D_3, C_4, B_4 \\ \succeq \\ C_1, B_2 \\ \succeq \\ D_1, C_2, C_3, B_3 \\ \succeq \\ B_1 \\ \succeq \\ A \end{array}$$

However, if we reduce the example to the hatched part of the graph, then the value of A is $\frac{13}{19}$. So, $v(A)$ is better than $v(B_1)$ and $v(D_1)$, but also than $v(C_1)$ (A becomes better than its defender).

Example 8 (Isolated cycle) *Consider the following graph reduced to an isolated cycle:*



.

A generic valuation gives $v(A) = v(B) =$ fixpoint of g^2.

If we use the instance proposed by [BH01], $v(A)$ and $v(B)$ are solutions of the following second degree equation: $x^2 + x - 1 = 0$.

So, we obtain: $v(A) = v(B) = \frac{-1+\sqrt{5}}{2} \approx 0.618$ (the inverse of the golden ratio again).

3.2.2.2 Global approach (with tuples)

We now consider a second approach for the valuation step, called the global approach. Here, the key idea is that the value of A must describe the subgraph whose root is A. So, we want to memorise the length of each branch leading to A in a tuple (for an attack branch, we have an odd integer, and for a defence branch, we have an even integer).

In this approach, the main constraint is that we must be able to identify the branches leading to the argument and to compute their lengths. This is very easy in the case of an acyclic graph. We therefore introduce first a global gradual valuation for acyclic graphs. Then, in the next sections, we extend our proposition to the case of graphs with cycles, and we study the properties of this global gradual valuation.

3.2.2.2.1 Gradual valuation with tuples for acyclic graphs

First, in order to record the lengths of the branches leading to the arguments, we use the notion of tuples and we define some operations on these tuples:

Definition 27 (Tuple) *A tuple is a sequence of integers. The tuple $\underbrace{(0, \ldots, 0, \ldots)}_{\infty}$ will be denoted by 0^∞. The tuple $\underbrace{(1, \ldots, 1, \ldots)}_{\infty}$ will be denoted by 1^∞.*

Notation 2 *$\mathcal{T}$ denotes the set of the tuples built with positive integers.*

Definition 28 (Operations on the tuples) *We have two kinds of operations on tuples:*

- *the concatenation of two tuples is defined by the function $\star : \mathcal{T} \times \mathcal{T} \to \mathcal{T}$ such that*

$$\begin{array}{rcl} 0^{\infty} \star t = t \star 0^{\infty} & = & t \text{ for } t \neq () \\ (x_1, \ldots, x_n, \ldots) \star (x'_1, \ldots, x'_n, \ldots) & = & \mathtt{Sort}(x_1, \ldots, x_n, \ldots, x'_1, \ldots, x'_n, \ldots) \end{array}$$

  `Sort` *being the function which orders a tuple by increasing values.*

- *the addition of a tuple and an integer is defined by the function $\oplus : \mathcal{T} \times \mathbb{N} \to \mathcal{T}$ such that*

$$\begin{array}{rcl} 0^{\infty} \oplus k & = & (k) \\ () \oplus k & = & () \\ (x_1, \ldots, x_n) \oplus k & = & (x_1 + k, \ldots, x_n + k) \\ (x_1, \ldots, x_n, \ldots) \oplus k & = & (x_1 + k, \ldots, x_n + k, \ldots) \text{ if } (x_1, \ldots, x_n, \ldots) \neq 0^{\infty} \end{array}$$

Note that we allow infinite tuples, among other reasons, because they are needed later in order to compute the ordering relations described in Section 3.2.2.2.4 on page 49 (when the graph will be cyclic).

The operations on the tuples have the following properties:

Proposition 12 (Properties of $\star$ and $\oplus$)
The concatenation $\star$ is commutative and associative.
For any tuple t and any integers k and k', $(t \oplus k) \oplus k' = t \oplus (k + k')$.
For any integer k and any tuples t and t' different from 0^{∞}[10]*, $(t \star t') \oplus k = (t \oplus k) \star (t' \oplus k)$.*

In order to valuate the arguments, we split the set of the lengths of the branches leading to the argument in two subsets, one for the lengths of defence branches (even integers) and the other one for the lengths of attack branches (odd integers). This is captured by the notion of tupled values:

Definition 29 (Tupled value) *A* tupled value *is a pair of tuples $vt = [vt_p, vt_i]$ with:*

- *vt_p is a tuple of even integers ordered by increased values; this tuple is called the even component of vt;*
- *vt_i is a tuple of odd integers ordered by increased values; this tuple is called the odd component of vt.*

Notation 3 *$\mathcal{V}$ denotes the subset of $\mathcal{T} \times \mathcal{T}$ of all tupled values (so, $\forall vt \in \mathcal{V}$, vt is a pair of tuples satisfying Definition 29).*

Using this notion of tupled-values, we can define the *computation process* of the gradual valuation with tuples[11] in the case of acyclic graphs.

Definition 30 (Valuation with tuples for acyclic graphs) *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system without cycles. A valuation with tuples is a function $v : \mathcal{A} \to \mathcal{V}$ such that:*

If $A \in \mathcal{A}$ is a leaf then

$$v(A) = [0^{\infty}, ()].$$

If $A \in \mathcal{A}$ has direct attackers denoted by $B_1, \ldots, B_n, \ldots$ then

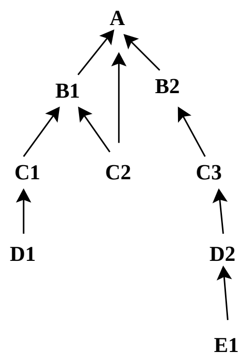
$$v(A) = [v_p(A), v_i(A)] \text{ with: } \left| \begin{array}{l} v_p(A) = (v_i(B_1) \oplus 1) \star \ldots \star (v_i(B_n) \oplus 1) \star \ldots \\ v_i(A) = (v_p(B_1) \oplus 1) \star \ldots \star (v_p(B_n) \oplus 1) \star \ldots \end{array} \right.$$

[10] Otherwise it is false : $(0^{\infty} \star (p)) \oplus k = (p + k)$, whereas $(0^{\infty} \oplus k) \star ((p) \oplus k) = (k) \star (p + k) = (k, p + k)$.
[11] This definition is different from the definition given in [CLS03c]. The ideas are the same but the formalisation is different.

Notes: The choice of the value $[0^\infty, ()]$ for the leaves is justified by the fact that the value of an argument memorises all the lengths of the branches leading to the argument. Using the same constraint, either $v_p(A)$ or $v_i(A)$ may be empty but not both[12].

Note also that the set of the direct attackers of an argument can be infinite (this property will be used when we take into account an argumentation graph with cycles).

Example 9 *On this graph, the valuation with tuples gives the following results:*



On this graph $\mathcal{G}$, we have:

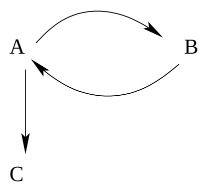
- $v(D_1) = v(C_2) = v(E_1) = [0^\infty, ()]$,
- $v(C_1) = v(D_2) = [(), (1)]$,
- $v(C_3) = [(2), ()]$,
- $v(B_1) = [(2), (1)]$,
- $v(B_2) = [(), (3)]$,
- $v(A) = [(2, 4), (1, 3)]$.

3.2.2.2.2 Study of cycles

Handling cycles raises some important issues: the notion of branch is not always useful in a cycle (for example, in an unattacked cycle like in Examples 8 on page 40 and 10), and when this notion is useful, the length of a branch can be defined in different ways.

Let us consider different examples:

Example 10 (Unattacked cycle) *The graph is reduced to an unattacked cycle $A - B - A$ which attacks the argument C:*



The notion of branch is useless in this case, because there is no leaf in the graph.

There are two possibilities:

- *First, one can consider that the cycle is like an infinite branch; so A (resp. B) is the root of one branch whose length is ∞. But the parity of the length of this branch is undefined, and it is impossible to say if this branch is an attack branch or a defence branch.*
- *The second possibility is to consider that the cycle is like an infinity of branches; so A (resp. B) is the root of an infinity of attack branches and defence branches whose lengths are known and finite.*

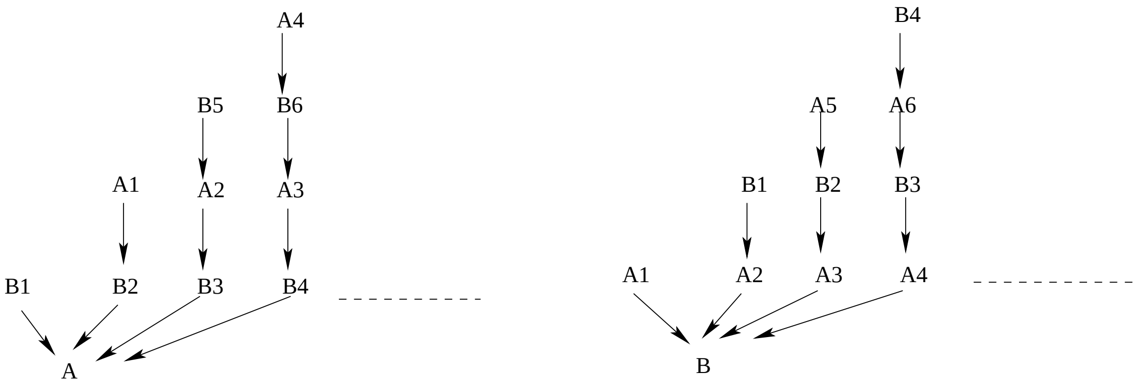
---

[12] The proof is the following:.

- If A is not a leaf, at least one of the tuples is not empty, because there exists at least one branch whose length is > 0 leading to A (see Definitions 28 on the previous page and 30 on the preceding page).
- And, if A is a leaf, there also exists at least one defence branch because the path from A to A is allowed and its length is 0 (in fact, there are an infinity of such paths – see Definition 15 on page 32) and no attack branch leading to the leaf (see Definition 30 on the preceding page).

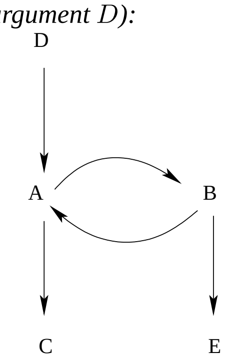
So, the value of a leaf is $[0^\infty, ()]$, and it is impossible that $v_p(A) = v_i(A) = ()$.

The second possibility means that the cycle may have two representations which are acyclic but also infinite graphs (one with the root A and the other one with the root B). This is a rewriting process of the cycle:



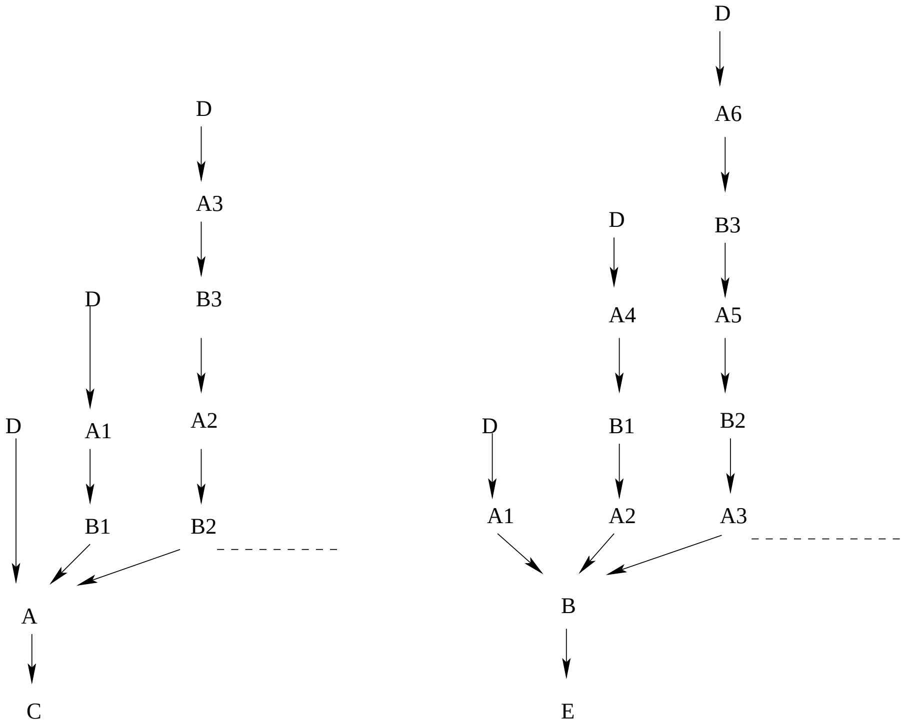
The A_i and B_i must be new arguments created during the rewriting process of the cycle.

Example 11 (Attacked cycle) *The cycle $A - B - A$ is attacked by at least one argument which does not belong to the cycle (here, the attacker is the unattacked argument D):*



In this case, the notion of branch is useful because there exists one leaf in the graph, but the difficulty is to compute the length of this branch. As in Example 10 on the preceding page, we can consider either that there is only one infinite branch (so, it is impossible to know if this branch is an attack or a defence branch), or that there is an infinity of attack branches and defence branches whose lengths are known and finite.

In the second case, the graph can be rewritten into the following structures:



The A_i and B_i must be new arguments created during the rewriting process of the graph.

From the previous examples, we have chosen to manage a cycle as an infinity of attack branches and defence branches whose lengths are known and finite because we would like to be able to apply Definition 30 on page 41 in all cases (acyclic graphs and graphs with cycles). However, we need a rewriting process of the graph with cycles into an acyclic graph. There are two different cases, one for the unattacked cycles and one for the attacked cycles:

Definition 31 (Rewriting of an unattacked cycle) *Let $\mathcal{C} = A_0 - A_1 - \ldots - A_{n-1} - A_0$ an unattacked cycle. The graph $\mathcal{G}$ which contains $\mathcal{C}$ is rewritten as follows:*

1. *the cycle $\mathcal{C}$ is removed,*
2. *and replaced by the infinite acyclic graphs, one for each A_i, $i = 0 \ldots n-1$:*

$$
\begin{array}{ccccccc}
 & & A_i & & & & \\
\nearrow & \nearrow & \ldots & \nwarrow & \nwarrow & \nwarrow & \ldots \\
A_{i_1}^{1} & A_{i_1}^{2} & \ldots & A_{i_1}^{n-1} & A_{i_1}^{n} & A_{i_1}^{n+1} & \ldots \\
 & \uparrow & \ldots & \uparrow & \uparrow & \uparrow & \ldots \\
 & A_{i_2}^{2} & \ldots & A_{i_2}^{n-1} & A_{i_2}^{n} & A_{i_2}^{n+1} & \ldots \\
 & & \ldots & \ldots & \ldots & \ldots & \ldots \\
 & & & \uparrow & \uparrow & \uparrow & \ldots \\
 & & & A_{i_{n-1}}^{n-1} & A_{i_{n-1}}^{n} & A_{i_{n-1}}^{n+1} & \ldots \\
 & & & & \uparrow & \uparrow & \ldots \\
 & & & & A_{i_n}^{n} & A_{i_n}^{n+1} & \ldots \\
 & & & & & \uparrow & \ldots \\
 & & & & & A_{i_{n+1}}^{n+1} & \ldots
\end{array}
$$

3. *the edges between each of the A_i and an argument which does not belong to $\mathcal{C}$ are kept.*

Example 10 on page 42 – Unattacked cycle (cont'd) *The graph $\mathcal{G}$ containing the unattacked cycle $A - B - A$ and the argument C, which is attacked by A, is rewritten as follows:*

$$\begin{array}{cccc} & & C & \\ & & \uparrow & \\ & & A & \\ \nearrow & \nearrow & \nwarrow & \dots \\ A_1^1 & A_1^2 & A_1^3 & \dots \\ & \uparrow & \uparrow & \dots \\ & A_2^2 & A_2^3 & \dots \\ & & \uparrow & \dots \\ & & A_3^3 & \dots \end{array} \qquad \begin{array}{cccc} & & B & \\ \nearrow & \nearrow & \nwarrow & \dots \\ B_1^1 & B_1^2 & B_1^3 & \dots \\ & \uparrow & \uparrow & \dots \\ & B_2^2 & B_2^3 & \dots \\ & & \uparrow & \dots \\ & & B_3^3 & \dots \end{array}$$

where the A_k^l and B_k^l are new arguments.

Definition 32 (Rewriting of an attacked cycle) *Let $\mathcal{C} = A_0 - A_1 - \dots - A_{n-1} - A_0$ an attacked cycle, the direct attacker of each A_i is denoted B_i, if it exists. The graph $\mathcal{G}$ which contains $\mathcal{C}$ is rewritten as follows:*

1. *the cycle $\mathcal{C}$ is removed,*

2. *and replaced by the infinite acyclic graphs, one for each A_i $i = 0 \dots n-1$:*

$$\begin{array}{ccccccc} & & A_i & & & & \\ \nearrow & \nearrow & \dots & \nwarrow & \nwarrow & \nwarrow & \dots \\ B_i & A_{i_1}^{1} & \dots & A_{i_1}^{n-1} & A_{i_1}^{n} & A_{i_1}^{n+1} & \dots \\ & \uparrow & \dots & \uparrow & \uparrow & \uparrow & \dots \\ & B_{(i-1+n)\ mod\ n} & \dots & A_{i_2}^{n-1} & A_{i_2}^{n} & A_{i_2}^{n+1} & \dots \\ & & \dots & \dots & \dots & \dots & \dots \\ & & & \uparrow & \uparrow & \uparrow & \dots \\ & & & A_{i_{n-1}}^{n-1} & A_{i_{n-1}}^{n} & A_{i_{n-1}}^{n+1} & \dots \\ & & & \uparrow & \uparrow & \uparrow & \dots \\ & & & B_{(i+1)\ mod\ n} & A_{i_n}^{n} & A_{i_n}^{n+1} & \dots \\ & & & & \uparrow & \uparrow & \dots \\ & & & & B_i & A_{i_{n+1}}^{n+1} & \dots \\ & & & & & \uparrow & \dots \\ & & & & & B_{(i-1+n)\ mod\ n} & \dots \end{array}$$

   (the branches leading to B_k exist iff B_k exists[13]).

3. *the edges between each of the A_i and an argument which does not belong to $\mathcal{C}$ are kept.*

4. *the edges between each of the B_i and an argument which does not belong to $\mathcal{C}$ are kept.*

Example 11 on page 43 – Attacked cycle (cont'd) *The graph $\mathcal{G}$ containing the cycle $A - B - A$ attacked in A by the argument D and with the argument C (resp. E) attacked by A (resp. B) is rewritten as follows:*

[13] The operator mod is the modulo function.

$$
\begin{array}{cccc}
 & \multicolumn{2}{c}{C} & \\
 & \multicolumn{2}{c}{\uparrow} & \\
 & \multicolumn{2}{c}{A} & \\
\nearrow & \uparrow & \nwarrow & \ldots \\
D & A_1^2 & A_1^4 & \ldots \\
 & \uparrow & \uparrow & \ldots \\
 & A_2^2 & A_2^4 & \ldots \\
 & \uparrow & \uparrow & \ldots \\
 & D & A_3^4 & \ldots \\
 & & \uparrow & \ldots \\
 & & A_4^4 & \ldots \\
 & & \uparrow & \ldots \\
 & & D & \ldots
\end{array}
\qquad\qquad
\begin{array}{cccc}
 & \multicolumn{2}{c}{E} & \\
 & \multicolumn{2}{c}{\uparrow} & \\
 & \multicolumn{2}{c}{B} & \\
\nearrow & \uparrow & \nwarrow & \ldots \\
B_1^1 & B_1^3 & B_1^5 & \ldots \\
\uparrow & \uparrow & \uparrow & \ldots \\
D & B_2^3 & B_2^5 & \ldots \\
 & \uparrow & \uparrow & \ldots \\
 & B_3^3 & B_3^5 & \ldots \\
 & \uparrow & \uparrow & \ldots \\
 & D & B_4^5 & \ldots \\
 & & \uparrow & \ldots \\
 & & B_5^5 & \ldots \\
 & & \uparrow & \ldots \\
 & & D & \ldots
\end{array}
$$

where the A_k^l and B_k^l are new arguments.

Note: If there exist several cycles in a graph, we have two cases.

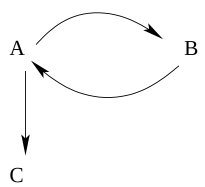
- If they are not interconnected, we rewrite each cycle, and the valuation of the resulting graph after rewriting does not depend on the order of cycles we select to rewrite because the valuation process only uses the length of the branches.
- If they are interconnected, they are considered as a metacyle which is in turn attacked or unattacked and the previous methodology can be used leading to a more complex rewriting process which is not formalized here (see details and examples in [CLS05d]).

3.2.2.2.3 A gradual valuation with tuples for general graphs

Using the definitions given in Sections 3.2.2.2.1 on page 40 and 3.2.2.2.2 on page 42, the gradual valuation with tuples given by Definition 30 on page 41 is applicable for arbitrary graphs *after the rewriting process*.

Let us apply the rewriting process and Definition 30 on page 41 on different examples.

Example 10 on page 42 – Unattacked cycle (cont'd)
Consider the following graph:



The rewriting of this graph has been given in Section 3.2.2.2.2 on page 42.

Definition 30 on page 41 produces:

$$v_p(A) = (v_i(A_1^1) \oplus 1) \star \ldots \star (v_i(A_1^n) \oplus 1) \star \ldots$$

$$v_i(A) = (v_p(A_1^1) \oplus 1) \star \ldots \star (v_p(A_1^n) \oplus 1) \star \ldots$$

Applying Definition 30 on page 41 for different arguments in the rewritten graph produces the following equalities:

- $v(A_n^n) = [0^\infty, ()]$ *for each* $n \geq 1$
- $v(A_{n-1}^n) = [(), (1)]$ *for each* $n \geq 2$
- $v(A_n^m) = [v_p(A_{n+2}^m) \oplus 2, v_i(A_{n+2}^m) \oplus 2]$ *for each n* ≥ 1 *and* $m \geq n+2$

So, using the above equalities in the formulae giving $v_p(A)$ *and* $v_i(A)$*, we define two sequences of tuples : a sequence* $(x_k, k \geq 1)$ *of infinite tuples of even integers, and a sequence* $(y_k, k \geq 1)$ *of infinite tuples of odd integers*

$$x_k = (2) \star (v_i(A_{2k-1}^{2k+1}) \oplus 1) \star \ldots \star (v_i(A_{2k-1}^n) \oplus 1) \star \ldots$$

$$y_k = (1) \star (v_p(A_{2k-1}^{2k+1}) \oplus 1) \star \ldots \star (v_p(A_{2k-1}^n) \oplus 1) \star \ldots$$

From the results stated in Property 12 on page 41, it is easy to prove that $v_p(A) = x_1$ *and for each* $k \geq 1$, $x_k = (2) \star (x_{k+1} \oplus 2)$.

Similarly, $v_i(A) = y_1$ *and for each* $k \geq 1$, $y_k = (1) \star (y_{k+1} \oplus 2)$.

These equations enable to prove that :

> *For each even integer* p *with* $p > 0$*,* p *belongs to each tuple* $x_i, i \geq 1$.
> *For each odd integer* p*,* p *belongs to each tuple* y_i*,* $i \geq 1$.

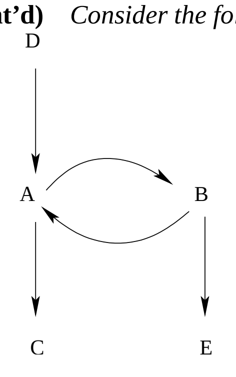
The proof is done by induction on p.

So, $v(A) = v(B) = [(2, 4, 6, \ldots), (1, 3, 5, \ldots)]$.
Then, $v(C) = [(2, 4, 6, \ldots), (3, 5, 7, \ldots)]$.

Note that all the above results can be readily extended to an unattacked cycle of length n, $n \geq 2$.

Proposition 13 (Properties of unattacked cycles)
For each unattacked cycle, for each argument A *of the cycle,* $v(A) = [(2, 4, 6, \ldots), (1, 3, 5, \ldots)]$.

Example 11 on page 43 – Attacked cycle (cont'd) *Consider the following graph:*



The rewriting of this graph has been given in Section 3.2.2.2.2 on page 42.

Definition 30 on page 41 produces:

$$v_p(A) = (v_i(D) \oplus 1) \star (v_i(A_1^2) \oplus 1) \star \ldots \star (v_i(A_1^{2n}) \oplus 1) \star \ldots$$

$$v_i(A) = (v_p(D) \oplus 1) \star (v_p(A_1^2) \oplus 1) \star \ldots \star (v_p(A_1^{2n}) \oplus 1) \star \ldots$$

and also

$v(D) = [0^\infty, ()]$

$v(A_n^n) = [(), (1)]$ *for each* $n \geq 2$

As done in the treatment of Example 10 on page 42, the formulae giving $v_p(A)$ *and* $v_i(A)$ *can be rewritten in order to bring to light some interesting sequences of tuples.*

$$x'_k = (v_i(A_{2k-1}^{2k}) \oplus 1) \star \ldots \star (v_i(A_{2k-1}^{2(k+p)}) \oplus 1) \star \ldots$$

$$y'_k = (1) \star (v_p(A_{2k-1}^{2k}) \oplus 1) \star \ldots \star (v_p(A_{2k-1}^{2(k+p)}) \oplus 1) \star \ldots$$

Then, it is easy to prove that $v_p(A) = x'_1$ *and for each* $k \geq 1$, $x'_k = (x'_{k+1} \oplus 2)$.

Similarly, $v_i(A) = y'_1$ *and for each* $k \geq 1$, $y'_k = (1) \star (y'_{k+1} \oplus 2)$.

The first equation enables to prove that x'_1 *is the empty tuple*[14].

The second equation has already been solved and produces $y'_1 = (1, 3, 5, \ldots)$.

So, $v(A) = [(), (1, 3, 5, \ldots)]$. *For* B, *we can reason as for* A, *and we have* $v(B) = [(2, 4, 6, \ldots), ()]$. *Then,* $v(C) = [(2, 4, 6, \ldots), ()]$, $v(E) = [(), (3, 5, 7 \ldots)]$.

Notation: *in order to simplify the writing, we will not repeat the values inside the tuples (we will just indicate under each value how many times it appears). For example:*

$$[(2, 4, 4, 6, 6, 6, 8, 8, 8, 8 \ldots), (3, 5, 5, 7, 7, 7, 9, 9, 9, 9 \ldots)]$$

will be denoted by

$$[(2, \underbrace{4}_{2}, \underbrace{6}_{3}, \underbrace{8}_{4}, \ldots), (3, \underbrace{5}_{2}, \underbrace{7}_{3}, \underbrace{9}_{4}, \ldots)]$$

Conclusion about cycles Cycles are expensive since all the values obtained are infinite.

In [CLS05d], we introduce an algorithm for computing these tupled values. It uses a process of value propagation and is parametrized by a maximum "number of runs through a cycle". This number will be used in order to stop the propagation mechanism and to obtain finite (thus incomplete) tupled values.

[14] The proof is the following:.

- x'_1 contains only even integers.
- For each k, $x'_k \neq 0^\infty$ since x'_k is the result of the addition of a tuple and an integer.
- If x'_1 is not empty, let e_1 denote the least even integer present in x'_1. As $x'_1 = x'_2 \oplus 2$, x'_2 is not empty and e_2 will denote the least integer present in x'_2. We have $e_1 = e_2 + 2$. So, we are able to build a sequence of positive even integers $e_1, e_2, \ldots$, which is strictly decreasing. That is impossible. So, $x'_1 = ()$.

3.2.2.2.4 Comparison of tupled values

In this section, we define the comparison relation between arguments (so, between some particular tupled values), using the following idea: an argument A is better than an argument B iff A has a better defence (for it) and a lower attack (against it).

The first idea is to use a lexicographic ordering on the tuples. This lexicographic ordering denoted by $\leq_{lex\infty}$ on $\mathcal{T}$ is defined by:

Definition 33 (Lexicographic ordering on tuples) *Let $(x_1, \ldots, x_n, \ldots)$ and $(y_1, \ldots, y_m, \ldots)$ be 2 finite or infinite tuples $\in \mathcal{T}$. $(x_1, \ldots, x_n, \ldots) <_{lex\infty} (y_1, \ldots, y_m, \ldots)$ iff $\exists i \geq 1$ such that:*

- *$\forall j < i$, $x_j = y_j$ and*
- *y_i exists and:*
  - *either the tuple $(x_1, \ldots, x_n, \ldots)$ is finite with a number of elements equal to $i - 1$ (so, x_i does not exist),*
  - *or x_i exists and $x_i < y_i$.*

$(x_1, \ldots, x_n, \ldots) =_{lex\infty} (y_1, \ldots, y_m, \ldots)$ iff the tuples contain the same number $p \in \mathbb{N} \cup \{\infty\}$ of elements and $\forall i$, $1 \leq i \leq p$, $x_i = y_i$.

So, we define: $(x_1, \ldots, x_n, \ldots) \leq_{lex\infty} (y_1, \ldots, y_m, \ldots)$ iff
$(x_1, \ldots, x_n, \ldots) =_{lex\infty} (y_1, \ldots, y_m, \ldots)$ or $(x_1, \ldots, x_n, \ldots) <_{lex\infty} (y_1, \ldots, y_m, \ldots)$.

The ordering $<_{lex\infty}$ is a generalization of the classical lexicographic ordering (see [Xuo92]) to the case of infinite tuples. This ordering is complete but not well-founded (there exist infinite sequences which are strictly non-increasing: $(0) <_{lex\infty} (0, 0) <_{lex\infty} \ldots <_{lex\infty} (0, \ldots, 0, \ldots) <_{lex\infty} \ldots <_{lex\infty} (0, 1)$).

Since the even values and the odd values in the tupled value of an argument do not play the same role, we cannot use a classical lexicographic comparison. So, we compare tupled values in two steps:

- The "first step" *compares the number of attack branches and the number of defence branches* of each argument. So, we have two criteria (one for the defence and the other for the attack). These criteria are aggregated using a *cautious method*: we conclude if one of the arguments has more defence branches (it is better according to the defence criterion) and less attack branches than the other argument (it is also better according to the attack criterion). Note that we conclude positively only when *all* the criteria agree: if one of the arguments has more defence branches (it is better according to the defence criterion) and more attack branches than the other argument (it is worse according to the attack criterion), the arguments are considered to be incomparable.
- Else, the arguments have the same number of defence branches and the same number of attack branches, and a "second step" *compares the quality of the attacks and the quality of the defences* using the length of each branch. This comparison is made with a lexicographic principle (see Definition 33) and gives two criteria which are again aggregated using a cautious method. In case of disagreement, the arguments are considered to be incomparable.

Let us consider some examples:

- $[(2), (1)]$ is better than $[(2), (1, 1)]$ because there are less attack branches in the first tupled value than in the second tupled value, the numbers of defence branches being the same (first step).
- $[(2), (1)]$ is incomparable with $[(2, 2), (1, 1)]$ because there are less defence branches and less attack branches in the first tupled value than in the second tupled value (first step).

- $[(2),(3)]$ is better than $[(2),(1)]$ because there are weaker attack branches in the first tupled value than in the second tupled value (the attack branch of the first tupled value is longer than the one of the second tupled value), the defence branches being the same (second step, using the lexicographic comparison applied on even parts then on odd parts of the tupled values).
- $[(2),(3)]$ is better than $[(4),(3)]$ because there are stronger defence branches in the first tupled value than in the second tupled value (the defence branch is shorter in the first tupled value than in the second tupled value), the attack branches being the same (second step).
- $[(2),(1)]$ is incomparable with $[(4),(3)]$ because there are worse attack branches and better defence branches in the first tupled value than in the second tupled value (second step).

The comparison of arguments is done using Algorithm 1 which implements the principle of a double comparison (first quantitative, then qualitative) with two criteria (one defence criterion and one attack criterion) using a cautious method.

Algorithm 1: Comparison of two tupled values

% **Description of the parameters:** v, w: 2 tupled values %
% **Notations:** %
% $|v_p|$ (resp. $|w_p|$): number of elements in the even component of v (resp. w) %
% if v_p (resp. w_p) is infinite then $|v_p|$ (resp. $|w_p|$) is taken equal to ∞ %
% $|v_i|$ (resp. $|w_i|$): number of elements in the odd component of v (resp. w) %
% if v_i (resp. w_i) is infinite then $|v_i|$ (resp. $|w_i|$) is taken equal to ∞ %
% As usual, $\succ$ will denote the strict relation associated with $\succeq$ defined by: %
% $v \succ w$ iff $v \succeq w$ and not($w \succeq v$). %

begin
- **if** $v = w$ **then** $v \succeq w$ AND $w \succeq v$ — % Case 1 %
- **else**
  - **if** $|v_i| = |w_i|$ *AND* $|v_p| = |w_p|$ **then**
    - % lexicographic comparisons between v_p and w_p and between v_i and w_i %
    - **if** $v_p \leq_{lex\infty} w_p$ *AND* $v_i \geq_{lex\infty} w_i$ **then** $v \succ w$ — % case 2 %
    - **else**
      - **if** $v_p \geq_{lex\infty} w_p$ *AND* $v_i \leq_{lex\infty} w_i$ **then** $v \prec w$ — % case 3 %
      - **else** $v \not\succeq w$ AND $v \not\preceq w$ — % Incomparable tupled values. case 4 %
  - **else**
    - **if** $|v_i| \geq |w_i|$ *AND* $|v_p| \leq |w_p|$ **then** $v \prec w$ — % case 5 %
    - **else**
      - **if** $|v_i| \leq |w_i|$ *AND* $|v_p| \geq |w_p|$ **then** $v \succ w$ — % case 6 %
      - **else** $v \not\succeq w$ AND $v \not\preceq w$ — % Incomparable tupled values. Case 7 %

end

Algorithm 1 defines a partial preordering on the set $v(\mathcal{A})$:

Proposition 14 (Partial preordering) *Algorithm 1 defines a partial preordering $\succeq$ on the set $v(\mathcal{A})$.*
The tupled value $[0^\infty, ()]$ is the only maximal value of the partial preordering $\succeq$.
The tupled value $[(), 1^\infty]$ is the only minimal value of the partial preordering $\succeq$.

Notation: the partial preordering $\succeq$ on the set $v(\mathcal{A})$ induces a partial preordering on the arguments (the partial preordering on $\mathcal{A}$ will be denoted like the partial preordering on $v(\mathcal{A})$): $A \succeq B$ if and only if $v(A) \succeq v(B)$[15].

[15] We will also use the notation $B \preceq A$ defined by: $B \preceq A$ iff $A \succeq B$.

In order to present the underlying principles satisfied by the global valuation, we first consider the different ways for modifying the defence part or the attack part of an argument:

Definition 34 (Adding/removing a branch to an argument)
Let A be an argument whose tupled value is $v(A) = [v_p(A), v_i(A)]$ with $v_p(A) = (x_1^p, \ldots, x_n^p)$ and $v_i(A) = (x_1^i, \ldots, x_m^i)$ ($v_p(A)$ or $v_i(A)$ may be empty but not simultaneously).

Adding (resp. removing) a defence branch to A is defined by:

$v_p(A)$ becomes $\texttt{Sort}(x_1^p, \ldots, x_n^p, x_{n+1}^p)$ where x_{n+1}^p is the length of the added branch (resp. $\exists j \in [1..n]$ such that $v_p(A)$ becomes $(x_1^p, \ldots, x_{j-1}^p, x_{j+1}^p, \ldots, x_n^p)$).

And the same thing on $v_i(A)$ for adding (resp. removing) an attack branch to A.

Definition 35 (Increasing/decreasing the length of a branch of an argument)
Let A be an argument whose tupled value is $v(A) = [v_p(A), v_i(A)]$ with $v_p(A) = (x_1^p, \ldots, x_n^p)$ and $v_i(A) = (x_1^i, \ldots, x_m^i)$ ($v_p(A)$ or $v_i(A)$ may be empty but not simultaneously).

Increasing (resp. decreasing) the length of a defence branch of A is defined by:

$\exists j \in [1..n]$ such that $v_p(A)$ becomes $(x_1^p, \ldots, x_{j-1}^p, x_j'^p, x_{j+1}^p, \ldots, x_n^p)$ where $x_j'^p > x_j^p$ (resp. $x_j'^p < x_j^p$) and the parity of $x_j'^p$ is the parity of x_j^p.

And the same thing on $v_i(A)$ for increasing (resp. decreasing) an attack branch to A.

Definition 36 (Improvement/degradation of the defences/attacks)
Let A be an argument whose tupled value is $v(A) = [v_p(A), v_i(A)]$ ($v_p(A)$ or $v_i(A)$ may be empty but not simultaneously). We define:

An improvement (resp. degradation) of the defence *consists in*

- *adding a defence branch to A if initially $v_p(A) \neq 0^\infty$ (resp. removing a defence branch of A);*
- *or decreasing (resp. increasing) the length of a defence branch of A;*
- *or removing the only defence branch leading to A (resp. adding a defence branch leading to A if initially $v_p(A) = 0^\infty$);*

An improvement (resp. degradation) of the attack *consists in*

- *adding (resp. removing) an attack branch to A;*
- *or decreasing (resp. increasing) the length of an attack branch of A.*

Proposition 15 (Underlying principles) *Let v be a valuation with tuples (Definition 30 on page 41) associated with Algorithm 1 on the preceding page, v respects the following principles:*

P1′ *The valuation is maximal for an argument without attackers and non maximal for an argument which is attacked (whether it is defended or not).*

P2′ *The valuation of an argument takes into account all the branches which are rooted in this argument.*

P3′ *The improvement of the defence or the degradation of the attack of an argument leads to an increase of the value of this argument.*

P4′ *The improvement of the attack or the degradation of the defence of an argument leads to a decrease of the value of the argument.*

Example 7 on page 39 (cont'd) *With the valuation with tuples, we obtain:*

- $v(E_1) = v(D_2) = v(D_3) = v(C_4) = v(B_4) = [0^\infty, ()]$,
- $v(D_1) = v(C_2) = v(C_3) = v(B_3) = [(), (1)]$,
- $v(C_1) = v(B_2) = [(2), ()]$,
- $v(B_1) = [(2), (3)]$,
- $v(A) = [(2, 4), (1, 3, 3)]$.

So, we have:

$$\begin{array}{ccc} E_1, D_2, D_3, C_4, B_4 & & \\ \succ & & \\ C_1, B_2 & & E_1, D_2, D_3, C_4, B_4 \\ \succ & \textit{but also} & \succ \\ B_1 & & A \\ \succ & & \\ D_1, C_2, C_3, B_3 & & \end{array}$$

A is incomparable with almost all the other arguments (except with the leaves of the graph).
Similarly, on the hatched part of the graph, we obtain the following results:

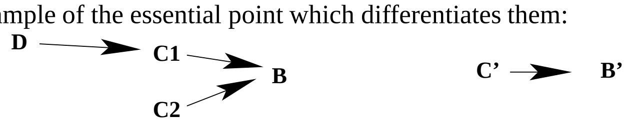
$$E_1, D_2 \succ C_1 \succ B_1 \succ A \succ D_1, C_2$$

A is now comparable with all the other arguments (in particular, A is "worse" than its defender C_1 and than its direct attacker B_1).

3.2.2.3 Main differences between "local" and "global" valuations

[CLS03c] give a comparison of these approaches with some existing approaches ([Dun95, JV99b, BH01]), and also a comparison of the "local" approaches and the "global" approach. The improvement of the global approach proposed in [CLS05d] does not modify the main results of this comparison.

Let us give again here an example of the essential point which differentiates them:



In the local approach, B' is better than B (since B' suffers one attack whereas B suffers two attacks).

In the global approach, B is better than B' (since it has at least a defence whereas B' has none). In this case, C_1 loses its negative status of attacker, since it is in fact "carrying a defence" for B.

The following table synthesises the results about the different proposed valuations:

global approach

arguments having only attack branches	$\preceq$	arguments having attack branches and defence branches	$\preceq$	arguments having only defence branches	$\preceq$	arguments never attacked

local approach

<table>
<tr><td>arguments having several unattacked direct attackers</td><td>$\preceq$</td><td>arguments having only one unattacked direct attacker</td><td>$\preceq$</td><td>arguments having only one attacked direct attacker (possibly defended)</td><td rowspan="2">$\preceq$</td><td rowspan="2">arguments never attacked</td></tr>
<tr><td colspan="5">arguments having several attacked direct attackers (possibly defended)</td></tr>
</table>

The difference between the local approaches and the global approach is also illustrated by the following property:

Proposition 16 (Independence of branches in the global approach)
Let A be an argument having the following direct attackers:

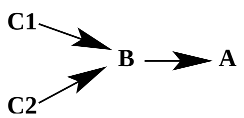
- A_1 *whose value is* $v(A_1) = [(a^1_{p_1}, \ldots, a^1_{p_{m_1}}), (a^1_{i_1}, \ldots, a^1_{i_{m_1}})]$,
- $\ldots$,
- A_n *whose value is* $v(A_n) = [(a^n_{p_1}, \ldots, a^n_{p_{m_n}}), (a^n_{i_1}, \ldots, a^n_{i_{m_n}})]$.

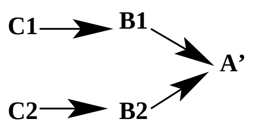
Let A' be an argument having the following direct attackers:

- $A^1_{p_1}$ *whose value is* $v(A^1_{p_1}) = [(a^1_{p_1})()]$,
- $\ldots$,
- $A^1_{p_{m_1}}$ *whose value is* $v(A^1_{p_{m_1}}) = [(a^1_{p_{m_1}})()]$,
- $A^1_{i_1}$ *whose value is* $v(A^1_{i_1}) = [()(a^1_{i_1})]$,
- $\ldots$,
- $A^1_{i_{m_1}}$ *whose value is* $v(A^1_{i_{m_1}}) = [()(a^1_{i_{m_1}})]$,
- $\ldots$,
- $A^n_{p_1}$ *whose value is* $v(A^n_{p_1}) = [(a^n_{p_1})()]$,
- $\ldots$,
- $A^n_{p_{m_n}}$ *whose value is* $v(A^n_{p_{m_n}}) = [(a^n_{p_{m_n}})()]$,
- $A^n_{i_1}$ *whose value is* $v(A^n_{i_1}) = [()(a^n_{i_1})]$,
- $\ldots$,
- $A^n_{i_{m_n}}$ *whose value is* $v(A^n_{i_{m_n}}) = [()(a^n_{i_{m_n}})]$.

Then $v(A) = v(A')$.

This property illustrates the "independence" of branches during the computation of the values in the global approach, even when these branches are not graphically independent. On the following example, A and A' have the same value $[(2,2)()]$ though they are the root of different subgraphs:





This property is not satisfied by the local approach since, using the underlying principles of the local approach (see Property 9 on page 38), the value of the argument A must be at least as good as (and sometimes better than[16]) the value of the argument A' (A having one direct attacker, and A' having two direct attackers).

[16]With the valuation proposed by [BH01], we obtain: $v(A) = \frac{3}{4}$ and $v(A') = \frac{1}{2}$.

3.2.2.4 Conclusion about valuation step

For unipolar argumentation systems, one of my work has been the graduality introduction in the two main related issues of argumentation systems:

- the valuation of arguments,
- the acceptability of arguments.

Regarding the first issue which is concerned by the current section, we have defined two formalisms introducing an interaction-based gradual valuation of arguments.

- First, a generic gradual valuation which covers existing proposals (for example [BH01, JV99b]). This approach is essentially "local" since it computes the value of the argument only from the value of its direct attackers.
- Then, an approach based on a labelling which takes the form of a pair of tuples; this labelling memorises the structure of the graph representing the interactions (the "attack graph"), associating each branch with its length (number of the edges from the leaf to the current node) in the attack graph (if the length of the branch is an even integer, the branch is a defence branch for the current node, otherwise the branch is an attack branch for the current node). This approach is said to be "global" since it computes the value of the argument using the whole attack graph influencing the argument.

We have shown that each of these valuations induces a preordering on the set of the arguments, and we have brought to light the main differences between these two approaches.

These valuations will be used for the selection of the arguments (see Section 3.2.3).

3.2.3 Acceptability

In this section, we now shift to the selection step and introduce graduality in the notion of acceptability.

The basic idea is to select an argument depending on the non-selection of its direct attackers. Following this idea, we propose two different methods:

- The first method consists in refining the classical partition issued from Dung's collective acceptability; this refinement may be achieved using the gradual valuations defined in Section 3.2.2 on page 35.
- The second method takes place in an individual acceptability and consists in defining a new acceptability using only the gradual valuations defined in Section 3.2.2 on page 35.

This work has been done with Claudette CAYROL and has been published in [CLS05d].

3.2.3.1 Different levels of collective acceptability

Under a given semantics, and following Dung, the acceptability of an argument depends on its membership to an extension under this semantics. We consider three possible cases[17]:

- the argument can be *uni-accepted*, when it belongs to all the extensions of this semantics,
- or the argument can be *exi-accepted*, when it belongs to at least one extension of this semantics,
- or the argument can be *not-accepted* when it does not belong to any extension of this semantics.

[17] The terminology used in this section is also used in the domain of nonmonotonic reasoning (see [PL92] and Chapter 1 on page 9): the word *uni* comes from the word *universal* which is a "synonym" of the word *skeptical*, and the word *exi* comes from the word *existential* which is a "synonym" of the word *credulous*. We have chosen to use the words *uni* and *exi* because they recall the logical quantifiers $\forall$ (*for all*) and $\exists$ (*exists at least one*).

However, these three levels seem insufficient. For example, what should be concluded in the case of two arguments A and B which are exi-accepted and such that $A\mathcal{R}B$ or $B\mathcal{R}A$?

So, we introduce a new definition which takes into account the situation of the argument w.r.t. its attackers. This refines the class of the exi-accepted arguments under a given semantics S.

Definition 37 (Cleanly-accepted argument) *Consider $A \in \mathcal{A}$, A is* cleanly-accepted *if and only if A belongs to at least one extension of S and $\forall B \in \mathcal{A}$ such that $B\mathcal{R}A$, B does not belong to any extension of S.*

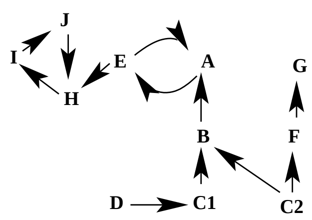
Thus, we capture the idea that an argument will be better accepted, if its attackers are not-accepted.

Proposition 17 *Consider $A \in \mathcal{A}$ and a semantics S such that each extension for S is conflict-free. If A is uni-accepted then A is cleanly-accepted. The converse is false.*

The notion of cleanly-accepted argument refines the class of the exi-accepted arguments. For a semantics S and an argument A, we have the following states:

- A can be *uni-accepted*, if A belongs to all the extensions for S (so, it will also be cleanly-accepted);
- or A can be *cleanly-accepted* (so, it is by definition also exi-accepted); note that it is possible that the argument is also uni-accepted;
- or A can be *only-exi-accepted*, if A is not cleanly-accepted, but A is exi-accepted;
- or A is *not-accepted* if A does not belong to any extension for S.

Example 12 *Consider the following argumentation system.*



There are two preferred extensions $\{D, C_2, A, G\}$ and $\{D, C_2, E, G, I\}$. So, for the preferred semantics, the acceptability levels are the following:

- *D, C_2 and G are uni-accepted,*
- *I is cleanly-accepted but not uni-accepted,*
- *A and E are only-exi-accepted,*
- *B, C_1, F, H and J are not-accepted.*

Note that, in all the cases where there is only one extension, the first three levels of acceptability coincide[18]. This is the case:

- Under the preferred semantics, when there is no even cycle (see [Dou02]).
- Under the grounded semantics (another semantics proposed by Dung – see [Dun95, Dou02] – which has only one extension).

Looking more closely, we can prove the following result:

Proposition 18 *Under the stable semantics, the class of the uni-accepted arguments coincides with the class of the cleanly-accepted arguments.*

Then, using a result issued from the work of [DBC01, DBC02] and reused by [Dou02] which shows that, when there is no odd cycle, all the preferred extensions are stable[19], we apply Property 18 and we obtain the following consequence:

[18] If there is only one extension then the fact that A belongs to all the extensions is equivalent to the fact that A belongs to at least one extension. Moreover, with only one extension containing A, all the attackers of A do not belong to an extension. So, A is cleanly-accepted.

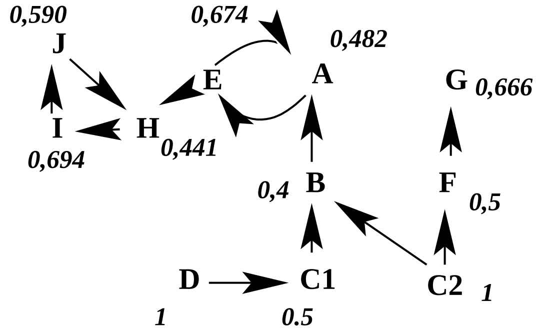
[19] This corresponds to the consistent argumentation system proposed by [Dun95].

Consequence 1 *Under the preferred semantics, when there is no odd cycle, the class of the uni-accepted arguments coincides with the class of the cleanly-accepted arguments.*

Finally, the exploitation of the gradual interaction-based valuations (see Section 3.2.2 on page 35) allows us to define new levels of collective acceptability.

Let v be a gradual valuation and let $\succeq$ be the associated preordering (partial or complete) on $\mathcal{A}$. This preordering can be used inside each acceptability level (for example, the level of the exi-accepted arguments) in order to identify arguments which are better accepted than others.

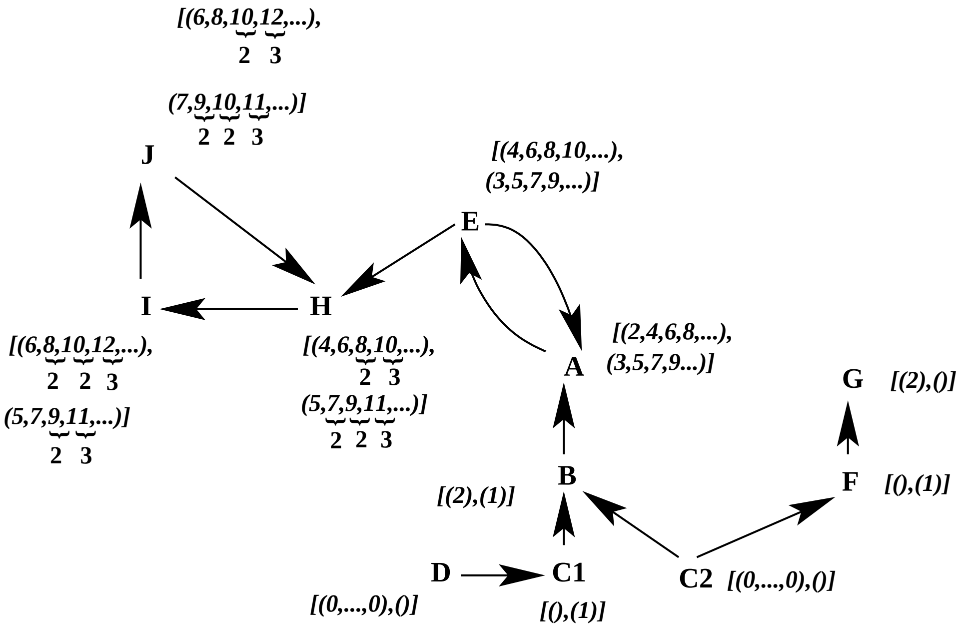
Example 12 on the preceding page (cont'd) *Two different gradual valuations are applied on the same graph:*



Besnard & Hunter's (2001) valuation

With the instance of the generic valuation proposed by [BH01] (see Section 3.2.2.1 on page 35), we obtain the following comparisons:

$$D, C_2 \succ I \succ E \succ G \succ J \succ C_1, F \succ A \succ H \succ B$$



Valuation with tuples

With the global valuation with tuples presented in Section 3.2.2.2 on page 40, we obtain the following comparisons:

$$D, C_2 \succ G \succ B \succ F, C_1$$
$$D, C_2 \succ A \succ E$$
$$D, C_2 \succ H \succ E$$
$$D, C_2 \succ I$$
$$D, C_2 \succ J$$

So, all the arguments belonging to a cycle are incomparable with G, B, F, C_1 and, even between them, there are few comparison results.

If we apply the preordering induced by a valuation without respecting the acceptability levels defined in this section, counter-intuitive situations may happen. In Example 12 on page 55, we obtain:

- With the valuation of [BH01] and under the preferred semantics, $E \succ G$ despite the fact that G is uni-accepted and E is only-exi-accepted.
- With the valuation with tuples and under the preferred semantics, $H \succ E$ despite the fact that E is only-exi-accepted and H is not-accepted.

These counter-intuitive situations illustrate the difference between the acceptability definition and the valuation definitions (even if both use the interaction between arguments, they do not use it in the same way).

3.2.3.2 Towards a gradual individual acceptability

The individual acceptability is based on the comparison of an argument with its attackers.

The first proposal has been to select an argument if and only if it does not have any attacker (see [EGFK93b]).

This has later been extended by [AC98] where, using a preference relation between arguments (an intrinsic valuation), an argument is accepted if and only if it is preferred to each of its attackers.

Following this proposal, we propose the same mechanism but with the *interaction-based valuation*.

Given v a gradual valuation, the preordering induced by v can be directly used in order to compare, from the acceptability point of view, an argument and its attackers[20]. This defines a new class of acceptable arguments: well-defended arguments.

Definition 38 (Well-defended argument) *Consider $A \in \mathcal{A}$, A is* well-defended (for v) *if and only if $\forall B \in \mathcal{A}$ such that $B \mathcal{R} A$, $B \not\succ A$.*

Thus, we capture the idea that an argument will be better accepted if it is at least as good as its direct attackers (or incomparable with them in the case of a partial ordering). The set of well-defended arguments will depend on the valuation used.

Using this new notion, the set of arguments is partitioned in three classes:

- the first class contains the arguments which are not attacked,
- the second class contains the arguments which are attacked but are well-defended,
- the third class contains the other arguments (attacked and not well-defended).

[20] This idea is also used in the notion of "defeat" proposed by [BC02]. So, there is a link between a "well-defended argument" and an argument which is not "attacked" in the sense of [BC02] by its direct attackers. Note that, in the work of [BC02], the valuation is an extra knowledge added in the argumentation system. In contrast, here, the v-preference is extracted from the attack graph.

Note that the set of well-defended arguments corresponds to the union of the two first classes. A further refinement uses the gradual valuation inside each of the classes as in Section 3.2.3.1 on page 54.

In Example 12 on page 55 presented in Section 3.2.3.1 on page 54, the well-defended arguments are:

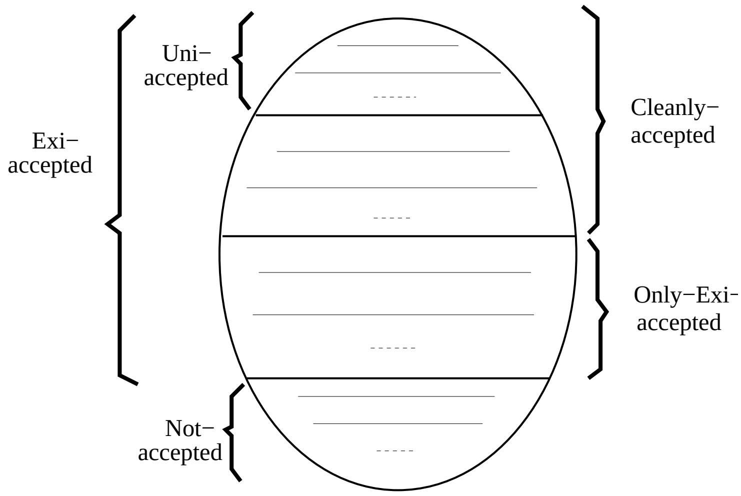
- D, C_2, G, H and A (A is incomparable with B but better than E) for the valuation with tuples,
- though with the valuation of [BH01] the well-defended arguments are D, C_2, G, I and E (E is better than A).

Note also that, as in the semantics of [Dun95], Definition 38 on the previous page considers the attackers one by one. It is not suitable for a valuation which handles the "direct attack" as a whole (as the valuation of [BH01] – see the counterexamples presented in Section 3.2.3.3).

3.2.3.3 Compatibility between acceptability and gradual valuation

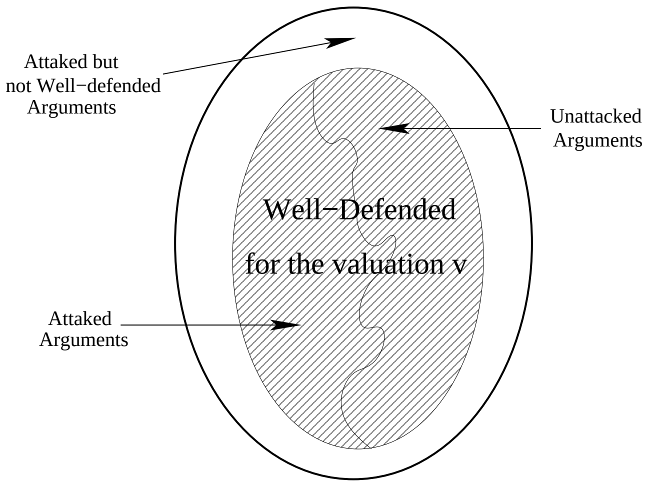
Following the previous sections, the set of arguments can be partitioned in two different ways:

- First, given a semantics S and a gradual valuation v, it is possible to use the partition issued from [Dun95] which we have refined:



Refinement of each level with the gradual valuation v

- Second, given a gradual valuation v, it is possible to use the partition induced by the notion of well-defended arguments:



A very natural and interesting question is: “is it possible to find a semantics S and a gradual valuation v such that the associated partitions have some compatibilities?”

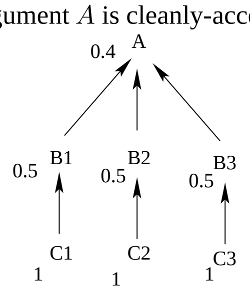
The following examples show that the class of well-defended arguments does not correspond to the class of cleanly-accepted arguments (in some cases, some uni-accepted arguments are even not well-defended).

3.2.3.3.1 Examples showing the non-compatibility in the general case

We give examples for each usual valuation (the global valuation with tuples and 2 instances of the generic local valuation: [BH01, JV99b]) and for the most classical semantics for acceptability (preferred semantics and stable semantics of [Dun95]).

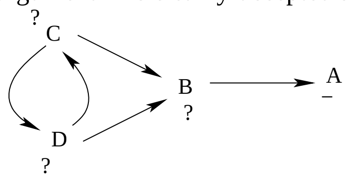
Cleanly-accepted argument but not well-defended: There are 3 examples (each using a distinct valuation: one for the global valuation and two for the two well-known instances of the local valuation):

- the argument A is cleanly-accepted but it is not well-defended:



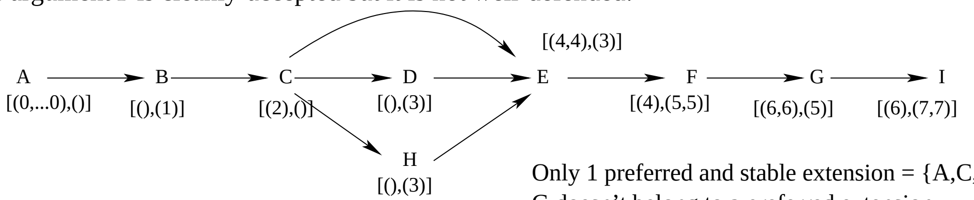
Only 1 preferred and stable extension = { C1, C2, C3, A}
B1, B2, B3 do not belong to a preferred extension
Bi $\succ$ A forall i = 1, 2, 3

- the argument A is cleanly-accepted but it is not well-defended:



2 preferred and stable extensions : {C,A} and {D,A}
B doesn’t belong to a preferred extension
B $\succ$ A

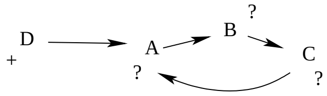
- the argument I is cleanly-accepted but it is not well-defended:



Only 1 preferred and stable extension = {A,C,F,I}
G doesn’t belong to a preferred extension
G $\succ$ I

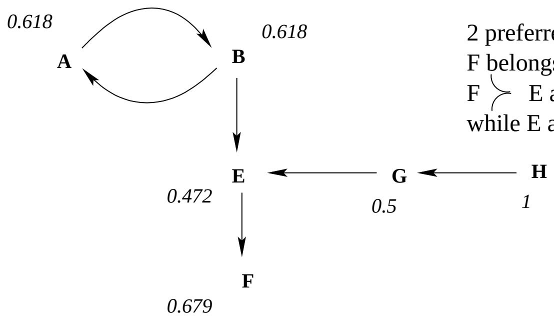
Well-defended argument but not cleanly-accepted: Similarly, for the same three valuations, we have:

- the argument C is well-defended but it is not cleanly-accepted:



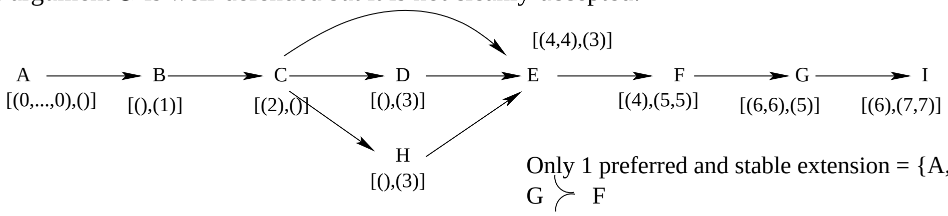
Only 1 preferred and stable extension : {D, B}
C $\succeq$ B
C doesn’t not belong to a preferred or a stable extension

- the argument F is well-defended but it is not cleanly-accepted:



2 preferred and stable extensions = {A,H,E} and {B,H,F}
F belongs to a preferred and stable extension
F $\succ$ E and E belongs also to a preferred and stable extension
while E attacks F

- the argument G is well-defended but it is not cleanly-accepted:



Only 1 preferred and stable extension = {A,C,F,I}
G $\succ$ F
G doesn't belong to a preferred or stable extension

3.2.3.3.2 Particular cases leading to compatibility

In the context of an argumentation system with a finite relation $\mathcal{R}$ without cycles[21], the stable and the preferred semantics provide only one extension and the levels of uni-accepted, exi-accepted, cleanly-accepted coincide.

In this context, there are at least two particular cases leading to compatibility.

First case: It deals with the global valuation with tuples.

Theorem 1 *Let $\mathcal{G}$ be the graph associated with $\langle\mathcal{A},\mathcal{R}\rangle$, $\langle\mathcal{A},\mathcal{R}\rangle$ being an argumentation system with a finite relation $\mathcal{R}$ without cycles and satisfying the following condition: $\exists\ A \in \mathcal{A}$ such that*

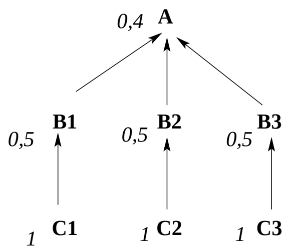
- *$\forall X_i$, leaf of $\mathcal{G}$, $\exists$ only one path from X_i to A, $X_i^1 - \ldots - X_i^{l_i} - A$ with $X_i^1 = X_i$ and l_i the length of this path (if l_i is even, this path is a defence branch for A, else it is an attack branch),*
- *all the paths from X_i to A are root-dependent in A,*
- *$\forall A_i \in \mathcal{A}$, $\exists X_j$ a leaf of $\mathcal{G}$ such that A_i belongs to a path from X_j to A.*

Let v be a valuation with tuples. Let S be a semantics $\in$ {preferred, stable}.

1. *$\forall B \in \mathcal{A}$, $B \neq A$, B (exi, uni, cleanly) accepted for S iff B well-defended for v.*
2. *If A is (exi, uni, cleanly) accepted for S then A is well-defended for v (the converse is false).*
3. *If A is well-defended for v and if all the branches leading to A are defence branches for A then A is (exi, uni, cleanly) accepted for S.*

Note that Theorem 1 is, in general, not satisfied by a local valuation. See the following counterexample for the valuation of [BH01]:

[21] So, $(\mathcal{A},\mathcal{R})$ is well-founded.



The graph satisfies the condition stated in Theorem 1 on the facing page. The set of well-defended arguments is $\{C_1, C_2, C_3\}$ (so, A is not well-defended). Nevertheless, $\{C_1, C_2, C_3, A\}$ is the preferred extension.

Second case: This second case concerns the generic local valuation:

Theorem 2 *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system with a finite relation $\mathcal{R}$ without cycles. Let S be a semantics $\in$ {preferred, stable}. Let v be a generic local valuation satisfying the following condition $(*)$:*

$$(\forall i = 1 \dots n, g(x_i) \geq x_i) \Rightarrow (g(h(x_1, \dots, x_n)) \geq h(x_1, \dots, x_n)) \qquad (*)$$

$\forall A \in \mathcal{A}$, A (exi, uni, cleanly) accepted for S iff A well-defended for v.

This theorem is a direct consequence of the following lemma:

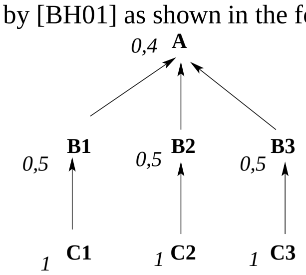
Lemma 1 *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system with a finite relation $\mathcal{R}$ without cycles. Let S be a semantics $\in$ {preferred, stable}. Let v be a generic local valuation satisfying the condition $(*)$.*

(i) *If A is exi-accepted and A has only one direct attacker B then $A \succeq B$.*

(ii) *If B is not-accepted and B has only one direct attacker C then $C \succeq B$.*

Remark: The condition $(*)$ stated in Theorem 2 is:

- false for the local valuation proposed by [BH01] as shown in the following graph:



  We know that $g(x) = \frac{1}{1+x}$ and $h(x_1, \dots, x_n) = \Sigma_{i=1}^{n} x_i$ (see Property 11 on page 38). We get:
  - $\forall i = 1 \dots 3$, $x_i = v(B_i) = 0.5$,
  - $\forall i = 1 \dots 3$, $g(x_i) = 0.66$, so $g(x_i) \geq x_i$,
  - and nevertheless $g(h(x_1, x_2, x_3)) = v(A) = 0.4 \not\geq h(x_1, x_2, x_3) = 1.5$.
- false for the local valuations defined with h such that $\exists n > 1$ with $h(x_1, \dots, x_n) > \mathcal{M}\mathtt{ax}(x_1, \dots, x_n)$ (for all the functions g strictly non-increasing): see the previous graph where $h(x_1, x_2, x_3) = 1.5$ and $\mathcal{M}\mathtt{ax}(x_1, x_2, x_3) = 0.5$.
- true for the local valuations defined with $h = \mathcal{M}\mathtt{ax}$ (for all the functions g): if $h = \mathcal{M}\mathtt{ax}$ then $g(h(x_1, \dots, x_n)) = g(\mathcal{M}\mathtt{ax}(x_1, \dots, x_n)) = g(x_j)$, x_j being the maximum of the x_i; and, by assumption, $g(x_i) \geq x_i$, $\forall x_i$, so in particular for x_j; so, we get:
$$g(h(x_1, \dots, x_n)) = g(x_j) \geq x_j = \mathcal{M}\mathtt{ax}(x_1, \dots, x_n) = h(x_1, \dots, x_n).$$

3.2.3.4 Conclusion on acceptability

In this paper, we have introduced graduality in the two main related issues of argumentation systems:

- the valuation of arguments,
- the acceptability of arguments.

Regarding the second issue which is concerned by the current section, two distinct approaches have been proposed:

- First, in the context of the collective acceptability of [Dun95]: three levels of acceptability (uni-accepted, exi-accepted, not-accepted) were already defined. More graduality can be introduced in the collective acceptability using the notion of *cleanly-accepted* arguments (those whose direct attackers are not-accepted).
- Then, in the context of individual acceptability: using the previously defined gradual valuations, the new notion of *well-defended* arguments has been introduced (those which are preferred to their direct attackers in the sense of a given gradual valuation v).

The first concept induces a refinement of the level of exi-accepted in two sublevels (cleanly-accepted arguments and only-exi-accepted arguments). The gradual valuation allows graduality inside each level of this collective acceptability.

The second concept induces two new levels of acceptability (well-defended arguments and not-well-defended arguments). The gradual valuation also allows graduality inside each level of this individual acceptability.

Regarding our initial purpose of introducing graduality in the definition of acceptability, we have adopted a basic principle:

- acceptability is strongly related to the interactions between arguments (represented on the graph of interactions),
- and an argument is all the more acceptable if it is preferred to its direct attackers.

Then, we have followed two different directions. One is based on a refinement of an existing partition and remains in the framework of Dung's work. The other one is based on the original concept of "being well-defended", and deserves further investigation, in particular from a computational point of view.

3.2.4 Merging in unipolar argumentation

In a multi-agent setting, argumentation can also be used to represent (part of) some information exchange processes, like negotiation, or persuasion (see for example [Mac79, WK95, Gor95, PJ96, AMP00, AP02, AP04]). For instance, a negotiation process between two agents about whether some belief must be considered as true given some evidence can be modelled as a two-player game where each move consists in reporting an argument which attacks arguments given by the opponent.

In this section, we also consider argumentation in a multi-agent setting, but from a very different perspective. Basically, our purpose is to characterize the set of arguments acceptable by a group of agents, when the data furnished by each agent consist solely of an (abstract) argumentation system from Dung's theory.

At a first glance, a simple approach for achieving this goal consists in voting on the acceptable sets provided by each agent: a set of arguments is considered acceptable by the group if and only if it is acceptable for "sufficiently many" agents from the group (where the meaning of "sufficiently many" refers to different *voting* methods). No merging at all is required here. By means of example, we show that our merging-based approach leads to results which are much more expected than those furnished by a direct vote on the (sets of) arguments acceptable by each agent.

Our approach is more sophisticated. It follows a three-step process: first, each argumentation system is expanded into a partial system over the set of all arguments considered by the group of agents (reflecting that some agents may easily ignore arguments pointed out by other agents, as well as how such arguments interact with her own ones); then, merging is used on the expanded systems as a way to solve the possible conflicts between them, and a set of argumentation systems which are as close as possible to the whole profile is generated; finally, the last step consists in selecting the acceptable arguments at the group levels from the set of argumentation systems.

In order to reach this goal, we first introduce a notion of *partial argumentation system*, which extends Dung's argumentation system so as to represent *ignorance* concerning the attack relation. This is necessary in our setting since all the agents participating in the merging process are not assumed to share the same global set of arguments. Accordingly, the argumentation system furnished by each agent is first expanded into a partial argumentation system, and all such partial systems are built over the same set of arguments, those pointed out by at least one agent. Of course, there exist many different ways to incorporate a new argument into an argumentation system. Each agent can have her own expansion policy. We mention some possible policies, and focus on one of them, called the *consensual expansion*: when incorporating a new argument into her own system, an agent is ready to conclude that this argument attacks (resp. is attacked by) another argument whenever all the other agents who are aware of both arguments agree with this attack; otherwise, she concludes that she ignores whether an attack takes place or not.

Once all the expansions of the input argumentation systems have been computed, the proper merging step can be achieved; it consists in computing all the argumentation systems over the global set of arguments which are "as close as possible" to the partial systems generated during the last stage. Closeness is characterized by a notion of distance between an argumentation system and a profile of partial systems, induced from a primitive notion of distance between partial systems and an aggregation function. Several primitive distances and aggregation functions can be used; we mainly focus on the edit distance (which is, roughly speaking, the number of insertions/deletions of attacks needed to turn a given system into another one), and consider sum, max and leximax as aggregation functions.

Like the input of the overall merging process, the result of the merging step is a set of argumentation systems. However, while the first one reflects different points of view (since each system is provided by a specific agent), the second set expresses some uncertainty on the merging due to the presence of conflicts. The last step of the process consists in defining the acceptable arguments for the group under the uncertainty provided by this set of argumentation systems. Once again, several sensible definitions are given. We show that the sets of arguments considered acceptable when the input is the set of argumentation systems primarily furnished by the agents may drastically differ from the sets of arguments considered acceptable after the merging step, and by means of example, we show that the latter ones are more in accordance with the intuition.

This work has been done collaborating with researchers at CRIL Laboratory (Lens) and LERIA Laboratory (Angers). All the proofs of the propositions given in this section can be found in [CMDK$^+$07].

3.2.4.1 Some examples

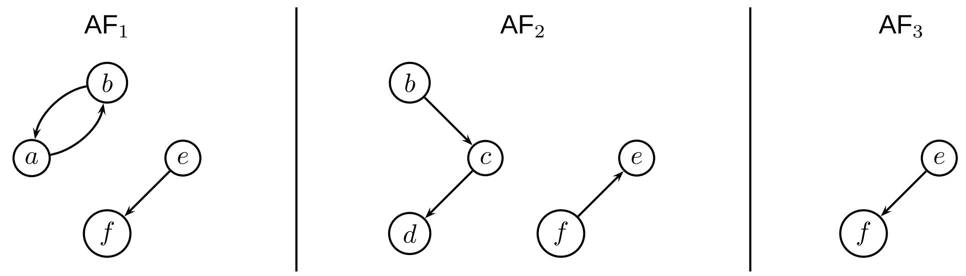
Given a profile (*i.e.*, a vector) $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ of n AFs (with $n \geq 1$) where each $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle$ represents the data given by Agent i, our purpose is to determine the subsets of $\bigcup_i \mathcal{A}_i$ which are acceptable by the group of n agents. Voting is one way to achieve this goal.

3.2.4.1.1 Voting is not enough

Indeed, a simple approach to address the problem consists in considering a set of arguments acceptable for the group when it is acceptable for "sufficiently many" agents of the group. The voting method under consideration makes precise what "sufficiently many" means: it can be, for instance, simple majority. Let us illustrate such an approach on an example:

Example 13 *Consider the three following argumentation systems:*

- $\mathsf{AF}_1 = \langle \{a, b, e, f\}, \{(a, b), (b, a), (e, f)\} \rangle$,
- $\mathsf{AF}_2 = \langle \{b, c, d, e, f\}, \{(b, c), (c, d), (f, e)\} \rangle$,
- $\mathsf{AF}_3 = \langle \{e, f\}, \{(e, f)\} \rangle$.



Whatever the chosen semantics (among Dung's ones), c does not belong to any extension of AF_2. As c is not known by the two other agents, it cannot be considered as acceptable by the group whatever the voting method (under the reasonable assumption that it is a choice function based on extensions, i.e., only subsets of an extension of an AF_i are eligible as acceptable sets). However since c (resp. a) is not among the arguments reported by the first agent and the third one (resp. the second and the third ones), it can be sensible to assume that the three agents agree on the fact that a attacks b, b attacks a and b attacks c. Indeed, this assumption is compatible with any of the three argumentation systems reported by the agents. Under this assumption, it makes sense to consider $\{c\}$ credulously acceptable for the group given that c is considered defended by a against b by Agent 1 and there is no conflicting evidence about it in the AFs provided by the two other agents.

As this example illustrates it, our claim is that, in general, voting is not a satisfying way to aggregate the data furnished by the different agents under the form of argumentation systems. Two problems arise:

Problem 1 Voting makes sense only if all agents consider the same set of arguments $\mathcal{A}$ at start (otherwise, the set $2^{\mathcal{A}}$ of alternatives is not common to all agents). However, it can be the case that the sets of arguments reported by the agents differ from one another.

Problem 2 Voting relies only on the selected extensions: the attack relations (from which extensions are characterized) are not taken into consideration any more once extensions have been computed. This leads to much significant information being set aside which could be exploited to define the sets of acceptable arguments at the group level.

3.2.4.1.2 Union is not merging (in general)

In order to solve both problems, a simple approach (at a first glance) consists in forming the union of the argumentation systems AF_1, ..., AF_n, *i.e.*, considering the argumentation system denoted $\mathsf{AF} = \bigcup_{i=1}^{n} \langle \mathcal{A}_i, \mathcal{R}_i \rangle$ and defined by $\mathsf{AF} = \langle \bigcup_{i=1}^{n} \mathcal{A}_i, \bigcup_{i=1}^{n} \mathcal{R}_i \rangle$. Unfortunately, such a merging approach to argumentation systems cannot be taken seriously. Let us illustrate it on our running example:

Example 13 on the previous page (cont'd) *The resulting AF is* $\bigcup_{i=1}^{3} \mathsf{AF}_i = \langle \{a, b, c, d, e, f\}, \{(a, b), (b, a), (b, c), (c, d), (e, f), (f, e)\} \rangle$.

Example 13 on the preceding page shows that the union approach to merging argumentation systems suffers from a major problem: it solves conflicts by giving to the explicit attack information some undue prominence to implicit non-attack information. Thus, when a pair of arguments (like, say (f, e)) does not belong to the attack relation furnished by an agent (say, Agent 1) while both arguments (f and e) belong to the set of arguments she points out, the meaning is that for Agent 1, argument f *does not attack* argument e. Imagine now that in the considered profile of argumentation systems, 999 agents report the same system as Agent 1, and the 1000^{th} agent is Agent 2. In the resulting argumentation system considered at the group level, assuming that union is used as a merging operator, it will be the case that f attacks e while 999 agents over 1000 believes that it is not the case!

3.2.4.2 Partial Argumentation Systems

The example introduced in the previous section has illustrated that different cases must be taken into account:

- an argument exists in the argumentation system AF_1 of one of the agents and does not exist in the argumentation system AF_2 of at least another agent;
- an interaction between two arguments exists in the argumentation system AF_1 of one agent and does not exist in the argumentation system AF_2 of at least another agent.

In the first case, the new argument can be added to AF_2 but the question is what to do for the interactions between this new argument and the other arguments of AF_2.

In the second case, things are different: if an interaction between two arguments a and b exists in a system AF_1 and not in another system AF_2, even when a and b are in AF_2, we cannot add the interaction in AF_2 (that Agent 2 did not include this attack in AF_2 is on purpose). Indeed, if an interaction is not present in an AF, it means that this interaction *does not exist* for the corresponding agent. The consequence of this is the necessity to discriminate among several cases whenever an argument a has to be added to an AF. Let b be an argument of the AF under consideration, three cases must be considered:

- the agent believes that the interaction (a, b) exists (attack);
- the agent believes that the interaction (a, b) does not exist (non-attack);
- the agent does not know whether the interaction (a, b) exists (ignorance).

The first two cases express the fact that the knowledge of the agent is sufficient for computing the new interaction concerning a. The third case expresses that the agent is not able to compute the new interaction concerning a and the arguments she pointed out (several reasons can explain it, especially a lack of information, or a lack of computational resources).

Handling these different kinds of information within a uniform setting calls for an extension[22] of the notion of argumentation systems, that we call *partial argumentation systems*.

Definition 39 (Partial argumentation system (PAF)) *A* (finite) partial argumentation system *over $\mathcal{A}$ is a quadruple* $\mathsf{PAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{I}, \mathcal{N} \rangle$ *where*

- *$\mathcal{A}$ is a finite set of arguments,*
- *$\mathcal{R}, \mathcal{I}, \mathcal{N}$ are binary relations on $\mathcal{A}$:*
  - *$\mathcal{R}$ is the* attack relation,
  - *$\mathcal{I}$ is called the* ignorance relation *and is such that $\mathcal{R} \cap \mathcal{I} = \varnothing$,*
  - *and $\mathcal{N} = (\mathcal{A} \times \mathcal{A}) \setminus (\mathcal{R} \cup \mathcal{I})$ is called the* non-attack relation.

$\mathcal{N}$ is deduced from $\mathcal{A}$, $\mathcal{R}$ and $\mathcal{I}$, so a partial argumentation system can be fully specified by $\langle \mathcal{A}, \mathcal{R}, \mathcal{I} \rangle$. We use both notations in the following.

Each AF is a particular PAF for which the set $\mathcal{I}$ is empty (we say that such an AF is equivalent to the associated PAF). In an AF, the $\mathcal{N}$ relation also exists even if it is not given explicitly ($\mathcal{I} = \varnothing$ and $\mathcal{N} = \mathcal{A} \times \mathcal{A} \setminus \mathcal{R}$). So, an AF could also be denoted by $\langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$.

Each PAF over $\mathcal{A}$ can be viewed as a compact representation of a set of AFs over $\mathcal{A}$, called its *completions*:

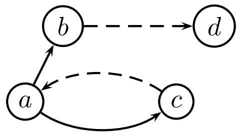
[22] In [CLS05c] and Section 3.3 on page 109, a new binary relation on the arguments is also introduced in Dung's argumentation system : however, this new relation represents a notion of support between arguments. Clearly enough, this is unrelated with the relation introduced here representing the ignorance about the attack between arguments.

Definition 40 (Completion of a PAF) *Let* $\mathsf{PAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{I} \rangle$*. Let* $\mathsf{AF} = \langle \mathcal{A}, S \rangle$*.* AF *is a* completion *of* PAF *if and only if* $\mathcal{R} \subseteq S \subseteq \mathcal{R} \cup \mathcal{I}$*.*

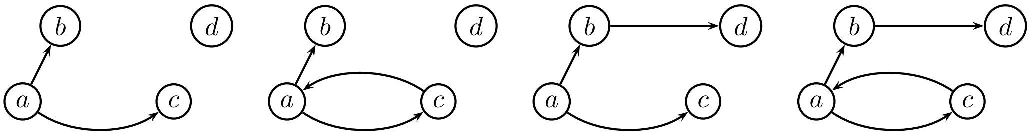
The set of all completions of PAF *is denoted* $\mathcal{C}(\mathsf{PAF})$*.*

Example 14 *The partial argumentation system* $\mathsf{PAF} = \langle \mathcal{A} = \{a, b, c, d\}$*,* $\mathcal{R} = \{(a, b), (a, c)\}$*,* $\mathcal{I} = \{(c, a), (b, d)\}$*,* $\mathcal{N} = \{(a, a), (b, b), (c, c), (d, d), (b, a), (b, c), (c, b), (a, d), (d, a), (d, b), (c, d), (d, c)\}\rangle$ *is illustrated on the following figure (solid arrows represent the attack relation and dotted arrows represent the ignorance relation; non-attack relations are not represented explicitly as in the AF case):*

PAF



The completions of this PAF are:



Now, **Problem 1** can be addressed by first associating each argumentation system AF_i with a corresponding PAF_i so that all PAF_i are about the same set of arguments $\bigcup_{i=1}^{n} \mathcal{A}_i$. To this end, we introduce the notion of *expansion* of an AF:

Definition 41 (Expansion of an AF) *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of* n *AFs such that* $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i, \mathcal{N}_i \rangle$*. Let* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ *be an argumentation system. An* expansion *of* AF *given* $\mathcal{P}$ *is any PAF* $\mathtt{exp}(\mathsf{AF}, \mathcal{P})$ *defined by* $\langle \mathcal{A} \cup \bigcup_i \mathcal{A}_i, \mathcal{R}', \mathcal{I}', \mathcal{N}' \rangle$ *such that* $\mathcal{R} \subseteq \mathcal{R}'$ *and* $(\mathcal{A} \times \mathcal{A}) \setminus \mathcal{R} \subseteq \mathcal{N}'$*.* $\mathtt{exp}$ *is referred to as an* expansion function*.*

In order to be general enough, this definition does not impose many constraints on the resulting PAF: what is important is to preserve the attack and non-attack relations from the initial AF while extending its set of arguments. Many policies can be used to give rise to expansions of different kinds, reflecting the various attitudes of agents in light of "new" arguments; for instance, if a is any argument considered by Agent i at the start and a "new" argument b has to be incorporated, Agent i can (among other things):

- always reject b (e.g., adding (b, b) to her relation $\mathcal{R}'_i$),
- always accept b (adding (a, b), (b, a) and (b, b) to her non-attack relation $\mathcal{N}'_i$),
- just express her ignorance about b (adding (a, b), (b, a) and (b, b) to her ignorance relation $\mathcal{I}'_i$).

Each agent may also compute the exact interaction between a and b when the attack relation is not primitive but defined from more basic notions (as in the approach by Elvang-Gøransson et al., see e.g., [EGFK93a, EGFK93b, EGH95]). Note that if she has limited computational resources, Agent i can compute exact interactions as far as she can, then express ignorance for the remaining ones.

In the following, we specifically focus on *consensual expansions*. Intuitively, the consensual expansion of an argumentation system $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ given a profile of such systems is obtained by adding a pair of arguments (a, b) (where at least one of a, b is not in $\mathcal{A}$) into the attack (resp. the non-attack relation) *provided that all other agents of the profile* *who know the two arguments agree on the existence of the attack*[23] (resp. *the non-attack*); otherwise, it is added to the ignorance relation.

This expansion policy is sensible as soon as each agent has a minimum level of confidence in the other agents: if a piece of information conveyed by one agent is not conflicting with the information stemming from the other agents, every agent of the group is ready to accept it.

Definition 42 (Consensual expansion) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ be a profile of n AFs such that $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle$. Let $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$ be an argumentation system. Let $conf(\mathcal{P}) = (\bigcup_i \mathcal{R}_i) \cap (\bigcup_i \mathcal{N}_i)$ be the set of interactions for which a conflict exists within the profile. The* consensual expansion *of* AF *over $\mathcal{P}$ is the tuple denoted by $\mathtt{exp}_C$ = $\langle \mathcal{A}', \mathcal{R}', \mathcal{I}', \mathcal{N}' \rangle$ with:*

- $\mathcal{A}' = \mathcal{A} \cup \bigcup_i \mathcal{A}_i$,
- $\mathcal{R}' = \mathcal{R} \cup ((\bigcup_i \mathcal{R}_i \setminus conf(\mathcal{P})) \setminus \mathcal{N})$,
- $\mathcal{I}' = conf(\mathcal{P}) \setminus (\mathcal{R} \cup \mathcal{N})$,
- $\mathcal{N}' = (\mathcal{A}' \times \mathcal{A}') \setminus (\mathcal{R}' \cup \mathcal{I}')$.

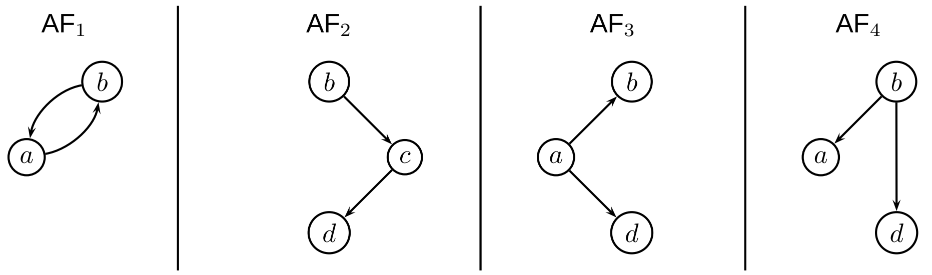
The next proposition states that, as expected, the consensual expansion of an argumentation system over a profile is an expansion:

Proposition 19 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ be a profile of n AFs such that $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle$. Let $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$ be an argumentation system. The consensual expansion $\mathtt{exp}_C$ of* AF *over $\mathcal{P}$ is an expansion of* AF *over $\mathcal{P}$ in the sense of Definition 41 on the facing page.*

The consensual expansion is among the most cautious expansions one can define since it leads to adding a pair of arguments in the attack (or the non-attack relation) associated with an agent only when all the other agents agree on it.

Example 15 *Consider the profile consisting of the following four argumentation systems:*

- $\mathsf{AF}_1 = \langle \mathcal{A}_1 = \{a, b\}, \mathcal{R}_1 = \{(a, b), (b, a)\} \rangle$,
- $\mathsf{AF}_2 = \langle \mathcal{A}_2 = \{b, c, d\}, \mathcal{R}_2 = \{(b, c), (c, d)\} \rangle$,
- $\mathsf{AF}_3 = \langle \mathcal{A}_3 = \{a, b, d\}, \mathcal{R}_3 = \{(a, b), (a, d)\} \rangle$,
- $\mathsf{AF}_4 = \langle \mathcal{A}_4 = \{a, b, d\}, \mathcal{R}_4 = \{(b, d), (b, a)\} \rangle$.

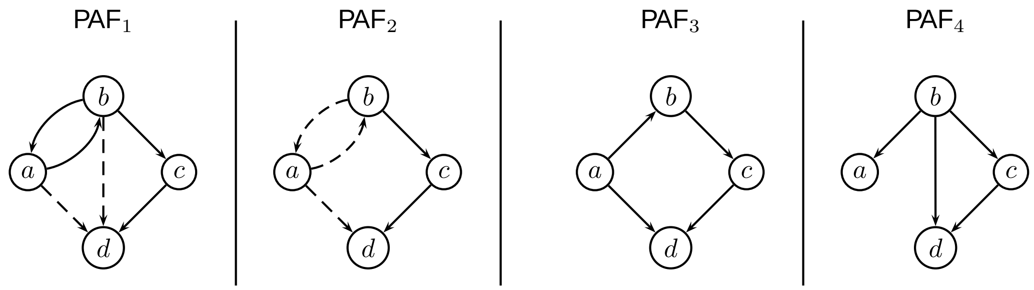


For each i, the consensual expansion PAF_i of AF_i is given by:

- $\mathsf{PAF}_1 = \langle \quad \{a, b, c, d\}, \quad \{(a, b), (b, a), (b, c), (c, d)\}, \quad \{(a, d), (b, d)\} \quad \rangle$,
- $\mathsf{PAF}_2 = \langle \quad \{a, b, c, d\}, \quad \{(b, c), (c, d)\}, \quad \{(a, b), (b, a), (a, d)\} \quad \rangle$,
- $\mathsf{PAF}_3 = \langle \quad \{a, b, c, d\}, \quad \{(a, b), (a, d), (b, c), (c, d)\}, \quad \varnothing \quad \rangle$,

[23] *i.e.*, if $a, b \in \mathcal{A}_i$, then $(a, b) \in \mathcal{R}_i$.

- $\mathsf{PAF}_4 = \langle \quad \{a, b, c, d\}, \quad \{(b, d), (b, a), (b, c), (c, d)\}, \quad \varnothing \quad \rangle$.



When the expansion policies considered by each agent are the same one $\mathtt{exp}$, for any profile $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ we shall often note $\mathtt{exp}(\mathcal{P})$ the profile of PAFs $\langle \mathtt{exp}(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}(\mathsf{AF}_n, \mathcal{P}) \rangle$.

3.2.4.3 Merging Operators

In order to deal with **Problem 2**, we propose to merge interactions instead of sets of acceptable arguments. The goal is to characterize the argumentation systems which are as close as possible to the given profile of argumentation systems, taken as a whole.

A way to achieve this consists in defining a notion of "distance" between an AF and a profile of AFs, or more generally between a PAF and a profile of PAFs. This calls for a notion of pseudo-distance between two PAFs, and a way to combine such pseudo-distances:

Definition 43 (Pseudo-distance) *A* pseudo-distance d *between PAFs over $\mathcal{A}$ is a mapping which associates a non-negative real number to each pair of PAFs over $\mathcal{A}$ and satisfies the properties of symmetry ($d(x, y) = d(y, x)$) and minimality ($d(x, y) = 0$ if and only if $x = y$).*

d is a distance if it satisfies also the triangular inequality ($d(x, z) \leq d(x, y) + d(y, z)$).

Definition 44 (Aggregation function) *An* aggregation function *is a mapping $\otimes$ from $(\mathbb{R}+)^n$ to $(\mathbb{R}+)$ (strictly speaking, it is a family of mappings, one for each n), that satisfies*

- *if $x_i \geq x'_i$, then $\otimes(x_1, \ldots, x_i, \ldots, x_n) \geq \otimes(x_1, \ldots, x'_i, \ldots, x_n)$* **(non-decreasingness)**
- *$\otimes(x_1, \ldots, x_n) = 0$ if $\forall i$, $x_i = 0$* **(minimality)**
- *$\otimes(x) = x$* **(identity)**

The merging of a profile of AFs is defined as a set of AFs:

Definition 45 (Merging of n AFs) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of n AFs. Let d be any pseudo-distance between PAFs, let $\otimes$ be an aggregation function, and let $\mathtt{exp}_1, \ldots, \mathtt{exp}_n$ be n expansion functions. The* merging *of $\mathcal{P}$ is the set of AFs*

$$\Delta_d^{\otimes}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle, \langle \mathtt{exp}_1, \ldots, \mathtt{exp}_n \rangle) =$$

$$\{\mathsf{AF} \text{ over } \bigcup_i \mathcal{A}_i \mid \mathsf{AF} \text{ minimizes } \otimes_{i=1}^n d(\mathsf{AF}, \mathtt{exp}_i(\mathsf{AF}_i, \mathcal{P}))\}.$$

In order to avoid heavy notations, we shall sometimes identify the resulting set of AFs $\{\mathsf{AF}'_1, \ldots, \mathsf{AF}'_k\}$ with the profile $\langle \mathsf{AF}'_1, \ldots, \mathsf{AF}'_k \rangle$ (or any other permutation of it).

Thus, merging a profile of AFs $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ is a two-step process:

expansion: An expansion of each AF_i over $\mathcal{P}$ is first computed. Note that considering expansion functions specific to each agent is possible. What is important is that $\mathtt{exp}_i(\mathsf{AF}_i, \mathcal{P})$ is a PAF over $\mathcal{A} = \bigcup_i \mathcal{A}_i$.

fusion: The AFs over $\mathcal{A}$ that are selected as the result of the merging process are the ones that best represent $\mathcal{P}$ (*i.e.*, that are the "closest" to $\mathcal{P}$ w.r.t. the aggregated distances).

In the following, we assume that each agent uses consensual expansion. In order to lighten the notations, we remove $\langle \mathtt{exp}_1, \ldots, \mathtt{exp}_n \rangle$ from the list of parameters of merging operators.

Note that it would be possible to refine Definition 45 on the preceding page so as to include integrity constraints into the picture. This can be useful if there exists some (unquestionable) knowledge about the expected result (some attacks between arguments which have to hold for the group). It is then enough to look only to the AFs which satisfy the constraints, similarly to what is done in propositional belief base merging (see e.g., [KP02]). In contrast to the belief base merging scenario, constraints on the *structure* of the candidate AFs can also be set. In particular, considering only acyclic AFs can prove valuable since (1) such AFs are well-founded, (which implies that only one extension has to be considered whatever the underlying semantics – among Dung's ones), and (2) this extension (which turns out to be the grounded one, see [Dun95]) can be computed in time polynomial in the size of the AF (while computing a single extension is intractable for the other semantics in the general case – under the standard assumptions of complexity theory – see [DBC02]).

Now, many pseudo-distances between PAFs and many aggregation functions can be used, giving rise to many merging operators. Usual aggregation functions include the sum Σ, the max $\mathcal{M}\mathtt{ax}$ and the leximax $\mathcal{L}\mathtt{eximax}$[24] but using non-symmetric functions is also possible (this may be particularly valuable if some agents are more important than others). Some aggregation functions (like the sum) enable the merging process to take into account the number of agents believing that an argument attacks or not another argument:

Example 13 on page 63 (cont'd) *Two agents over three agree with the fact that e attacks f and f does not attack e. It may prove sensible that the group agrees with the majority.*

The choice of the aggregation function is very important for tuning the operator behaviour with the expected one. For example, sum is a possible choice in order to solve conflicts using majority. Otherwise, the leximax function can prove more valuable if the aim is to behave in a more consensual way, trying to define a result close to the AF of each agent of the group. The distinction between majority and arbitration operators as considered in propositional belief base merging [KP02] also applies here.

In the following, we focus on the *edit distance* between PAFs:

Definition 46 (Edit distance) *Let $\mathsf{PAF}_1 = \langle \mathcal{A}, \mathcal{R}_1, \mathcal{I}_1, \mathcal{N}_1 \rangle$ and $\mathsf{PAF}_2 = \langle \mathcal{A}, \mathcal{R}_2, \mathcal{I}_2, \mathcal{N}_2 \rangle$ be two PAFs over $\mathcal{A}$.*

- *Let a, b be two arguments $\in \mathcal{A}$. The* edit distance between PAF_1 and PAF_2 over a, b *is the mapping $de_{a,b}$ such that:*
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 0$ *if and only if* $(a, b) \in \mathcal{R}_1 \cap \mathcal{R}_2$ *or* $\mathcal{I}_1 \cap \mathcal{I}_2$ *or* $\mathcal{N}_1 \cap \mathcal{N}_2$*,*
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 1$ *if and only if* $(a, b) \in \mathcal{R}_1 \cap \mathcal{N}_2$ *or* $\mathcal{N}_1 \cap \mathcal{R}_2$*,*
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 0.5$ *otherwise.*
- *The* edit distance between PAF_1 and PAF_2 *is given by*

$$de(\mathsf{PAF1}, \mathsf{PAF2}) = \Sigma_{(a,b) \in \mathcal{A} \times \mathcal{A}} de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}).$$

The edit distance between two PAFs is the (minimum) number of additions/deletions which must be made to render them identical. Ignorance is treated as halfway between attack and non-attack.

It is easy to show that:

[24] When applied to a vector of n real numbers, the $\mathcal{L}\mathtt{eximax}$ function gives the list of those numbers sorted in a decreasing way. Such lists are compared w.r.t. the lexicographic ordering induced by the standard ordering on real numbers.

Proposition 20 *The edit distance de between PAFs is a distance.*

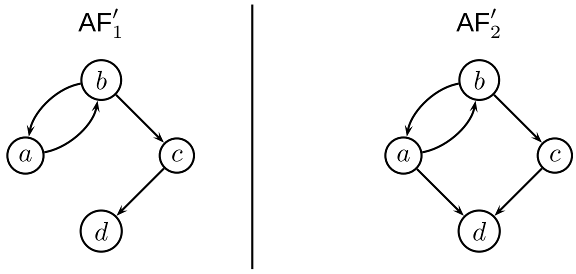
Let us now illustrate the notion of edit distance as well some associated merging operators on Example 15 on page 67.

Example 15 on page 67 (cont'd) *We consider the following argumentation system* $\mathsf{AF}'_1 = \langle\{a, b, c, d\}, \{(a, b), (b, a), (b, c), (c, d)\}\rangle$.
The edit distance between AF'_1 *and each of the PAFs* PAF_1, PAF_2, PAF_3, PAF_4 *obtained by consensual expansion from the profile* $\langle\mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4\rangle$ *is:*

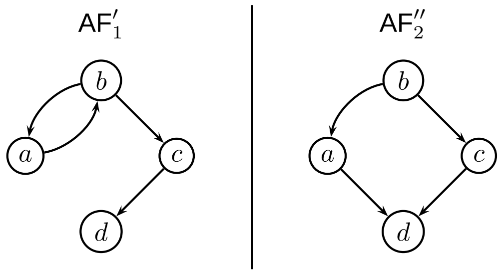
- $de(\mathsf{AF}'_1, \mathsf{PAF}_1) = 1$,
- $de(\mathsf{AF}'_1, \mathsf{PAF}_2) = 1.5$,
- $de(\mathsf{AF}'_1, \mathsf{PAF}_3) = 2$,
- $de(\mathsf{AF}'_1, \mathsf{PAF}_4) = 2$.

Taking the sum as the aggregation function, we obtain: $\Sigma_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 6.5$.
Taking the max, we obtain: $\mathcal{M}\mathtt{ax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 2$.
Taking the leximax, we obtain: $\mathcal{L}\mathtt{eximax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = (2, 2, 1.5, 1)$.
By computing such distances for all candidate AFs (i.e., all AFs over $\{a, b, c, d\}$*), we can compute the result of the merging;*
$\Delta_{de}^{\Sigma}(\langle\mathsf{AF}_1, \ldots, \mathsf{AF}_4\rangle)$ *is the set containing the two following AFs:*

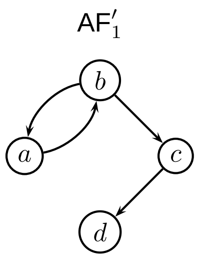
- $\mathsf{AF}'_1 = \langle\{a, b, c, d\}, \{(a, b), (b, a), (b, c), (c, d)\}\rangle$,
- $\mathsf{AF}'_2 = \langle\{a, b, c, d\}, \{(a, b), (b, a), (b, c), (a, d), (c, d)\}\rangle$.



$\Delta_{de}^{\mathcal{M}\mathtt{ax}}(\langle\mathsf{AF}_1, \ldots, \mathsf{AF}_4\rangle)$ *is the set containing* AF'_1 *and* $\mathsf{AF}''_2 = \langle\{a, b, c, d\}, \{(b, a), (b, c), (a, d), (c, d)\}\rangle$.



$\Delta_{de}^{\mathcal{L}\mathtt{eximax}}(\langle\mathsf{AF}_1, \ldots, \mathsf{AF}_4\rangle)$ *is the singleton containing* AF'_1.



The discrepancies between the merging obtained with the various aggregation operators can be explained in the following way:

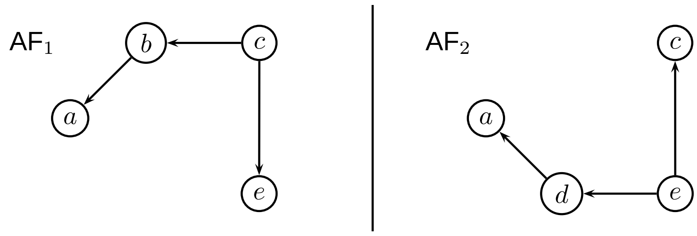
- AF'_1 *is the most consensual AF obtained as it is almost equidistant from each PAF;*
- AF'_2 *is much closer to* PAF_1, PAF_2 *and* PAF_3 *than to* PAF_4*, thus it is selected with the sum as an aggregation operator but it is too far from* PAF_4 *for being selected with the* $\mathcal{M}\mathtt{ax}$ *or* $\mathcal{L}\mathtt{eximax}$ *operators.*
- AF''_2 *is nearly equidistant from all four PAFs of the profile but less consensual than* AF'_1*, thus it is selected neither with* Σ *nor with* $\mathcal{L}\mathtt{eximax}$ *but only with* $\mathcal{M}\mathtt{ax}$ *as it is not far from any of the given PAFs.*

Having AF'_1 *in all mergings - whatever the aggregation function chosen - seems very intuitive. Indeed, whenever an attack (or a non-attack) is present in the (weak) majority of the initial AFs, it is also in* AF'_1*. This is not the case for the two others AFs belonging to the above mergings.*

Here is another simple example:

Example 16 *Consider the two following argumentation systems:*

- $\mathsf{AF}_1 = \langle\{a,b,c,e\}, \{(b,a),(c,b),(c,e)\}\rangle$
- $\mathsf{AF}_2 = \langle\{a,d,e,c\}, \{(d,a),(e,d),(e,c)\}\rangle$



Note that the attack from c to e is known by Agent 1 but not by Agent 2 and the attack from e to c is known by Agent 2 but not by Agent 1. This illustrates the fact that the agents do not share the same attack relation.

AF_1 *has a unique preferred extension:* $\{c,a\}$*.* AF_2 *has a unique preferred extension:* $\{e,a\}$*.*

The consensual expansions of AF_1 *and* AF_2 *are respectively:*

- $\mathsf{PAF}_1 = \langle\{a,b,c,d,e\}, \{(b,a),(c,b),(c,e),(d,a),(e,d)\}, \varnothing\rangle$,
- $\mathsf{PAF}_2 = \langle\{a,b,c,d,e\}, \{(d,a),(e,d),(e,c),(b,a),(c,b)\}, \varnothing\rangle$.

The result of merging the profile $\langle\mathsf{AF}_1,\mathsf{AF}_2\rangle$ *with* de *and* $\otimes = \mathcal{M}\mathtt{ax}$ *(or* $\otimes = \mathcal{L}\mathtt{eximax}$*) is:*

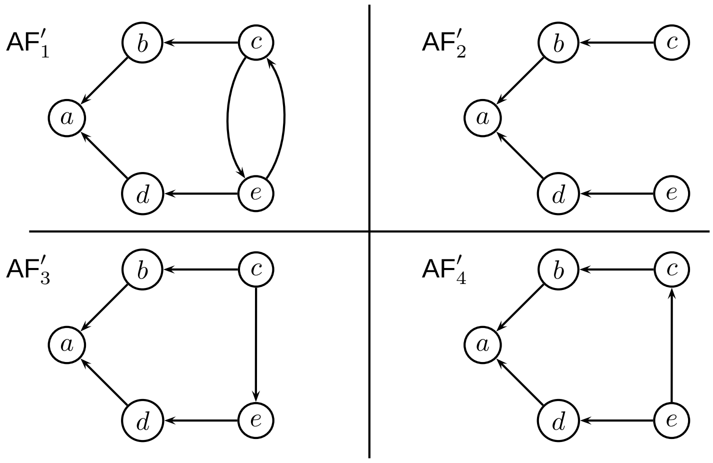
$$\Delta_{de}^{\mathcal{M}\mathtt{ax}}(\langle\mathsf{AF}_1,\mathsf{AF}_2\rangle) = \Delta_{de}^{\mathcal{L}\mathtt{eximax}}(\langle\mathsf{AF}_1,\mathsf{AF}_2\rangle) = \{\mathsf{AF}'_1,\mathsf{AF}'_2\} \textit{ with}$$

- $\mathsf{AF}'_1 = \langle\{a,b,c,d,e\},\{(b,a),(c,b),(c,e),(d,a),(e,d),(e,c)\}\rangle$,
- $\mathsf{AF}'_2 = \langle\{a,b,c,d,e\},\{(b,a),(c,b),(d,a),(e,d)\}\rangle$.

Using the sum as an aggregation function, two additional AFs are generated:

$$\Delta_{de}^{\Sigma}(\langle\mathsf{AF}_1,\mathsf{AF}_2\rangle) = \{\mathsf{AF}'_1,\mathsf{AF}'_2,\mathsf{AF}'_3,\mathsf{AF}'_4\}, \textit{ with}$$

- $\mathsf{AF}'_3 = \langle\{a,b,c,d,e\},\{(b,a),(c,b),(c,e),(e,d),(d,a)\}\rangle$,
- $\mathsf{AF}'_4 = \langle\{a,b,c,d,e\},\{(b,a),(c,b),(e,c),(e,d),(d,a)\}\rangle$.



Each of the resulting mergings contains an argumentation system from which argument a can be derived, as it is the case in AF_1 and AF_2. Using the sum as an aggregation function leads to the most consensual result here since it preserves the initial AFs of the different agents. Indeed, AF'_3 is equivalent to PAF_1 and AF'_4 is equivalent to PAF_2.

3.2.4.4 Some Properties

Let us now present some properties of consensual expansions and merging operators based on the edit distance, showing them as interesting choices.

3.2.4.4.1 Properties of PAFs and consensual expansions

Intuitively speaking, a natural requirement on any AF resulting from a merging is that it preserves all the information which are shared by the agents participating in the merging process, and more generally, all the information on which the agents participating in the merging process do not disagree.

In order to show that our merging operators satisfy those requirements, one first needs the notions of clash-free part and of common part of a profile of PAFs:

Definition 47 (Clash-free part of a profile of PAFs) *Let $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ be a profile of PAFs. The* clash-free part *of $\mathcal{P}$ is denoted by $CFP(\mathcal{P})$ and is defined by:*

$$CFP(\mathcal{P}) = \langle \bigcup_i \mathcal{A}_i, \bigcup_i \mathcal{R}_i \setminus \bigcup_i \mathcal{N}_i, \mathcal{I}_{CFP}, \bigcup_i \mathcal{N}_i \setminus \bigcup_i \mathcal{R}_i \rangle$$

where $\mathcal{I}_{CFP} = (\bigcup_i \mathcal{A}_i \times \bigcup_i \mathcal{A}_i) \setminus ((\bigcup_i \mathcal{R}_i \setminus \bigcup_i \mathcal{N}_i) \cup (\bigcup_i \mathcal{N}_i \setminus \bigcup_i \mathcal{R}_i))$.

The clash-free part of a profile of PAFs represents the pieces of information (attack / non-attack) that are not questioned by any other agent. As they are not the source of any disagreement, they are expected to be included in each AF resulting from the merging process.

Example 15 on page 67 (cont'd) *With $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$,*

$$CFP(\mathcal{P}) = \quad \langle \{a, b, c, d\}, \quad \{(b, c), (c, d)\} \quad \{(a, b), (b, a), (a, d), (b, d), (a, c), (c, a)\} \rangle.$$

Note that with $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathsf{exp}_C(\mathsf{AF}_4, \mathcal{P}) \rangle$, $CFP(\mathsf{exp}_C(\mathcal{P})) = \langle \{a, b, c, d\}, \{(b, c), (c, d)\}, \{(a, b), (b, a), (a, d), (b, d)\} \rangle$ (now (a, c) and (c, a) are non-attacks); so $CFP(\mathcal{P}) \neq CFP(\mathsf{exp}_C(\mathcal{P}))$.

Definition 48 (Common part of a profile of PAFs) *Let $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ be a profile of PAFs. The* common part *of $\mathcal{P}$ is denoted by $CP(\mathcal{P})$ and is defined by: $CP(\mathcal{P}) = \langle \bigcap_i \mathcal{A}_i, \bigcap_i \mathcal{R}_i, \bigcap_i \mathcal{I}_i, \bigcap_i \mathcal{N}_i \rangle$.*

The common part of a profile of PAFs is a much more demanding notion than the clash-free one. It represents the pieces of information on which all the agents agree. There is no doubt that those pieces of information must hold in any consensual view of the group's opinion, so the common part of the profile must be included in each AF of the result of the merging process.

Example 15 on page 67 (cont'd) *With* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$, $CP(\mathcal{P}) = \langle \{b\}, \varnothing, \varnothing, \{(b,b)\} \rangle$.

We have the following easy property:

Proposition 21 *Let* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *be a profile of PAFs. The* common part *of* $\mathcal{P}$ *is pointwise included into the clash-free part of* $\mathcal{P}$*, i.e.:*

- $\bigcap_i \mathcal{R}_i \subseteq \bigcup_i \mathcal{R}_i \setminus \bigcup_i \mathcal{N}_i$;
- $\bigcap_i \mathcal{I}_i \subseteq \mathcal{I}_{CFP}$;
- $\bigcap_i \mathcal{N}_i \subseteq \bigcup_i \mathcal{N}_i \setminus \bigcup_i \mathcal{R}_i$.

The common part of a profile of n PAFs (resp. AFs) is not always a PAF (resp. an AF). Contrastingly, the clash-free part of a profile of n PAFs is a PAF (however, the clash-free part of a profile of n AFs is not always an AF).

There exists an interesting particular case: if the various PAFs of the profile are based on the same set of arguments and if for each ordered pair of arguments (a, b) such that (a, b) belongs to the ignorance relation in one PAF, this pair belongs to the attack relation for another PAF of the profile and to the non-attack relation for at least a third PAF of the profile, then the clash-free part of the profile and its common part are identical:

Proposition 22 *Let* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *be a profile of* n *PAFs over the same set of arguments* $\mathcal{A}$*. Consider the clash-free part of* $\mathcal{P}$ *denoted by* $CFP(\mathcal{P}) = \langle \mathcal{A}_{CFP}, \mathcal{R}_{CFP}, \mathcal{I}_{CFP}, \mathcal{N}_{CFP} \rangle$ *and the common part of* $\mathcal{P}$ *denoted by* $CP(\mathcal{P}) = \langle \mathcal{A}_{CP}, \mathcal{R}_{CP}, \mathcal{I}_{CP}, \mathcal{N}_{CP} \rangle$*. If* $\bigcup_i \mathcal{I}_i \subseteq conf(\mathcal{P}) = (\bigcup_i \mathcal{R}_i) \cap (\bigcup_i \mathcal{N}_i)$*, we have:*

- $\mathcal{A}_{CFP} = \mathcal{A}_{CP}$,
- $\mathcal{R}_{CFP} = \mathcal{R}_{CP}$,
- $\mathcal{N}_{CFP} = \mathcal{N}_{CP}$.

This result is interesting since this situation always holds (by definition) if consensual expansion is used as an expansion policy by each agent.

Example 15 on page 67 (cont'd) *With* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$ *and* $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_2, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_3, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_4, \mathcal{P}) \rangle$*, we have:*

- $CFP(\mathsf{exp}_C(\mathcal{P})) = \langle \{a,b,c,d\}, \{(b,c),(c,d)\}, \{(a,b),(b,a),(a,d),(b,d)\}, \{(a,a), (b,b), (c,c), (d,d), (a,c), (c,a), (d,a), (d,b), (d,c), (c,b)\} \rangle$
- $CP(\mathsf{exp}_C(\mathcal{P})) = \langle \{a,b,c,d\}, \{(b,c),(c,d)\}, \varnothing, \{(a,a),(b,b),(c,c),(d,d),(a,c),(c,a),(d,a),(d,b),(d,c),(c,b)\} \rangle$.

A valuable property of any consensual expansion over a profile of AFs is that it preserves the clash-free part of the profile:

Proposition 23 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs. For each* i*, we have:*

- $\mathcal{A}_{CFP(\mathcal{P})} = \mathcal{A}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$,
- $\mathcal{R}_{CFP(\mathcal{P})} \subseteq \mathcal{R}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$,
- $\mathcal{N}_{CFP(\mathcal{P})} \subseteq \mathcal{N}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$.

Now, concordance between AFs can be defined as follows:

Definition 49 (Concordance) *Let* $\mathsf{AF}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$, $\mathsf{AF}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$ *be two AFs.* AF_1, AF_2 *are said to be* concordant *if and only if* $\forall (a,b) \in (\mathcal{A}_1 \cap \mathcal{A}_2) \times (\mathcal{A}_1 \cap \mathcal{A}_2)$, $(a,b) \in \mathcal{R}_1$ *if and only if* $(a,b) \in \mathcal{R}_2$. *Otherwise they are said to be* discordant.
Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs.* $\mathcal{P}$ *is said to be* concordant *if and only if all its AFs are pairwise concordant. Otherwise it is said to be* discordant.

Of course, concordance is related to the set $conf(\mathcal{P})$ representing clashes between attack and non-attack relations in the different AFs of the profile:

Proposition 24 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of argumentation systems.* $\mathcal{P}$ *is concordant if and only if* $conf(\mathcal{P}) = \bigcup_i \mathcal{R}_i \cap \bigcup_i \mathcal{N}_i$ *is empty.*

When a profile of AFs is concordant, its clash-free part is the union of its elements, and the converse also holds:

Proposition 25 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs.* $\mathcal{P}$ *is concordant if and only if* $CFP(\mathcal{P}) = \bigcup_i \mathsf{AF}_i$.

Proposition 26 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs.* $\mathcal{P}$ *is concordant if and only if*

$$\mathtt{exp}_C(\mathcal{P}) = \langle \mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P}) \rangle \text{ is reduced to } \langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle$$

(i.e., each of the n *elements of the vector is* $\bigcup_i \mathsf{AF}_i$*).*

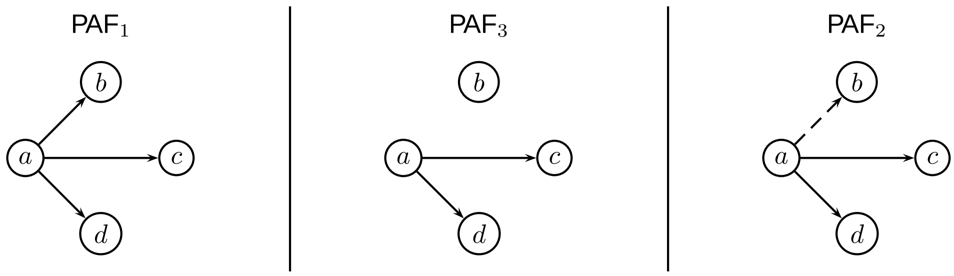
Note that $\bigcup_i \mathsf{AF}_i$ may appear into $\mathtt{exp}_C(\mathcal{P})$, even if $\mathcal{P}$ is discordant. This is illustrated by the following example:

Example 17 *Consider the profile* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *consisting of the following three AFs:*

- $\mathsf{AF}_1 = \langle \{a,b,c\}, \{(a,b),(a,c)\} \rangle$,
- $\mathsf{AF}_2 = \langle \{a,b,c\}, \{(a,c)\} \rangle$,
- $\mathsf{AF}_3 = \langle \{a,d\}, \{(a,d)\} \rangle$.

The profile $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *is discordant and* $\mathtt{exp}_C(\mathcal{P}) = \langle \mathsf{PAF}_1, \mathsf{PAF}_2, \mathsf{PAF}_3 \rangle$ *is such that:*

- $\mathsf{PAF}_1 = \langle \{a,b,c,d\}, \{(a,b),(a,c),(a,d)\}, \varnothing \rangle$ *(*$= \bigcup_i \mathsf{AF}_i$*),*
- $\mathsf{PAF}_2 = \langle \{a,b,c,d\}, \{(a,c),(a,d)\}, \varnothing \rangle$,
- $\mathsf{PAF}_3 = \langle \{a,b,c,d\}, \{(a,c),(a,d)\}, \{(a,b)\} \rangle$.



The following proposition states that whenever the presence of an attack (a,b) does not clash with a profile of AFs, such an attack is present in all the corresponding PAFs obtained by consensual expansion if and only if it is present in one of the input AFs.

Proposition 27 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs. Let* (a,b) *be a pair of arguments such that* $a, b \in \bigcup_i \mathcal{A}_i$ *and* $\nexists \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$ *such that* $(a,b) \in (\mathcal{R}_i \setminus \mathcal{R}_j) \cup (\mathcal{R}_j \setminus \mathcal{R}_i)$.
$\exists \mathsf{AF}_l \in \mathcal{P}$ *such that* $(a,b) \in \mathcal{R}_l$ *if and only if* $\forall \mathsf{AF}_k \in \mathcal{P}$, $(a,b) \in \mathcal{R}'_k$ *with* $\mathcal{R}'_k$ *denoting the attack relation of the PAF* $\mathtt{exp}_C(\mathsf{AF}_k, \mathcal{P})$.

A notion of compatibility of a profile of PAFs over the same set of arguments can also be defined:

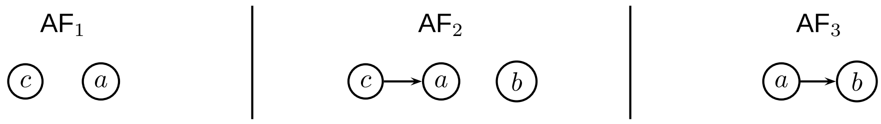
Definition 50 (Compatibility) *Let $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ be a profile of PAFs over a set of arguments $\mathcal{A}$. PAF_1, ..., PAF_n are said to be* compatible *if and only if they have at least one common completion. Otherwise they are said to be* incompatible.

Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. Let $\mathtt{exp}$ *be an expansion function.* $\mathsf{AF}_1, \ldots, \mathsf{AF}_n$ *are said to be* compatible *given* $\mathtt{exp}$ *if and only if* $\mathtt{exp}(\mathsf{AF}_i, \mathcal{P})$, $\forall i = 1 \ldots n$, *are said to be compatible. Otherwise they are said to be* incompatible.

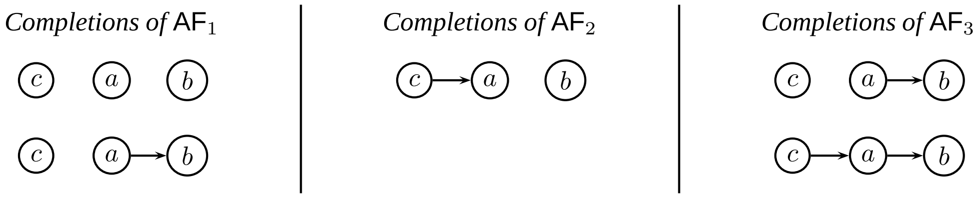
There is a clear link between concordance and compatibility in the case of the consensual expansion applied to a profile of AFs:

Proposition 28 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. $\mathcal{P}$ is concordant if and only if $\mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P})$ are compatible.*

Example 18 *Consider the following argumentation systems AF_1, AF_2 and AF_3.*



The completions of their respective consensual expansions PAF_1, PAF_2 and PAF_3 are:



AF_1 and AF_2 are discordant and incompatible given $\mathtt{exp}_C$. AF_3 and AF_1 are concordant and compatible given $\mathtt{exp}_C$.

3.2.4.4.2 Properties of merging operators

Let us now give some properties of merging operators, focusing on those based on the edit distance:

Proposition 29 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. Assume that the expansion function used for each agent is the consensual one. If $\mathcal{P}$ is concordant then $\Delta^{\otimes}_{de}(\mathcal{P}) = \{\bigcup_i \mathsf{AF}_i\}$.*

Now we show an expected property: that the clash-free part of any profile $\mathcal{P}$ is included in each AF from the merging of $\mathcal{P}$ when the edit distance is used.

Proposition 30 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of argumentation systems. Assume that the expansion function used for each agent is the consensual one. For any aggregation function $\otimes$, we have that : $\forall \mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle)$:*

- $\mathcal{A}_{CFP(\mathcal{P})} \subseteq \mathcal{A}$,
- $\mathcal{R}_{CFP(\mathcal{P})} \subseteq \mathcal{R}$,
- $\mathcal{N}_{CFP(\mathcal{P})} \subseteq \mathcal{N}$.

As a direct corollary of Propositions 21 on page 73 and 30 on the previous page, we get that:

Corollary 1 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of argumentation systems. Assume that the expansion function used for each agent is the consensual one. For any aggregation function $\otimes$, we have that: $\forall \mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle)$:*

- $\mathcal{A}_{CP(\mathcal{P})} \subseteq \mathcal{A}$,
- $\mathcal{R}_{CP(\mathcal{P})} \subseteq \mathcal{R}$,
- $\mathcal{N}_{CP(\mathcal{P})} \subseteq \mathcal{N}$.

When sum is used as the aggregation function and all AFs are over the same set of arguments, the merging of a profile can be characterized in a concise way, thanks to the notion of majority graph. Intuitively the majority graph of a profile of AFs over the same set of arguments is the PAF obtained by applying the strict majority rule to decide whether a attacks b or not, for every ordered pair (a, b) of arguments. Whenever there is no strict majority, an ignorance edge is generated.

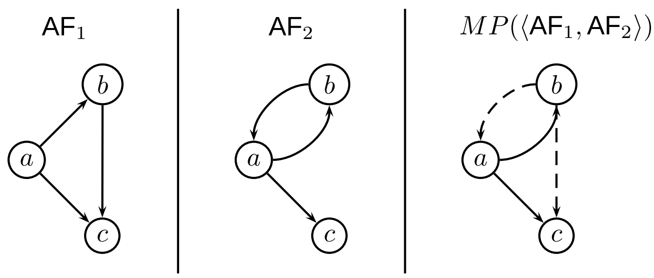
Definition 51 (Majority PAF) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set $\mathcal{A}$ of arguments. The* majority PAF *$MP(\mathcal{P})$ of $\mathcal{P}$ is the triple $\langle \mathcal{R}, \mathcal{N}, \mathcal{I} \rangle$ such that $\forall a, b \in \mathcal{A}$:*[25]

- $(a, b) \in \mathcal{R}$ *if and only if* $\#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{R}_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{N}_i\})$;
- $(a, b) \in \mathcal{N}$ *if and only if* $\#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{N}_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{R}_i\})$;
- $(a, b) \in \mathcal{I}$ *otherwise.*

The next proposition states that, as expected, the majority PAF of a profile of AFs over the same set of arguments is a PAF:

Proposition 31 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set $\mathcal{A}$ of arguments. The majority PAF $MP(\mathcal{P})$ of $\mathcal{P}$ is a PAF.*

Example 19 *Consider* $\mathsf{AF}_1 = \langle \{a, b, c\}, \{(a, b), (b, c), (a, c)\} \rangle$, $\mathsf{AF}_2 = \langle \{a, b, c\}, \{(a, b), (b, a), (a, c)\} \rangle$.



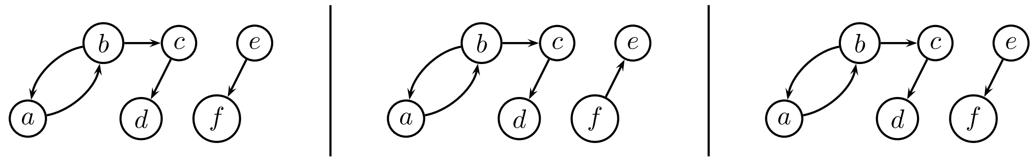
We have $MP(\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle) = \langle \{a, b, c\}, \{(a, b), (a, c)\}, \{(b, c), (b, a)\}, \{(a, a), (b, b), (c, c), (c, a), (c, b)\} \rangle$.

Proposition 32 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set $\mathcal{A}$ of arguments. $\Delta^{\Sigma}_{de}(\mathcal{P}) = \mathcal{C}(MP(\mathcal{P}))$.*

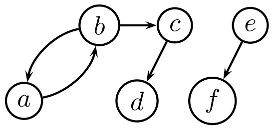
Let us illustrate the previous proposition on Example 13 on page 63:

Example 13 on page 63 (cont'd) *The consensual expansions of* AF_1, AF_2 *and* AF_3 *are respectively:*

[25] For any set S, $\#(S)$ denotes the cardinality of S.



So, the majority PAF of $\langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *is:*



Using the edit distance and sum as the aggregation function, this PAF also represents the result of the merging in the sense that the latter is the set of all completions of this PAF.

Computing the majority PAF of a profile of AFs over the same set of arguments amounts to *voting on the attack relations* associated to each AF. As explained in Section 3.2.4.1 on page 63, this can prove more suited to our goal than the approach which consists in voting directly on the acceptable sets of arguments for each agent. The previous proposition shows that such a simple voting approach corresponds to a specific merging operator in our framework (but many other operators, especially arbitration ones, can also be used).

3.2.4.5 Acceptability for Merged AFs

Starting from a profile of AFs (over possibly different sets of arguments), a merging operator enables the computation of a set of AFs (this time, over the same set of arguments) which are the best candidates to represent the AFs of the group (a kind of "consensus").

There is an important epistemic difference between those two sets of AFs, the first one reflects different points of view given by different agents (and it can be the case that two distinct agents give the same AF), while the second set expresses some uncertainty on the merging due to the presence of conflicts.

Let us recall that the main goal of this paper is to characterize the sets of arguments acceptable by the whole group of agents. In order to achieve it, it remains to define some mechanisms for exploiting the resulting set of AFs. This calls for a notion of joint acceptability.

Definition 52 (Joint acceptability) *A* joint acceptability relation *for a profile* $\langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ *of AFs, denoted by* $Acc_{\langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle}$*, is a total function from* $2^{\bigcup_i \mathcal{A}_i}$ *to* $\{true, false\}$ *which associates each subset* E *of* $\bigcup_i \mathcal{A}_i$ *with true if* E *is a jointly acceptable set for* $\langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ *and with false otherwise.*

For instance, a joint acceptability relation for a profile $\langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ can be defined by the acceptability relations Acc_{AF_i} (based themselves on some semantics and some selection principles), which can coincide for every AF_i (but this is not mandatory) and a voting method $V : \{true, false\}^n \mapsto \{true, false\}$:

$$Acc_{\langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle}(E) = V(Acc_{\mathsf{AF}_1}(E), \dots, Acc_{\mathsf{AF}_n}(E)).$$

Here are some instances of Definition 52 based on voting methods:

Definition 53 (Acceptabilities for profiles of AFs) *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \dots, \mathsf{AF}_n \rangle$ *be a profile of* n *AFs over the same set of arguments* $\mathcal{A}$*. Let* Acc_{AF_i} *be the (local) acceptability relation associated with* AF_i*. If* $n = 1$*, then we define* $Acc_{\langle \mathsf{AF}_1 \rangle} = Acc_{\mathsf{AF}_1}$*. Otherwise, for any subset* S *of* $\mathcal{A}$*, we say that:*

- S *is* skeptically jointly acceptable for $\mathcal{P}$ *if and only if* S *is included in at least one acceptable set for each* AF_i*:*
  $\forall \mathsf{AF}_i \in \mathcal{P}, \exists E_i$ *such that* $Acc_{\mathsf{AF}_i}(E_i)$ *is true and* $S \subseteq E_i$*.*

- S *is* credulously jointly acceptable for $\mathcal{P}$ *if and only if* S *is included in at least one acceptable set for at least one* AF_i*:*
$$\exists \mathsf{AF}_i \in \mathcal{P}, \exists E_i \text{ such that } Acc_{\mathsf{AF}_i}(E_i) \text{ is true and } S \subseteq E_i.$$
- S *is* jointly acceptable by majority for $\mathcal{P}$ *if and only if* S *is included in at least one acceptable set for at least a weak majority of* AF_i*:*
$$\#(\{\mathsf{AF}_i \mid \exists E_i \text{ such that } Acc_{\mathsf{AF}_i}(E_i) \text{ is true and } S \subseteq E_i\}) \geq \frac{n}{2}.$$

Obviously enough, when none of the local acceptabilities Acc_{AF_i} is trivial (*i.e.*, equivalent to the constant function $false$) for the profile under consideration, we have that any set of arguments which is skeptically jointly acceptable is also jointly acceptable by majority, and that any set of arguments which is jointly acceptable by majority is also credulously jointly acceptable.

Note that skeptical (resp. credulous) joint acceptability does not require that the skeptical (resp. credulous) inference principle is at work for defining local acceptabilities Acc_{AF_i}, which remain unconstrained.

Focusing on the preferred semantics together with credulous local acceptabilities, let us consider again some previous examples:

Example 16 on page 71 (cont'd) *Using the edit distance and* $\otimes = \mathcal{L}\texttt{eximax}$ *(or* $\mathcal{M}\texttt{ax}$*) as the aggregation function, we get two AFs* AF'_1 *and* AF'_2 *in the merging.*

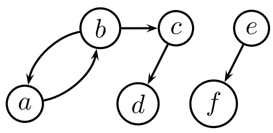
If the local acceptability relations are based on credulous inference from preferred extensions, we have:

- $Acc_{\mathsf{AF}'_1}(E) = $ *true if and only if* $E \subseteq \{c, d\}$ *or* $E \subseteq \{b, e\}$*;*
- $Acc_{\mathsf{AF}'_2}(E) = $ *true if and only if* $E \subseteq \{a, c, e\}$.

$\{c\}$ *and* $\{e\}$ *are skeptically jointly acceptable and* $\{b, e\}$,$\{c, d\}$ *and* $\{a, c, e\}$ *(and their subsets) are credulously (and by majority) jointly acceptable for the merging.*

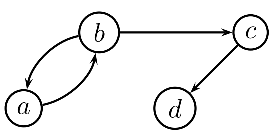
Using this method, the argument a *can still be derived credulously, contrariwise to what happens when the union of the two AFs* AF_1 *and* AF_2 *is considered.*

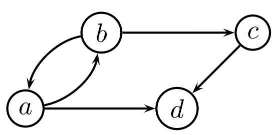
Example 13 on page 63 (cont'd) *Using the edit distance and the sum as the aggregation function, we get one AF in the merging, denoted* AF*:*



AF *has two preferred extensions :* $\{a, c, e\}$ *and* $\{b, d, e\}$*. So,* $Acc_{\mathsf{AF}}(E) = $ *true if and only if* $E \subseteq \{a, c, e\}$ *or* $E \subseteq \{b, d, e\}$*. The three joint acceptability relations coincide here (as there is only one AF in the result). The sets* $\{a, c, e\}$ *and* $\{b, d, e\}$ *(and their subsets) are credulously, skeptically and by majority, jointly acceptable for the merging, which is a more sensible result that the one obtained using a voting method on the derived arguments of the initial AFs (as explained in Section 3.2.4.1 on page 63).*

Example 15 on page 67 (cont'd) *Using the edit distance and the sum as the aggregation function, we get two AFs in the merging:*





The preferred extensions for these 2 AFs coincide (they are $\{a, c\}$ and $\{b, d\}$). As the preferred extensions for the 2 AFs are the same ones, the three relations of joint acceptability coincide here. Thus, the sets $\{a, c\}$ and $\{b, d\}$ (and their subsets) are skeptically, credulously and by majority jointly acceptable for the merging.

It is interesting to compare the joint acceptability relation for the input profile $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ with the joint acceptability relation for the merging $\Delta_d^{\otimes}(\mathcal{P})$. Unsurprisingly, both predicates are not logically connected (*i.e.*, none of them implies the other one), even in the case when the two joint acceptability relations are based on the same notion of local acceptability (for instance, considering a set of arguments E as acceptable for an AF when it is included in at least one of its preferred extensions) and the same voting method (for instance, the simple majority rule).

Thus, it can be the case that new jointly acceptable sets are obtained after merging while they were not jointly acceptable at start:

Proposition 33 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set of arguments $\mathcal{A}$. The set of all jointly acceptable sets for the profile $\mathcal{P}$ is not necessarily equal to the set of all jointly acceptable sets for the merging of $\mathcal{P}$.*

A counter-example is given by Example 15 on page 67.

When each local acceptability relation corresponds exactly to the collective acceptability proposed by Dung (for a given semantics and $\forall \mathsf{AF}_i$, $Acc_{\mathsf{AF}_i}(E) =$ *true* if and only if E is an extension of AF_i for this semantics), the following remarks can be done:

- If a set of arguments is included in *one* of the acceptable sets for an agent, it is not necessarily included into one of the acceptable sets of any AF from the merging (and it also holds for singletons). The converse is also true.
- More surprisingly, even if a set of arguments is included into *each* acceptable set for an agent, it is not guaranteed to be included into an acceptable set of an AF from the merging. Conversely, if a set of arguments is included into every acceptable set of the AFs from the merging, it is not guaranteed to be included into an acceptable set for one of the agents. Intuitively, this can be explained by the fact that if an argument is accepted by all agents *for bad reasons* (for instance, because they lack information about attacks on it), it can be rejected by the group after the merging. More formally, this is due to the fact that nothing ensures that one of the initial AFs will belong to the result of the merging and also to the fact that acceptability is nonmonotonic (in the sense that adding a single attack (a, b) in an AF may drastically change its extensions – see Section 3.2.5 on the next page).

3.2.4.6 Conclusion and perspectives on merging

We have presented a framework for deriving sensible information from a collection of argumentation systems *à la* Dung. Our approach consists in merging such systems. The proposed framework is general enough to allow for the representation of many different scenarios. It is not assumed that all agents must share the same sets of arguments. No assumption is made concerning the meaning of the attack relations, so that such relations may differ not only because agents have different points of view on the way arguments interact but more generally may disagree on what an interaction is. Each agent may be associated to a specific expansion function, which enables for encoding many attitudes when facing a new argument. Many different distances between PAFs and many different aggregation functions can be used to define argumentation systems which best represent the whole group.

By means of example, we have shown that our merging-based approach leads to results which are much more expected than those furnished by a direct vote on the (sets of) arguments acceptable by each agent. We have also shown that union cannot be taken as a valuable merging operator in the general case. We have investigated formally some properties of the merging operators which we point out. Among other results, we have shown that merging operators based on the edit distance preserve all the information on which all the agents participating in the merging process agree, and more generally, all the information on which the agents participating in the merging process do not disagree. We have also shown that the merging operator based on the edit distance and the sum as aggregation function is closely related to the merging approach which consists in voting on the attack relations when the input profile gathers argumentation systems over the same set of arguments. Finally, we have proved that in the general case, the derivable sets of arguments when joint acceptability concerns the input profile may drastically differ from the the derivable sets of arguments when joint acceptability concerns the profile obtained after the merging step.

This work can be continued in several directions:

Merging PAFs. Our framework can be extended to PAFs merging (instead of AFs). This enables us to take into account agents with incomplete belief states regarding the attack relation between arguments. Expansions of PAFs can be defined in a very similar way to expansions of AFs (what mainly changes is the way ignorance is handled). As PAFs are more expressive than AFs, an interesting issue for further research is to define acceptability for PAFs.

Attacks strengths. Assume that each attack believed by Agent i is associated to a numerical value reflecting the strength of the attack according to the agent, *i.e.*, the degree to which Agent i believes that a attacks b. It is easy to take into account those values by modifying slightly the definition of the edit distance over an ordered pair of arguments (for instance, viewing such values as weights once normalized within $[0, 1]$). Another possibility regarding attack strengths is, from unweighted attack relations, to generate a weighted one, representing different degrees of accordance in the group. For instance, each attack (a, b) in the majority PAF of a profile $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ can be labelled by the ratio $\frac{\#(\{i \in 1 \ldots n | (a,b) \in \mathcal{R}_i\})}{n}$ and similarly for the non-attack relation (this leads to consider both the attack and the non-attack relations of the majority PAF as fuzzy relations). Corresponding acceptability relations remain to be defined. This notion of gradual attack is already used in another context for argumentation (see [RMF$^+$08] for revision of argumentation systems).

Merging audiences. In [BCDD07], an extension of the notion of AF, called valued AF — VAF for short —, has been proposed in order to take advantage of values representing the agent's preferences in the context of a given audience. A further perspective of our work concerns the merging of such VAFs.

3.2.5 Revision in unipolar argumentation

When an agent receives a new piece of information, she must adapt its beliefs; this adaptation is not always easy because it may imply to drop some previous knowledge. Choosing the better way to adapt itself to its environment is a very old problem for human being, this is, perhaps, a reason why belief change theory has been so largely studied in the artificial intelligence community. The seminal work of Alchourrón, Gärdenfors and Makinson (AGM) [AGM85] has settled a formal framework for reasoning about belief change and introduced the concept of "belief revision". Belief revision consists in answering the question of what remains of the old beliefs after the arrival of a new piece of information. In this paper, we transpose this question into argumentation theory, and study the case of the arrival of a new argument into an argumentation system.

When a new argument is added to a set of arguments together with its interactions with the initial set of arguments, the outcome of the argumentation system may change. In this paper, we study the impact of this addition on the set of initial extensions. This leads us to characterize the possible revision operations with respect to the change they induce on the outcome. This study has two main applications, the first one concerns the computation, while the second one belongs to the field of dialogue strategies. On the first hand, the interest for computational processing is that knowledge about the kind of revision that is done may help to deduce what are the changes in the extensions. For instance, knowing that the revision is conservative allows us to deduce that the revision will not change the previous extensions. On the other hand, knowing the impact of adding an argument may help choosing the good one in order to achieve a given goal. For instance, in order to make a dialogue more open, an argument inducing an "expansive revision"[26] must be added (see Section 3.2.5.5 on page 89).

This section is organized as follows. Subsection 3.2.5.1 on the facing page recalls the basic concepts in revision theory. Subsection 3.2.5.2 on page 82 settles a definition of revision in argumentation. In this section, *we restrict our study to the case of adding one argument having only one interaction with an initial argument*. So, the research reported here is a first step towards a study of general revision operators. A typology of revision in argumentation is proposed, based on the impact of the revision under the set of extensions. A particular property for the revision operator is to keep the added argument in each extension. It is called "classical" and the cases when the revision operator is classical are

[26] The precise definition of this notion is given in Definition 57 on page 88.

described in Subsection 3.2.5.3 on page 85. Subsection 3.2.5.4 on page 87 is dedicated to the study of two particular revision operators, namely the "decisive" and the "expansive" revision operators. A last section discusses the related approaches in the literature.

This work has been done with Claudette CAYROL and Florence DUPIN DE SAINT-CYR. All the proofs and several important lemmas are given in [CdSCLS08].

3.2.5.1 Basic concepts in revision theory

In the field of belief change theory, the paper of AGM [AGM85] has introduced the concept of "belief revision". Belief revision consists in answering the question of what remains of the old beliefs after the arrival of a new piece of information. Beliefs are represented by sentences of a formal language. AGM have defined three types of belief change, namely contraction, expansion and revision. Expansion consists only in adding information without checking its consistency with previous beliefs. Contraction is an operation designed for removing information. Revision consists in adding information while preserving consistency. This last operation is the most interesting one since inconsistency leads to un-exploitable information. The main interest of AGM's work is the definition of a set of postulates which should hold for any rational revision operator. As noticed in [Som94] these postulates are founded on three principles:

- a consistency principle (the result should be consistent),
- a minimum change principle (as few beliefs as possible should be modified),
- priority to the new piece of information principle (the new piece of information should hold after the revision process).

More formally, a revision operator associates to a set of deductively closed formulae K (encoding the initial beliefs[27]) and to a formula p (encoding a new piece of information), another set of beliefs denoted by $K * p$. In order to be "rational" the operator $*$ should satisfy the following AGM postulates:

K*1 $K * p = Th(K * p)$.

K*2 $p \in K * p$.

K*3 $K * p \subseteq Th(K \cup \{p\})$.

K*4 If $\neg p \notin K$, then $Th(K \cup \{p\}) \subseteq K * p$.

K*5 $\bot \in K * p$ if and only if $p \leftrightarrow \bot$.

K*6 If $p \leftrightarrow q$ then $K * p = K * q$.

K*7 $K * (p \wedge q) \subseteq Th((K * p) \cup \{q\})$.

K*8 If $\neg q \notin K * p$ then $Th((K * p) \cup \{q\}) \subseteq K * (p \wedge q)$.

K*1 ensures that the result of the revision is deductively closed. **K*2** imposes that the new piece of information should belong to the revised beliefs. **K*3** implies that beliefs after revision should not contain more information than what can be logically derived from K and the new piece of information p. **K*4** together with **K*3** means that when the new piece of information is not contradictory with the old beliefs then revision is simply an expansion. **K*5** says that the revised beliefs set is inconsistent if and only if the new piece of information is itself inconsistent. **K*6** expresses that belief revision is syntax-independent. These first six postulates are the basic revision postulates and the last two express change minimality. **K*7** implies that revising by a conjunction $p \wedge q$ should not contain more information that what can be logically derived from the revision of K by p together with the piece of information q. **K*8** means that, when revising K by $p \wedge q$, every logical deduction from q and $K * p$ should be kept as soon as q is not contradictory with $K * p$.

[27] If BC is a set of formulae encoding these beliefs then $K = Th(BC)$ where Th is the deductive closure operator.

Note that in the following we are going to limit our study to the case of **K*2**. And we call "classical" an operator which satisfies **K*2**. However this precise postulate may not always be suitable in the argumentation system, this is developed in Section 3.2.5.3 on page 85.

A last recall about belief change is the distinction between belief revision and belief update (this was first established in [Win88]). The difference is in the nature of the new piece of information: either it is completing the knowledge of the world or it informs that there is a change in the world. More precisely, update is a process which takes into account a physical evolution of the system while revision is a process taking into account an epistemic evolution, it is the knowledge about the world that is evolving. In this paper, we suppose that we rather face a revision problem: the agent was not aware of some argument that suddenly appears, it means that the world has not changed but the awareness of the agent has evolved.

3.2.5.2 Revision in argumentation

First, we introduce a formal definition of revision in argumentation. The outcome of a revision process is the set of extensions under a given semantics. Then, by considering how the set of extensions is modified under the revision process, we propose a typology of different revisions.

3.2.5.2.1 Definition

Informally, a revision occurs when a new argument is presented. Note that the case of adding a new argument which is not connected to $\langle \mathcal{A}, \mathcal{R} \rangle$ is trivial. It has only to be added to each preferred extension. Indeed, revision is more interesting when the new argument interacts with previous ones. In this paper, which reports a preliminary study on revision in argumentation, we restrict revision to the addition of *exactly* one argument Z that has *exactly* one interaction, $Z\mathcal{R}X$ or $X\mathcal{R}Z$, where X belongs to $\mathcal{A}$.

In the following, we identify an argumentation system $\langle \mathcal{A}, \mathcal{R} \rangle$ with its associated attack graph $\mathcal{G}$. We write $X \in \mathcal{G}$ instead of "X is an argument represented by a node of $\mathcal{G}$". The set of extensions of $\langle \mathcal{A}, \mathcal{R} \rangle$ is denoted by $\mathcal{E}$ (with E_1, …, E_n denoting the extensions).

Revising $\langle \mathcal{A}, \mathcal{R} \rangle$ consists in adding an argument Z which attacks (or is attacked by) an argument X of $\mathcal{A}$. The revision process produces a new system represented by a graph $\mathcal{G}'$ and a new set of extensions $\mathcal{E}'$ (with E'_1, …, E'_p denoting the extensions).

$$(\mathcal{G}, \mathcal{E}) \quad \xrightarrow[i = (Z,X) \text{ or } i = (X,Z)]{\text{revision with } Z \text{ and } i} \quad (\mathcal{G}', \mathcal{E}')$$

Definition 54 *Let $\mathcal{G}$ be an attack graph. Let s be a semantics. Let $X \in \mathcal{G}$, $Z \notin \mathcal{G}$ and i be a pair of arguments (either (X, Z) or (Z, X)). Let $\mathcal{G}'$ be the graph obtained from $\mathcal{G}$ by adding the node Z and the edge i. The* revision operator Θ *maps $(Z, i, \mathcal{G}, s)$ to $\mathcal{E}'$ which is the set of extensions of $\mathcal{G}'$ under the semantics s.*

Let us mention several results which will be useful in the following. As we revise with only one new argument having only one interaction with an already existing argument, it is easy to prove that:

Proposition 34

- *If the new interaction is (Z, X), Z is not attacked in $\mathcal{G}'$.*
- *If the new interaction is (X, Z), Z attacks no argument of $\mathcal{G}'$.*
- *The revision process introduces no cycle in $\mathcal{G}'$.*

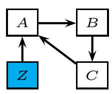
As defined above, revising an argumentation system may change the set of extensions. Given a semantics, the modifications are more or less important. It depends on the kind of interaction which is added and more precisely on the status of the argument X involved in that interaction. In the next section, we propose a typology of different kinds of revision according to how the set of extensions is modified. The next step will be to characterize each kind of revision by providing conditions on the interaction i.

3.2.5.2.2 Typology of revisions

Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system and $\mathcal{E}$ the set of extensions of $\langle \mathcal{A}, \mathcal{R} \rangle$ under a given semantics s. Different situations may be encountered in the general case. $\mathcal{E}$ may be empty (implying that s is the stable semantics), may be reduced to a singleton $\{E_1\}$ (where E_1 may be empty), or may contain more than one extension $\{E_1, \ldots, E_n\}$. The situation with only one non-empty extension is convenient for the determination of the status of an argument. In contrast, when several extensions exist, different choices are available. We have first considered revisions such that $\mathcal{G}'$ has a unique non-empty extension, while it was not the case for $\mathcal{G}$. Such a revision is called **decisive**.

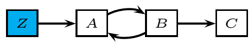
Example 20

1. *Under the stable semantics, with* $i = (Z, A)$



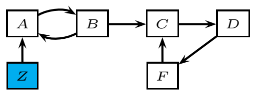
Before revision $\mathcal{E} = \varnothing$,
after revision $\mathcal{E}' = \{\{Z, B\}\}$

2. *Under the grounded semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{\}\}, \mathcal{E}' = \{\{Z, B\}\}$

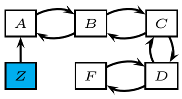
3. *Under the preferred semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{A\}, \{B, D\}\}$,
$\mathcal{E}' = \{\{Z, B, D\}\}$

A weaker requirement is the decrease of the number of choices. A revision such that $\mathcal{G}'$ has strictly less extensions than $\mathcal{G}$, but still has at least two, is called **selective**. Note that selective revision does not make sense under the grounded semantics, since there is always a unique grounded extension.

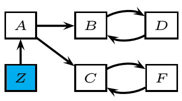
Example 21 *Under the preferred (or stable) semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{A, C, F\}, \{A, D\}, \{B, D\}, \{B, F\}\}$,
$\mathcal{E}' = \{\{Z, C, F\}, \{Z, B, D\}, \{Z, B, F\}\}$

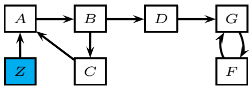
An opposite point of view enables to consider revisions which raise ambiguity, by increasing the number of extensions. This is the case for instance when $\mathcal{G}$ has at least one non-empty extension and $\mathcal{G}'$ has strictly more extensions than $\mathcal{G}$. A slightly different situation occurs when $\mathcal{G}$ has no extension or an empty one, while $\mathcal{G}'$ has more than one extension. In that case, revision brings some information, but is not decisive. Such revisions are called **questioning**. As for selective revision, questioning revision does not make sense under the grounded semantics.

Example 22 *Under the preferred (or stable) semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, B, C\}, \{Z, B, F\}, \{Z, D, C\}, \{Z, D, F\}\}$

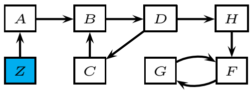
Under the stable semantics, with $i = (Z, A)$



$\mathcal{E} = \varnothing$,
$\mathcal{E}' = \{\{Z, B, F\}, \{Z, B, G\}\}$

Pursuing along the previous line, we consider revisions removing every extension, thus leading to a kind of decisional dead-end. A revision such that $\mathcal{G}'$ has no extension, while $\mathcal{G}$ had at least one, is called **destructive**. Note that destructive decision makes sense only under the stable semantics.

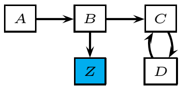
Example 23 *Under the stable semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{A, D, F\}, \{A, D, G\}\}$,
$\mathcal{E}' = \varnothing$

So far, the considered revisions have an impact on the number of extensions. Now, we are interested in revisions which modify the content of extensions, without modifying the number of extensions. The most interesting situation occurs when each extension of $\mathcal{G}'$ strictly includes one extension of $\mathcal{G}$, the number of extensions being the same. Such revisions are called **expansive**.

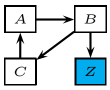
Example 24 *Under the preferred (or stable) semantics, with* $i = (B, Z)$



$\mathcal{E} = \{\{A, C\}, \{A, D\}\}$,
$\mathcal{E}' = \{\{Z, A, C\}, \{Z, A, D\}\}$

When nothing is changed, that is $\mathcal{E} = \mathcal{E}'$, the revision is called **conservative**.

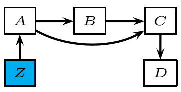
Example 25 *Under the preferred semantics, with* $i = (B, Z)$



$\mathcal{E} = \{\{\}\}, \mathcal{E}' = \{\{\}\}$

Otherwise, it may happen that some extensions (and sometimes all of them) are altered. This is called an **altering** revision. It is the case for instance when each extension of $\mathcal{G}'$ has a non-empty intersection with (but does not include) an extension of $\mathcal{G}$.

Example 26 *Under the grounded semantics, with* $i = (Z, A)$



$\mathcal{E} = \{\{A, D\}, \mathcal{E}' = \{\{Z, B, D\}\}$

The above discussion can be summarized on the following table.

<table>
<tr><th>$\mathcal{E}' =$
$\mathcal{E} =$</th><th>$\varnothing$</th><th>$\{\{\}\}$</th><th>$\{E'_1\}$</th><th>$\{E'_1, \ldots, E'_p\}$
$p \geq 2$</th></tr>
<tr><td>$\varnothing$</td><td>conservative</td><td>#1</td><td rowspan="2">decisive</td><td rowspan="2">questioning</td></tr>
<tr><td>$\{\{\}\}$</td><td>#2</td><td>conservative</td></tr>
<tr><td>$\{E_1\}$</td><td rowspan="2">destructive</td><td>#3</td><td>conservative
expansive
altering</td><td>questioning</td></tr>
<tr><td>$\{E_1,$
$\ldots,$
$E_n\}$
$n \geq 2$</td><td>#4</td><td>decisive</td><td>$n < p$:
questioning
$n > p$:
selective
$n = p$:
conservative
expansive
altering</td></tr>
</table>

With $E_i \neq \varnothing$ and $E'_i \neq \varnothing$. Each cell of the table contains the name of the corresponding revision. It can be checked that cells with #i correspond to situations which cannot occur:

#1 and #2 The only acceptability semantics in which an argumentation system may have no extension is the stable semantics. However, with the stable semantics, an argumentation system cannot have an empty extension when its set of arguments is not empty. And, by assumption, the cases #1 and #2 correspond to argumentation systems with non-empty sets of arguments (because at least X belongs to $\mathcal{G}$ and X and Z belong to $\mathcal{G}'$). So these cases cannot occur for all the acceptability semantics used in this section.

#3 Under the stable semantics, this case cannot occur for the same reason as that given previously (cases #1 and #2).

Under the grounded semantics, as $\mathcal{G}$ has one non-empty extension, there exists at least one unattacked argument W; so, if the added interaction is (X, Z), W is always unattacked and $\mathcal{G}'$ has always one non-empty extension; and, if the added interaction is (Z, X), then Z is unattacked and it belongs to the grounded extension of $\mathcal{G}'$; so, $\mathcal{G}'$ cannot have an empty extension.

Under the preferred semantics, if the added interaction is (Z, X), Z is unattacked and it belongs to the preferred extensions of $\mathcal{G}'$; so these preferred extensions are not empty. And, if the added interaction is (X, Z), then Z does not attack the arguments of E_1; so these arguments also belong to a preferred extension of $\mathcal{G}'$ and the preferred extensions are not empty.

In conclusion, this case cannot occur for all the acceptability semantics used in this section.

#4 This case could appear only with the preferred semantics (because with the grounded semantics there exists only one extension, and with the stable semantics, an extension cannot be empty since the set of arguments is not empty). If the added interaction is (Z, X), Z can "remove" or "create" extensions, but it belongs to each of them (because it is unattacked), so $\mathcal{G}'$ cannot have an empty extension. And if the added interaction is (X, Z), Z does not attack the arguments of E_i, $\forall i$, so these arguments belong to the preferred extensions of $\mathcal{G}'$ and $\mathcal{G}'$ cannot have an empty extension. Thus, this case cannot occur.

3.2.5.3 Classical revision in argumentation

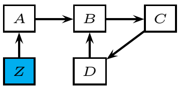
Revising a knowledge base consists in changing its beliefs in a minimal way in order to take into account a new piece of information considered as "prior" (according to AGM **K* 2** postulate). However, the revision operators defined above do not ensure at all that the new argument is accepted in the new graph extensions. In this section, we study when this property (called "classical") holds for a given revision operator.

Definition 55 *The revision Θ is* classical *iff $\mathcal{G}'$ has at least one extension and the added argument Z belongs to each extension of $\mathcal{G}'$.*

Proposition 35 *If the added interaction is (Z, X), then the revision is classical under the grounded and the preferred semantics.*

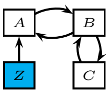
Moreover, if $\mathcal{G}$ has no odd-length cycle then the revision is also classical under the stable semantics.

Example 27 *Under the grounded semantics:*

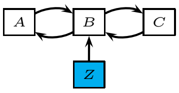


$\mathcal{E} = \{\{A, C\}\}, \mathcal{E}' = \{\{Z\}\}$

Example 28 *Let us compute the extensions of the two following graphs under the preferred semantics:*

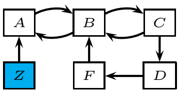


$\mathcal{E} = \{\{A, C\}, \{B\}\}$,
$\mathcal{E}' = \{\{Z, B\}, \{Z, C\}\}$



$\mathcal{E} = \{\{A, C\}, \{B\}\}$,
$\mathcal{E}' = \{\{Z, A, C\}\}$

Example 29 *Under the stable semantics:*



$\mathcal{E} = \{\{A, C, F\}, \{B, D\}\}$,
$\mathcal{E}' = \{\{Z, C, F\}, \{Z, B, D\}\}$

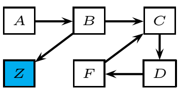
The condition that $\mathcal{G}$ should not have an odd-length cycle ensures the existence of at least one stable extension after adding Z. It is a sufficient but not necessary condition.

Example 27 (cont'd) *Before revision the stable extension is $\{A, C\}$, and after revision there is no stable extension.*

Proposition 36 *If the interaction is (X, Z) such that X is attacked by each extension of $\mathcal{G}$ then the revision is classical under the grounded and the preferred semantics.*

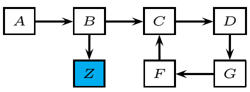
Moreover, if $\mathcal{G}$ has at least one stable extension then the revision is also classical under the stable semantics.

Example 30 *Under the grounded semantics:*



$\mathcal{E} = \{\{A\}\}, \mathcal{E}' = \{\{Z, A\}\}$

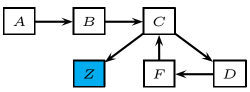
Example 31 *Under the preferred or the stable semantics:*



$\mathcal{E} = \{\{A, C, G\}, \{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, A, C, G\}, \{Z, A, D, F\}\}$

Note that, under the grounded semantics, X must be attacked and the fact that X is not in the only extension E of $\mathcal{E}$ does not ensure that the revision is classical :

Example 32 *Under the grounded semantics:*



$E = E' = \{A\}$.
So, $\mathcal{E} = \mathcal{E}' = \{\{A\}\}$.

C is not in E, and nevertheless Z does not belong to E'.

As said before, revising a graph by one argument and (one interaction) does not systematically lead to accept this argument. Hence, classicality is not the only property that is worth being studied for revision operators.

3.2.5.4 Case study

In this section, we study two cases: the decisive revision and the expansive revision. In the first case, after the revision there is only one extension (so it is easy to take a decision); in the second case, the number of extensions remains unchanged but each new extension is a superset of an extension which existed before the revision.

3.2.5.4.1 Decisive revision

Decisive revision makes possible a decision: before this revision, in $\mathcal{G}$ there is either no acceptable set of arguments (no possible conclusion), or too many acceptable sets of arguments (so too many possible conclusions), and after this revision there is only one acceptable set of arguments in $\mathcal{G}'$.

Definition 56 *The revision Θ is* decisive *iff Θ applied to $\mathcal{G}$, with $\mathcal{E} = \varnothing$, or $\mathcal{E} = \{\{\}\}$, or $\mathcal{E} = \{E_1, \ldots, E_n\}$, $n \geq 2$, the result of Θ is $\mathcal{G}'$ with $\mathcal{E}' = \{E'\}$, $E' \neq \varnothing$.*

Proposition 37 *If a revision is decisive then the added interaction is (Z, X). A decisive revision is classical.*

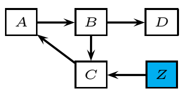
Example 20 on page 83 (cont'd)

1. *Under the stable semantics, Example 20.1 illustrates the decisive revision with $\mathcal{E} = \varnothing$ and $\mathcal{E}' = \{\{Z, B\}\}$.*
2. *Under the grounded semantics, Example 20.2 illustrates the decisive revision with $\mathcal{E} = \{\{\}\}$ and $\mathcal{E}' = \{\{Z, B\}\}$.*
3. *Under the preferred semantics, Example 20.3 illustrates the decisive revision with $\mathcal{E} = \{\{A\}, \{B, D\}\}$ and $\mathcal{E}' = \{\{Z, B, D\}\}$.*

Theorem 3 *Under the grounded semantics, if the added interaction is (Z, X) and $\mathcal{E} = \{\{\}\}$, then the revision is decisive.*

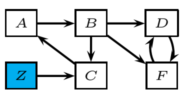
Theorem 4 *Under the preferred semantics, if the added interaction is (Z, X), $\mathcal{E} = \{\{\}\}$ and there is no even-length cycle in $\mathcal{G}$, then the revision is decisive.*

Example 33 *Under the preferred semantics:*



$\mathcal{E} = \{\{\}\}, \mathcal{E}' = \{\{Z, A, D\}\}$

Note that, if even-length cycles exist in the graph, the revision may induce several extensions; this revision would be a questioning one:



$\mathcal{E} = \{\{\}\}$,
$\mathcal{E}' = \{\{Z, A, D\}, \{Z, A, F\}\}$

For this reason, we have considered graphs without even-length cycle in Theorem 4.

Note: under the stable semantics, we have not found any characterization theorem for the decisive revision.

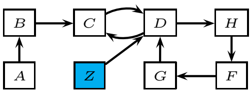
3.2.5.4.2 Expansive revision

A revision is said “expansive” when it does nothing but to add new arguments in the existing extensions.

Definition 57 *The revision $\ominus$ is* expansive *iff $\mathcal{G}$ and $\mathcal{G}'$ have the same number of extensions and each extension of $\mathcal{G}'$ strictly includes an extension of $\mathcal{G}$.*

Proposition 38 *The expansive revision is classical.*

Example 34 *Under the grounded semantics:*

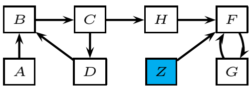


$\mathcal{E} = \{\{A\}\}$,
$\mathcal{E}' = \{\{Z, A, C, H, G\}\}$

Example 24 on page 84 gives also an illustration of the expansive revision under preferred and stable semantics.

Theorem 5 *Under the grounded semantics with $\mathcal{E} = \{E\}$, if the added interaction is (Z, X), $X \notin E$ and $E \neq \varnothing$, then the revision is expansive.*

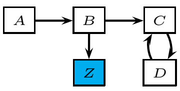
Example 35 *Under the grounded semantics:*



$\mathcal{E} = \{\{A, C\}\}$,
$\mathcal{E}' = \{\{Z, A, C, G\}\}$

Theorem 6 *Under the grounded semantics, if the added interaction is (X, Z), $X \notin E$ and E attacks X, then the revision is expansive and $\mathcal{E}' = \{E \cup \{Z\}\}$.*

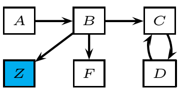
Example 36 *Under the grounded semantics:*



$\mathcal{E} = \{\{A\}\}, \mathcal{E}' = \{\{Z, A\}\}$

Theorem 7 *Under the stable semantics, if the added interaction is (X, Z), if $\mathcal{E} \neq \varnothing$, and $\forall i \geq 1$, $X \notin E_i$, then the revision is expansive and $\forall i$, $E'_i = E_i \cup \{Z\}$.*

Example 37 *Under the stable semantics:*



$\mathcal{E} = \{\{A, C, F\}, \{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, A, C, F\}, \{Z, A, D, F\}\}$

Note that, in an acyclic graph, Theorem 6 may be applied under the stable semantics. It is a particular case of Theorem 7.

Theorem 8 *Under the preferred semantics, if the added interaction is (X, Z), and $\forall i \geq 1$, E_i attacks X, then the revision is expansive and $\forall i$, $E'_i = E_i \cup \{Z\}$.*

Example 36 (cont’d) *Under the preferred semantics, $\mathcal{E} = \{\{A, C\}, \{A, D\}\}$ and $\mathcal{E}' = \{\{Z, A, C\}, \{Z, A, D\}\}$*

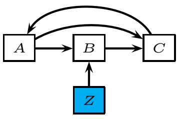
Note that when the initial graph is acyclic, Theorem 6 may be applied under the preferred semantics. It is a particular case of Theorem 8.

If the interaction is (Z, X), weaker results can be obtained. In that case, the revision is not expansive in the sense that $\mathcal{G}'$ may have more extensions than $\mathcal{G}$, however, adding Z to an extension of $\mathcal{G}$ yields an extension of $\mathcal{G}'$.

Proposition 39 *Under the stable semantics, if the added interaction is* (Z, X)*, and* $\forall i \geq 1, X \notin E_i$*, then* $\forall i$*,* $E_i \cup \{Z\}$ *is a stable extension of* $\mathcal{G}'$*.*

However, other stable extensions may appear in $\mathcal{G}'$, see Example 38 for instance. So the revision is not expansive.

Example 38 *Under the stable semantics:*



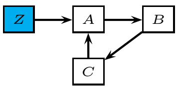
$\mathcal{E} = \{\{A\}\}$,
$\mathcal{E}' = \{\{Z, A\}, \{Z, C\}\}$

Note that in the particular case when the initial graph is acyclic, Theorem 5 on the preceding page may be applied under the stable semantics. And the obtained result is stronger than the result proposed by Property 39. Another interesting property is:

Proposition 40 *Under the preferred semantics, if the added interaction is* (Z, X) *and* $\forall i \geq 1$*,* $X \notin E_i$*, then* $\forall i$*,* $E'_i = E_i \cup \{Z\}$ *is admissible in* $\mathcal{G}'$*.*
Moreover, if there is no odd-length cycle in $\mathcal{G}$*,* $\forall i$*,* $E'_i = E_i \cup \{Z\}$ *is a preferred extension in* $\mathcal{G}'$*.*

Example 38 also shows that other preferred extensions may appear in $\mathcal{G}'$. And the following example illustrates the first part of Property 40.

Example 39 *Under the preferred semantics:*



$\mathcal{E} = \{\{\}\}$
$\{\} \cup \{Z\}$ *is admissible in* $\mathcal{G}'$
but $\mathcal{E}' = \{\{Z, B\}\}$*.*

Note that in the particular case when the initial graph is acyclic, Theorem 5 on the preceding page may also be applied under the preferred semantics and gives a stronger result than the one proposed by Property 40.

3.2.5.5 Discussion and future works in revision

In this section, we transpose the basic question of revision into argumentation theory. We propose a study of the impact of the arrival of a new argument on the outcome of an argumentation system. The term "revision" is used by analogy with traditional belief revision. However, there are two main differences.

- The basic underlying formalism is different: in standard belief revision, logical formulae are used for knowledge representation whereas, in this section, an argumentation system represents the current knowledge. In the first case, the outcome is a new set of logical formulae, whereas, in the second case, the outcome is a new set of accepted arguments.

- Revision is a task in knowledge representation which is strongly related to concepts such as inference and consistency. The postulates for standard belief revision (AGM) are built on a consistency notion, since it aims at incorporating a new piece of information while preserving consistency. Moreover, “revision” has also been studied in the framework of nonmonotonic theories [WvdH97]. Argumentation theory is linked to nonmonotonicity, but postulates for nonmonotonic theories are also based on consistency and inference notions that are not explicitly present in our framework. So, these postulates are not suited for our problem. Some of the belief revision postulates can be transposed (this is the case for what we call classical revision), but other principles must be proposed.

Our work is a preliminary step towards a formal revision in argumentation systems. And it departs from previous work relating argumentation and revision. Indeed, we have chosen to remain at an abstract level in this preliminary study. We do not consider knowledge from which arguments and interactions could be built. More precisely, there are many approaches that deal with adding new pieces of information within an argumentation system. The point of view adopted in this family of works is different because of the status of the new piece of information that is added. For instance, Wassermann [Was99], as well as [FGS04, PC05], define under which conditions, expressed in terms of arguments, unjustified beliefs should become accepted. The approach of [PG00] studies the properties of knowledge revision under the argumentation point of view, *i.e.*, the problem is to generate a knowledge base in which each piece of information is justified by "good" arguments.

Very recently, [RMF$^+$08] have proposed a warrant-prioritized revision operation, which consists in adding an argument to a theory in such a way that this argument is warranted afterwards. Even if the underlying ideas are similar, this work differs from our approach in at least two points:

- First, in [RMF$^+$08], arguments are given a structure through the subargument relation, and properties such as minimality, consistency and atomicity. And the definition of warranted arguments relies upon an evaluation of argumentation lines. In contrast, our approach remains at the most abstract level, and our sets of accepted arguments are computed with the well-known extension-based semantics.

- Secondly, the warrant-prioritized argument revision is designed in order to satisfy the AGM postulate **K*2**, since the added argument must be warranted in the revised theory. Our work follows another direction. We propose an extensive theoretical study of the impact of an addition on the outcome of an abstract argumentation system, which enables us to define several kinds of revision.

Note that other crucial cognitive tasks linked to belief change theory have already been transposed in the field of argumentation, see for instance the work on merging presented in [CMDK$^+$07] and in Section 3.2.4 on page 62.

A promising application of our work could be to design dialogue strategies. Most of the works about dialogue strategies consider that a dialogue is defined by a protocol giving the set of legal moves and that a strategy selects exactly one move (the move which must be done next). For instance, [BC98] proposes a selection strategy leading to more cooperative dialogues. Other approaches propose dialogue games for answering queries such as: does a given initial argument belong to some extension? In that case, a strategy helps to choose which argument must be defeated in order that the initial argument should be accepted. [AM02] have proposed heuristics that select the less attackable arguments in a persuasion dialogue. In a similar way, [RPRS08] have proposed an optimal strategy in order to win a debate based on the probability of success of the argument and on the cost of this argument for the agent. [Hun04], with a more global approach, has defined a strategy which builds an optimal subtree of arguments maximizing the resonance with the agent goals and minimizing their cost.

Our approach takes another point of view. We do not define any protocol and we do not restrict to a dialogue type. Given a set of arguments which may interact, we are interested in the outcome of the argumentation system, that is the set of extensions under a given semantics. In other words, we study the impact of an argument with respect to the structural change induced on the set of extensions. We do not focus on a particular argument that should be accepted at the end. We just want to act as to modify the form of that outcome (by doing an expansive revision, or a decisive revision for instance). The work reported in this paper enables us to choose the right way of revising (which argument must be affected by the revision, with which kind of interaction) in order to obtain the new outcome. This is why we plan to focus more on strategies for directing a dialogue than on strategies for taking part in it. For instance, if a dialogue arbitrator wants the debate to be more open then she should rather force the next speaker to use arguments appropriate for an expansive revision. If she wants the debate to be more focused then only arguments appropriate for a selective (and even decisive) revision should be accepted.

In order to continue this work, the following directions seem to be of interest:

1. generalize our revision operation to the adding of one argument with several interactions and to the adding of a subgraph of arguments;

2. restate other existing standard belief revision postulates and study the postulates for revision in nonmonotonic systems, in the case where arguments are built from knowledge bases and the outcome of the argumentation system is a set of formulae;
3. since decisive revision seems to be a "good" kind of revision, it would be interesting to investigate the question "How to make the *minimal change*[28] to a given argumentation system so that it has a unique non-empty extension?".

3.2.6 Practical reasoning and unipolar argumentation

Practical reasoning (PR) [Raz78], is concerned with the generic question "what is the right thing to do for an agent in a given situation". In [Woo00], it has been argued that PR is a two steps process. The first step, often called *deliberation*, consists of identifying the desires of an agent. In the second step, called *means-end reasoning*, one looks for ways for achieving those desires, *i.e.* for actions or plans.

A desire is *justified* if it holds in the current state of the world, and is *feasible* if it has a plan for achieving it. The agent's intentions, *i.e.* what an agent decides to do, is a consistent subset of desires that are both justified and feasible.

What is worth noticing in most works on practical reasoning is the use of arguments for providing reasons for choosing or discarding a desire as an intention. Indeed, some argumentation-based systems for PR have been proposed in the literature [AK05, HvdT03, RA06]. However, in most of these works, the problem of PR is modelled in terms of at least two separate systems, each of them capturing one step of the process. Such an approach may suffer from a serious drawback. In fact, some desires that are not feasible may be accepted at the deliberation step to the detriment of other justified and feasible desires. Another limitation of those systems is that their properties are not investigated.

This section proposes the first argumentation system that computes the intentions of an agent in one step. The system is grounded on a recent work on *constrained* argumentation systems [CMDM06]. These last extend the well-known general system of Dung [Dun95] by adding constraints on arguments that need to be satisfied by the extensions returned by the system. Our system takes as input

- three categories of arguments: *epistemic* arguments that support beliefs, *explanatory* arguments that show that a desire holds in the current state of the world, and *instrumental* arguments that show that a desire is feasible,
- different conflicts among those arguments, and
- a particular constraint on arguments that captures the idea that for a desire to be pursued it should be both feasible and justified. This is translated by the fact that in a given extension each instrumental argument for a desire should be accompanied by at least an explanatory argument in favour of that desire and each explanatory argument for a desire should be accompanied by at least an instrumental argument for that desire.

The output of the system is different sets of arguments as well as different sets of intentions. The use of a constrained system makes it possible to compute directly the intentions from the extensions, and to return only useful information. Indeed, each extension will support "warranted" beliefs as well as desires that are both justified and feasible. The properties of this system are deeply investigated. In particular, we show that the results of such a system are safe, and satisfy the rationality postulates identified in [CA05], namely consistency and completeness.

The section is organized as follows. Subsection 3.2.6.1 on the following page introduces an example of practical reasoning. Subsection 3.2.6.2 on page 93 presents the logical language. Subsection 3.2.6.3 on page 94 studies the different types of arguments involved in a practical reasoning problem, and Subsection 3.2.6.4 on page 97 investigates the conflicts that may exist between them. Subsection 3.2.6.5 on page 101 presents the constrained argumentation system for PR. This system is then explained on the motivating example in Subsection 3.2.6.6 on page 103. The properties of the system are studied in Subsection 3.2.6.7 on page 105. Subsection 3.2.6.8 on page 107 compares our approach with existing systems of practical reasoning. All the proofs are given in [ADLS08, ADLS09].

This work has been done in collaboration with Leila AMGOUD (IRIT) and Caroline DEVRED (LERIA).

[28] In terms of number of edges to add or to remove and/or in terms of number of arguments to add or to remove.

3.2.6.1 Motivating example

Let us consider an example of practical reasoning. In this example, Paula is a PhD student. She has four desires and would like to know whether she can reach them or not and with which plans. The four desires are the following:

1. To go on a journey to central Africa (jca)
2. To finish a publication before going on a journey (fp)
3. To be a lecturer (lec)
4. To visit her friend Carla (vc) if Carla is back from her trip (cb)

What is worth noticing is that the three first desires are unconditional, whereas the third one depends on whether the friend is back or not.

In addition to the desires, Paula has some beliefs on the way of achieving a given desire. Namely:

1. In order to go to central Africa, she should get tickets (t) and should be vaccinated (vac)
2. In order to get tickets, Paula can either go to an agency (ag) or ask a friend who may bring them (fr)
3. In order to be vaccinated, she can either go to the hospital (hop) or go to a doctor (dr)
4. In order to finish the paper, she should work (w)
5. In order to visit her friend, she can go by car (gc) if it is in good state
6. In order to be a lecturer, she should finish her thesis (ft)

Paula has also another kind of beliefs representing integrity constraints or the current state of the world.

1. If Paula works, then she can neither pass to the agency nor go to the doctor
2. Actually, the car of Paula is in good state (gs)
3. Carla is not yet back from her trip ($\neg cb$)
4. The thesis of Paula is not finished ($\neg ft$)
5. Paula is not vaccinated and has not her tickets

From the above information, it is clear that the desire of becoming a lecturer is not yet feasible. The desire of visiting Carla is feasible since there is a plan for reaching it; however, according to the current state of the world, this desire is not justified. Indeed, for Paula to consider this desire, she should be in a state where Carla is back from her trip and this is not the case. Regarding the two first desire (i.e. jca and fp) things are different. Both desires are justified and feasible. However, the problem is that some ways of achieving these desires are conflicting.

Of course, it would be ideal if all the desires can become intentions. As our example illustrates, this may not always be the case. In this section we will answer the following questions: “which desires will become the *intentions* of the agent?” and “with *which plans*?”

3.2.6.2 Logical language

In this subsection we present the logical language that will be used throughout the section. Let $\mathcal{L}$ be a *propositional language*, and $\equiv$ be the classical equivalence relation.

From $\mathcal{L}$, a subset $\mathcal{D}$ is distinguished and is used for encoding *desires*. By desire we mean a state of affairs that an agent wants to reach. Elements of $\mathcal{D}$ are *literals*. We will write $d_1, \ldots, d_n$ to denote desires and the lowercase letters will denote formulae of $\mathcal{L}$.

From the above sets, *desire-generation* rules can be defined. A desire-generation rule expresses under which conditions an agent may adopt a given desire. A desire may come from beliefs. For instance, "if the weather is sunny, then I desire to go to the park". In this case, the desire of going to the park depends on my belief about the weather. A desire may also come from other desires. For example, if there is a conference in India, and I have the desire to attend it, then I desire also to attend the tutorials. In this example, the desire of attending the tutorials depends on my belief about the existence of a conference in India, and on my desire to attend that conference. Finally, a desire may be unconditional, this means that the desire depends on neither beliefs nor desires. These three sources of desires are captured by the following desire-generation rules.

Definition 58 (Desire-Generation Rules) *A* desire-generation rule *(or a* desire rule*) is an expression of the form*

$$b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow^{29} d_m$$

where b is a propositional formula of $\mathcal{L}$, and each d_i is an element of the set $\mathcal{D}$. Moreover, $\nexists d_i, d_j$ with $i, j \leq m$ such that $d_i \equiv d_j$.

$b \wedge d_1 \wedge \ldots \wedge d_{m-1}$ is called the body *of the rule (this body may be empty; this is the case of an unconditional desire), and d_m its* consequent*.*

The meaning of the rule is "if the agent *believes* b and *desires* $d_1, \ldots, d_{m-1}$, then the agent will *desire* d_m as well". Note that the same desire d_i may appear in the consequent of several rules. This means that the same desire may depend on different beliefs or desires.

A desire rule is consistent if it depends on consistent beliefs and on non contradictory desires.

Definition 59 (Consistent Desire Rule) *A desire rule $b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow d_m$ is* consistent *iff*

- $b \nvdash \bot$,
- $\nexists d_i, d_j$ *with* $i, j \leq m$ *such that* $d_i \equiv \neg d_j$.

Otherwise, the rule is said inconsistent*.*

An agent is also equipped with different *plans* provided by a given planning system. The generation of such plans is beyond the scope of this section. A plan is a way of achieving a desire. It is defined as a triple: i) a set of preconditions that should be satisfied before executing the plan, ii) a set of postconditions that hold after executing the plan, and iii) the desire that is reached by the plan. Formally:

Definition 60 (Plan) *A plan is a triple $\langle S, T, x\rangle$ such that*

- *S and T are* consistent *sets of propositional formulae of $\mathcal{L}$,*
- *$x \in \mathcal{D}$,*
- *$T \vdash x$ and $S \nvdash x$.*

[29] The symbol $\hookrightarrow$ is not the material implication.

Of course, there exists a link between S and T. But this link is not explicitly defined here because we are not interested by this aspect of the process. We just consider that the plan is given by a correct and sound planning system (for instance [GNT04, RN95]).

Note that the set of preconditions may be empty ($S = \varnothing$), for instance when the desire is already realized in the current state of the world. In this case, the plan is always activable.

In the remaining of the section, an agent is equipped with three *finite bases*:

1. a base $\mathcal{K} \neq \varnothing$ and $\mathcal{K} \neq \{\bot\}$ containing its *basic beliefs* about the environment (elements of $\mathcal{K}$ are propositional formulae of the language $\mathcal{L}$),
2. a base $\mathcal{B}_d$ containing its "consistent" desire rules,
3. a base $\mathcal{P}$ containing its plans.

Using $\mathcal{B}_d$, we can characterize the *potential desires* of an agent as follows:

Definition 61 (Potential Desires) *The set of* potential desires *of an agent is* $\mathcal{PD} = \{d_m | \exists b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow d_m \in \mathcal{B}_d\}$.

These are "potential" desires because, when the body of the rule is not empty, the agent does not know yet whether the antecedents (*i.e.* bodies) of the corresponding rules are true or not.

Example 40 *The following sentence can be translated into a consistent desire rule: "If the weather is beautiful and if I desire to relax then I desire to be in a park". The translation can be obtained using the following vocabulary:*

- wb *means "the weather is beautiful" (it is a belief),*
- r *means "I relax" (it is a desire),*
- bp *means "I am in a park" (it is a desire).*

Then we can have the consistent desire rule: $wb \wedge r \hookrightarrow bp$ *(and the desire* bp *is a potential desire).*

If we assume that we also have the following beliefs:

- *the park is near my house if I use the subway, line number 3 (denoted by* b_1*),*
- *the subway works and I have valid tickets (denoted by* b_2*),*
- *the park is open and the entrance is free (denoted by* b_3*)*

and there exists a way to obtain bp *using* b_1*,* b_2 *and* b_3 *then the following plan can be defined:* $\langle \{b_1, b_2, b_3\}, \{bp\}, bp \rangle$.

3.2.6.3 Typology of arguments

The aim of this section is to present the different kinds of arguments involved in a practical reasoning problem. Three categories of arguments are distinguished. The first category justifies/attacks beliefs of the knowledge base $\mathcal{K}$, while the two others justify the adoption of the potential desires in $\mathcal{PD}$. Note that the arguments will be denoted with lowercase Greek letters.

3.2.6.3.1 Justifying beliefs

The first category of arguments is that studied in argumentation literature, especially for handling inconsistency in knowledge bases. Indeed, arguments are built from a knowledge base in order to support or to attack potential conclusions or inferences. These arguments are called *epistemic* in [Har04a]. In our application, such arguments are built from the base $\mathcal{K}$. In what follows, we will use the definition proposed in [SL92].

Definition 62 (Epistemic Argument) *Let $\mathcal{K}$ be a knowledge base. An* epistemic argument α *is a pair* $\alpha = \langle H, h\rangle$ *s.t:*

1. $H \subseteq \mathcal{K}$,
2. H *is consistent,*
3. $H \vdash h$ *and*
4. H *is minimal (for set $\subseteq$) among the sets satisfying conditions 1, 2, 3.*

The support *of the argument is given by the function* $\mathtt{SUPP}(\alpha) = H$, *whereas its* conclusion *is returned by* $\mathtt{CONC}(\alpha) = h$. $\mathcal{A}_b$ *stands for the set of all epistemic arguments that can be built from the base* $\mathcal{K}$.

Example 40 on the facing page (cont'd) *Assume that $\mathcal{K}_{40}$ contains the following beliefs:*

- *the weather is beautiful (wb),*
- *if the weather is beautiful then I do not take an umbrella ($wb \to \neg u$),*
- *if the weather is beautiful then I take a hat ($wb \to h$).*

With $\mathcal{K}_{40}$, one can build at least the following epistemic arguments:

- $\alpha_1 = \langle\{wb, wb \to \neg u\}, \neg u\rangle$,
- $\alpha_2 = \langle\{wb, wb \to h\}, h\rangle$.

Note that $\langle\{wb, wb \to \neg u, wb \to h\}, h\rangle$ is not an epistemic argument because it is not minimal.

3.2.6.3.2 Justifying desires

A desire may be pursued by an agent only if it is *justified* and *feasible*. Thus, there are two kinds of *reasons* for adopting a desire:

- the conditions underlying the desire hold in the current state of world. Such reasons will be called *explanatory arguments*.
- there is a plan for reaching the desire. Such reasons will be called *instrumental arguments*.

The definition of the first kind of arguments involves two bases: the belief base $\mathcal{K}$ and the base of desire rules $\mathcal{B}_d$. In what follows, we will use a tree-style definition of arguments [Vre97]. Before presenting that definition, let us first introduce some functions that will be used throughout the section.

The functions $\mathtt{BELIEFS}(\delta)$, $\mathtt{DESIRES}(\delta)$, $\mathtt{CONC}(\delta)$ and $\mathtt{SUB}(\delta)$ return respectively, for a given explanatory argument δ, the beliefs used in δ, the desires supported by δ, the conclusion and the set of sub-arguments of the argument δ.

Definition 63 (Explanatory Argument) *Let $\langle\mathcal{K}, \mathcal{B}_d\rangle$ be two bases.*

- *If* $\exists \hookrightarrow d \in \mathcal{B}_d$ *then* $\longrightarrow d$ *is an* explanatory argument[30] *(δ) with*
  - $\texttt{BELIEFS}(\delta) = \varnothing$,
  - $\texttt{DESIRES}(\delta) = \{d\}$,
  - $\texttt{CONC}(\delta) = d$,
  - $\texttt{SUB}(\delta) = \{\delta\}$.
- *If* α *is an epistemic argument, and* $\delta_1, \ldots, \delta_m$ *are explanatory arguments, and* $\exists\ \texttt{CONC}(\alpha) \wedge \texttt{CONC}(\delta_1) \wedge \ldots \wedge \texttt{CONC}(\delta_m) \hookrightarrow d \in \mathcal{B}_d$ *then* $\alpha, \delta_1, \ldots, \delta_m \longrightarrow d$ *is an* explanatory argument *(δ) with*
  - $\texttt{BELIEFS}(\delta) = \texttt{SUPP}(\alpha) \cup \texttt{BELIEFS}(\delta_1) \cup \ldots \cup \texttt{BELIEFS}(\delta_m)$,
  - $\texttt{DESIRES}(\delta) = \texttt{DESIRES}(\delta_1) \cup \ldots \cup \texttt{DESIRES}(\delta_m) \cup \{d\}$,
  - $\texttt{CONC}(\delta) = d$,
  - $\texttt{SUB}(\delta) = \{\alpha\} \cup \texttt{SUB}(\delta_1) \cup \ldots \cup \texttt{SUB}(\delta_m) \cup \{\delta\}$.

$\mathcal{A}_d$ stands for the set of all explanatory arguments that can be built from $\langle \mathcal{K}, \mathcal{B}_d \rangle$ respecting the fact that their DESIRES *set is consistent*[31].

This tree-style definition explains why the set $\texttt{DESIRES}(\delta)$ contains the desire d, conclusion of δ, but also, in the case of a conditional desire, *all the desires* used for justifying d (*i.e.* desires used by δ_1, ..., desires used by δ_m).

Note that the fact that rules in $\mathcal{B}_d$ are consistent is not sufficient to ensure the consistency of the set DESIRES of an explanatory argument.

One can easily show that this set DESIRES of an explanatory argument is a subset of $\mathcal{PD}$ and the set BELIEFS is a subset of the knowledge base $\mathcal{K}$.

Proposition 41 *Let $\delta \in \mathcal{A}_d$.*

- $\texttt{BELIEFS}(\delta) \subseteq \mathcal{K}$.
- $\texttt{DESIRES}(\delta) \subseteq \mathcal{PD}$.

Note that the same desire may be supported by several explanatory arguments since a desire may be the consequent of different desire rules.

The last category of arguments claims that "a desire may be pursued since it has a plan for achieving it". The definition of this kind of arguments involves the belief base $\mathcal{K}$, the base of plans $\mathcal{P}$, and the set $\mathcal{PD}$.

Definition 64 (Instrumental Argument) *Let $\langle \mathcal{K}, \mathcal{P}, \mathcal{PD} \rangle$ be three bases, and $d \in \mathcal{PD}$. An* instrumental *argument is a pair $\pi = \langle \langle S, T, x \rangle, d \rangle$ where*

- $\langle S, T, x \rangle \in \mathcal{P}$,
- $S \subseteq \mathcal{K}$,
- $x \equiv d$.

$\mathcal{A}_p$ stands for the set of all instrumental arguments that can be built from $\langle \mathcal{K}, \mathcal{P}, \mathcal{PD} \rangle$. The function CONC *will return for an argument π the desire d. Similarly, the functions* PLAN, Prec *and* Postc *will return respectively the plan $\langle S, T, x \rangle$ of the argument, the preconditions S of the plan, its postconditions T.*

[30] Note that the long arrow $\longrightarrow$ represents an explanatory argument whereas the short arrow $\rightarrow$ represents the classical logical implication.
[31] We do not accept cases where contradictory desires are necessary for justifying another desire.

The second condition of the above definition says that the preconditions of the plan hold in the current state of the world. In other words, the plan can be executed. Indeed, it may be the case that the base $\mathcal{P}$ contains plans whose preconditions are not true. Such plans cannot be executed and their corresponding instrumental arguments do not exist.

Example 40 on page 94 (cont'd) *Using the belief wb, the consistent desire rules $\hookrightarrow r$ and $wb \wedge r \hookrightarrow bp$ and the plan $\langle\{b_1, b_2, b_3\}$, $\{bp\}$, $bp\rangle$, we can define some new arguments:*

- $\alpha_3 = \langle\{wb\}, wb\rangle$ *(epistemic argument),*
- $\delta_1 = \longrightarrow r$ *(explanatory argument),*
- $\delta_2 = \alpha_3, \delta_1 \longrightarrow bp$ *(explanatory argument),*
- $\pi_1 = \langle\langle\{b_1, b_2, b_3\}$, $\{bp\}$, $bp\rangle, bp\rangle$ *(instrumental argument).*

In what follows, $\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p$. Note that $\mathcal{A}$ is *finite* since the three initial bases ($\mathcal{K}$, $\mathcal{B}_d$ and $\mathcal{P}$) are finite.

3.2.6.4 Interactions between arguments

Arguments built from a knowledge base cannot generally be considered separately in an inference problem. Indeed, an argument constitutes a reason for believing, or adopting a desire. However, it is not a proof that the belief is true, or in our case that the desire should be adopted. The reason is that an argument can be attacked by other arguments. In this section, we will investigate the different kinds of conflicts among the arguments identified in the previous section.

3.2.6.4.1 Conflicts among epistemic arguments

An argument can be attacked by another argument for three main reasons: i) they have contradictory conclusions (this is known as *rebuttal*), ii) the conclusion of an argument contradicts a premise of another argument (*assumption attack*), iii) the conclusion of an argument contradicts an inference rule used in order to build the other argument (*undercutting*).

Since the base $\mathcal{K}$ is built around a propositional language, it has been shown in [AC02a] that the notion of assumption attack is sufficient to capture conflicts between epistemic arguments.

Definition 65 *Let α_1, $\alpha_2 \in \mathcal{A}_b$. α_1 $\mathcal{R}_b$ α_2 iff $\exists h \in \mathtt{SUPP}(\alpha_2)$ such that $\mathtt{CONC}(\alpha_1) \equiv \neg h$.*

Example 40 on page 94 (cont'd) *Assume that the following beliefs are added to $\mathcal{K}_{40}$:*

- *the wind is very strong (wvs),*
- *if the wind is very strong then my hat is likely to fly away ($wvs \to hf$),*
- *if my hat is likely to fly away then I do not take my hat ($hf \to \neg h$).*

Then a new epistemic argument can be built: $\alpha_4 = \langle\{wvs, wvs \to hf, hf \to \neg h, wb\}, \neg(wb \to h)\rangle$.

Then, using $\alpha_2 = \langle\{wb, wb \to h\}, h\rangle$, the following conflict exists: $\alpha_4 \mathcal{R}_b \alpha_2$.

Note that the assumption attack is a binary relation that *is not symmetric*. Moreover, one can show that there are no self-defeating arguments.

Proposition 42 $\nexists \alpha \in \mathcal{A}_b$ *such that α $\mathcal{R}_b$ α.*

In [Cay95], the argumentation system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$ has been applied for handling inconsistency in a knowledge base, say $\mathcal{K}$. In this particular case, a full correspondence has been established between the stable extensions of the system and the maximal consistent subsets of the base $\mathcal{K}$. Before presenting formally the result, let us introduce some useful notations:

- Let $E \subseteq \mathcal{A}_b$, $\texttt{Base}(E) = \bigcup H_i$ such that $\langle H_i, h_i \rangle \in E$.
- Let $T \subseteq \mathcal{K}$, $\texttt{Arg}(T) = \{\langle H_i, h_i \rangle$ is an epistemic argument $| H_i \subseteq T\}$.

Proposition 43 ([Cay95]) *Let E be a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.*

- $\texttt{Base}(E)$ *is a maximal (for set inclusion) consistent subset of $\mathcal{K}$.*
- $\texttt{Arg}(\texttt{Base}(E)) = E$.

Proposition 44 ([Cay95]) *Let T be a maximal (for set inclusion) consistent subset of $\mathcal{K}$.*

- $\texttt{Arg}(T)$ *is a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.*
- $\texttt{Base}(\texttt{Arg}(T)) = T$.

A direct consequence of the above result is that if the base $\mathcal{K}$ is not reduced to $\perp$, then the system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$ has at least one non-empty stable extension.

Proposition 45 *The argumentation system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$ has non-empty stable extensions.*

3.2.6.4.2 Conflicts among explanatory arguments

Explanatory arguments may also be conflicting. Indeed, two explanatory arguments may be based on two contradictory desires. This kind of conflict is captured by the following relation:

Definition 66 *Let δ_1, $\delta_2 \in \mathcal{A}_d$.*
$\delta_1 \ \mathcal{R}_d \ \delta_2$ iff $\exists d_1 \in \texttt{DESIRES}(\delta_1)$, $d_2 \in \texttt{DESIRES}(\delta_2)$ such that $d_1 \equiv \neg d_2$.

Example 40 on page 94 (cont'd) *With the additional belief "My work is late" (denoted by l) in $\mathcal{K}_{40}$ and using the new consistent desire rule $l \hookrightarrow \neg r$, the following arguments can be built:*

- $\alpha_5 = \langle \{l\}, l \rangle$ *(epistemic argument),*
- $\delta_3 = \alpha_5 \longrightarrow \neg r$ *(explanatory argument).*

So, using $\delta_1 = \longrightarrow r$, the following conflicts exist: $\delta_1 \mathcal{R}_d \delta_3$ and $\delta_3 \mathcal{R}_d \delta_1$.

Proposition 46 *The relation $\mathcal{R}_d$ is symmetric and irreflexive.*

It can also be checked that any two explanatory arguments having conflicting desires are conflicting in the sense of the relation $\mathcal{R}_d$. Formally:

Proposition 47 *Let d_1, $d_2 \in \mathcal{PD}$. If $d_1 \equiv \neg d_2$, then $\forall \delta_1, \delta_2 \in \mathcal{A}_d$ s.t. (1) $\exists \delta'_1 \in \texttt{SUB}(\delta_1)$ with $\texttt{CONC}(\delta'_1) = d_1$, and (2) $\exists \delta'_2 \in \texttt{SUB}(\delta_2)$ with $\texttt{CONC}(\delta'_2) = d_2$, then $\delta_1 \ \mathcal{R}_d \ \delta_2$.*

Note that, from the definition of an explanatory argument, its set DESIRES cannot be inconsistent. However, its set BELIEFS may be inconsistent. The union of the beliefs of two explanatory arguments may also be inconsistent. Later in the section, we will show that it is useless to explicit these kinds of conflict, since they are captured by conflicts between explanatory arguments and epistemic ones (see Property 50 on page 100 and Property 51 on page 101).

3.2.6.4.3 Conflicts among instrumental arguments

Two plans may be conflicting for four main reasons:

1. incompatibility of their preconditions (indeed, both plans cannot be executed at the same time),
2. incompatibility of their postconditions (the execution of both plans will lead to contradictory states of the world),
3. incompatibility between the postconditions of a plan and the preconditions of the other (this means that the execution of a plan will prevent the execution of the second plan in the future),
4. incompatibility of their supporting desires (indeed, plans for achieving contradictory desires are conflicting; their execution will in fact lead to a contradictory state of the world).

The above reasons are captured in the following definition of attack among instrumental arguments.

Definition 67 *Let $\pi_1, \pi_2 \in \mathcal{A}_p$ and $\pi_1 \neq \pi_2$. $\pi_1 \ \mathcal{R}_p \ \pi_2$ iff:*

- $\texttt{Prec}(\pi_1) \wedge \texttt{Prec}(\pi_2) \models \perp$, *or*
- $\texttt{Postc}(\pi_1) \wedge \texttt{Postc}(\pi_2) \models \perp$, *or*
- $\texttt{Postc}(\pi_1) \wedge \texttt{Prec}(\pi_2) \models \perp$ *or* $\texttt{Prec}(\pi_1) \wedge \texttt{Postc}(\pi_2) \models \perp$

It is clear from the above definition that $\mathcal{R}_p$ is symmetric and irreflexive[32].

Proposition 48 *The relation $\mathcal{R}_p$ is symmetric and irreflexive.*

Example 40 on page 94 (cont'd) *Using the instrumental argument $\pi_1 = \langle\langle\{b_1, b_2, b_3\}, \{bp\}, bp\rangle, bp\rangle$ and the new instrumental argument $\pi_2 = \langle\langle\{b_4, b_5, b_6\}, \{st, nl\}, nl\rangle, nl\rangle$ defined with:*

- *$b_4 =$ I have a car, $b_5 =$ the subway failed, $b_6 =$ I must go to the meeting,*
- *$st =$ I am very stressed and tired, $nl =$ I don't want to be late to my meeting (it is a desire),*

There exists a conflict between π_1 and π_2 because b_5 (subway failed) and b_2 (the subway works and I have valid tickets) are in contradiction. So $\pi_1 \mathcal{R}_p \pi_2$ and $\pi_2 \mathcal{R}_p \pi_1$.

From the above definition, one can show that if two plans realize conflicting desires, then their corresponding instrumental arguments are conflicting too.

Proposition 49 *Let d_1, $d_2 \in \mathcal{PD}$. If $d_1 \equiv \neg d_2$, then $\forall \pi_1, \pi_2 \in \mathcal{A}_p$ s.t. $\texttt{CONC}(\pi_1) = d_1$ and $\texttt{CONC}(\pi_2) = d_2$, then $\pi_1 \ \mathcal{R}_p \ \pi_2$.*

In this section, we have considered only *binary conflicts* between plans, and consequently between their corresponding instrumental arguments. However, in every-day life, one may have for instance three plans such that any pair of them is not conflicting, but the three together are incompatible. For simplicity reasons, in this section we suppose that we do not have such conflicts.

[32] The fact that the postconditions of a plan are inconsistent with its preconditions is not considering as a conflict. In this case, after the execution of the plan, we must have an update mechanism which will modify the beliefs. It is also for this reason that there is no conflict between epistemic arguments and instrumental arguments concerning the postconditions (see Definition 68 on the following page).

3.2.6.4.4 Conflicts among mixed arguments

In the previous sections we have shown how arguments of the same category can interact with each other. In this section, we will show that arguments of different categories can also interact. Indeed, epistemic arguments play a key role in ensuring the acceptability of explanatory or instrumental arguments. Namely, an epistemic argument can attack both types of arguments. The idea is to invalidate any belief used in an explanatory or an instrumental argument. An explanatory argument may also conflicts with an instrumental argument when this last achieves a desire whose negation is among the desires of the explanatory argument.

Definition 68 *Let $\alpha \in \mathcal{A}_b$, $\delta \in \mathcal{A}_d$, $\pi \in \mathcal{A}_p$.*

- $\alpha \ \mathcal{R}_{bd} \ \delta$ *iff* $\exists h \in \texttt{BELIEFS}(\delta)$ *s.t.* $h \equiv \neg\texttt{CONC}(\alpha)$.
- $\alpha \ \mathcal{R}_{bp} \ \pi$ *iff* $\exists h \in \texttt{Prec}(\pi)$, *s.t.* $h \equiv \neg\texttt{CONC}(\alpha)$.
- $\delta \ \mathcal{R}_{pdp} \ \pi$ *and* $\pi \ \mathcal{R}_{pdp} \ \delta$ *iff* $\texttt{CONC}(\pi) \equiv \neg d$ *with* $d \in \texttt{DESIRES}(\delta)$[33].

A trivial consequence of this definition is the following link between $\mathcal{R}_b$ and $\mathcal{R}_{bd}$:

Consequence 2 *Let $\alpha_1 \in \mathcal{A}_b$, $\alpha_2 \in \mathcal{A}_b$ and $\delta \in \mathcal{A}_d$ such that $\alpha_1 \in \texttt{SUB}(\delta)$. If $\alpha_2 \mathcal{R}_b \alpha_1$ then $\alpha_2 \mathcal{R}_{bd} \delta$.*

Example 40 on page 94 (cont'd) *In $\mathcal{K}_{40}$, the following beliefs are added:*

- $b_5 \to \neg b_2$ *(if the subway fails then it does not work!),*
- ca *(there are clouds),*
- $ca \to \neg wb$ *(if there are clouds then the weather is not beautiful),*

and the following new desire-generation rule is added to $\mathcal{B}_d$: $l \hookrightarrow \neg bp$ (if my work is late than I desire not to be in the park).
So we can add two epistemic and one explanatory arguments:

- $\alpha_6 = \langle\{b_5, b_5 \to \neg b_2\}, \neg b_2\rangle$,
- $\alpha_7 = \langle\{ca, ca \to \neg wb\}, \neg wb\rangle$,
- $\delta_4 = \alpha_5 \longrightarrow \neg bp$.

These new arguments have some interactions with the other ones:

- $\alpha_6 \mathcal{R}_{bp} \pi_1$
- $\alpha_7 \mathcal{R}_{bd} \delta_2$ *(note that $\alpha_7 \mathcal{R}_b \alpha_3$ and $\alpha_3 \in \texttt{SUB}(\delta_2)$ which illustrates Consequence 2),*
- $\delta_4 \mathcal{R}_{pdp} \pi_1$ *and* $\pi_1 \mathcal{R}_{pdp} \delta_4$.

Moreover, as already said, the set of beliefs of an explanatory argument may be inconsistent. In such a case, the explanatory argument is attacked (in the sense of $\mathcal{R}_{bd}$) for sure by an epistemic argument. Formally:

Proposition 50 *Let $\delta \in \mathcal{A}_d$. If $\texttt{BELIEFS}(\delta) \vdash \bot$, then $\exists \alpha \in \mathcal{A}_b$ such that $\alpha \ \mathcal{R}_{bd} \ \delta$.*

Similarly, when the beliefs of two explanatory arguments are inconsistent, it can be checked that there exists an epistemic argument that attacks at least one of the two explanatory arguments. Formally:

[33] Note that if $\delta_1 \mathcal{R}_{pdp} \pi_2$ and there exists δ_2 such that $\texttt{CONC}(\delta_2) = \texttt{CONC}(\pi_2)$ then $\delta_1 \mathcal{R}_d \delta_2$.

Proposition 51 *Let $\delta_1, \delta_2 \in \mathcal{A}_d$ with* $\texttt{BELIEFS}(\delta_1) \nvdash \bot$ *and* $\texttt{BELIEFS}(\delta_2) \nvdash \bot$*. If* $\texttt{BELIEFS}(\delta_1) \cup \texttt{BELIEFS}(\delta_2) \vdash \bot$*, then $\exists \alpha \in \mathcal{A}_b$ such that $\alpha\ \mathcal{R}_{bd}\ \delta_1$, or $\alpha\ \mathcal{R}_{bd}\ \delta_2$.*

Conflicts may also exist between an instrumental argument and an explanatory one since the beliefs of the explanatory argument may be conflicting with the preconditions of the instrumental one. Here again, we'll show that there exists an epistemic argument that attacks at least one of the two arguments. Formally:

Proposition 52 *Let $\delta \in \mathcal{A}_d$ and $\pi \in \mathcal{A}_p$ with* $\texttt{BELIEFS}(\delta) \nvdash \bot$*. If* $\texttt{BELIEFS}(\delta) \cup \texttt{Prec}(\pi) \vdash \bot$ *then $\exists \alpha \in \mathcal{A}_b$ such that $\alpha\ \mathcal{R}_{bd}\ \delta$, or $\alpha\ \mathcal{R}_{bp}\ \pi$.*

Later in the section, it will be shown that the three above propositions are sufficient for ignoring these conflicts (between two explanatory arguments, and between an explanatory argument and an instrumental one). Note also that explanatory arguments and instrumental arguments are not allowed to attack epistemic arguments. In fact, a desire cannot invalidate a belief. Let us illustrate this issue by an example borrowed from [Tho00]. An agent thinks that it will be raining, and that when it is raining, she gets wet. It is clear that this agent does not desire to be wet when it is raining. Intuitively, we should get one extension $\{rain, wet\}$. The idea is that if the agent believes that it is raining, and she will get wet if it rains, then she should believe that she will get wet, regardless of her likings. To do otherwise would be to indulge in *wishful thinking*.

3.2.6.5 Argumentation system for PR

The notion of constraint forms the backbone of constrained argumentation systems. In a practical reasoning context, it encodes the link between the justification of a desire and the plan for achieving it. The basic idea is the following: as already said, for a desire to be pursued, it should be both justified (i.e. supported by an explanatory argument) and feasible (i.e. supported by an instrumental argument). Thus, explanatory arguments that are not accompanied by instrumental arguments for their conclusions will not be considered. Similarly, instrumental arguments that cannot be accompanied by explanatory arguments in favour of their desires will also be discarded. This constraint is formalized as follows:

Definition 69 (Constraint for PR) *Let $\mathcal{A} = \mathcal{A}_d \cup \mathcal{A}_p$ be a set of arguments and $\mathcal{L}_\mathcal{A}$ be the propositional language defined using $\mathcal{A}$ as the set of propositional variables. A* constraint for PR *is a constraint C on arguments of $\mathcal{A}$ such that:*

$$\begin{aligned} C \quad = \quad & \Big(\bigwedge_{\pi_i \in \mathcal{A}_p} (\pi_i \Rightarrow (\bigvee_{\delta_j \in \{\delta \in \mathcal{A}_d \mid \texttt{CONC}(\pi_i) \equiv \texttt{CONC}(\delta)\}} \delta_j))\Big) \\ & \wedge \\ & \Big(\bigwedge_{\delta_k \in \mathcal{A}_d} (\delta_k \Rightarrow (\bigvee_{\pi_l \in \{\pi \in \mathcal{A}_p \mid \texttt{CONC}(\delta_k) \equiv \texttt{CONC}(\pi)\}} \pi_l))\Big) \end{aligned}$$

with the convention: $(\bigvee_{x \in X} x) = \bot$ if $X = \varnothing$.

A constrained argumentation system for PR is defined as follows:

Definition 70 (Constrained argumentation system for PR) *A* constrained argumentation system *for practical reasoning is the triple $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ with:*

- $\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p$,
- $\mathcal{R} = \mathcal{R}_b \cup \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{bd} \cup \mathcal{R}_{bp} \cup \mathcal{R}_{pdp}$,
- *C a constraint on arguments defined on $\mathcal{A}$ as in Definition 69.*

In what follows, $E_1, \ldots, E_n$ will denote the C-extensions of $\mathsf{CoAF_{PR}}$ under a given a semantics. Due to the constraint C, each extension E_i contains, among the instrumental arguments, only the ones for which there exists at least one explanatory argument in the same set for their conclusions. Similarly, an extension contains, among the explanatory arguments, only the ones for which we can find at least one instrumental argument in favour of their conclusions. This means that the constraint makes it possible to filter the content of the extensions and to keep only useful information.

Remark 1 *An important remark concerns the notion of defence: this notion has two different semantics in PR context. When we consider only epistemic or explanatory arguments, the defence corresponds exactly to the notion defined in Dung's argumentation systems and in its constrained extension: an argument α attacks the attacker of another argument β; so α "reinstates" β; without the defence, we cannot keep β in an admissible set. Things are different with instrumental arguments: when an instrumental argument attacks another instrumental argument, this attack is always symmetric (so, each instrumental argument defends itself against an instrumental argument). In this case, it would be sufficient to take into account the notion of conflict-free in order to identify the plans which belong to an admissible set*[34]*. However, in order to keep an homogeneous definition of admissibility, the notion of defence is also used for instrumental arguments knowing that it is without impact when conflicts from an instrumental argument are concerned.*

Another important remark concerns the existence of C-extensions, namely the preferred ones.

Remark 2 *The empty set is always a C-admissible set for the practical system $\mathsf{CoAF_{PR}}$. Indeed, $\varnothing$ is admissible (as shown by Dung in [Dun95]) and all π_i and δ_k variables are false in $\widehat{\varnothing}$, so $\widehat{\varnothing} \vdash C$)*[35]*. Thus, the argumentation system $\mathsf{CoAF_{PR}}$ has a C-preferred extension.*

Remember that the purpose of a practical reasoning problem is to compute the intentions to be pursued by an agent, *i.e.* the desires that are both justified and feasible. These intentions are defined as follows:

Definition 71 (Set of intentions) *Let $\langle \mathcal{K}, \mathcal{B}_d, \mathcal{P} \rangle$ be three bases and $\mathsf{CoAF_{PR}}$ its corresponding constrained system. Let $E_1, \ldots, E_n$ the C-extensions of $\mathsf{CoAF_{PR}}$ under a given semantics.*

A set $\mathcal{I} \subseteq \mathcal{PD}$ is a set of intentions *of $\mathsf{CoAF_{PR}}$ under the given semantics iff there exists a C-extension E_i such that for each $d \in \mathcal{I}$,*

1. *there exists $\pi \in \mathcal{A}_p \cap E_i$ such that $d = \mathtt{CONC}(\pi)$*
2. *and $\bigcup_{\alpha_j \in E_i \cap \mathcal{A}_b} \mathtt{SUPP}(\alpha_j) \not\vdash d$.*

The first condition corresponds to the fact that an intention must be a justified *and* feasible desire. And the idea behind the second condition is that desires that already hold in the current state of the world are discarded. Indeed, an intention is a desire that is not yet achieved and that an agent could realize. However, note that the system not only returns the intentions to be pursued, but also infers the desires that are already achieved.

Different intention sets may be returned by our $\mathsf{CoAF_{PR}}$. Indeed, each extension gives a set of intentions, the state of the world which justifies these intentions and the plans which can realize them. The exact set that an agent decides to pursue is merely a decision problem as argued in [AP07]. This choice is beyond the scope of this section. Recall that the aim of this section is only to identify the different possibilities for an agent.

[34]This property can be extended to $\langle \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{pdp} \rangle$ because this subpart of $\mathsf{AF_{PR}}$ is symmetric in the sense of [CMDM05]; in this case, the admissibility is equivalent to the conflict-free notion.

[35]This is due to the particular form of the constraint for practical reasoning. This is not true for all constraints (see Section 3.2.1 on page 30 and [CMDM06]).

3.2.6.6 Motivating example revisited

Let us now analyze the motivating example and compute the intention set(s) of Paula. We will start by presenting the three bases encoding the problem. The first base is $\mathcal{B}_d$ which contains the desire generation rules.

$$\mathcal{B}_d = \{\hookrightarrow jca, \hookrightarrow fp, \hookrightarrow lec, cb \hookrightarrow vc\}$$

Recall that jca stands for "a journey to Central Africa", fp stands for "finishing a paper", lec means "becoming a lecturer", cb stands for "Carla is back", and vc stands for "visiting Carla". In this case, the set of potential desires is $\mathcal{PD} = \{jca, fp, lec, vc\}$.

Paula has four different plans for reaching her desire of going to central Africa:

- $p_1 = \langle\{\neg t, \neg vac\}, \{t, vac, ag, dr, jca, \neg w\}, jca\rangle$,
- $p_2 = \langle\{\neg t, \neg vac\}, \{t, vac, fr, dr, jca, \neg w\}, jca\rangle$,
- $p_3 = \langle\{\neg t, \neg vac\}, \{t, vac, ag, hop, jca, \neg w\}, jca\rangle$, and
- $p_4 = \langle\{\neg t, \neg vac\}, \{t, vac, fr, hop, jca\}, jca\rangle$.

She has also one plan for reaching each of fp, lec and vc.

- $p_5 = \langle\{\}, \{fp, w\}, fp\rangle$,
- $p_6 = \langle\{ft\}, \{lec\}, lec\rangle$, and
- $p_7 = \langle\{gs\}, \{vc\}, vc\rangle$.

Thus, the base $\mathcal{P}$ of plans is the following:

$$\mathcal{P} = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}$$

In addition to these two bases, Paula has the following knowledge base that gathers her beliefs about the current state of the world.

$$\mathcal{K} = \{w \to (\neg ag \wedge \neg dr), gs, \neg cb, \neg ft, \neg vac, \neg t\}$$

Note that this base is consistent in this particular example. Consequently, all the epistemic arguments that can be built from this base are not conflicting. Thus, the relation $\mathcal{R}_b$ is empty. Similarly, the two relations $\mathcal{R}_{bd}$ and $\mathcal{R}_{bp}$ are empty. Thus, these arguments are useless in this case, that's why we will not give the content of the set $\mathcal{A}_b$.

Let us now define the explanatory arguments. There are exactly three such arguments:

- $\delta_1 : \longrightarrow jca$
- $\delta_2 : \longrightarrow fp$
- $\delta_3 : \longrightarrow lec$

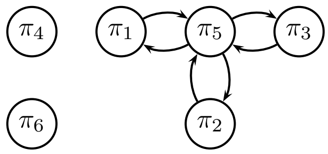
The desire vc, i.e. visiting Carla, is not justified since there is no explanatory argument in its favour. The reason is that the condition cb is not satisfied. Indeed, the base $\mathcal{K}$ contains $\neg cb$. Thus, $\mathcal{A}_d = \{\delta_1, \delta_2, \delta_3\}$. It can be checked that the three arguments are not conflicting, thus $\mathcal{R}_d = \varnothing$.

Regarding the feasibility of the desires, it is clear that the desire lec is not feasible since its plan requires that the thesis should be finished, and this actually not the case. The three other desires are however feasible. Their instrumental arguments are as follows:

- $\pi_1 : \langle p_1, jca\rangle$

- $\pi_2 : \langle p_2, jca \rangle$
- $\pi_3 : \langle p_3, jca \rangle$
- $\pi_4 : \langle p_4, jca \rangle$
- $\pi_5 : \langle p_5, fp \rangle$
- $\pi_6 : \langle p_7, vc \rangle$

Thus, $\mathcal{A}_p = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$. Some of these arguments are conflicting, thus the relation $\mathcal{R}_p$ is not empty and is depicted in figure below.



In this simple example, the relation $\mathcal{R}_{pdp}$ is empty since there are no contradictory desires. The constrained argumentation system is thus $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_p, C \rangle$ where C is the constraint on arguments of $\mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p$. In this example, the constraint C is:

$$
\begin{array}{rll}
C = & (& (\pi_1 \Rightarrow \delta_1) \\
& & \wedge(\pi_2 \Rightarrow \delta_1) \\
& & \wedge(\pi_3 \Rightarrow \delta_1) \\
& & \wedge(\pi_4 \Rightarrow \delta_1) \\
& & \wedge(\pi_5 \Rightarrow \delta_2) \\
& & \wedge(\pi_6 \Rightarrow \bot)) \\
& \wedge(& (\delta_1 \Rightarrow (\pi_1 \vee \pi_2 \vee \pi_3 \vee \pi_4) \\
& & \wedge(\delta_2 \Rightarrow \pi_5) \\
& & \wedge(\delta_3 \Rightarrow \bot))
\end{array}
$$

Note the particular cases of δ_3 and π_6: for δ_3 (resp. π_6) there is no corresponding instrumental (resp. explanatory) argument.

The system $\mathsf{AF}_{\mathsf{PR}}$ has two stable and preferred extensions:

- $E_1 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \delta_3, \pi_1, \pi_2, \pi_3, \pi_4, \pi_6\}$
- $E_2 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \delta_3, \pi_4, \pi_5, \pi_6\}$

Note that the above extensions contain the explanatory argument δ_3 in favour of the desire lec even if this desire is not feasible. Similarly, they contain the instrumental argument π_6 while the desire vc is not justified. If now, we apply the system $\mathsf{CoAF}_{\mathsf{PR}}$, then we will get two C-stable and C-preferred extensions:

- $E'_1 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \pi_1, \pi_2, \pi_3, \pi_4\}$
- and $E'_2 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \pi_4, \pi_5\}$.

Note that the C-stable extensions contain only useful information.

Now that the C-extensions are defined, we are able to define Paula's sets of intentions. She has two sets of intentions under the stable or preferred semantics:

- $\mathcal{I}_1 = \{jca\}$
- $\mathcal{I}_2 = \{jca, fp\}$

The choice of the exact set is a decision problem and is beyond the scope of this section. For instance, one may think that since the two desires may be satisfied, it is natural to assume that Paula will choose the second set. Consequently, she should choose the plans π_4 and π_5. Assume now that Paula is very cautious, and she does not want to miss her journey to central Africa. In this case, we can easily imagine that she chooses the set I_1 since she has four plans for reaching this desire, and if for any reason one of them fails, she can still satisfy her desire by another plan.

3.2.6.7 Properties of the system

The aim of this section is to study the properties of the proposed argumentation system for PR: $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. At some places, we will refer by $\mathsf{AF}_{\mathsf{PR}}$ to the corresponding basic argumentation system $\langle \mathcal{A}, \mathcal{R} \rangle$ (i.e. the system without the constraint C).

The first results concern the extensions of the system, and are mainly direct consequences of results got in [CMDM06]. The first property establishes a link between C-admissible sets and C-preferred extensions, and shows the impact of applying constraints on the notion of admissibility.

Proposition 53 *Let $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. Let Ω be the set of C-admissible sets of $\mathsf{CoAF}_{\mathsf{PR}}$.*

1. *Let $E \in \Omega$. There exists a C-preferred extension E' of $\mathsf{CoAF}_{\mathsf{PR}}$ s.t. $E \subseteq E'$.*
2. *Let $\mathsf{CoAF}_{\mathsf{PR}}{}' = \langle \mathcal{A}, \mathcal{R}, C' \rangle$ s.t. $C' \models C$. Let Ω' be the set of C'-admissible sets of $\mathsf{CoAF}_{\mathsf{PR}}{}'$. The inclusion $\Omega' \subseteq \Omega$ holds.*

The two following properties show that the constrained argumentation system is more general than a classical argumentation system. However, they may coincide in some circumstances.

Proposition 54 *Let $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. For each C-preferred extension E of $\mathsf{CoAF}_{\mathsf{PR}}$, there exists a preferred extension E' of $\mathsf{AF}_{\mathsf{PR}}$ such that $E \subseteq E'$.*

This property is illustrated by the example of Section 3.2.6.6 on page 103: $E_1 \subseteq E_1'$ and $E_2 \subseteq E_2'$, with E_1, E_2 being the C-preferred extensions and E_1', E_2' being the preferred extensions.

Proposition 55 *Let $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ such that C is a valid formula on $\mathcal{A}$. Then the preferred extensions of $\mathsf{AF}_{\mathsf{PR}}$ are the C-preferred extensions of $\mathsf{CoAF}_{\mathsf{PR}}$.*

Recall that $\mathsf{AF}_{\mathsf{PR}} = \langle \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_b \cup \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{bd} \cup \mathcal{R}_{bp} \cup \mathcal{R}_{pdp} \rangle$, an important property shows that the set $\mathcal{A}_b$ of epistemic arguments in a given stable extension of $\mathsf{AF}_{\mathsf{PR}}$ is itself a stable extension of the system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$. Knowing that the argumentation system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$ is intended to handle inconsistency in the knowledge base $\mathcal{K}$, the following result shows that stable extensions of $\mathsf{AF}_{\mathsf{PR}}$ are "complete" w.r.t. epistemic arguments. This means also that explanatory and instrumental arguments have no impact on the status of beliefs, and that wishful thinking is avoided.

Proposition 56 *If E is a stable extension of $\mathsf{AF}_{\mathsf{PR}}$, then the set $E \cap \mathcal{A}_b$ is a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.*

We show also that the basic argumentation system $\mathsf{AF}_{\mathsf{PR}}$ for PR has always stable extensions.

Proposition 57 *The system $\mathsf{AF}_{\mathsf{PR}}$ has at least one non-empty stable extension.*

One can show that if an explanatory argument belongs to a stable extension of $\mathsf{AF}_{\mathsf{PR}}$, then all its sub-arguments belong to that extension.

Proposition 58 *Let $\delta \in \mathcal{A}_d$. Let E_i be a stable extension of $\mathsf{AF}_{\mathsf{PR}}$. If $\delta \in E_i$, then $\mathtt{SUB}(\delta) \subseteq E_i$.*

This means that the beliefs on which this explanatory argument is built are “warranted”. Similarly, we can show that if an instrumental argument belongs to a stable extension then all its preconditions are supported by this extension.

Proposition 59 *Let $\pi \in \mathcal{A}_p$. Let E be a stable extension of $\mathsf{AF}_{\mathsf{PR}}$. If $\pi \in E$, then $\mathtt{Prec}(\pi) \subseteq \bigcup_{\alpha_j \in E \cap \mathcal{A}_b} \mathtt{SUPP}(\alpha_j)$.*

The above result is also true for epistemic arguments. In a previous section, we have shown that an explanatory argument may be based on contradictory beliefs. We have also shown that such an argument is attacked by an epistemic argument. In what follows, we will show that the situation is worse since such an argument is attacked by each stable extension of the system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$. That's why such arguments will be discarded.

Proposition 60 *Let $\delta \in \mathcal{A}_d$. If $\mathtt{BELIEFS}(\delta) \vdash \bot$, then $\forall E_i$ with E_i is a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$, $\exists \alpha \in E_i$ such that $\alpha \mathcal{R}_{bd} \delta$.*

A direct consequence of the above result is that such explanatory argument (with contradictory beliefs) will never belong to a stable extension of the system $\mathsf{AF}_{\mathsf{PR}}$.

Proposition 61 *Let $\delta \in \mathcal{A}_d$ with $\mathtt{BELIEFS}(\delta) \vdash \bot$. Under the stable semantics, the argument δ is rejected in $\mathsf{AF}_{\mathsf{PR}}$.*

Since an explanatory argument with contradictory beliefs is rejected in $\mathsf{AF}_{\mathsf{PR}}$, then it will be also rejected in $\mathsf{CoAF}_{\mathsf{PR}}$.

Proposition 62 *Let $\delta \in \mathcal{A}_d$ with $\mathtt{BELIEFS}(\delta) \vdash \bot$. Under the stable semantics, δ is a rejected argument in $\mathsf{CoAF}_{\mathsf{PR}}$.*

Besides in Property 51 on page 101, we have shown that when two explanatory arguments are based on contradictory beliefs, then at least one of the two arguments is attacked by an epistemic argument. Again, we will show that one of the two arguments is attacked by each stable extension of the system $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.

Proposition 63 *Let δ_1, $\delta_2 \in \mathcal{A}_d$ with $\mathtt{BELIEFS}(\delta_1) \nvdash \bot$ and $\mathtt{BELIEFS}(\delta_2) \nvdash \bot$.
If $\mathtt{BELIEFS}(\delta_1) \cup \mathtt{BELIEFS}(\delta_2) \vdash \bot$, then $\forall E_i$ with E_i is a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$, $\exists \alpha \in E_i$ such that $\alpha \mathcal{R}_{bd} \delta_1$, or $\alpha \mathcal{R}_{bd} \delta_2$.*

We go further, and we show that the two arguments cannot be accepted at the same time, *i.e.* they cannot belong to the same stable extension at the same time. This guarantees that the system proposed here returns safe results.

Proposition 64 *Let δ_1, $\delta_2 \in \mathcal{A}_d$ with $\mathtt{BELIEFS}(\delta_1) \nvdash \bot$ and $\mathtt{BELIEFS}(\delta_2) \nvdash \bot$.
If $\mathtt{BELIEFS}(\delta_1) \cup \mathtt{BELIEFS}(\delta_2) \vdash \bot$, then $\nexists E$ with E a C-stable extension of $\mathsf{CoAF}_{\mathsf{PR}}$ such that $\delta_1 \in E$ and $\delta_2 \in E$.*

Similarly, some conflicts between explanatory and instrumental arguments were discarded. We have shown in Property 52 on page 101 that in such a case, at least one of the two arguments will be attacked by an epistemic argument. Here we will show that the explanatory argument cannot be accepted at the same time with the instrumental one. One of them will be for sure rejected in the system.

Proposition 65 *Let $\delta \in \mathcal{A}_d$ and $\pi \in \mathcal{A}_p$ with $\mathtt{BELIEFS}(\delta) \nvdash \bot$. If $\mathtt{BELIEFS}(\delta) \cup \mathtt{Prec}(\pi) \vdash \bot$ then $\forall E_i$ with E_i is a stable extension of $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$, $\exists \alpha \in E_i$ such that $\alpha \mathcal{R}_{bd} \delta$, or $\alpha \mathcal{R}_{bp} \pi$.*

Proposition 66 *Let $\delta \in \mathcal{A}_d$ and $\pi \in \mathcal{A}_p$ with $\mathtt{BELIEFS}(\delta) \nvdash \bot$. If $\mathtt{BELIEFS}(\delta) \cup \mathtt{Prec}(\pi) \vdash \bot$ then $\nexists E$ with E a C-stable extension of $\mathsf{CoAF}_{\mathsf{PR}}$ such that $\delta \in E$ and $\pi \in E$.*

The next results are of great importance. They show that the proposed argumentation system for PR satisfies the "consistency" rationality postulate identified in [CA05]. Indeed, we show that each stable extension of our system supports a consistent set of desires and a consistent set of beliefs.

The following notations will be used: Let $E \subseteq \mathcal{A}$.

$$\begin{aligned}
\mathtt{Bel}(E) &= \Big(\bigcup_{\alpha_i \in E \cap \mathcal{A}_b} \mathtt{SUPP}(\alpha_i)\Big) \cup \Big(\bigcup_{\delta_j \in E \cap \mathcal{A}_d} \mathtt{BELIEFS}(\delta_j)\Big) \cup \Big(\bigcup_{\pi_k \in E \cap \mathcal{A}_p} \mathtt{Prec}(\pi_k)\Big) \\
\mathtt{Des}(E) &= \Big(\bigcup_{\delta_j \in E \cap \mathcal{A}_d} \mathtt{DESIRES}(\delta_j)\Big) \cup \Big(\bigcup_{\pi_k \in E \cap \mathcal{A}_p} \mathtt{CONC}(\pi_k)\Big)
\end{aligned}$$

Theorem 9 *(Consistency) Let* $\mathsf{CoAF_{PR}}$ *be a constrained argumentation system for PR, and* $E_1, \ldots, E_n$ *its* C*-stable extensions.* $\forall E_i$*,* $i = 1, \ldots, n$*, it holds that:*

1. *The set* $\mathtt{Bel}(E_i) = \mathtt{Bel}(E_i \cap \mathcal{A}_b)$.
2. *The set* $\mathtt{Bel}(E_i)$ *is a maximal (for set inclusion) consistent subset of* $\mathcal{K}$*.*
3. *The set* $\mathtt{Des}(E_i)$ *is consistent.*

As direct consequence of the above result, a set of intentions is consistent. Formally:

Theorem 10 *Under the stable semantics, each set of intentions of* $\mathsf{CoAF_{PR}}$ *is consistent.*

We have also shown that our system satisfies the rationality postulate concerning the closeness of the extensions [CA05]. Namely, we have shown that the set of arguments that can be built from the beliefs, desires, and plans involved in a given stable extension, is that extension itself. Before giving this result, let us first introduce some notations: Let E_i be a C-stable extension of $\mathsf{CoAF_{PR}}$. Let $\mathcal{A}s$ be the set of all (epistemic, explanatory and instrumental) arguments that can be built from $\mathtt{Bel}(E_i)$, $\mathtt{Des}(E_i)$, the plans involved in building arguments of E_i, and the base $\mathcal{B}_d$.

Theorem 11 *(Closeness) Let* $\mathsf{CoAF_{PR}}$ *be a constrained argumentation system for PR, and* $E_1, \ldots, E_n$ *its* C*-stable extensions.* $\forall E_i$*,* $i = 1, \ldots, n$*, it holds that:*

- $\mathtt{Arg}(\mathtt{Bel}(E_i)) = E_i \cap \mathcal{A}_b$.
- $\mathcal{A}s = E_i$.

In fact, this shows that every "good" argument is included in a stable extension. Thus, each desire that deserves to be pursued will be returned in an intention set.

3.2.6.8 Related Works

Recently, a number of attempts have been made to use formal models of argumentation as a basis for practical reasoning. Some of these models (e.g. [Amg03, AK05, HvdT04]) are instantiations of the *abstract* argumentation system of Dung [Dun95]. Others (e.g. [KM03, SGC04]) are based on an encoding of argumentative reasoning in logic programs. Finally, there are frameworks based on completely new theories of practical reasoning and persuasion (e.g. [ABCM04, TP05]). Our framework builds on the former, and is therefore a contribution towards formalizing practical reasoning using abstract argumentation systems.

In [Amg03], an argumentation system for generating consistent plans from a given set of desires and planning rules has been presented. This was later extended with another argumentation system that generates the desires themselves in [AK05]. For that purpose, a notion of "desire generation rules" has been introduced. These rules are meant to generate desires from beliefs. Thus, our desire generation rules are more general since we allow the generation of desires not only from beliefs, but also from other desires. Another problem with the work proposed in [AK05] arises because desires and beliefs are not correctly distinguished in the antecedent and consequent of the desire generation rules. This may lead to incorrect inferences where an agent may conclude beliefs on the basis of yet-unachieved desires, hence exhibiting a form of wishful thinking. Our approach resolves this by distinguishing between beliefs and desires in the rules, and refining the notion of attack among explanatory arguments accordingly. The problem of the logical language has been fixed in [RA06]. In that work, the authors considered three separate systems: one for reasoning about beliefs, one for generating justified desires, and finally one for generating feasible desires. The three systems are related with each others by attacks. Indeed, arguments supporting beliefs may attack both explanatory arguments and instrumental ones. However, explanatory arguments do not conflict with the instrumental ones. Once the results of the three systems are known, the intentions of an agent are computed. The main drawback of this approach is the following: it may be the case that two desires, say d_1 and d_2, are supported by two conflicting explanatory arguments, however d_1 is not feasible since there is no plan for reaching it. What happens is that the system may discard the desire d_2 since its explanatory argument is stronger than the one in favour of d_1. However, when computing the set of intentions, d_1 will neither be considered since it is not feasible. Thus, we lose both desires even if it was possible to achieve d_2 since it is both justified and feasible. In summary, handling separately the three types of arguments may lead to undesirable situations.

Hulstijn and van der Torre [HvdT04], on the other hand, have a notion of "desire rule," which contains only desires in the consequent. But their approach is still problematic. It requires that the selected goals[36] are supported by goal trees[37] which contain both desire rules and belief rules that are deductively consistent. This consistent deductive closure again does not distinguish between desire literals and belief literals (see Proposition 2 in [HvdT04]). This means that one cannot both believe $\neg p$ and desire p. In our framework, on the other hand, the distinction enables us to have an acceptable belief argument for believing $\neg p$ and, at the same time, an acceptable explanatory argument for desiring p.

Other researchers in AI like Atkinson and Bench Capon [ABCM04] are more interested in studying the different argument schemes that one may encounter in practical reasoning. Their starting point was the following practical syllogism advocated by the philosopher Walton.

- G is a goal/desire for agent X
- Doing action A is sufficient for agent X to carry out G
- Then, agent X ought to do action A

The above syllogism, which would apply to the means-end reasoning step, is in essence already an argument in favour of doing action A. However, this does not mean that the action is warranted, since other arguments (called counter-arguments) may be built or provided against the action. The authors have defined different variants of this syllogism as well as different ways of attacking it. However, it is not clear how all these arguments can be put together in order to answer the critical question of PR "what is the right thing to do in a given situation?" In our approach this question is answered. It is worth mentioning that most of the schemes and attacks suggested in [ABCM04] are already captured in our constrained system. For instance, to the above syllogism the following critical questions are associated:

1. Are there alternative ways of realizing G?
2. Is it possible to do A?
3. Does the agent has other goals that can be taken into account?
4. Are there other consequences of doing A which should be taken into account?

The first question amounts to find the different instrumental arguments for the desire G and to take all of them into account in the reasoning, i.e. when computing the set of intentions. The second question amounts to verify whether we are in a state of the world where A can be executed. In our approach this is captured by the preconditions of the plans. The third question is also captured in our approach. Indeed, we start with the set of all potential desires of the agent, and then we select the ones that will become its intentions. The last question is captured in our system by the postconditions of the plans and with the beliefs in the base $\mathcal{K}$.

3.2.6.9 Conclusion on argumentation and PR

The section has tackled the problem of practical reasoning, which is concerned with the question "what is the best thing to do at a given situation"? The approach followed here for answering this question is based on argumentation theory, in which choices are explained and justified by arguments. The contribution of this section is two-fold:

[36] Similar to our justified desires
[37] Similar to our explanatory arguments.

- To the best of our knowledge, this section proposes the first argumentation system that computes the intentions in one step, *i.e.* by combining desire generation and planning. This avoids undesirable results encountered by previous proposals in the literature.

- The second contribution of the section consists of studying deeply the properties of argumentation-based PR.

This work can be extended in different ways.

- First, we are currently working on relaxing the assumption that the attack relation among instrumental arguments is binary. Indeed, it may be the case that more than two plans may be conflicting while each pair of them is compatible.

- Another urgent extension would be to introduce preferences to the system. The idea is that beliefs may be pervaded with uncertainty, desires may not have equal priorities, and plans may have different costs. Thus, taking into account these preferences will help to reduce the intention sets into more relevant ones.

- In [CLS05e, KP01], it has been shown that an argument may not only be attacked by other arguments, but may also be supported by arguments. It would be interesting to study deeply the impact of such a relation between arguments in the context of PR.

- Finally, an interesting area of future work is investigating the proof theories of this system. The idea is to answer the question "is a given potential desire a possible intention of the agent ?" without computing the whole preferred extensions.

3.3 Bipolar argumentation

In most existing argumentation systems, only one kind of interaction is considered between arguments. It is the so-called attack relation. However, recent studies on argumentation [KP01, Ver99, Ver03, ACLS04a] have shown that another kind of interaction may exist between the arguments. Indeed, an argument can attack another argument, but it can also support another one. This suggests a notion of bipolarity, *i.e.* the existence of two independent kinds of information which have a diametrically opposed nature and which represent repellent forces.

Bipolarity has been widely studied in different domains such as knowledge and preference representation [Bou94, TP94, LVW02, BDKP02]. Indeed, in [BDKP02] two kinds of preferences are distinguished: the *positive* preferences representing what the agent really wants, and the *negative* ones referring to what the agent rejects. This distinction has been supported by studies in cognitive psychology which have shown that the two kinds of preferences are completely independent and are processed separately in the mind. Another application where bipolarity is largely used is that of decision making. In [ABP05, DF05], it has been argued that when making decision, one generally takes into account some information in favour of the decisions and other pieces of information against those decisions.

In [DP06], a nomenclature of three types of bipolarity has been proposed using particular characteristics like *exclusivity* (can a piece of information be at the same time positive and negative), *duality* (can negative information be computed using positive information), *exhaustivity* (can information be neither positive, nor negative), computation of positive and negative information *on the same data*, computation of positive and negative information *with the same process*, *existence of a consistency constraint* between positive and negative information.

The first type of bipolarity proposed by [DP06] (*symmetric univariate bipolarity*) expresses the fact that the negative feature is a reflection of the positive feature (so, they are mutually exclusive and a single bipolar univariate scale is enough for representing them).

The second one (*dual bivariate bipolarity*) expresses the fact that we need two separate scales in order to represent both features, although they stem from the same data (so, an information can be positive and negative at the same time and there is no exclusivity). However a duality must exist between both features.

And the third one (*heterogeneous bipolarity*) expresses the fact that both features do not stem from the same data though there is some minimal consistency requirement between both features.

In this section, we focus on the use of bipolarity in the particular domain of argumentation. and we are only concerned by the use of bipolarity at the interaction level (a more complete study of bipolarity in each step of the argumentation process is proposed in [ACLS04a, ACLSL08b]).

At this level, the main point is the definition of the interactions between arguments. As already said, due for instance to the presence of inconsistency in knowledge bases, arguments may be conflicting. Indeed, in all argumentation systems, an attack relation is considered in order to capture the conflicts.

However, most logical theories of argumentation assume that: if an argument a_1 attacks an argument a_3 and a_3 attacks an argument a_2, then a_1 supports a_2. In this case, the notion of support does not have to be formalized in a way really different from the notion of attack. It is the case of the basic argumentation system defined by Dung, in which only one kind of interaction is explicitly represented by the *attack* relation. In this context, the support of an argument a by another argument b can be represented only if b defends a in the sense of [Dun95]. So, support and attack are *dependent* notions. It is a parsimonious strategy, but it is not a correct description of the process of argumentation.Let us take several examples for illustrating the difference between "defence" and "support":

Example 41 *We want to begin a hike. We prefer a sunny weather, then a sunny and cloudy one, then a cloudy but not rainy weather, in this order. We will cancel the hike only if the weather is rainy. But clouds could be a sign of rain. We look at the sky early in the morning. It is cloudy. The following exchange of informal arguments occurs between Tom, Ben and Dan:*

t_1 *Today we have time, we begin a hike.*

b *The weather is cloudy, clouds are sign of rain, we had better cancel the hike.*

t_2 *These clouds are early patches of mist, the day will be sunny, without clouds, so the weather will be not cloudy (and we can begin the hike).*

d *These clouds are not early patches of mist, so the weather will be not sunny but cloudy; however these clouds will not grow, so it will not rain (and we can begin the hike).*

In this exchange, we can identify the following path of conflicts between arguments: argument d attacks argument t_2 which attacks argument b which in turn attacks argument t_1. So, with Dung's system, argument t_2 is a defender of argument t_1, and argument d is a defeater of argument t_1. Nevertheless, arguments t_2 and d support the hike project. So, the idea of a chain of arguments and counter-arguments in which we just have to count the links and take the even ones as defeaters and the odd ones as supporters is an oversimplification. So, the notion of defence proposed by [Dun95] is not sufficient to represent support.

The following example also illustrates the need for a new kind of interaction between arguments; the following arguments are exchanged during the meeting of the editorial board of a newspaper:

Example 42

a**:** *Assuming agreement and no right of censorship, information I concerning X will be published.*

b_1**:** *X is the prime minister who may use the right of censorship.*

c_0**:** *We are in democracy and even a prime minister cannot use the right of censorship.*

c_1**:** *I believe that X has resigned. So, X is no longer the prime minister.*

d**:** *The resignation has been announced officially yesterday on TV Channel 1.*

b_2**:** *I is private information so X denies publication.*

e**:** *I is an important information concerning X's son.*

c_2**:** *Any information concerning a prime minister is public information.*

repetition of c_1 and d: ...

c_3**:** *But I is of national interest, so I cannot be considered as private information.*

In this example, some conflicts appear: for instance, b_1 (resp. b_2) is in conflict with a. But we may also 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 "dialogue-like speech act": a new piece of information is really given and it is given *after* the production of the argument c_1. So taking d into account leads either to modify c_1, or to find a more intuitive solution for representing the interaction between d and c_1. In this case, we adopt an incremental point of view, considering that pieces of information given by different agents enable them to provide more and more arguments. We do not want to revise already advanced arguments. In contrast, we intend to represent as much as possible all the kinds of interaction between these arguments.

The last example shows how a notion of support between two arguments can be formalized with a logical representation of the structure of the arguments.

Example 43 *A murder has been performed and the suspects are Liz, Mary and Peter. The following pieces of information have been gathered:*

The type of murder suggests us that the killer is a female. The killer is certainly small. Liz is tall and Mary and Peter are small. The killer has long hair and uses a lipstick. A witness claims that he saw the killer who was tall. Moreover, we are told that the witness is short-sighted, so he is no more reliable.

We use the following propositional symbols: sm (the killer is small), fem (the killer is a female), $mary$ (the killer is Mary), $lglip$ (the killer has long hair and uses a lipstick), wit (the witness is reliable), bl (the witness is short-sighted).

Here, an argument takes the form of a set of premises which entails a conclusion. So the following arguments can be formed: a_1 in favour of $mary$ (with premises $\{sm, fem, (sm \wedge fem) \to mary\}$), a_2 in favour of $\neg sm$ (with premises $\{wit, wit \to \neg sm\}$), a_3 in favour of $\neg wit$ (with premises $\{bl, bl \to \neg wit\}$), a_4 in favour of fem (with premises $\{lglip, lglip \to fem\}$).

a_3 attacks a_2 which attacks a_1. So a_3 defends a_1 against a_2.

Moreover, a_4 confirms the premise fem of a_1. So, a_4 supports a_1 (in the sense that a_4 strengthens a_1). Contrastedly, a_3 defends a_1 against a_2 means that a_3 weakens the attack on a_1 brought by a_2. So, on one side, a_1 gets a support and on the other side a_1 suffers a weakened attack.

The above examples show that the argumentation process uses arguments and counter-arguments, support and attack relations, but not always in the same way. The arguments which are available in a dynamic argumentation process rely upon premises which are not always pieces of evidence. If we accept that a new fact can undermine one of the premises (thus forming an attack), we must also accept that a new fact can enforce, or confirm a premise (thus forming a support interaction).

Following all these remarks, and in order to formalize realistic examples, a more powerful tool than the abstract argumentation system proposed by Dung is needed. In particular, we are interested in modelling situations where two *independent* kinds of interactions are available: a positive and a negative one (see for example in the medical domain the work [KP01]). So, following [KP01, Ver03], we present a new argumentation system: an abstract bipolar argumentation system.

The section is organized as follows. Subsection 3.3.1 on the next page introduces the formal definitions of an abstract Bipolar Argumentation System (BAF for Bipolar Argumentation Framework). Then, we consider the fundamental problem of determining which arguments (or sets of arguments) can be considered as acceptable. The formal way to handle this problem is to define argumentation semantics. Subsection 3.3.2 on page 113 introduces extension-based acceptability semantics for a BAF. These new semantics rely upon criteria which make explicitly use of both support and attack relations. In Subsection 3.3.3 on page 116, another way to define extension-based semantics for a BAF is followed. First, a transformation of a BAF into a Dung's meta-argumentation system is given. The support relation is used to form meta-arguments (called coalitions) in such a way that at the meta-level only conflict interactions may appear. Extensions of a BAF can then be defined from Dung's extensions of the meta-system. Subsection 3.3.4 on page 119 addresses the question of labelling-based semantics in a BAF. Some labelling functions are proposed for a BAF. Subsection 3.3.5 on page 122 is devoted to the related issues and to some concluding remarks.

Note that the main part of this work has been done with Claudette CAYROL and the proofs of the properties given in this section can be found in the associated original papers (see [ACLS04b, ACLSL08a, CLS04, CLS05c, CLS05e, CLS07a, MCLS05]). A complete synthesis of this work is to appear in [CLS09].

3.3.1 Definition of BAF

An abstract bipolar argumentation system is an extension of the basic abstract argumentation system introduced by [Dun95] in which a new kind of interaction between arguments is represented by the *support*[38] relation[39]. This new relation is assumed to be totally independent of the attack relation (*i.e.* it is not defined using the attack relation). So, we have a bipolar representation of the interactions between arguments.

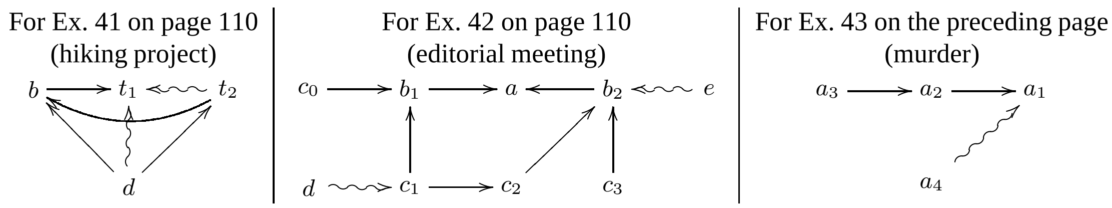
Definition 72 *An* abstract bipolar argumentation system *(BAF)* $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ *consists of:*

- *a set $\mathcal{A}$ of arguments,*
- *a binary relation $\mathcal{R}_{att}$ on $\mathcal{A}$ called the* attack relation
- *and another binary relation $\mathcal{R}_{sup}$ on $\mathcal{A}$ called the* support relation.

These binary relations must verify the following consistency constraint: $\mathcal{R}_{att} \cap \mathcal{R}_{sup} = \varnothing$[40].
Consider a_i and $a_j \in \mathcal{A}$, $a_i \mathcal{R}_{att} a_j$ (resp. $a_i \mathcal{R}_{sup} a_j$) means that a_i attacks (resp. supports) a_j. Let $a \in \mathcal{A}$, ${\mathcal{R}_{att}}^{-}(a)$ (resp. ${\mathcal{R}_{sup}}^{-}(a)$) denotes the set of attackers (resp. supporters) of a.

In the following, we assume that $\mathcal{A}$ represents the set of arguments proposed by rational agents at a given time, so we will assume that $\mathcal{A}$ is finite.

A BAF can be represented by a directed graph $\mathcal{G}^b$ called the *bipolar interaction graph*, with two kinds of edges, one for the attack relation ($\rightarrow$) and another one for the support relation ($\rightsquigarrow$). See for instance the following representations:



In the following, we abstract from the structure of the arguments and we consider arbitrary independent relations $\mathcal{R}_{\text{att}}$ and $\mathcal{R}_{\text{sup}}$.

Definition 73 *Let* `BAF` $= \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ *be a bipolar argumentation system and $\mathcal{G}^b$ be the associated interaction graph. Let $a, b \in \mathcal{A}$. A* path from a to b in $\mathcal{G}^b$ *is a sequence $(a_1, \dots, a_n)$ of elements of $\mathcal{A}$ s.t. $n \geq 2$, $a = a_1$, $b = a_n$, $a_1 \mathcal{R} a_2, \dots, a_{n-1} \mathcal{R} a_n$, with $\mathcal{R} = \mathcal{R}_{att}$ or $\mathcal{R}_{sup}$. Such a path has length $n - 1$.*

Note that if $n = 2$ and $a = b$ then the path is a loop *and if the relation $\mathcal{R}$ used in the loop is $\mathcal{R}_{att}$ then a is said* self-attacking.

[38] Note that the term "support" refers to a relation between two arguments and not a relation between premises and conclusion, as in Toulmin [Tou58].

[39] If the support relation is removed, we retrieve Dung's system.

[40] In the context of the argumentation, this consistency constraint is essential: it does not seem rational to advance an argument which simultaneously attacks and supports the same other argument.

The use of bipolarity suggests new kinds of interaction between arguments: in Example 42 on page 110, 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 cautious point of view, such arguments cannot appear together in a same extension. In order to address this problem, a new kind of attack has been introduced [CLS05a, CLS05b] which combines a sequence of supports with a direct attack.

Definition 74 *Let a, $b \in \mathcal{A}$. There is a* sequence of supports for b by a *(or for short a* supports *b) iff there exists a sequence $(a_1, \ldots, a_n)$ of elements of $\mathcal{A}$ s.t. $n \geq 2$, $a = a_1$, $b = a_n$, $a_1 \mathcal{R}_{sup} a_2, \ldots, a_{n-1} \mathcal{R}_{sup} a_n$.*

Definition 75 *A* supported attack *for an argument b by an argument a is a sequence (a, x, b) of arguments of $\mathcal{A}$ s.t. a supports x*[41] *and $x \mathcal{R}_{att} b$.*

In Example 42 on page 110, there is a supported attack for b_1 by d.
Then, taking into account attacks and sequences of supports leads to the following definitions applying to sets of arguments:

Definition 76 *Let $S \subseteq \mathcal{A}$, let $a \in \mathcal{A}$. S* set-attacks *a iff there exists a supported attack or a direct attack for a from an element of S. S* set-supports *a iff there exists a sequence of supports for a from an element of S.*

The above definitions are illustrated on the following example:

Example 44 *Consider the following graph:*

$$\begin{tikzcd}
 & a \arrow[r, squiggly] & b \arrow[r, squiggly] & c \arrow[r] & d & j \\
g \arrow[ur] & h \arrow[ur] & i \arrow[r, squiggly] & e \arrow[u] & f \arrow[u, squiggly] \arrow[r] & k \arrow[u]
\end{tikzcd}$$

In this graph, the paths $a - b - c - d$ and $i - c$ correspond to supported attacks. The set $\{a, h\}$ set-attacks d and b and set-supports b and c.

3.3.2 Extension-based semantics for acceptability

In Dung's framework, the *acceptability* of an argument depends on its membership to some sets, called acceptable sets or extensions. These extensions are characterised by particular properties. It is a collective acceptability. Following Dung's methodology, we propose characteristic properties that a set of arguments must satisfy in order to be an output of the argumentation process, in a bipolar framework. We recall that such a set of arguments must be in some sense coherent and must enable to win a dispute. Maximality for set-inclusion is also often required.

Considering a BAF $\langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ and using the notion of "set-attack" and "set-support" given by Definition 76, we first investigate the notion of coherence, then we propose new semantics for acceptability in bipolar argumentation systems.

3.3.2.1 Managing the conflicts

In the basic argumentation system, whatever the considered semantics, selected acceptable sets of arguments are constrained to be coherent in the sense that they must be conflict-free. In a bipolar argumentation system, the concept of coherence can be extended along two different lines:

- forbidding not only direct attacks but also supported attacks enforces a kind of *internal* coherence: we do not accept a set S of arguments which set-attacks one of its elements (this is a generalization of Dung's notion of conflict-free).

[41] In the sense of Definition 74.

- extending the consistency constraint between support and attack relations leads to define a kind of *external* coherence: we do not accept a set S of arguments which set-attacks *and* set-supports the same argument.

Definition 77 *Let $S \subseteq \mathcal{A}$. S is* +conflict-free[42] *iff $\nexists a, b \in S$ s.t. $\{a\}$ set-attacks b.*
S is safe[43] *iff $\nexists b \in \mathcal{A}$ s.t. S set-attacks b and either S set-supports b, or $b \in S$.*

Example 44 on the previous page (cont'd) *The set $\{h, b\}$ is not +conflict-free (there is a direct attack). The set $\{b, d\}$ is not +conflict-free since d suffers a supported attack from b. Contrastedly, $\{a, h\}$ and $\{b, f\}$ are +conflict-free.*
The set $\{a, h\}$ is not safe since a supports b and h attacks b. The set $\{b, f\}$ is not safe since d suffers a supported attack from b and f supports d. Contrastedly, $\{g, i, h\}$ is safe.

Note that the notion of safe set encompasses the notion of +conflict-free set:

Proposition 67 ([CLS05b]) *Let $S \subseteq \mathcal{A}$. If S is safe, then S is +conflict-free. If S is +conflict-free and closed for $\mathcal{R}_{sup}$ then S is safe.*

Example 44 on the preceding page (cont'd) *The set $\{g, h, i, e\}$ is +conflict-free and closed for $\mathcal{R}_{sup}$. So it is safe.*

3.3.2.2 New acceptability semantics

According to the methodology proposed by [Dun95], two notions play an important role in the definition of extension-based semantics: the notion of coherence, and the notion of defence (that is for short attack against attack). In a BAF, several notions of coherence, and two kinds of attack (direct and supported) are available. So several extensions of the notion of defence could be proposed. However, we have chosen to restrict to the classical defence, for the following reasons. First, the purpose of this section is to present some principles governing bipolar frameworks, rather than an exhaustive survey. Secondly, most of the works talking about bipolarity consider that a support does not have the same strength as an attack. In that sense, an argument can be considered as defended if and only if its direct attackers are directly attacked.

The above remark is illustrated by the following example: $\begin{tikzcd} a_1 \arrow[r] & a_2 \arrow[l, bend right] & a_3 \arrow[l, squiggly] \end{tikzcd}$

There is a supported attack for a_1 by a_3 and no attack for a_3. However, a_1 directly attacks a_2 and it seems sufficient to reinstate a_1.

Let us recall the definition of defence given in [Dun95] and recalled in 3.2.1 on page 30.

Definition 78 *Let $S \subseteq \mathcal{A}$. Let $a \in \mathcal{A}$. S* defends *a iff $\forall b \in \mathcal{A}$, if $b\mathcal{R}_{att}a$ then $\exists c \in S$ s.t. $c\mathcal{R}_{att}b$.*

In the following, the concept of admissibility is first extended. The idea is to reinforce the coherence of the admissible sets. Then, extensions under the preferred semantics will be defined as maximal (for $\subseteq$) admissible sets of arguments.

Three different definitions for admissibility can be given, from the most general one to the most specific one. First, a direct translation of Dung's definition gives the definition of d-admissibility ("d" means "in the sense of Dung"). Taking into account external coherence leads to s(afe)-admissibility. Finally, external coherence can be strengthened by requiring that an admissible set is closed for $\mathcal{R}_{\text{sup}}$. So, we obtain the definition of c(losed)-admissibility.

Definition 79 *Let $S \subseteq \mathcal{A}$.*
S is d-admissible *iff S is +conflict-free and defends all its elements.*
S is s-admissible *iff S is safe and defends all its elements.*
S is c-admissible *iff S is +conflict-free, closed for $\mathcal{R}_{sup}$ and defends all its elements.*

From the above definitions, it follows that each c-admissible set is s-admissible, and each s-admissible set is d-admissible.

[42] This notation means that checking if a set is +conflict-free needs to consider more conflicts than with the basic notion of conflict-free suggested by Dung.
[43] This definition is inspired by [Ver03] and by the notion of a controversial argument given in [Dun95].

Definition 80 *A set $S \subseteq \mathcal{A}$ is a d-preferred (resp. s-preferred, c-preferred) extension iff S is maximal for $\subseteq$ (or for short $\subseteq$-maximal) among the d-admissible (resp. s-admissible, c-admissible) subsets of $\mathcal{A}$.*

Example 41 on page 110 (cont'd) *In this case, the three semantics give the same result: $\{d, t_1\}$ is the unique d-preferred, s-preferred and c-preferred extension.*

Example 44 on page 113 (cont'd) *The set $\{g, h, i, e, f, d, j\}$ is the unique c-preferred extension.*

Example 45 *Consider the BAF represented by $a \rightsquigarrow b \leftarrow h$. The set $\{a, h\}$ is the unique d-preferred extension. There are two s-preferred extensions $\{a\}$ and $\{h\}$. And there is only one c-preferred extension $\{h\}$.*

One of the most important issues with regard to extensions concerns their existence. The existence of d-preferred (resp. s-preferred, c-preferred) extensions is guaranteed since the empty set is d-admissible (resp. s-admissible, c-admissible), and each d-admissible (resp. s-admissible, c-admissible) is included in a d-preferred (resp. s-preferred, c-preferred) extension. Note that analogous definitions for admissibility could be proposed using a stronger notion of defence (a stronger defence would be defined for instance by replacing *attack* with *set-attack* in Definition 78 on the facing page).

Considering another well-known semantics, the stable semantics, nice results can be obtained if we keep the basic definition of a stable extension, but replace *attack* with *set-attack*. It is a straightforward way to extend the stable semantics in a BAF.

Definition 81 *S is a* stable extension *iff S is +conflict-free and $\forall a \notin S$, S set-attacks a.*

In the following, we restrict to acyclic BAF, in the sense that the associated interaction graph is acyclic. In Dung's basic framework, it has been proved that, in the case of an acyclic attack graph, there is always a unique stable (which is also preferred) extension. So, Definition 81 ensures the existence of a unique stable extension in an acyclic BAF[44]. However, the unique stable extension is not always safe.

Example 45 (cont'd) *The set $\{a, h\}$ is the unique stable extension, and it is not safe.*

Indeed, the following result can be proved:

Proposition 68 ([CLS05b]) *Let S be a stable extension. S is safe iff S is closed for $\mathcal{R}_{sup}$.*

The following results enable to characterize d-preferred, s-preferred and c-preferred extensions when the BAF is acyclic:

Proposition 69 ([CLS05b]) *Let S be the unique stable extension of an acyclic BAF.*

1. *S is also the unique d-preferred extension.*
2. *The s-preferred extensions and the c-preferred extensions are subsets of S.*
3. *Each s-preferred extension which is closed for $\mathcal{R}_{sup}$ is c-preferred.*
4. *If S is safe, then S is the unique c-preferred and the unique s-preferred extension.*
5. *If $\mathcal{A}$ is finite, each c-preferred extension is included in a s-preferred extension.*
6. *If S is not safe, the s-preferred extensions are the subsets of S which are $\subseteq$-maximal among the s-admissible sets.*
7. *If S is not safe, and $\mathcal{A}$ is finite, there is only one c-preferred extension.*

[44] We instantiate Dung's AF with the relation set-attacks and the resulting graph is still acyclic.

Example 45 on the previous page (cont'd) *$\{h\}$ is the only s-preferred extension which is also closed for $\mathcal{R}_{sup}$. So, $\{h\}$ is the unique c-preferred extension.*

Example 46 *Consider the BAF represented by:*

$$\begin{tikzcd} a_1 \arrow[r, squiggly] \arrow[dr] & a_2 \arrow[r, squiggly] & b \\ & c & h \arrow[l, squiggly] \arrow[u] \end{tikzcd}$$

$\{a_1, a_2, h\}$ is the only d-preferred extension. $\{a_1, a_2\}$ and $\{h\}$ are the only two s-preferred extensions. None of them is closed for $\mathcal{R}_{sup}$. $\varnothing$ is the unique c-preferred extension. If we add an isolated argument a_3 (for which no interaction exists with the other available arguments), then we obtain: $\{a_1, a_2, a_3, h\}$ is the only d-preferred extension. $\{a_1, a_2, a_3\}$ and $\{h, a_3\}$ are the only two s-preferred extensions. None of them is closed, and $\{a_3\}$ is the unique c-preferred extension.

The above discussion enables to draw the following conclusions. In the particular case of an acyclic BAF, two semantics present nice features: the stable semantics and the c-preferred semantics. If we are interested in internal coherence only, we will have to determine the unique stable extension, which is also the unique d-preferred extension. If we are interested in a more constrained concept of coherence, we will compute the unique c-preferred extension.

3.3.3 Turning a bipolar system into a Dung meta-system

The extension-based acceptability semantics introduced in Section 3.3.2 on page 113 rely upon criteria which make explicitly use of support and attack relations, through the concept of supported attack. Here, we follow another way to define extension-based semantics for a BAF. First, a transformation of a BAF into a Dung's meta-argumentation system is given. This meta-argumentation system consists only of a set of meta-arguments (called coalitions), and a conflict relation between these meta-arguments. The attack relation of the initial BAF will appear only at the meta-level. As a consequence, a meta-argument will gather arguments which are not in conflict. The support relation of the initial BAF will not appear at the meta-level, but will be used to gather arguments in a coalition. The idea is that a meta-argument makes sense only if its members are somehow related by the support relation. So, the two fundamental principles governing the definition of a coalition are: the *Coherence principle* (there is no direct attack between two arguments of a same coalition) and the *Support principle* (if two arguments belong to a same coalition, they must be somehow, directly or indirectly, related by the support relation).

3.3.3.1 The concept of coalition

Consider $\mathtt{BAF} = \langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ represented by the graph $\mathcal{G}^b$. $\mathcal{G}^b_{\text{sup}}$ will denote the partial graph representing the partial system $\langle \mathcal{A}, \mathcal{R}_{\text{sup}} \rangle$[45]. $\mathtt{AF}$ will denote the partial argumentation system $\langle \mathcal{A}, \mathcal{R}_{\text{att}} \rangle$ associated with $\mathtt{BAF}$ and represented by the partial graph denoted by $\mathcal{G}^b_{\text{att}}$.

Definition 82 *$C \subseteq \mathcal{A}$ is a* coalition *of* $\mathtt{BAF}$ *iff: (i) The subgraph of $\mathcal{G}^b_{sup}$ induced by C is connected; (ii) C is conflict-free*[46] *for* $\mathtt{AF}$*; (iii) C is $\subseteq$-maximal among the sets satisfying* **(i)** *and* **(ii)***.*

Note that when $\mathcal{R}_{\text{att}}$ is empty, the coalitions are exactly the connected components[47] of the partial graph $\mathcal{G}^b_{\text{sup}}$.

Proposition 70 ([CLS07b]) *An argument which is not self-attacking is in at least one coalition.*

[45] We consider that the reader knows the basic concepts of graph theory (chain, connexity,…). See for instance [Ber73] for a background on graph theory.

[46] In the basic sense proposed by Dung.

[47] Let $G = (V, E)$ be a graph. Let $S \subseteq V$. S is a *connected component of* G iff the subgraph of G induced by S is connected and there exists no $S' \subseteq V$ s.t. $S \subset S'$ and the subgraph of G induced by S' is connected.

Example 47 *Consider the BAF:*

$$\begin{tikzcd}
a \arrow[d, squiggly] & c \arrow[l] \arrow[r] \arrow[d, squiggly] & e \arrow[r, squiggly] \arrow[d, squiggly] & f \arrow[r, squiggly] & g \arrow[r, squiggly] & h \\
b \arrow[ur, squiggly] & d & i \arrow[u, bend right]
\end{tikzcd}$$

The coalitions are: $C_1 = \{b, c, d\}$, $C_2 = \{i\}$, $C_3 = \{a, b\}$, $C_4 = \{e, f, g, h\}$

The following result shows that coalitions can be restated in terms of connected components of an appropriate subgraph. By the way, it gives a constructive way for computing coalitions.

Proposition 71 ([CLS07b]) *$C \subseteq \mathcal{A}$ is a coalition of* `BAF` *iff: (i) There exists $S \subseteq \mathcal{A}$ $\subseteq$-maximal conflict-free for* `AF` *s.t. C is a connected component of the subgraph of $\mathcal{G}^b_{sup}$ induced by S and (ii) C is $\subseteq$-maximal among the subsets of $\mathcal{A}$ satisfying* **(i)**.

Proposition 71 suggests a procedure for computing the coalitions of `BAF`:

Step 1: Consider `AF` and determine the maximal conflict-free sets for `AF`.

Step 2: For each set of arguments S_i obtained at Step 1, determine the connected components of the subgraph of $\mathcal{G}^b_{\text{sup}}$ induced by S_i.

Step 3: Keep the $\subseteq$-maximal sets obtained at Step 2.

The notion of conflict-free set is related to the notion of independent set:

Proposition 72 ([CLS07b]) *Let $S \subseteq \mathcal{A}$. S is conflict-free for* `AF` *iff S is an independent subset of $\mathcal{A}$ in the graph $\mathcal{G}^b_{att}$.*

So, S is $\subseteq$-maximal conflict-free for `AF` iff S is a $\subseteq$-maximal independent set of vertices in the graph $\mathcal{G}^b_{\text{att}}$ and Step 1 of the computational procedure consists in determining all the $\subseteq$-maximal independent subsets of $\mathcal{G}^b_{\text{att}}$. Remark that the time complexity of the best algorithms providing all the $\subseteq$-maximal independent sets is exponential. Note also that there exist several algorithms in the literature for finding all the $\subseteq$-maximal independent sets (see for instance the work of J.M. Nielsen [Nie02]). We also know that:

- For Step 2, a depth-first exploration of a graph provides the connected components in linear time $\mathcal{O}$(number of vertices + number of edges).
- And for Step 3, maximization with respect to $\subseteq$ is also an exponential process.

3.3.3.2 A meta-argumentation system

Let $C(\mathcal{A})$ denote the set of coalitions of `BAF`. We define a conflict relation on $C(\mathcal{A})$ as follows.

Definition 83 *Let C_1 and C_2 be two coalitions of* `BAF`*. C_1* C-attacks *C_2 iff there exists an argument a_1 in C_1 and an argument a_2 in C_2 s.t. $a_1 \mathcal{R}_{att} a_2$.*

It can be proved that:

Proposition 73 ([CLS07b]) *Let C_1 and C_2 be two distinct coalitions of* `BAF`*. If $C_1 \cap C_2 \neq \varnothing$ then C_1 C-attacks C_2 or C_2 C-attacks C_1.*

So a new argumentation system `CAF` $= \langle C(\mathcal{A}), \text{C-attacks}\rangle$ can be defined, referred to as the coalition system associated with `BAF`.

Example 47 (cont'd) *In this example,* `CAF` *can be represented by (by abusing notations, $\rightarrow$ represents the attack relation in* `BAF` *and also the C-attack relation in* `CAF`*):*

$$C_3 \longleftarrow C_1 \longrightarrow C_4 \longleftarrow C_2$$

Dung's definitions apply to `CAF`, and it can be proved that:

Proposition 74 ([CLS07b]) *Let $\{C_1, \ldots, C_p\}$ be a finite set of distinct coalitions. $\{C_1, \ldots, C_p\}$ is conflict-free for* `CAF` *iff $C_1 \cup \ldots \cup C_p$ is conflict-free for* `AF`*.*

So, `CAF` is a "meta-argumentation" system with a set of "meta-arguments" (the set of coalitions $C(\mathcal{A})$) and a "meta-attack" relation on these coalitions (the C-attacks relation). A coalition gathers arguments which are close in some sense and can be produced together. However, as coalitions may conflict, following Dung's methodology, preferred and stable extensions of `CAF` can be computed. Such extensions will contain coalitions which are collectively acceptable. The last step consists in gathering the elements of the coalitions of an extension of `CAF`. By this way, the best groups of arguments (w.r.t. the given interaction relations) will be selected.

Definition 84 *Let $S \subseteq \mathcal{A}$. S is a* Cp-extension *(Cp means "Coalition-preferred") of* `BAF` *iff there exists $\{C_1, \ldots, C_p\}$ a preferred extension of* `CAF` *s.t. $S = C_1 \cup \ldots \cup C_p$.*
S is a Cs-extension *(Cs means "Coalition-stable") of* `BAF` *iff there exists $\{C_1, \ldots, C_p\}$ a stable extension of* `CAF` *s.t. $S = C_1 \cup \ldots \cup C_p$.*

When the only preferred extension of `CAF` is the empty set, we define the empty set as the unique Cp-extension of `BAF`.

Example 47 on the previous page (cont'd) *There is only one preferred extension of* `CAF`*, which is also stable: $\{C_1, C_2\}$. So, $S = \{b, c, d, i\}$ is the Cp-extension (and also the Cs-extension) of* `BAF`*.*

Some nice properties of Dung's basic framework are preserved:

- A `BAF` has always a (at least one) Cp-extension. It is a consequence of Definition 84.
- In contrast, there does not always exist a Cs-extension of `BAF`. The reason is that there may be no stable extension of `CAF`.
- Each Cs-extension is also a Cp-extension. The converse is false.
- There cannot exist two Cp-extensions s.t. one strictly contains the other one. It follows from Definitions 82 on page 116 and 84.

However, other properties are lost. A Cp-extension is not always admissible for `AF`, and a Cs-extension is not always a stable extension of `AF`:

Example 48 *Consider the BAF represented by:*

$$a \longrightarrow b \leftsquigarrow c \rightsquigarrow d \longrightarrow e$$

The coalitions are: $C_1 = \{a\}$, $C_2 = \{b, c, d\}$, $C_3 = \{e\}$. And the associated CAF can be represented by:

$$C_1 \longrightarrow C_2 \longrightarrow C_3$$

There is only one preferred extension of `CAF`*, which is also stable: $\{C_1, C_3\}$. So, $S = \{a, e\}$ is the Cp-extension (and also the Cs-extension) of* `BAF`*. We have $d\mathcal{R}_{att}e$, but a does not defend e against d (neither by a direct attack, nor by a supported attack, though a attacks an element of the coalition which attacks e). So, S is not admissible for* `AF`*. S does not contain c, but there is no attack (no supported attack) of an element of S against c. So, S is not a stable extension of* `AF`*.*

Note that a coalition is considered as a whole and its members cannot be used separately in an attack process. Example 48 suggests that admissibility is lost due to the size of the coalition $\{b, c, d\}$, and that it would be more fruitful to consider two independent coalitions $\{c, b\}$ and $\{c, d\}$. A new formalization of coalitions in terms of conflict-free maximal support paths has been proposed in [CLS07b]. However, it does not enable to recover Dung's properties.

Note that the lost of admissibility in Dung's sense is not surprising: admissibility is lost because it takes into account "individual" attack and defence, whereas, with meta-argumentation and coalitions, "collective" attack and defence are considered.

3.3.4 Labellings in bipolar systems

This section addresses the question of labelling-based semantics in a BAF. A labelling-based semantics relies upon a set of labels and is defined by specifying the criteria for assigning labels to arguments. An example of labelling-based semantics in a basic argumentation system is given by [JV99b] with the robust semantics. More generally, several approaches have been proposed for valuing the arguments in a classical argumentation system (for example [KAEF95, Par97, PS97, BH01, CLS03d, Amg99, Ver96]). In some of them, the value of an argument depends on its interactions with the other arguments; in other ones, it depends on an intrinsic strength of the argument.

Besides, Karacapilidis & Papadias [KP01] have proposed a labelling approach for a bipolar graph representing a decision-making debate. However, they consider only two labels: active and inactive.

In this section, we propose a limited use of the notion of labelling-based semantics for a BAF: we show how bipolar interactions can be used for defining valuations over the set of arguments, *i.e.* functions which assign a value to each argument of the BAF (a further step would be to use such a valuation in order to select arguments, that is to completely define labelling-based semantics in a BAF, in an analogous way as what has been done in [CLS05d] and Sections 3.2.2 on page 35 and 3.2.3 on page 54 for basic argumentation systems).

3.3.4.1 Labellings in HERMES

Karacapilidis & Papadias [KP01] propose an argumentation web-tool for decision making in a medical domain. This argumentation system, named HERMES, permits the expression and the weighting of arguments, verifies the coherence of preferences between arguments and values the arguments. The basic elements of this system are: *issues* (questions whose answer is open for discussion[48]), *positions* which express the support for, or the opposition to a solution, to another position, or to a constraint (a position gives an information for the discussion) and *constraints* which express a preference between two positions (so, it is a comparison tool on the set of positions). HERMES can label the solutions and the positions by the status "active" or "inactive". At the end of the discussion, the "active" positions (resp. "inactive") are accepted (resp. rejected). An "active" solution is a recommended choice among the other solutions concerning a same issue. Different labellings are proposed in HERMES. They are recursive: the label of an element e depends on the labels of the elements which are linked to e in the discussion graph. In HERMES, the discussion graph is acyclic, the value of a position p depends only on the *active* positions which are linked to p, and the value of a position is always binary, even when preference constraints are taken into account.

Labelling 1: A position is active if and only if there is neither support, nor attack on this position, or if it is supported by an active position.

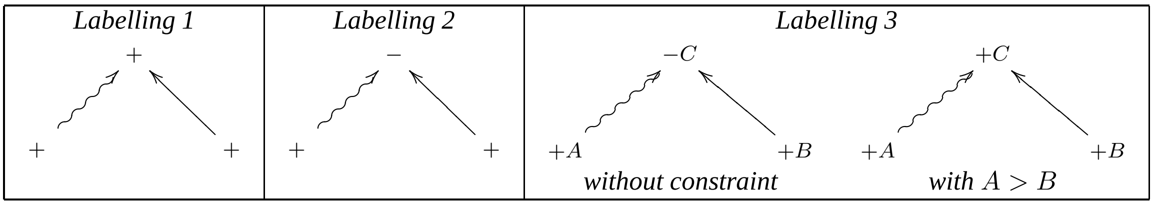
Labelling 2: A position is active if and only if it is not attacked by an active position.

Note that the labelling 1 and 2 do not permit to take into account in the same time the supports and the attacks on a position.

Labelling 3: A position is active if and only if there is neither support, nor attack on this position, or if its score is strictly positive. The score of a position p is defined by: $\Sigma_i w(p_i) - \Sigma_j w(p_j)$ with p_i the *active* positions which support p and p_j the *active* positions which attack p. Each position has the same initial weight and taking into account the preferences between positions modifies the relative weights of the positions.

Example 49 *An active (resp. inactive) position will be denoted by + (resp. –).*

[48] For example: "if the patient Y has the pathology X, what is the appropriate treatment?" An issue is a set of *solutions*. Examples of solutions are surgical operation or use of medicines.



3.3.4.2 Gradual bipolar valuation

The approach presented here (see [CLS05a]) has the following features: the valuation process takes place before the selection process; the valuation process makes use of a rich set of values and not only two as in HERMES (so, it is called a gradual valuation); the value assigned to an argument takes into account all the direct attackers and supporters of this argument (it is not the case in HERMES in which the value of an argument only depends on the *active* positions); so it is called a local valuation.

This proposition extends the works [JV99b, BH01, CLS05d] to bipolar argumentation systems as defined in Section 3.3.1 on page 112. It follows the same principles as those already described in [CLS03d] augmented with new principles corresponding to the "support" information. So, the three underlying principles for a gradual interaction-based local valuation are:

- **P1**: The value of an argument depends on the values of its direct attackers and of its direct supporters.
- **P2**: If the quality of the support (resp. attack) increases then the value of the argument increases (resp. decreases).
- **P3**: If the quantity of the supports (resp. attacks) increases then the quality of the support (resp. attack) increases.

The value of an argument is obtained with the composition of several functions:

- one for aggregating the values of all the direct attackers; this function computes the value of the "attack";
- one for aggregating the values of all the direct supporters; this function computes the value of the "support";
- one for computing the effect of the attack and of the support on the value of the argument.

In the respect of the previous principles, we assume that there exists a completely ordered set $\mathcal{V}$ with a minimum element $\mathcal{V}_{\text{min}}$ and a maximum element $\mathcal{V}_{\text{max}}$. The following formal definition for a gradual local valuation can be given.

Definition 85 *Let $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ be a bipolar argumentation system. Let $a \in \mathcal{A}$ with ${\mathcal{R}_{att}}^{-}(a) = \{b_1, \ldots, b_n\}$ and ${\mathcal{R}_{sup}}^{-}(a) = \{c_1, \ldots, c_p\}$.*
A local gradual valuation on $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ is a function $v : \mathcal{A} \to \mathcal{V}$ s.t.:

$$v(a) = g(h_{sup}(v(c_1), \ldots, v(c_p)), h_{att}(v(b_1), \ldots, v(b_n))) \text{ with}$$

the function h_{att} (resp. h_{sup}): $\mathcal{V}^* \to \mathcal{H}_{att}$ (resp. $\mathcal{V}^* \to \mathcal{H}_{sup}$)[49] valuing the quality of the attack (resp. support) on a, and the function g: $\mathcal{H}_{sup} \times \mathcal{H}_{att} \to \mathcal{V}$ with $g(x, y)$ increasing on x and decreasing on y. The function h, $h = h_{att}$ or h_{sup}, must satisfy:

1. *if $x_i \geq x'_i$ then $h(x_1, \ldots, x_i \ldots, x_n) \geq h(x_1, \ldots, x'_i \ldots, x_n)$,*
2. *$h(x_1, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_n)$,*

[49] $\mathcal{V}^*$ denotes the set of the finite sequences of elements of $\mathcal{V}$, including the empty sequence. $\mathcal{H}_{\text{att}}$ and $\mathcal{H}_{\text{sup}}$ are ordered sets.

3. $h() = \alpha \leq h(x_1, \ldots, x_n) \leq \beta$, *for all* $x_1, \ldots, x_n$[50].

Definition 85 on the preceding page produces a generic local gradual valuation. Let us give two instances of the generic definition, to illustrate the different principles.

- A first instance is defined by $\mathcal{H}_{\text{att}} = \mathcal{H}_{\text{sup}} = \mathcal{V} = [-1, 1]$ interval of the real line, $h_{\text{att}}(x_1, \ldots, x_n) = h_{\text{sup}}(x_1, \ldots, x_n) = \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)$, and $g(x, y) = \frac{x-y}{2}$ (so, we have $\alpha = -1$, $\beta = 1$ and $g(\alpha, \alpha) = 0$).
- Another one is defined by $\mathcal{V} = [-1, 1]$ interval of the real line, $\mathcal{H}_{\text{att}} = \mathcal{H}_{\text{sup}} = [0, \infty[$ interval of the real line, $h_{\text{att}}(x_1, \ldots, x_n) = h_{\text{sup}}(x_1, \ldots, x_n) = \Sigma_{i=1}^{n} \frac{x_i+1}{2}$, and $g(x, y) = \frac{1}{1+y} - \frac{1}{1+x}$ (so, we have $\alpha = 0$, $\beta = \infty$ and $g(\alpha, \alpha) = 0$).

The following table shows the values computed with both instances on some simple examples:

Example	with 1^{st} instance	with 2^{nd} instance
No attack, no support: a	$v(a) = 0$	$v(a) = 0$
Direct attack: $a \longrightarrow b$	$v(b) = -0.5$	$v(b) = -0.33$
Direct support: $a \rightsquigarrow b$	$v(b) = 0.5$	$v(b) = 0.33$
Defence: $a \longrightarrow b \longrightarrow c$	$v(c) = -0.25$	$v(c) = -0.25$
Sequence of supports: $a \rightsquigarrow b \rightsquigarrow c$	$v(c) = 0.75$	$v(c) = 0.4$
Supported attack: $a \rightsquigarrow b \longrightarrow c$	$v(c) = -0.75$	$v(c) = -0.4$

Example 41 on page 110 (cont'd) *With the first (resp. second) instance,* $v(t_1) = \frac{1}{4}$ *(resp.* $\frac{37}{154}$*).*

Example 45 on page 115 (cont'd) *With the first and the second instances,* $v(b) = 0$*. In this case, there is a perfect equilibrium*[51] *between support and attack.*

A local gradual valuation defined as above satisfies the following properties [CLS05a]:

- If $\mathcal{R}_{\text{att}}{}^{-}(a) = \mathcal{R}_{\text{sup}}{}^{-}(a) = \varnothing$ then $v(a) = g(\alpha, \alpha)$.
- If $\mathcal{R}_{\text{att}}{}^{-}(a) \neq \varnothing$ and $\mathcal{R}_{\text{sup}}{}^{-}(a) = \varnothing$ then $v(a) = g(\alpha, y) \leq g(\alpha, \alpha)$ for $y \geq \alpha$.
- If $\mathcal{R}_{\text{att}}{}^{-}(a) = \varnothing$ and $\mathcal{R}_{\text{sup}}{}^{-}(a) \neq \varnothing$ then $v(a) = g(x, \alpha) \geq g(\alpha, \alpha)$ for $x \geq \alpha$.

And we have the following comparative scale[52]:

$$\mathcal{V}_{\text{min}} \leq \underset{(\text{for } y \geq \alpha)}{g(\alpha, y)} \leq g(\alpha, \alpha) \leq \underset{(\text{for } x \geq \alpha)}{g(x, \alpha)} \leq \mathcal{V}_{\text{max}}$$

Moreover the valuation proposed in Definition 85 on the facing page satisfies the principles **P1** to **P3** (see [CLS05a] for a more detailed discussion).

[50] So, α is the minimal value for an attack (resp. a support) – *i.e.* there is no attack (resp. no support) –, and β is the maximal value for an attack (resp. a support).

[51] Note that it is not necessarily the case, and an appropriate choice of the function g enables to give more importance to the attack than to the support.

[52] Using this scale, the values $\leq$ (resp. $\geq$) to $g(\alpha, \alpha)$ are considered as negative (resp. positive) ones even if $g(\alpha, \alpha) \neq 0$.

3.3.5 Related issues and conclusion on bipolarity

In this section, an extension of [Dun95]'s abstract argumentation system has been proposed in order to take into account two kinds of interaction between arguments modelled with a support relation and an attack relation. In this abstract BAF, two issues have been considered:

- taking into account bipolarity for defining acceptability semantics: either by enforcing the coherence of the admissible sets, or by turning a BAF into a meta-argumentation system using the concept of coalition;
- taking into account bipolar interactions for proposing gradual labellings for the arguments.

Note that I have also done some other works in the bipolar framework (see for instance, the handling of controversial arguments in a BAF; this work has been realized with Claudette CAYROL and Caroline DEVRED – see [CDLS06a]).

Concerning the three main points evoked in this section (acceptability, coalitions and valuation in a bipolar framework), one can found in the literature some other works about these points.

3.3.5.1 Acceptability and bipolarity

Deflog [Ver03]: DEFLOG argumentation system enables to express a support or an attack between sentences in the language, with a new sentence using specific connectors (one for each kind of interaction). Examples of sentences (with $\rightarrow$ for the attack relation and $\rightsquigarrow$ for the support relation) are: a, b, $(a \rightsquigarrow b)$, $(a \rightarrow b)$, $(c \rightsquigarrow (a \rightsquigarrow b))$, $(d \rightarrow (a \rightsquigarrow b))$. In DEFLOG, the notions of sequence of supports and of supported attacks can be retrieved but at the language level (between sentences). Moreover, the notion of conflict-free set proposed in DEFLOG corresponds to the notion of safe set (no sentence which is, at the same time, supported and attacked by the set).

DEFLOG enables to define the dialectical interpretations (or extensions) of a given set of sentences S: an extension is built from a partition (J, D) of S such that J is conflict-free and attacks the sentences of D.

Note that the attack relation and the support relation are explicitly expressed in the sentences. So, one can have an extension of a set S s.t. some supported sentences by J do not belong to S. DEFLOG extensions correspond to [Dun95]'s stable extensions for DEFLOG theories that do not go beyond the expressiveness of Dung's argumentation systems, and note that a Dung's AF can always be expressed in DEFLOG. So in this precise sense, DEFLOG's extensions are a faithful generalization of Dung's stable extensions, allowing more expressiveness. Moreover, [Ver03] gives also a faithful generalization of Dung's preferred extensions.

Evidence-based argumentation [ON08]: In this work, the fundamental claim is that an argument cannot be accepted unless it is supported by evidence. So, special arguments are distinguished: the *prima-facie* arguments (which do not require any support to stand).

Arguments may be acceptable only if they are supported (indirectly) by *prima-facie* arguments; this is evidential support. Moreover, only supported arguments may attack other arguments.

Then, the notion of defence is rather complex: A set of arguments S defends an argument a if S provides evidential support for a and S invalidates each attack on a (either by a direct attack on the attacker of a or by rendering this attack unsupported).

Following our definitions, a BAF is an abstract system, where arguments may stand and attack with or without support. However, evidential reasoning as proposed by [ON08] could also be handled in a BAF in the following way: Given X a set of arguments (which are considered as *prima-facie* arguments in a given application), a notion of evidential support can be defined via a sequence of supports from an argument of X. Then, the notion of attack can be restricted so that attackers be elements of X, or receive evidential support from X. Finally, instead of choosing the classical definition for "S defends a" (as presented in Definition 78 on page 114), it can be required first that S provides support for a and secondly that for each supported attack on a, one argument of the sequence of supports is directly attacked by S.

3.3.5.2 Coalitions of arguments

Another way for defining acceptability semantics in a bipolar system is to turn a *bipolar* argumentation system into a *meta-argumentation system* . This transformation has the following characteristics: the support relation is used in order to identify "coalitions" (sets of arguments which can be used together without conflict and which are related by the support relation) and the attack relation is used in order to identify conflicts between coalitions and then to define new acceptability semantics as in Dung's framework.

The concept of coalition has already been related to argumentation.

Collective argumentation system [Boc03, NP06]: A collective argumentation system is an abstract system where the initial data are a set of arguments and a binary "attack" relation between *sets* of arguments. The key idea is the following: a set of arguments can produce an attack against other arguments, which is not reducible to attacks between particular arguments. That is in agreement with our notion of coalition, since in our work, a coalition is considered as a whole and its members cannot be used separately in an attack process. The proposal by Nielsen and Parsons is similar to Bochman's proposal. Both proposals take the attacks between sets of arguments as initial data, and define semantics by properties on subsets of arguments. However, Nielsen and Parsons propose an abstract system which allows sets of arguments to attack single arguments only, and they stick as close as possible to the semantics provided by Dung. In contrast, Bochman departs from Dung's methodology and give new specific definitions for stable and admissible sets of arguments. Our proposal essentially differs from collective argumentation in two points. First, we keep exactly Dung's construction for defining semantics, but we apply this construction in a meta-argumentation system (the coalition system). The second main difference lies in the meaning of a coalition: we intend to gather as many arguments as possible in a coalition, and a coalition cannot be broken in the defence process.

Generation of coalition structures in MAS [DJ04, Amg05]: In multi-agent systems (MAS), the coalition formation is a process in which independent and autonomous agents come together to act as a collective. A coalition structure (CS) is a partition of the set of agents into coalitions. Each coalition has a value (the utility that the agents in the coalition can jointly get minus the cost which this coalition induces for each agent). So the value of a CS is obtained by aggregating the values of the different coalitions in the structure. One of the main problems is to generate a preferred CS, that is a structure which maximizes the global value. Recently, [Amg05] has proposed an abstract system where the initial data are a set of coalitions equipped with a conflict relation. A preferred CS is a subset of coalitions which is conflict-free and defends itself against attacks. Coalitions may conflict for instance if they are non-disjoint or if they achieve a same task.

However, the generation of the coalitions is not studied in [Amg05]. So, one perspective is to apply our work to the formation of coalitions taking into account interactions between the agents. Arguments represent agents in that case. Indeed, it is very important to put together agents which want to cooperate ("supports" relation) and to avoid gathering agents who do not want to cooperate ("attacks" relation). Then, the concept of Cp-extension provides a tool for selecting the best groups of agents (w.r.t. the given interaction relations).

More generally, the work reported here is generic and takes place in abstract systems, since no assumption is made on the nature of the arguments. Arguments may have a logical structure such as a pair $\langle$explanation, conclusion$\rangle$, may just be positions advanced in a discussion, or may be agents interacting in a multi-agent system. All that we need is the bipolar interaction graph describing how the arguments under consideration are interrelated. We think that this generic work should stimulate discussion across boundaries.

3.3.5.3 Valuation and bipolarity

Most works about valuations of arguments take place in the basic framework. Some of them consider intrinsic valuations, which express to what extent an argument increases the confidence in the statement it promotes. Other approaches consider interaction-based valuations. These approaches usually differ in the set of values which are available.

However, very few works have been interested in valuations which handle both support and attack interactions. Most of these works have been developed for specific applications.

Medical applications: The most influential work has been proposed in HERMES system [KP01]. But there is no graduality (only two possible values with HERMES), and some parts of the interacting arguments are not taken into account for the computation of the value. See in Section 3.3.4.1 on page 119.

Valued maps of argumentations: The bipolar valuation in argumentation has been used for a collective annotation of documents.

Collective annotation models supporting exchange through discussion threads. A discussion thread is initiated by an annotation about a given document. Then, users can reply with annotations which confirm or refute the previous ones. Annotations are associated with a social validation which provides a synthetic view of the discussions. The purpose of this validation is to identify annotations which are globally confirmed by the discussion thread. It can also take into account an intrinsic value of the annotations.

In [CCCJ05, CCCJ07] a discussion thread is modelled by a BAF. The set of arguments contains the nodes of the thread. Pairs of the support (resp. attack) relation correspond to replies in the thread of the confirm (resp. refute) type. The social validation of a given annotation is computed with the local bipolar valuation. Moreover, the bipolar valuation procedure has been slightly modified in order to take into account an intrinsic value of each annotation.

3.4 Conclusion for Topic 3

Argumentation has been my main research topic for many years. It consists of at least five different aspects, all very rich in potentiality of research and existing results.

The first aspect is the building of arguments from an inconsistent information set. This point relates to an enlarged notion of consistency restoration, since arguments can be viewed as consistent subsets (even if there exist a number of methods for representing arguments among which several languages which are non logic-based). The second aspect is the identification of interactions between these arguments (generally we are only interested by binary interactions). I did not work on these two aspects because all my works in argumentation have used the following assumption: arguments and interactions are considered as given and the argumentation system is said to be abstract.

The two next aspects, valuation and selection, are the main pieces of my work and they have been studied for two kinds of argumentation systems: one with only one interaction representing conflicts (unipolar argumentation systems) and the other one with two interactions representing respectively conflicts and supports (bipolar argumentation systems).

Valuations can be obtained using different elements, for instance, preferences on the initial beliefs (the beliefs used for creating arguments). However my initial assumption (the argumentation system is abstract) prevents me from using this kind of information. I just have a directed graph in which vertices are arguments and edges representing binary interactions between arguments. So I have proposed valuations of arguments based on interactions for unipolar and bipolar argumentation systems. It is interesting to note that my proposition covers and even generalizes some other existing propositions.

For selection (called acceptability), I have also studied the two kinds of systems, unipolar and bipolar, following two methods. I have first used valuations for defining acceptability. And secondly, I have generalized Dung's extensions; Dung's idea was to select arguments by "consistent" subsets of arguments[53] satisfying some properties; so it is a collective acceptability (an argument is selected because it belongs to a set satisfying the given constraints). I have kept this idea of a collective acceptability and the list of the properties to satisfy, and I have enriched this list with new properties in order to take into account specificities of the studied systems, in particular bipolarity.

The last aspect of the argumentation process is essential in order to go back to Topic 1 "nonmonotonic inference", but I have not done anything yet along this line. This point will be also necessary for my research project, for comparing and for making an "axiomatisation" of argumentation processes (see Chapter 5 on page 151).

[53] This is not exactly consistency restoration because this selection is made at the argument level and not at the belief level (see Topics 1 and 2 – Chapters 1 on page 9 and 2 on page 19). Nevertheless, the idea is rather similar, since the complete set of argument is "inconsistent" because of conflicts/attacks between arguments. One could almost say, by abuse of language, that the approach suggested by Dung is "consistency restoration".

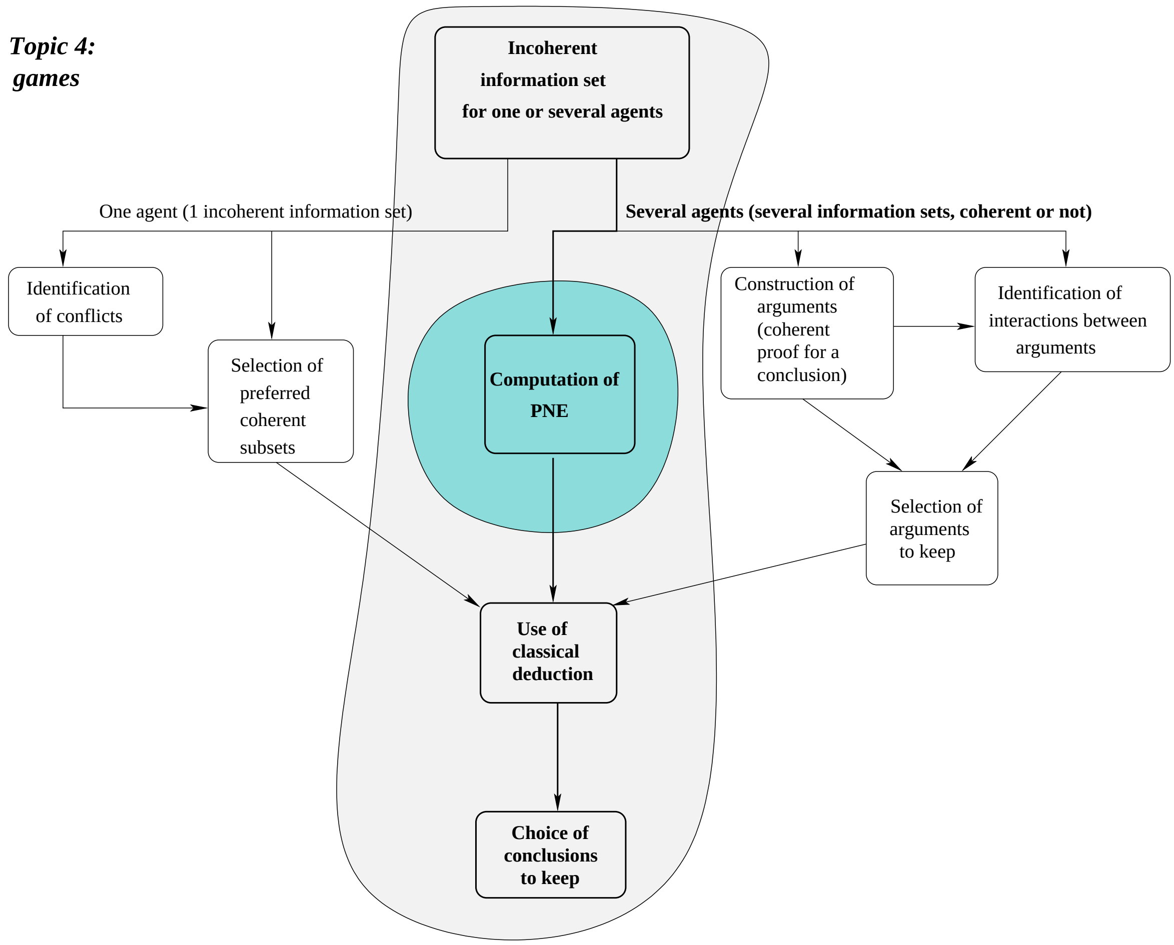
My works in argumentation also consist of the study of some mechanisms defined for nonmonotonic reasoning in the framework of argumentation, for instance, merging and revision. In the first case, I have proposed a merging mechanism of abstract unipolar argumentation systems using a new kind of interaction, “ignorance”, and therefore a new kind of argumentation systems, the “abstract partial argumentation systems”. In the second case, a study of the revision of an abstract unipolar argumentation system led us to identify and to characterize different types of revision.

The last points on which I am interested in this topic are close to “applications”, one is about the use of an argumentation system for practical reasoning and the other one is related to Topic 4 “games” (a translation of an argumentation system into a Boolean game); this work is presented in Chapter 4, Section 4.3.3 on page 145. There exist strong relations between Topics 3 and 4, because of the “exchange of information” which appears in argumentation processes and which is very close to the notion of game. The following chapter addresses this question.

Chapter 4

Topic 4: reasoning with games

In the following figure, Topic 4 is represented by the "clear surrounded" part and my works correspond to the "dark surrounded" part:



The study of games has been done in collaboration with Jérôme LANG, Élise BONZON, Bruno ZANUTTINI and Denis SIREYJOL and its starting point was the link with argumentation if we consider argumentation as an interaction process between agents.

Indeed, the main interest of Game Theory is the modelization of these interactions in many different domains (for instance, in Economy). This theory has been introduced by John Von Neumann and Oskar Morgenstern [vNM44], and John Nash [Nas50]. The basic assumptions of this framework are: each agent is a *player*, each player knows the other players, players are in *strategical interaction* because each player's decision depends on the other players' decisions, *players are rational* (*i.e.* each player tries to take the best decision for herself and knows that the other players make the same thing).

In this chapter, I do not give an exhaustive background of Game theory, but I just introduce some very useful basic concepts in Section 4.1: a partial taxonomy of games, different types of representation for games, and some solution concepts.

In Section 4.2 on page 130, I present some examples of links between argumentation and games.

Section 4.3 on page 132 describes a particular kind of games, Boolean games, which are interesting for us because:

- they use classical logic for modelling interactions between agents,
- they can be generalized using preference representations defined in Topic 1 "nonmonotonic inference",
- and they give a new method for exploiting the link between argumentation and games.

4.1 Basic concepts of Game Theory

Succinctly, a game is a set of players, a set of strategy profiles (*i.e.* a vector of strategies, one strategy for each player representing a possible choice for her) and a utility function which gives for each player her profit according to each strategy profile. Each strategy profile is called an issue of the game.

4.1.1 Game taxonomy

There exist at least 4 types of games, static games, dynamic games, cooperative games and non-cooperative games, knowing that a game can share several of these characteristics.

A game is **static** if players choose their strategies simultaneously.

A game is **dynamic** if it proceeds in **several steps**. If we assume that **games are dynamic with perfect information**[1], *i.e.* that all *past* players' choices are *observable* and *known* for all players, then it is possible for a player to directly and definitively modify the profits of the other players while making an intervention on previous steps of the game (this notion of perfect information does not make sense for static games).

A game is **cooperative** if players can make agreements. In this case, they make **coalitions**. In the opposite case, if players cannot make coalitions, the game is **non-cooperative**.

A cooperative game has **transferable utilities** if it is possible to add players' utilities and to distribute this sum to members of a coalition (there exists a "common currency" with which one can make transfers).

Non-cooperative games can be divided in two cases: zero-sum games and non-zero-sum games (in economy, this concept of zero-sum game is important, because it corresponds to the absence of production, or destruction, of products). **Zero-sum games** are games in which the "algebraical" sum of players' profits is a constant: what a player gains is necessarily lost by another player.

[1] which is not the same thing as complete information: in complete information, at *each step of the game*, each player knows her strategies, other players' strategies, profits of each strategy profiles and other players' motivations.

4.1.2 Some examples

The **prisoner's dilemma** is a famous example in Game Theory.

Example 50 *Two suspects are arrested by the police. The police has insufficient evidence for a conviction, and, having separated both prisoners, visit each of them to offer them the same deal.*

- *If one testifies (defects) for the prosecution against the other and the other remains silent, the betrayer goes free and the silent accomplice receives the full 10-year sentence.*
- *If both remain silent, both prisoners are sentenced to only six months in jail for a minor charge.*
- *If each betrays the other, each receives a five-year sentence.*

Each prisoner must choose to betray the other or to remain silent. Each one is assured that the other will not know about the betrayal before the end of the investigation.

This problem can be formalized by a two-player game, each player having two possible strategies: to cooperate with (denoted by T) or to defect from (i.e., betray) the other player (denoted by A).

- *The set of players is $N = \{1, 2\}$,*
- *Prisoner 1 has two possible strategies: $s_{1_1} = A$ and $s_{1_2} = T$. Idem for Player 2: $s_{2_1} = A$ and $s_{2_2} = T$.*
- *So there are 4 strategy profiles: AA, AT, TA and TT.*
- *And the utility functions are the following:*
  - $u_1(AA) = u_2(AA) = -5$,
  - $u_1(TT) = u_2(TT) = -0.5$,
  - $u_1(AT) = u_2(TA) = 0$,
  - $u_1(TA) = u_2(AT) = -10$.

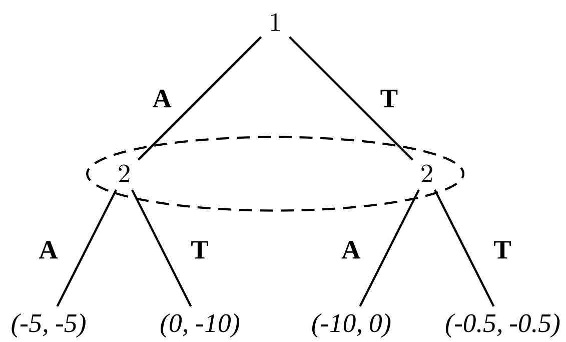
This game is static (only one step), non-cooperative and it is a non-zero-sum game: the two prisoners cannot communicate, so they cannot make a coalition, but what a prisoner gains is not always lost by the other one.

4.1.3 Game representation

There exist two possible but equivalent representations for strategical game: normal form or extensive form.

The **extensive form** of a game is a **decision tree** describing the possible strategies of each player at each step of the game. A node of this tree specifies the current player, a branch corresponds to a strategy profile and a leaf gives the profit of each player for the corresponding strategy profile. For instance, the extensive form of the game for the prisoner's dilemma is:

Example 50 (cont'd)



In the leaves, the first element of each pair represents Player 1's utility, and the second one represents Player 2's utility.
The dotted circle surrounding the two occurrences for Player 2 means that this player does not know in which case she is (she does not know if her accomplice has chosen to cooperate with her or to betray her).

This game has another extensive form in which the root-node corresponds to the second player. These two representations are equivalent.

The **normal form** of a game gives the set of players, the set of strategies of each player and the profits for each possible combination of strategies. This corresponds to a **matrix form** which associates with each strategy profile s a n-tuple giving the utility obtained by each player: $(u_1(s), u_2(s), \ldots, u_n(s))$. For instance, the normal form for the prisoner's dilemma is:

Example 50 on the previous page (cont'd)

1 \ 2	A	T
A	*(-5, -5)*	*(0, -10)*
T	*(-10, 0)*	*(-0.5, -0.5)*

4.1.4 Solution concepts

There exist many solution concepts but in this document, I just present Nash equilibria for static non-cooperative games.

Nash equilibrium, introduced by John Nash in 1950 [Nas50], is a fundamental solution concept in Game Theory. It describes an issue of the game in which no player wishes to modify her strategy while being given the strategy of each other player. So a Nash equilibrium is a strategy profile where no player may find it beneficial to deviate if it assumes that the other players will not deviate either. There exist several versions of this concept: pure-strategy Nash equilibrium (PNE) or mixed-strategy Nash equilibrium (using probabilities). In this document I use only the first version.

Example 50 on the preceding page (cont'd) *Strategy profile AA is a pure-strategy Nash equilibrium of prisoner's dilemma problem. Indeed, one can check that: $u_1(AA) \geq u_1(TA)$ and $u_2(AA) \geq u_2(AT)$.*

This is the only PNE of this game. Indeed, AT cannot be a PNE because AA is a better strategy for Player 2: $u_2(AA) > u_2(AT)$; it is the same thing for TA (because AA is a better strategy for Player 1: $u_1(AA) > u_1(TA)$) and for TT (because AT is a better strategy for Player 1: $u_1(TT) < u_1(AT)$).

Neither the existence, nor the unicity of a PNE are guaranteed.

4.2 Argumentation and Game Theory

The first links between argumentation and games have been identified by Dung himself in [Dun95]. Then they have been used by many people for defining proof theories for argumentation.

4.2.1 Dung's work

[Dun95] uses concepts of Game Theory developed by Von Neumann and Morgenstern in [vNM44] in order to present some links between argumentation and cooperative games.

Among these concepts, he has used the notion of "dominance": an issue s' dominates another one s if and only if there exists a non-empty coalition such that the profits of the members of the coalition are greater with s' than with s and in which the issue s is feasible.

[vNM44] has defined a NM-solution of a cooperative game as a set of issues S satisfying two postulates: (1) no issue of S dominates another issue of S, and (2) every issue which does not belong to S is dominated by at least one issue of S.

The "argumentative" version of this game is the following: in their own interest, players try to impose the coalition which optimizes their profits; each issue of the game is potentially an argument in favour of the coalition; moreover, an issue s attacks another one s' if s dominates s'; so one finds the notions of arguments and of attack relation.

From a cooperative game, Dung has proposed a particular argumentation system whose set of arguments is the set of the issues of the game and whose attack relation is the domination relation. And he has shown that the NM-solutions of a cooperative game are exactly the stable extensions of this particular argumentation system.

Then he has applied these results to the stable marriage problem which is a well-known cooperative game: consider a set of n men and a set of n women; each one has a strictly ordered list (representing preferences) containing all the members of the opposite sex; the stable marriage problem consists in finding the best way of marrying all the men and all the women in order to satisfy their preference criteria as well as possible.

In [Dun95], Dung has also identified some other links between argumentation and cooperative game, but I do not present them in this document.

Generally, Dung has been interested by the use of argumentation (particularly the notion of extensions) in order to study some kinds of games. Indeed, he has considered the possible issues of the game as arguments and he has used the concepts of extension and acceptability for extracting the solutions of the game.

4.2.2 Dialectical proof theory

The notion of game appears as soon as one defines a proof theory for argumentation problems.

The main idea is to "prove" that if an argument x is acceptable or not by the simulation of a "dialogue" between 2 agents: a proponent PRO which supports the acceptability of x and an opponent OPP which supports the non-acceptability of x. So this dialogue is a **two-player game**, in which the winner "has the last word". These ideas have been used in [JV99a] and have been refined in [Dou02].

This type of dialogue implies that each argument is significant, and this depends on the problem we want to solve. The *legal-move function* defines, at every step, what moves can be used to continue the dialogue given the previous moves. When the set of arguments returned by the legal-move function is empty, the dialogue cannot be continued.

It is also important to conclude the dialogue and to know who won: this is the role of the *winning criterion*. For instance, for the acceptability problem of an argument x, at least 2 winning criteria are possible:

1. there exists a *won dialogue* for PRO concerning x.

2. there exists a *winning strategy* for x, *i.e.* a way for PRO to defend x against all the attacks of OPP.

So if we want to decide the acceptability of an argument under a given semantics, we need to build a proof theory whose legal-move function and winning criterion characterize the chosen acceptability and semantics.

In [Dou02], this work has been done for the credulous acceptability under grounded and preferred semantics (and also in [JV99a] for other semantics which I do not present in this document).

On this subject, I have worked with Jérôme LANG, Sylvie DOUTRE and Denis SIREYJOL to identify some of the characteristics of a game corresponding to a dialectical proof theory for argumentation (see[Sir04]). This work has been applied to the dialectical proof theories proposed in [Dou02].

4.3 Boolean games

After my first works on games, I have tried to find a category of games which are close to logical languages and with which one can realize a nonmonotonic reasoning and go back to Topic 3 "argumentation". The main interest of this approach is the fact that the use of a logical language allows an important reduction of the size of the game. Indeed, for static games, extended forms and normal forms coincide and utility functions are usually represented explicitly, by listing the values for each combination of strategies. However, the number of utility values, that is, the number of possible combinations of strategies, is exponential in the number of players, which renders such an explicit way of representing the preferences of the players unreasonable when the number of players is not very small. This becomes even more problematic when the set of strategies available to an agent consists in assigning a value from a finite domain to each of a given set of variables (which is the case in many real-world domains). In this case, representing utility functions explicitly leads to a description whose size is exponential both in the number of agents ($n \times 2^n$ values for n agents each with two available strategies) and in the number of variables controlled by the agents ($2 \times 2^p \times 2^p$ values for two agents each controlling p Boolean variables). Thus, in all these cases, specifying players' preferences explicitly is clearly unreasonable, both because it would require exponential space, and because studying these games (for instance by computing solution concepts such as pure-strategy Nash equilibria) would require accessing all of these utility values at least once and would take time exponential in the numbers of agents and variables in all cases.

So we have chosen to study **Boolean games** introduced by Harrenstein, van der Hoek, Meyer and Witteveen in [HvdHMW01, Har04b]. They are two-player and zero-sum games in which Player 1's utility function (and Player 2's utility function which is its opposite) is represented by a formula in propositional logic, called the *Boolean form* of the game. This game is very compact to represent, but it is also very simple. So our first work has been to enrich it on different aspects.

Note that the main part of this work has been done by Elise BONZON (the subject of her PHD Thesis was "Modelization of interactions between rational agents: Boolean games"). I have co-supervised her work with Jérôme LANG, but the generalization of Boolean games to n players and to non-dichotomous preferences is completely Elise's work. Then, after her PHD Thesis, I have worked with her and Caroline Devred (LERIA, Angers) to the link between Boolean games and argumentation (see Section 4.3.3 on page 145).

4.3.1 Boolean games with n players

This is a generalization of the two-player version proposed by [HvdHMW01, Har04b].

Let $V = \{a, b, \ldots\}$ be a finite set of propositional variables, a Boolean game with n players on V is a n-player game in which each strategy of each player consists in assigning a truth value to all variables belonging to a subset of V. Each player's preferences are given by a propositional formula φ_i over variables of V.

Definition 86 *A* **Boolean game with n players** *is a 5-tuple* $(N, V, \pi, \Gamma, \Phi)$*, with*

- $N = \{1, 2, \ldots, n\}$ *the set of players (also called agents);*
- V *a set of propositional variables;*
- $\pi : N \mapsto V$ *a control assignment function which defines a partition of* V*;*
- $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ *a set of constraints, each* γ_i *being a satisfiable propositional formula of* $L_{\pi(i)}$[2]*;*
- $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ *a set of goals, each* φ_i *being a satisfiable propositional formula of* L_V*.*

A 4-tuple (N, V, π, Γ)*, with* N, V, π, Γ *defined as previously is called a* **Boolean pre-game**.

[2] The notation L_S denotes the subset of L defined on the set of propositional variables S, L being a propositional logical language.

The *control assignment function* π associates with each player all variables she controls. π_i denotes the *set of controlled variables* by Player i. Each variable is controlled by one and only one player. So, $\{\pi_1, \ldots, \pi_n\}$ is a partition of V.

Each γ_i represents the agent's constraints on the set of variables she controls. This representation choice respects agents' independence: each agent handles her variables, and her constraints which concern her variables, without depending on other agents.

The use of Boolean games allows a very compact representation of games. This point is illustrated by the following example which is a simplified variant of the prisoner's dilemma problem.

Example 51 *Consider n prisoners and only two kinds of sentences, freedom and jail*[3]*. So the deal proposed by the police is:*

- *Those which defect for the prosecution against the others go free and the ones which remain silent receive the full 10-year sentence.*
- *If everyone remains silent,everyone goes free.*

Each prisoner must choose to betray the other or to remain silent. Each one is assured that the other ones will not know about the betrayal before the end of the investigation.

For 3 prisoners, this problem can be formalized by a 3-player game, each player i having two possible strategies: to cooperate with (denoted by T_i, $i = 1, \ldots, n$) or to defect from (i.e., betray) the other players (denoted by $\neg T_i$, $i = 1, \ldots, n$). The normal form of this game is:

$3 : T_3$		
1 \ 2	T_2	$\neg T_2$
T_1	(1, 1, 1)	(0, 1, 0)
$\neg T_1$	(1, 0, 0)	(1, 1, 0)

$3 : \neg T_3$		
1 \ 2	T_2	$\neg T_2$
T_1	(0, 0, 1)	(0, 1, 1)
$\neg T_1$	(1, 0, 1)	(1, 1, 1)

The n-tuples give the result obtained by each player: (Player 1's result, Player 2's results, ...). The value 0 (resp. 1) means that the player loses (resp. wins).

So for n prisoners, we need a matrix with n dimensions, each dimension being equal to 2; so we need to specify 2^n n-tuples. However, this game can be very easily translated into a Boolean game $G = (N, V, \pi, \Gamma, \Phi)$ with:

- $N = \{1, 2, \ldots, n\}$,
- $V = \{T_1, \ldots, T_n\}$,
- $\forall i \in \{1, \ldots, n\}, \pi_i = \{T_i\}$,
- $\forall i \in \{1, \ldots, n\}, \gamma_i = \top$, *and*
- $\forall i \in \{1, \ldots, n\}, \varphi_i = (T_1 \wedge T_2 \wedge \ldots \wedge T_n) \vee \neg T_i$.

Definition 87 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean game. A* **strategy** *s_i for Player i of G is a π_i-interpretation satisfying γ_i. The* **set of strategies for Player** *i is represented by $S_i = \{s_i \in 2^{\pi_i} | s_i \models \gamma_i\}$.*

A **strategy profile** *s for G is a n-tuple $s = (s_1, \ldots, s_n)$, with for all i, $s_i \in S_i$. $S = S_1 \times \ldots \times S_n$ is the set of strategy profiles.*

[3] In Section 4.3.2 on page 135, the use of non-dichotomous preferences allows a more interesting translation of this example, with several kinds of sentences.

So a strategy for Player i is the assignment from true or false to the variables she controls and the constraint γ_i reduces the set of possible strategies for this player.

For a non-empty set of players (called **coalition**) $I \subseteq N$, the projection of s on I is defined by $s_I = (s_i)_{i \in I}$. If $I = \{i\}$, the projection of s on $\{i\}$ is denoted by s_i in place of $s_{\{i\}}$.

The following notations are usual in Game Theory. Let $s = (s_1, \ldots, s_n)$ and $s' = (s'_1, \ldots, s'_n)$ be two strategy profiles.

- s_{-i} denotes the *strategy profile s without Player i's strategy*: $s_{-i} = (s_1, s_2, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$. Similarly, if $I \subseteq N$, s_{-I} denotes $s_{N \setminus I}$.
- (s_{-i}, s'_i) denotes the *strategy profile s in which Player i's strategy has been replaced by that of profile s'*: $(s_{-i}, s'_i) = (s_1, s_2, \ldots, s_{i-1}, s'_i, s_{i+1}, \ldots, s_n)$.
- π_I represents the set of variables controlled by I, and $\pi_{-I} = \pi_{N \setminus I}$.
- If $I = \{i\}$, π_{-i} denotes the *set of variables controlled by all players except Player i*: $\pi_{-i} = V \setminus \pi_i$.
- $\{\pi_1, \ldots, \pi_n\}$ being a partition of V, a strategy profile s is an interpretation for V, *i.e.* $s \in 2^V$.
- π^{-1} represents the inverse function of π.
- The set of strategies of $I \subseteq N$ is denoted by $S_I = \times_{i \in I} S_i$.
- And the set of the goals of $I \subseteq N$ is denoted by $\Phi_I = \bigwedge_{i \in I} \varphi_i$.

All the previous notions are illustrated with the following example:

Example 52 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean game with*

- $V = \{a, b, c\}$, $N = \{1, 2\}$,
- $\pi_1 = \{a, b\}$ *and* $\pi_2 = \{c\}$.
- $\gamma_1 = \neg a \vee \neg b$, $\gamma_2 = \top$
- $\varphi_1 = (a \leftrightarrow b) \vee (\neg a \wedge b \wedge \neg c)$,
- $\varphi_2 = (\neg a \wedge b \wedge c) \vee (a \wedge \neg b)$,

Player 1 has three possible strategies: $s_{1_1} = a\overline{b}$, $s_{1_2} = \overline{a}b$, $s_{1_3} = \overline{a}\overline{b}$. Strategy ab does not satisfy Player 1's constraints, so it is not an acceptable strategy.

Player 2 has two possible strategies: $s_{2_1} = c$ or $s_{2_2} = \overline{c}$.

So there are 6 strategy profiles for G : $S = \{a\overline{b}c, a\overline{b}\overline{c}, \overline{a}bc, \overline{a}b\overline{c}, \overline{a}\overline{b}c, \overline{a}\overline{b}\overline{c}\}$.

With strategy profiles $\overline{a}\overline{b}c$, $\overline{a}b\overline{c}$ and $\overline{a}\overline{b}\overline{c}$ Player 1 wins, whereas, with $a\overline{b}c$, $a\overline{b}\overline{c}$ and $\overline{a}bc$ the winner is Player 2.

Player i's goal φ_i is a compact and dichotomous preference relation, corresponding to a binary utility function[4]: Player i is satisfied (so her utility is equal to 1) if and only if her goal φ_i is satisfied[5]. Otherwise, her utility is equal to 0. So goals $\{\varphi_i, i = 1, \ldots, n\}$ play the same role as utility functions.

Definition 88 *For Player i, the* **utility function** *induced by the goal of this player is denoted by $u_i : S \to \{0, 1\}$ and defined by:*

$$u_i(s) = \begin{cases} 0 \; if \; s \models \neg\varphi_i \\ 1 \; if \; s \models \varphi_i \end{cases}$$

[4] In Section 4.3.2 on the next page, we will present a way for removing the restriction about this binary preference.

[5] In the logical sense of this word.

So we have:

- s is at least as good as s' for i, denoted by $s \succeq_i s'$, if $u_i(s) \geq u_i(s')$, or equivalently if $s \models \neg\varphi_i$ implies $s' \models \neg\varphi_i$;
- s is strictly better than s' for i, denoted by $s \succ_i s'$, if $u_i(s) > u_i(s')$, or equivalently if $s \models \varphi_i$ and $s' \models \neg\varphi_i$.
- i is indifferent between s and s', denoted by $s \sim_i s'$, if $s \geq_i s'$ and $s' \geq_i s$, or equivalently if $s \models \varphi_i$ if and only if $s' \models \varphi_i$.

The next definition describes the notion of winning strategy for a player:

Definition 89 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean game, with $\Phi = \{\varphi_1, \ldots, \varphi_n\}$, and $N = \{1, \ldots, n\}$. The strategy s_i is a* **winning strategy** *for Player i if, whatever the other players' choices, i wins by choosing this strategy.*

$$\forall s_{-i} \in S_{-i}, (s_{-i}, s_i) \models \varphi_i$$

For Boolean games, the definition of PNE is exactly the same as the classical one given in Game Theory:

Definition 90 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean game, with $N = \{1, \ldots, n\}$ being the set of players. $s = (s_1, \ldots, s_n)$ is a* **pure-strategy Nash equilibrium** *(PNE) if and only if:*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, u_i(s) \geq u_i(s_{-i}, s'_i)$$

A simple characterization of PNE exists in the framework of Boolean games:

Proposition 75 *Let $s \in S$. s is a PNE of G if and only if for all $i \in N$:*

- *either $s \models \varphi_i$,*
- *or $s_{-i} \models \neg\varphi_i$.*

This characterization gives an easy way for computing PNE for Boolean games: a strategy profile s will be a PNE if and only if, for each Player i, either s satisfies Player i's goal, or Player i's goal cannot be satisfied if the other players keep their strategies.

4.3.2 Boolean games with non-dichotomous preferences

Binary utilities is a real loss of generality. So in this section I present a generalization of the n-player Boolean games, defined previously, in order to incorporate non-dichotomous preferences.

There exist different kinds of non-dichotomous preferences: they can be either numerical (one says *cardinal*), or *ordinal*. Knowing that the essential notion for our work is Nash equilibrium and that this notion can be defined using ordinal preferences, we have chosen to integrate ordinal preferences in Boolean games.

The second important point is the choice of a representation for these non-dichotomous preferences: we always need a compact representation. Among the different possibilities proposed in literature, I have chosen to present two cases: CP-nets and goals with priority.

The use of non-dichotomous preferences implies some modified definitions:

Definition 91 *Let L be a propositional language for a compact representation of preferences. A* **Boolean L-game** *is defined by a 5-tuple $G = (N, V, \pi, \Gamma, \Phi)$, with*

- *$N = \{1, \ldots, n\}$, V, π and Γ being defined as previously, and*

- $\Phi = \langle \Phi_1, \ldots, \Phi_n \rangle$. *For Player* i, Φ_i *is a compact representation in* L *of the preference relation* $\succeq_i$ *for Player* i *on* S. $Pref_G$ *denotes* $\langle \succeq_1, \ldots, \succeq_n \rangle$.

Definition 92 *Let* L *be a propositional language for a compact representation of preferences.*

Let $G = (N, V, \Gamma, \pi, \Phi)$ *be a Boolean* L*-game, and* $Pref_G = \langle \succeq_1, \ldots, \succeq_n \rangle$ *be the set of preferences (one for each player).*

There are two possible definitions for Nash equilibria:

$s = (s_1, \ldots, s_n)$ *is a* **weak pure-strategy Nash equilibrium (WPNE)** *if and only if:*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, (s'_i, s_{-i}) \not\succ_i (s_i, s_{-i}) \tag{4.1}$$

$s = (s_1, \ldots, s_n)$ *is a* **strong pure-strategy Nash equilibrium (SPNE)** *if and only if:*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, (s'_i, s_{-i}) \preceq_i (s_i, s_{-i}) \tag{4.2}$$

The set of strong (resp. weak) pure-strategy Nash equilibria is denoted by NE_{strong} *(resp.* NE_{weak}*).*

4.3.2.1 CP-nets

CP-nets define a so-called "graphical" representation language. This language is based on the comparison criterion *Ceteris Paribus*: if an agent expresses in natural language a preference such that "a round table will be better in the living room than a square table", she does not want to say that any round table would be better that any square table; she want to express the fact that, she prefers a round table to a square table if they do not significantly differ on their other characteristics. This is the *Ceteris Paribus* principle which leads to the following notion of independence:

Definition 93 *Let* X, Y *and* Z *be three non-empty sets forming a partition of* V. X *and* Y *are* **conditionally preferentially independent given** Z *if and only if* $\forall z \in D(Z)$, $\forall x_1, x_2 \in D(X)$ *and* $\forall y_1, y_2 \in D(Y)$ *we have:*

$$x_1 y_1 z \succeq x_2 y_1 z \text{ if and only if } x_1 y_2 z \succeq x_2 y_2 z$$

For a given value of Z, the preference relation on the values of X is the same whatever the values of Y.

This conditional preferential independence is used in the CP-nets introduced in [BBHP99] as a tool for compactly representing qualitative preference relations. CP-nets have been studied mainly in [Dom02], [BBD$^+$04a] and [BBD$^+$04b].

They can be used in the framework of Boolean games in order to represent players' preferences: each goal for each player will be a "propositionalized" CP-net (*i.e.* CP-net with binary variables). So $\forall x_i \in V$, $D(x_i) = \{x_i, \overline{x_i}\} = 2^{x_i}$ and $D(\{x_1 \ldots x_p\}) = 2^{\{x_1, \ldots, x_p\}}$. With this representation, an element of $D(x_i)$ corresponds to a $\{x_i\}$-interpretation, and an element of $D(\{x_1 \ldots x_p\})$ is a $\{x_1 \ldots x_p\}$-interpretation.

Definition 94 *For each variable* $X \in V$, *a set of* **parent variables** *is specified, and denoted by* $Pa(X)$. *These variables are those which influence the preferences of the agent about the different values for* X. *Formally,* X *and* $V \setminus (\{X\} \cup Pa(X))$ *are mutually conditionally preferentially independent given* $Pa(X)$.

The **conditional preference table** *(called CPT) describes the agent's preferences about the values of the variable* X, *with regard to combinations of values for the parent variables.*

For each combination of values for $Pa(X)$, $CPT(X)$ *specifies a* complete ordering *on* $D(X)$ *such that* $\forall x_i, x_j \in D(X)$ *either* $x_i \succ x_j$, *or* $x_j \succ x_i$.

Let $V = \{X_1, \ldots, X_n\}$ *be a set of variables.* $\mathcal{N} = \langle \mathcal{G}, \mathcal{T} \rangle$ *is a* **CP-net** *on* V, $\mathcal{G}$ *being a directed graph on* V, *and* $\mathcal{T}$ *being a set of conditional preference tables* $CPT(X_i)$ *for each* $X_i \in V$. *Each conditional preference table* $CPT(X_i)$ *is associated with a complete ordering* $\succ^i_p$, *with regard to* $p \in D(Pa(X_i))$.

Formally, the preference relation induced by the CP-net $\mathcal{N}$, represented by the induced preference graph, is defined by:

Definition 95 *The* **preference relation** *induced by the CP-net $\mathcal{N}$ is denoted by $\succ_{\mathcal{N}}$, and is defined by $\forall o, o' \in D(V)$, $o \succ_{\mathcal{N}} o'$ if and only if $\mathcal{N} \models o \succ o'$.*

All these notions are illustrated on the following example:

Example 53 *Consider the CP-net given by Figure 4.1 about my preferences for the dinner. Variables S and V correspond respectively to the soup and the wine. I strictly prefer to eat a fish soup (S_p) rather than a vegetable soup (S_l), and about wine, my preferences depend on the soup I eat: I prefer red wine (V_r) with vegetable soup ($S_l : V_r \succ V_b$) and white wine (V_b) with fish soup ($S_p : V_b \succ V_r$). So $D(S) = \{S_p, S_l\}$ and $D(V) = \{V_r, V_b\}$.*

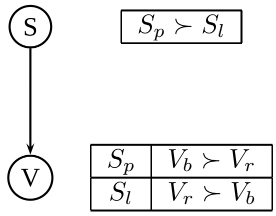


Figure 4.1: CP-net “My dinner”

Figure 4.2 represents the preference relation induced by this CP-net. The bottom element ($S_l \wedge V_b$) is the worst case and the top element ($S_p \wedge V_b$) is the best case.

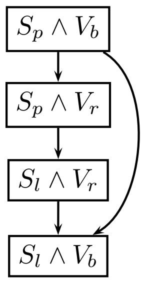


Figure 4.2: Preference graph induced by the CP-net “My dinner”

There is an arrow between the nodes $(S_p \wedge V_b)$ and $(S_l \wedge V_b)$ because we can compare these states, every other thing being equal.
So we can completely order the possible states (from the most preferred one to the least preferred one) :

$$(S_p \wedge V_b) \succ (S_p \wedge V_r) \succ (S_l \wedge V_r) \succ (S_l \wedge V_b)$$

This relation $\succ$ is the only ranking that satisfies this CP-net.

The introduction of CP-nets in Boolean games modifies some definitions:

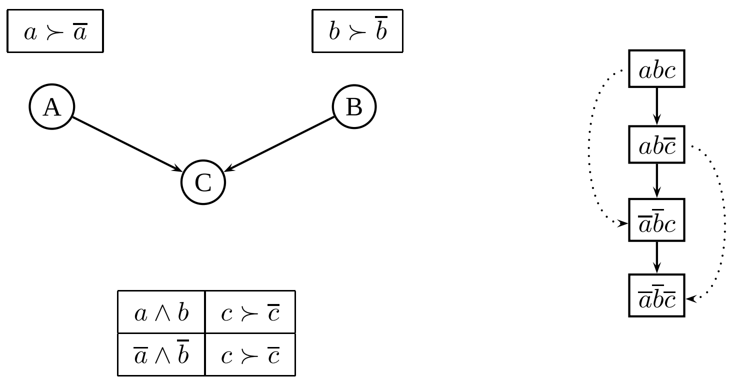
Definition 96 *The* **conditional preference table** *for a Boolean game (denoted by $CPT_i(X)$) describes Player i's preferences on the values for the variable X with regard to combinations of values for parent variables.*
For each combination of values p for $Pa_i(X)$, $CPT_i(X)$ specifies a complete ordering such that either $x \succ_{i,p} \overline{x}$, or $\overline{x} \succ_{i,p} x$.

Definition 97 *A* **Boolean CP-game** *is a 5-tuple $G = (N, V, \Gamma, \pi, \Phi)$, with $N = \{1, \dots, n\}$ a set of players, $V = \{x_1, \dots, x_n\}$ a set of variables, $\Phi = \langle \mathcal{N}_1, \dots, \mathcal{N}_n \rangle$, each $\mathcal{N}_i$ being a CP-net on V whose graph is denoted by $\mathcal{G}_i$ and for all $i \in N$, $\succeq_i = \succeq_{\mathcal{N}_i}$.*

The following example shows the possible use of these notions:

Example 54 *Consider the Boolean CP-game $G = (N, V, \Gamma, \pi, \Phi)$ with:*

- $N = \{1, 2\}$
- $V = \{a, b, c\}$, *with* $D(a) = \{a, \overline{a}\}$, $D(b) = \{b, \overline{b}\}$ *and* $D(c) = \{c, \overline{c}\}$.
- $\gamma_1 = \{a \leftrightarrow b\}$, $\gamma_2 = \top$,
- $\pi_1 = \{a, b\}$, $\pi_2 = \{c\}$,
- *Player* 1*'s goal is represented by the CP-net and the induced preference relation given in Figure 4.3,*
- *Player* 2*'s goal is represented by the CP-net and the induced preference relation given in Figure 4.4 on the next page,*



(a) (b)

Figure 4.3: CP-net and associated preordering for Player 1

Note that Player 1 *does not express her preferences on* c *for the states* $\overline{a}b$ *or* $a\overline{b}$*, these two states being impossible, but it is possible to build the preference relation by transitivity: states containing* ab *are always preferred to states containing* $\overline{a}\overline{b}$*.*

For computing Nash equilibria of this game, one only takes into account the strategy profiles appearing in the preference relations for both players: for instance, the strategy profile $\overline{a}b\overline{c}$ *does not satisfy Player* 1*'s constraints so it cannot be a PNE. On this example, the strong and weak PNE are:*

$$NE_{weak} = NE_{strong} = \{abc\}$$

These Boolean CP-games have some interesting properties, in particular if their CP-nets are acyclic.

Proposition 76 *In an acyclic Boolean CP-game, strong and weak Nash equilibria coincide.*

Proposition 77 *Let* $G = (N, V, \pi, \Gamma, \Phi)$ *be a Boolean CP-game such that graphs* $\mathcal{G}_i$ *are all identical (*$\forall i, j$, $\mathcal{G}_i = \mathcal{G}_j$*) and acyclic. This game* G *has one and only one PNE.*

We want to exploit this property by making all CP-nets identical. For that, we use the notions of graph union and equivalent game.

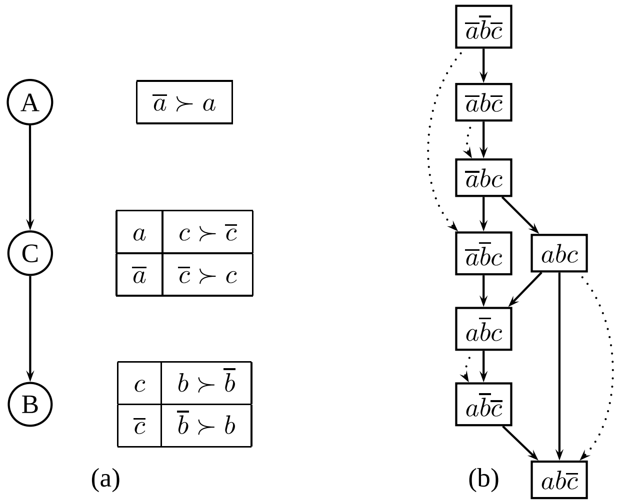


Figure 4.4: CP-net and associated preordering for Player 2

Definition 98 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean CP-game such that $\forall i \in N$, graphs $\mathcal{G}_i$ are all acyclic.*

For Player i, $\mathcal{G}_i = \langle V, Arc_i \rangle$, V being the set of the nodes of the graph[6]*, and Arc_i representing the set of the directed edges of the CP-net graph for i. The* **graph union** *of G is the graph $\mathcal{G} = \langle V, Arc_1 \cup \ldots \cup Arc_n \rangle$.*

Definition 99 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean CP-game such that $\forall i \in N$, graphs $\mathcal{G}_i$ are all acyclic.*

$G^* = (N, V, \pi, \Gamma, \Phi^*)$ is the **equivalent game by rewriting of** *G in which the CP-net graph of each player has been replaced by the graph union of G.*

The conditional preference tables are modified in order to correspond with the new graph, while giving the same preferences: if the edge (X, Y) is added to the graph, the conditional preference table of the variable Y will be the same table as previously for each value $x \in D(X)$. More formally, with $\succ_i^y$ denoting the associated relation with $CPT_i(Y)$ for Player i's CP-net in G, we have for G^*: $\forall x \in D(X), \succ_{i,x}^y = \succ_{i,\overline{x}}^y = \succ_i^y$.

Proposition 78 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean CP-game such that $\forall i \in N$, graphs $\mathcal{G}_i$ are all acyclic. Let $G^* = (N, V, \pi, \Gamma, \Phi^*)$ be the equivalent game by rewriting of G. G^* gives the same preorderings on strategy profiles as G. One says that G* **and** *G^** **are equivalent**.

The following example gives an illustration of all these concepts:

Example 55 *Consider the game $G = (N, V, \pi, \Gamma, \Phi)$ with:*

- $N = \{1, 2\}$
- $V = \{a, b, c\}$, *with* $D(a) = \{a, \overline{a}\}$, $D(b) = \{b, \overline{b}\}$ *and* $D(c) = \{c, \overline{c}\}$.
- $\gamma_1 = \gamma_2 = \top$,
- $\pi_1 = \{a, b\}, \pi_2 = \{c\}$,
- *Player 1's goal is given on Figure 4.5 on the following page,*
- *Player 2's goal is given on Figure 4.6 on the next page.*

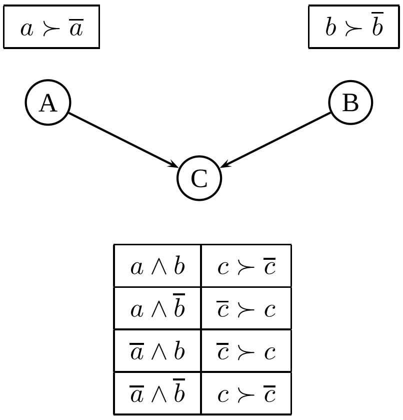


Figure 4.5: Player 1's CP-net

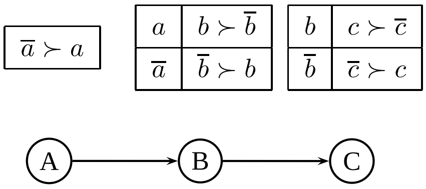
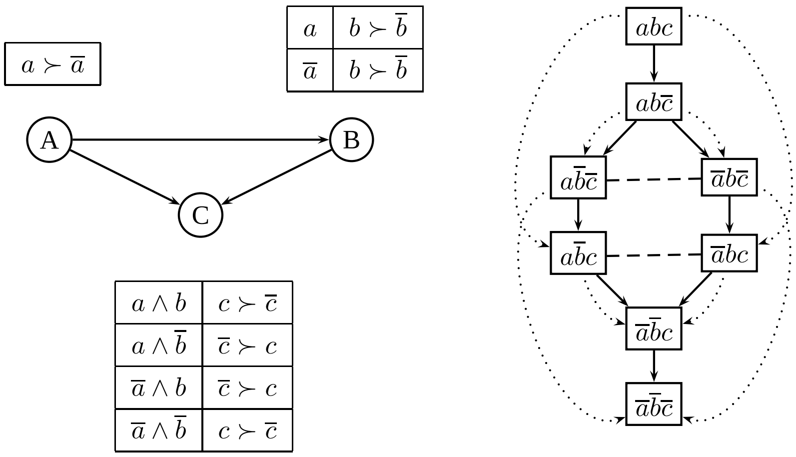


Figure 4.6: Player 2's CP-net



(a) (b)

Figure 4.7: For player 1: CP-net and associated preordering for the game G^* (equivalent by rewriting of G)

The graph union of G and the equivalent game by rewriting of G are computed, so the new goals for the players are given in Figures 4.7 on the facing page and 4.8.

On Figure 4.7 on the facing page, one can note that for computing PNE of this game, one needs to compare strategy profiles abc, $a\overline{b}c$, $\overline{a}bc$ and $\overline{a}\overline{b}c$ for Player 1, but $a\overline{b}c$ and $\overline{a}bc$ are incomparable. This "incomparability" is represented on the figure by a dashed line.

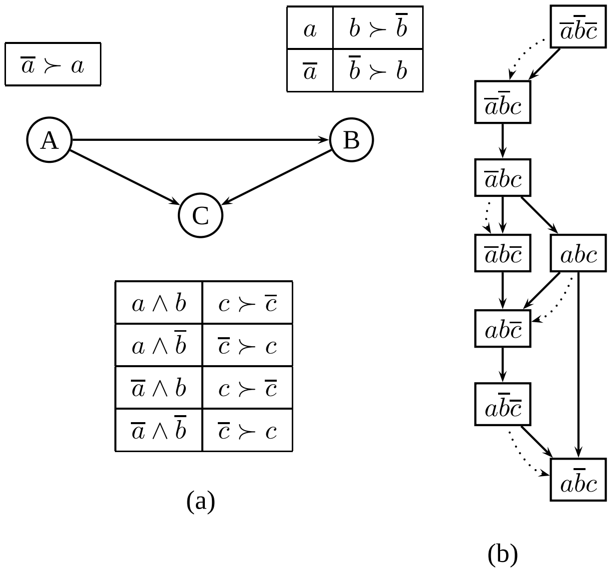


Figure 4.8: For player 2: CP-net and associated preordering for the game G^* (equivalent by rewriting of G)

The graph union is an acyclic graph, so we can apply Property 77 on page 138, and conclude that the game G has one and only one PNE: $NE = \{abc\}$.

4.3.2.2 Goals with priority

CP-nets do not allow to represent every kind of preferences, this language is not completely expressive. For instance, with a CP-net, it is impossible to express the following preferences: "My preferences are, by decreasing ordering, holidays at sea in summer, holidays at mountain in winter, holidays at mountain in summer, and then holidays at sea in winter". This preference relation compares states which are not identical "all other things being equal" ($xy \succ \overline{xy} \succ x\overline{y} \succ \overline{x}y$).

So I present in this section another representation language for non-dichotomous preferences: goals with priority. In this case, Players' preferences are expressed by a set of ordered goals with a priority relation. So we can reuse the priority relations presented in Chapter 1.1 on page 10 (for the selection mechanisms of consistent subbases in the framework of consistency restoration). These relations are:

- the preference relation called *"discrimin"* which is defined in [Bre89, DLP91, Gef92, BCD$^+$93] and which corresponds to the selection mechanism "incl",
- the preference relation called *"leximin"* which is defined in [DLP91, BCD$^+$93, Leh92] and which corresponds to the selection mechanism "card",
- and the preference relation called *"best-out"* which is defined in [DLP91, BCD$^+$93] and which corresponds to the selection mechanism "bo".

Using these relations, one can define the Boolean BP-games:

[6] This is the set of variables of the game.

Definition 100 *A* **Boolean BP-game** *is a 5-tuple $G = (N, V, \pi, \Gamma, \Phi)$, with $\Phi = (\Sigma_1, \ldots, \Sigma_n)$ being a collection of bases of goals with priority in which:*
$\Sigma_i = \langle \Sigma_i^1 \,; \ldots \,; \Sigma_i^p \rangle$, Σ_i^j *representing the stratum number j of Σ_i (the set of goals with priority j for Player i).*

We assume that the number of strata is the same for each player and we use the following notations:

- If G is a Boolean BP-game and if $c \in \{disc, lex, bo\}$, then $Pref_G^c = \langle \succeq_1^c, \ldots, \succeq_n^c \rangle$.
- $NE_{weak}^{disc}(G)$ (resp. $NE_{strong}^{disc}(G)$) represents the set of the weak PNE (resp. strong PNE) for $Pref_G^{disc}$. Note that $\succeq^{bo}$ and $\succeq^{lex}$ are complete preference relations; so, in these cases, strong and weak PNE coincide.

These notions are illustrated by the complete version of the prisoner's dilemma.

Example 50 on page 129 (cont'd) *The normal form for 2 prisoners is the following:*

1 \ 2	T_2	$\overline{T}_2$
T_1	*(-1/2, -1/2)*	*(-10, 0)*
$\overline{T}_1$	*(0, -10)*	*(-5, -5)*

There is one and only one PNE:

$$NE = \{\overline{T_1 T_2}\}$$

This game can be translated into a Boolean BP-game $G = (N, V, \pi, \Gamma, \Phi)$ with:

- $N = \{1, 2\}$,
- $V = \{T_1, T_2\}$,
- $\pi_1 = \{T_1\}$, $\pi_2 = \{T_2\}$,
- $\gamma_1 = \gamma_2 = \top$,
- $\Sigma_1 = \langle T_2 \,; \neg T_1 \rangle$,
- $\Sigma_2 = \langle T_1 \,; \neg T_2 \rangle$

Discrimin *Applying the discrimin criterion for each player, one finds the following complete relations:*

***Player** 1*

$$\begin{tikzcd} \overline{T_1}T_2 \arrow[d] \arrow[d, dotted, bend left] \\ T_1T_2 \arrow[d] \\ \overline{T_1T_2} \arrow[d] \arrow[d, dotted, bend right] \\ T_1\overline{T_2} \end{tikzcd}$$

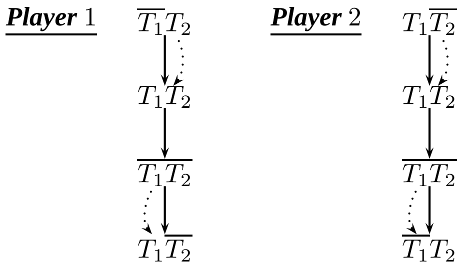
***Player** 2*

$$\begin{tikzcd} T_1\overline{T_2} \arrow[d] \arrow[d, dotted, bend left] \\ T_1T_2 \arrow[d] \\ \overline{T_1T_2} \arrow[d] \arrow[d, dotted, bend right] \\ \overline{T_1}T_2 \end{tikzcd}$$

Then the computation of the PNE (weak and strong) gives the following result:

$$NE_{weak}^{disc} = NE_{strong}^{disc} = \{\overline{T_1T_2}\}$$

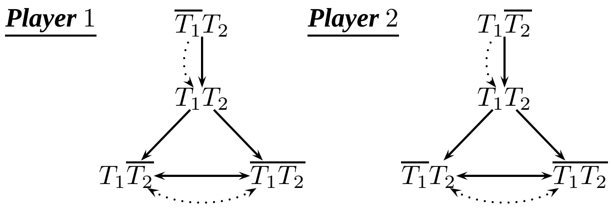
Leximin *Applying the leximin criterion gives exactly the same relations that those obtained with the discrimin criterion:*



So the PNE is the same:

$$NE^{lex} = \{\overline{T_1T_2}\}$$

Best out *With the best out criterion, we obtain:*



And we find:

$$NE^{bo} = \{T_1\overline{T_2}, \overline{T_1}T_2, \overline{T_1T_2}\}$$

The Boolean BP-games have some interesting properties.

First of all, there exist some inclusion links between the different sets of PNE:

Proposition 79 *Let $G = (N, V, \pi, \Gamma, \Phi)$ be a Boolean BP-game and $Pref^c_G = \langle \succeq^c_1, \ldots, \succeq^c_n \rangle$ be the set of preference relations on G using a criterion $c \in \{disc, lex, bo\}$.*

1. $NE^{disc}_{strong}(G) \subseteq NE^{lex}(G) \subseteq NE^{bo}(G)$
2. $NE^{lex}(G) \subseteq NE^{disc}_{weak}(G) \subseteq NE^{bo}(G)$

We can also *approximate* a Boolean BP-game considering only the k first strata for each player. The aim is double: to get a simpler game (for the computation of the PNE), and to increase the possibility to find a PNE taking into account strata with the greatest priority.

Definition 101 *Let $G = (N = \{1, \ldots, n\}, V, \pi, \Gamma, \Phi)$ be a Boolean BP-game. Let $k \in \{1, p\}$.*
*$G^{[1\to k]} = (N, V, \pi, \Gamma, \Phi^{[1\to k]})$ represents the k-**reducted Boolean game of** G in which players' goals are reducted to their k first strata: $\Phi^{[1\to k]} = \langle \Sigma_1^{[1\to k]}, \ldots, \Sigma_n^{[1\to k]} \rangle$.*

Proposition 80 *Let G be a Boolean BP-game, and let $c \in \{disc, lex, bo\}$. If s is a strong PNE (resp. weak PNE) for $Pref^c_{G^{[1\to k]}}$, then s is also a strong PNE (resp. weak PNE) for $Pref^c_{G^{[1\to k-1]}}$.*

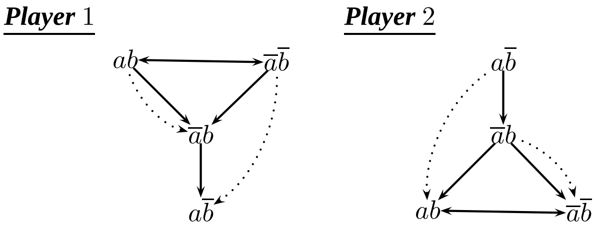
This is an interesting property for concluding on the absence of a PNE:

Proposition 81 *Let G be a Boolean BP-game, and let $G^{[1]}$ the 1-reducted Boolean game of G. Whatever the criterion used (discrimin, leximin or best-out), if the game $G^{[1]}$ does not have a PNE, then the game G does not have any more PNE.*

Example 56 *Consider the game $G = (N, V, \pi, \Gamma, \Phi)$ with:*

- $N = \{1, 2\}$,
- $V = \{a, b\}$,
- $\pi_1 = \{a\}, \pi_2 = \{b\}$,
- $\Sigma_1 = \langle a \to b \,; b \to a \rangle$,
- $\gamma_1 = \gamma_2 = \top$,
- $\Sigma_2 = \langle a \leftrightarrow \neg b \,; \neg b \rangle$

Applying the best out criterion at each player gives the following preference relations:



This game has no PNE.

Study the Boolean game $G^{[1]}$ in which players' goals are reducted to their first stratum; $G^{[1]} = \{N, V, \pi, \Gamma, \Phi^{[1]}\}$ with:

- $N = \{1, 2\}$,
- $V = \{a, b\}$,
- $\pi_1 = \{a\}, \pi_2 = \{b\}$,
- $\gamma_1 = \gamma_2 = \top$,
- $\Sigma_1^{[1]} = \{a \to b\}$,
- $\Sigma_2^{[1]} = \{a \leftrightarrow \neg b\}$

The normal form of $G^{[1]}$ is:

1 \ 2	$\overline{b}$	b
$\overline{a}$	*(1, 0)*	*(1, 1)*
a	*(0, 1)*	*(1, 0)*

This game has one PNE: $\overline{a}b$.

4.3.3 Back to argumentation

In this section, the leading idea consists in translating an argumentation system AF into a Boolean CP-game G, in order to use specific tools of Game Theory and some particular properties of Boolean CP-games for computing the extensions of AF.

This idea is born from the following facts:

- argumentation and games have many strong links, in particular the fact that both are able to represent interactions between rational agents;
- a static game should allow the representation of an abstract AF: in both cases, agents give their arguments (resp. strategies) without analysing those of the others agents, this analysis will be made afterwards that with the computation of extensions (resp. PNE);
- the graphical aspect of the CP-nets is similar to the graphical aspect of the argumentation (interaction graph).

The aim of this work is to establish a new link between argumentation and games. It is not to obtain more efficient algorithms for computing the extensions (there already exist many efficient algorithms defined in literature – see [DM01, CDM03]).

Note that the "constraint" of a Boolean CP-game is not useful for this translation, so we will use a simplified version of these Boolean CP-games (without the constraint Γ).

4.3.3.1 Translation of an AF into a CP-Boolean game

This transformation is done by Algorithm 2 on the following page. This algorithm assumes the existence of two others algorithms:

- ISCYCLIC which returns *true* if there exists at least one cycle in the argumentation graph[7],
- REMODDCYCLES for removing the odd-length cycles if there are some of them in the AF.[8].

The execution of these two algorithms can be viewed as a precompilation step of Algorithm 2 on the next page. The fact that the AF we translate does not contain odd-length cycle yields interesting properties (for the AF and for the Boolean CP-games).

Let AF be an argumentation system which does not contain odd-length cycles, the principles of Algorithm 2 on the following page are the following:

- each argument of AF is a variable of G;
- each variable is controlled by a different player (so we have as many players as variables);
- the CP-nets of all players are defined in the same way:
  - the graph of the CP-net is exactly the directed graph of AF;
  - the preferences over each variable v which is not attacked are $v \succ \overline{v}$ (if an argument is not attacked, we want to protect it),

[7] This algorithm is linear: (Step 1) removing all the vertices which do not have predecessors; (Step 2) iterating Step 1 until either all the remained vertices have at least one predecessor (there is a cycle in the initial graph), or the graph is empty (there is no cycle in the initial graph).

[8] This algorithm is polynomial: (Step 1) computation of the Boolean adjacency matrix corresponding to all the minimal odd-length paths of attack; it is sufficient to take the Boolean adjacency of the graph $\mathcal{M}$ ($\mathcal{M}(i,j) = 1$ if there is an edge from i to j in AF) and to compute $\mathcal{M}^{\text{olc}} = \mathcal{M}^1 + \mathcal{M}^3 + \ldots + \mathcal{M}^{2n-1}$ with $n = |\mathcal{A}|$ (the bound $2n-1$ is obtained using a general result given by graph theory: if a directed graph contains a path from a to b then there exists a simple path – a path in which each vertex appears only one – from a to b); (Step 2) removal of all the arguments for which the diagonal element of $\mathcal{M}^{\text{olc}}$ is 1; (Step 3) removal of all the edges having one removed argument as end point or as start point.

Algorithm 2: Translation of an argumentation system into a CP-Boolean game

begin

/* INPUT: $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ an argumentation system */
/* OUTPUTS: $G = (N, V, \pi, \Phi)$ a CP-Boolean game, AF after removal of odd-length cycles */
/* LOCAL VAR.: $i =$ current agent, $a =$ current arg. */

/* if necessary, removal of the odd-length cycles */
if ISCYCLIC(AF) **then** $\mathsf{AF} =$ REMODDCYCLES(AF)
/* no more odd-length cycle in AF */
/* computation of the CPTs for each argument */
for $a \in \mathcal{A}$ **do**
  if $\mathcal{R}^{-1}(a) = \varnothing$ **then** $CPT(a) = a \succ \overline{a}$
  /* unattacked argument */
  else
  /* case of the other arguments */
  $CPT(a) = \{\bigvee_{v \in \mathcal{R}^{-1}(a)} v : \overline{a} \succ a\} \cup \{\bigwedge_{v \in \mathcal{R}^{-1}(a)} \overline{v} : a \succ \overline{a}\}$

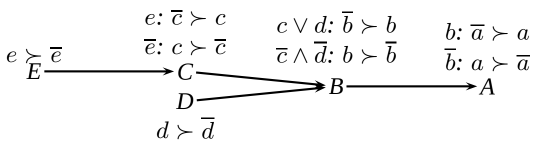
/* computation of the CP-net $\mathcal{N}$ */
$\mathcal{N} = \langle \mathsf{AF}, \cup_{a \in \mathcal{A}} CPT(a) \rangle$ /* it is the attack graph*/
/* after removal of odd-length cycles, */
/* associated with the CPT for each argument */
/* computation of N, V, π and Φ */
$i = 1$, $N = \varnothing$, $V = \mathcal{A}$ /* each argument is a variable */
for $a \in \mathcal{A}$ **do**
  $N = N \cup \{i\}$ /* an agent per argument */
  $\pi_i = \{a\}$ /* i controls only this argument */
  $\mathcal{N}_i = \mathcal{N}$ /* the same CP-net for each agent */
  $i = i + 1$
return ($G = (N, V, \pi, \langle \mathcal{N}_1, \ldots, \mathcal{N}_{|V|} \rangle)$, AF)

end

– the preferences over each variable v which is attacked by the set of variables $\mathcal{R}^{-1}(v)$ depends on these variables: if at least one variable $w \in \mathcal{R}^{-1}(v)$ is satisfied, v cannot be satisfied (so we have $\bigvee_{w \in \mathcal{R}^{-1}(v)} w : \overline{v} \succ v$); however, if all variables $w \in \mathcal{R}^{-1}(v)$ are not satisfied, v can be satisfied (and so $\bigwedge_{w \in \mathcal{R}^{-1}(v)} \overline{w} : v \succ \overline{v}$).

The construction of a CP-Boolean game G from an argumentation system AF is made in polynomial time (even if AF is cyclic and if we have to remove its odd-length cycles). This translation is illustrated on the following example:

Example 57 *Consider* $\mathsf{AF} = \langle \{a, b, c, d, e\}, \{(b, a), (c, b), (d, b), (e, c)\} \rangle$ *(*AF *is acyclic) and transform it in a CP-Boolean game* $G = (N, V, \pi, \Phi)$*. By applying Algorithm 2,* $V = \{a, b, c, d, e\}$ *and* $N = \{1, 2, 3, 4, 5\}$*, with* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$*,* $\pi_3 = \{c\}$*,* $\pi_4 = \{d\}$ *and* $\pi_5 = \{e\}$*. The following CP-net represents the preferences of all players:*



The following example shows the translation of a cyclic AF but with only even-length-cycles:

Example 58 *Consider* $\mathsf{AF} = \langle \{a, b\}, \{(a, b), (b, a)\} \rangle$*. By applying Algorithm 2,* $V = \{a, b\}$ *and* $N = \{1, 2\}$*, with* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$ *(*AF *is cyclic, but contains only even-length cycles). The following CP-net represents the preferences of all players:*

$$\begin{array}{l} b: \overline{a} \succ a \\ \overline{b}: a \succ \overline{a} \end{array} \quad A \rightleftarrows B \quad \begin{array}{l} a: \overline{b} \succ b \\ \overline{a}: b \succ \overline{b} \end{array}$$

And then we give an example of an AF with odd-length cycles:

Example 59 *Consider* $\mathsf{AF} = \langle\{a,b,c,d,e\},\{(a,b),(b,c),(c,d),(d,e),(e,c)\}\rangle$*. The initial* AF *is cyclic, and contains an odd-length cycle, which has to be removed. The final* AF *will contain only* a *and* b*. So, by applying Algorithm 2 on the preceding page,* $V = \{a,b\}$*,* $N = \{1,2\}$*, with* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$ *and the following CP-net represents the preferences of all players:*

Initial graph:
$$\begin{tikzcd} & & & e \arrow[dl] \\ a \arrow[r] & b \arrow[r] & c \arrow[dr] & \\ & & & d \arrow[uu] \end{tikzcd}$$

Final graph: $a \longrightarrow b$

CP net: $A \longrightarrow B$, $a \succ \overline{a}$, $\begin{array}{l} a: \overline{b} \succ b \\ \overline{a}: b \succ \overline{b} \end{array}$

There exists a link between preferred extensions of AF and PNE of G:

Proposition 82 *Let* $\mathsf{AF} = \langle\mathcal{A},\mathcal{R}\rangle$*. Let* $G = (N,V,\pi,\Phi)$ *be the CP-Boolean game and* AF' *be the argumentation system both obtained from* AF *by applying Algorithm 2 on the facing page.*
s *is a preferred extension of* AF' *iff* s *is a SPNE for*[9] G*.*

Example 57 on the preceding page (cont'd) G *has one SPNE* $\{ed\overline{c}\overline{b}a\}$ *and* AF *has only one preferred extension* $\{e,d,a\}$*.*

Example 58 on the facing page (cont'd) G *has two SPNEs* $\{a\overline{b},\overline{a}b\}$ *and* AF *has two preferred extensions* $\{a\}$*,* $\{b\}$*.*

Example 59 (cont'd) G *has one SPNE* $\{a\overline{b}\}$ *and the final* AF *(*after removal of odd-length cycles*) has 1 preferred extension* $\{a\}$*.*

4.3.3.2 Computation of preferred extensions

Since preferred extensions correspond exactly to SPNEs, the main properties about computation of SPNE in CP-Boolean games can be applied. The first interesting case concerns the acyclic argumentation systems:

Proposition 83 *Let* AF *be an argumentation system. Let* G *be the CP-Boolean game and* AF' *be the argumentation system both obtained from* AF *by applying Algorithm 2 on the facing page.*
If AF' *is acyclic,* AF' *has one and only one preferred extension which is computable in polynomial time using* G*.*

Of course, this proposition holds for the simple case of an initial acyclic argumentation system.

The computation of SPNE(s) for cyclic argumentation systems is much more complex. However, Algorithms 3 on page 149 and 4 on page 149 allow to compute such solution concept when the argumentation system contains even-length cycles.

These algorithms assume the existence of Algorithm COMPINTCYCLEFORPROP which returns the cycle (or one of the cycles if there are several) in a given set of variables which permits to reach as many variables as possible.[10] For instance, on the following graph:

[9] Recall that s denotes a V-interpretation, that is if $s = a\overline{b}c$ for example, this corresponds to the set $\{a,c\}$.

[10] This algorithm uses the notion of Boolean adjacency matrix as Algorithm REMODDCYCLES:

- computation of the Boolean adjacency matrix $\mathcal{M}^{\mathsf{ap}}$ corresponding to all minimal paths in the graph reduced to the given set of variables: $\mathcal{M}^{\mathsf{ap}} = \mathcal{M} + \mathcal{M}^2 + \mathcal{M}^3 + \ldots + \mathcal{M}^{2n}$ with $n = |V|$; $\mathcal{M}^{\mathsf{ap}}(i)$ will denote $(\mathcal{M}^{\mathsf{ap}}(i,1),\ldots,\mathcal{M}^{\mathsf{ap}}(i,n))$;
- $\mathtt{ToSee} = V$; $C = \varnothing$; $\mathtt{end}? = $ *false*;
- loop: while NOT($\mathtt{end}$?) do
  $v = \mathtt{top}(\mathtt{ToSee})$; $\mathtt{ToSee} = \mathtt{ToSee} \setminus \{v\}$;
  if ($\nexists w \in \mathtt{ToSee}$ s.t. $\mathcal{M}^{\mathsf{ap}}(v) \subset \mathcal{M}^{\mathsf{ap}}(w)$) then
  /* no var. permitting to reach more var. than v */

$$
\begin{tikzcd}
g \arrow[d] & & f \arrow[d] & e \arrow[l] & i \arrow[l] \arrow[r, bend right] & j \arrow[l, bend right] \\
a \arrow[r, bend right] & b \arrow[l, bend right] \arrow[r] & c \arrow[r] & d \arrow[u] & h \arrow[l]
\end{tikzcd}
$$

$\{a, b\}$ permits to reach variables a, b, c, d, e, f, and $\{i, j\}$ permits to reach variables i, j, c, d, e, f.

These cycles are more interesting than the other ones for the propagation of values over the graph (if they are the starting point of a propagation process then this propagation is more efficient).

Let $\mathcal{N}$ be the CP-net representing goals of players of a CP-Boolean game, the principles of Algorithms 3 on the next page and 4 on the facing page are:

- instantiation of all unattacked variables (which have no parents in $\mathcal{N}$ and are satisfied in the SPNE);
- propagation of these instantiations for as long as possible;
- once all feasible instantiations have been done, loop:
  - if all variables have been instantiated, the SPNE can be returned;
  - else, with Algorithm COMPINTCYCLEFORPROP, the more interesting cycle C remaining is computed (there is one, otherwise all variables would have been instantiated);
  - using the current state of the current SPNE, create two new SPNEs; the first one contains a variable of C instantiated to $true$, the second one contains this same variable instantiated to $false$
  - propagation of these instantiations for each one of these SPNEs as long as possible.

Example 60 *Using the following graph:*

$$
\begin{tikzcd}
g \arrow[d] & & f \arrow[d] & e \arrow[l] & i \arrow[l] \arrow[r, bend right] & j \arrow[l, bend right] \arrow[r] & k \arrow[r, bend right] & l \arrow[l, bend right] \\
a \arrow[r, bend right] & b \arrow[l, bend right] \arrow[r] & c \arrow[r] & d \arrow[u] & h \arrow[l]
\end{tikzcd}
$$

the steps of the computation process are:

- *g and h are instantiated to true (current state of SPNE $= gh$);*
- *then a and d are instantiated to false (current state of SPNE $= gh\overline{a}\overline{d}$);*
- *then b is instantiated to true (current state of SPNE $= gh\overline{a}\overline{d}b$);*
- *then c is instantiated to false (current state of SPNE $= gh\overline{a}\overline{d}b\overline{c}$);*
- *at this point the simple propagation stops; so we must compute the interesting cycles in the remaining set of variables (e, f, i, j, k, l) and the result is (i, j);*
- *the propagation process is restarted with the following current states of two SPNEs: $gh\overline{a}\overline{d}b\overline{c}i$ and $gh\overline{a}\overline{d}b\overline{c}\overline{i}$;*
- *...*
- *so, at the end of the propagation process, three SPNEs are obtained:*
  $$gh\overline{a}\overline{d}b\overline{c}i\overline{e}f\overline{j}k\overline{l},\ gh\overline{a}\overline{d}b\overline{c}\overline{i}e\overline{f}j\overline{k}l \text{ and } gh\overline{a}\overline{d}b\overline{c}i\overline{e}f\overline{j}\overline{k}l.$$
  These SPNEs correspond to the three preferred extensions $\{g, h, b, e, j, l\}$, $\{g, h, b, f, i, k\}$ and $\{g, h, b, i, f, l\}$.

The following proposition shows that Algorithms 3 on the next page and 4 on the facing page exactly compute the set of SPNEs of the CP-Boolean game.

Proposition 84 *Let G be a CP-Boolean game given by Algorithm 2 on page 146. Let SP be the set of strategy profiles of G given by Algorithms 3 on the facing page and 4 on the next page. $s \in SP$ iff s is a SPNE for G.*

$C = C \cup \{v\}$; `end?` $=$*true*;
$\forall w \in$ `ToSee` do if $\mathcal{M}^{\mathrm{ap}}(v) = \mathcal{M}^{\mathrm{ap}}(w)$ then $C = C \cup \{w\}$;
else if `ToSee` is empty then `end?` $=$*true*;

- Return C

Algorithm 3: Computation of SPNEs of a CP-Boolean game obtained from an argumentation system

```
begin
    /* INPUTS: a CP-Boolean game G = (N, V, π, Φ), where Φ = ⟨𝒩_1, ..., 𝒩_n⟩ */
    /* OUTPUTS: a set of SPNEs SP */
    /* LOCAL VARIABLES: v = current variable, In = (resp. Out =) set of variables instantiated to true (resp. false),
    R = set of variables remaining to be instantiated */

    In = ∅, Out = ∅, R = V                                                    /* Initialization */
    /* Instantiation of all variables without parents */
    for v ∈ R do  if Pa(v) = ∅ then  R = R \ {v}, In = In ∪ {v}
    /* propagation by a recursive process */
    return CompSPNERec(G, R, In, Out)
end
```

Algorithm 4: CompSPNERec: Recursive computation of SPNEs of a CP-Boolean game obtained from an argumentation system

```
begin
    /* INPUTS: a CP-Boolean game G = (N, V, π, Φ),
        R = set of variables remaining to be instantiated,
        In = set of variables already instantiated to true,
        Out = set of variables already instantiated to false*/
    /* OUTPUTS: a set of SPNEs SP */
    /* LOCAL VARIABLES: v = current variable, n = cardinal of R, C = set of variables forming a cycle*/

    if R = ∅ then
        /* all variables are instantiated: a SPNE is found */
        return {(In \overline{Out})}
    else
        n = |R|   /* n = number of variables remaining to be instantiated */
        for v ∈ R do
            /* simple propagation process */
            if Pa(v) ⊆ Out then
                /* all parents are instantiated to false */
                In = In ∪ {v}, R = R \ {v}
            else
                if (Pa(v) ∩ In) ≠ ∅ then
                    /* at least 1 parent instantiated to true */
                    Out = Out ∪ {v}, R = R \ {v}
        if n = |R| then
            /* none variable instantiated in For instruction */
            C = CompIntCycleForProp(G, R)
            v = Top(C)
            return(CompSPNERec(G, R \ {v}, In ∪ {v}, Out) ∪ CompSPNERec(G, R \ {v}, In, Out ∪ {v}))
        else
            /* at least 1 variable instantiated in For instruction */
            return CompSPNERec(G, R, In, Out)
end
```

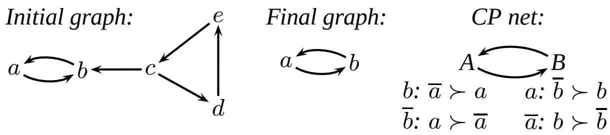
4.3.3.3 Managing odd-length cycles

Of course, the removal of odd-length cycles has an important influence on the computation of the SPNE(s) and this point could be considered as problematic in some cases if one does not agree with our initial assumption: in general, an odd-length cycle may be considered as a paradox. Of course, we know that some odd-length cycles make sense, in particular when they are not strict odd-length cycles. But the work presented in this paper is preliminary and the removal of this kind of cycles gives a interesting translation which guarantees some important properties (the treatment of odd-length cycles will be the subject of a future work.

Example 61 *Consider* $\mathsf{AF} = \langle\{b, c, d, e\}, \{(b, c), (c, d), (d, e), (e, c)\}\rangle$*. The initial* AF *is cyclic, and it contains an odd-length cycle which will be removed and the final AF will contain only* b*. So, by applying Algorithm 2 on page 146,* $V = \{b\}$*,* $N = \{1\}$*, with* $\pi_1 = \{b\}$ *and the following CP-net represents the preferences of all players:*

So, G has one SPNE $\{b\}$ *which corresponds to the preferred extension of the final* AF*. However, it does not correspond to the preferred extension of the initial* AF *which was the set* $\{b, d\}$*. If we consider that the odd-length cycle is a paradox, so that its arguments are not significant, we can consider that* $\{b\}$ *is a more realistic extension than* $\{b, d\}$ *(unfortunately, this approach is not accepted by the main semantics for acceptability).*

Example 62 *Consider* $\mathsf{AF} = \langle\{a, b, c, d, e\}, \{(a, b), (b, a), (c, b), (c, d), (d, e), (e, c)\}\rangle$*. The initial* AF *is cyclic, and it contains an odd-length cycle which will be removed and the final AF will contain only* a *and* b*. By applying Algorithm 2 on page 146,* $V = \{a, b\}$*,* $N = \{1, 2\}$*, with* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$ *and the following CP-net represents the preferences of all players:*



So, G has two SPNEs $\{a\overline{b}, \overline{a}b\}$ *which correspond to the two preferred extensions of the final* AF*. However, they do not correspond to the preferred extension of the initial* AF *which was only the set* $\{a\}$*. In this case, to take into account the extension* $\{b\}$ *means that the attack* $c \to b$ *is considered as not significant (because an odd-length cycle is a paradox and its arguments are not significant; so, they cannot be able to provide a realistic attack against other arguments).*

4.4 Conclusion on Topic 4

In this chapter, I have presented Boolean games and I have used them for simulating interactions between rational agents.

These games give another method for handling inconsistency which is close to the principles of consistency restoration: from an inconsistent initial information set one can define consistent information subsets on which a classical deduction can be applied. Using these games, a new link between argumentation and games can be established.

Indeed, these games are an interesting tool for describing agents' preferences under the form of more or less complex Boolean formulae (from simple propositional formulae to bases of formulae with priority or CP-nets). So in this framework there are several agents which share an inconsistent information set (the set of all players' goals); then the computation of solution concepts of a such game implies the identification of the consistent and "preferred" subsets of goals satisfied by players, which has exactly the same finality as consistency restoration.

And so this type of game can be used for modelling an argumentation system (AF) and so computing its extensions via the computation of solution concepts of the game (PNE). The aim of this work essentially is the definition of a new link between argumentation and games (the algorithms obtained by this way for computing preferred extensions are not more efficient than the ones already defined in literature). At this point of my work, this modelization is only partial because it does not take into account AFs with odd-length cycles. But this limitation does not seem really problematic for us since odd-length cycles in an AF are very often considered as paradoxes and since such argumentation systems do not always have interesting properties. Nevertheless the work realized on this subject should be pursued.

Chapter 5

Research project

My main works concern argumentation and I would like to continue on this topic which seems to still offer many potentialities.

Indeed, the complete argumentation process can be divided in different aspects:

1. construction of "arguments" from information sets,
2. identification of "interactions" (positive or negative) between these arguments,
3. "valuation" of arguments or interactions using different mechanisms (for instance, preferences on the initial information sets),
4. selection of "the best" arguments,
5. and "conclusion" using selected arguments (the conclusion depending on the problem for which argumentation was realized).

All these steps have been studied by many people and I myself have worked on Steps 3 and 4.

Existing argumentation systems can be divided into two groups:

- The first group of systems, called *abstract* argumentation systems, ignores completely the first and the last steps of an argumentation process, and focuses only on the notion of acceptability considering that arguments and interactions are given; all my works on argumentation belong to this framework but they are not alone (see for instance [Dun95, Cam06]).
- In the systems of the second group, all the five ingredients developed above are formally defined. However, such systems are generally defined for one particular application like, for instance, handling inconsistency in knowledge bases or explaining and making decisions. So, these systems may be called *concrete* argumentation systems and some examples of such systems can be found in [SL92, PS97, AC02b, BH01] ...

Consequently, an essential problem appears: how to classify all these systems? Different comparative studies have been carried out in order to show the links, similarities, and the differences between systems of each group.

For instance, in [BG07], for the abstract systems, different acceptability semantics have been compared on the basis of a collection of criteria.

Regarding the second group of systems, since such systems are developed for particular uses, thus only systems that are devoted to the same application are compared. For instance, in the ASPIC project (see http://aspic.acl.icnet.uk), one may find a comparison of argumentation systems developed for handling inconsistency in knowledge bases, and in [AP08], a comparison of argumentation-based decision systems has been carried out. Generally, these comparisons are only based on the outcome of the systems. Unfortunately, such comparisons may be considered as incomplete since the compared systems may use different languages, different definitions for an argument, different kinds of attack between arguments, different notions of acceptability for the arguments ... Thus, comparing two argumentation systems only on the basis of the outcome may be misleading.

Moreover, despite the huge literature on argumentation theory and its applications, there are unfortunately several questions that remain without clear answers:

- What are the main criteria that are needed for comparing two argumentation systems of the same group?
- What are the criteria that can be used for comparing two argumentation systems that are devoted to the same application?
- How can two argumentation systems developed for two distinct applications be compared?
- How can two systems pertaining to two distinct groups be compared?
- What is the most appropriate class of argumentation systems for a particular given application?

So my project would be an interdisciplinary project, combining expertise in Artificial Intelligence and Cognitive Psychology. It aims to answer the above questions. For that purpose, the following main steps are mandatory:

- To unify all existing argumentation systems by developing a general framework that takes into account all the ingredients involved in an argumentation process.
- To define rationality postulates that should be satisfied by any argumentation system.
- To define a collection of criteria that will be used to evaluate and compare systems.
- To validate those criteria theoretically and experimentally by psychological studies.

5.1 Background

The background concerning this project can be partitioned into 3 parts, which justify the interest of the proposed project:

Existence of some comparison criteria and rationality postulates for argumentation in A. I. These studies cover two distinct approaches. First, some studies concern an axiomatisation of specific argumentation systems. However, it is worth mentioning that very few works exist on axiomatizing argumentation systems. The only attempt that has been done concerns rule-based systems (strict or defeasible rules): in [CA07], it has been argued that any such a system should satisfy at least two main postulates (consistency and closure postulates). These postulates refer to the sets of assertions which can be accepted as conclusions of the argumentation process. Things are different with the second approach: many different studies propose comparison criteria for some argumentation systems taking into account only some particular notions in these systems. For instance, in the case of abstract argumentation systems, one can find the following works concerning the acceptability semantics: [Cam06] has compared different semantics (grounded, stable, semi-stable, complete, preferred) using the "reinstatement" criterion and a labelling process on the arguments; [JV99b] have also proposed a comparison of different semantics using a similar labelling process; [BG07] have proposed several evaluation criteria and have compared most of existing semantics with these criteria.

Existence of links between argumentation and nonmonotonic reasoning in A. I. (the importance of these links for our project is due to the existence of some rationality postulates for nonmonotonic reasoning in A. I. – see Chapter 1 on page 9) These links between argumentation and nonmonotonic reasoning have already been established from a theoretical point of view for some particular semantics or particular types of arguments [Dun95, Cay95, BDKT97, NOC08, TKI08]. Moreover, some works (see ASPIC project and [EGW08]) investigate this issue from a practical point of view as we plan to do. At this background, we can also add all the works concerning the definition of rationality postulates in nonmonotonic reasoning [Gab85, KLM90, LM92, GM94]; these postulates were already validated for nonmonotonic reasoning by some psychological experiments (see the last item of the following point of this background).

Existence of some comparison criteria and rationality postulates for argumentation in cognitive psychology and philosophy Argumentation does not only appear in the A.I. domain; this is also a research domain in philosophy and cognitive psychology. In these frameworks, there already exist some works proposing sets of rules or characteristics of argumentation processes. These sets could be also significant in A.I. (see for instance [Rip98, OCH08, HO07, Tha03]). Another interesting aspect relates to the previous point: as works on the rationality postulates for nonmonotonic inference were achieved, an experimental validation of these postulates has been carried out under the form of psychological experiments (see [NBR02, BBN04, BBN05]). It thus seems interesting to perform to the same experiment within the framework of argumentation.

5.2 Originality

Three main aspects can be pointed out for illustrating the originality of our project:

A global vision of an argumentation process in AI Some of the existing works take place at the language level and take advantage of the structure of the arguments in order to define postulates. Such postulates generally concern the set of justified assertions. Other existing studies abstract from the nature of arguments and try to propose evaluation criteria for the semantics, considering that the semantics is totally independent of any property of arguments at the language level. The originality of our project is to consider both levels, since building a complete argumentation system requires different steps, from the definition of arguments to the extraction of justified conclusions.

A multidisciplinary project involving psychology and computer science Another originality of our project is about the role of psychologists. In previous collaborations between AI and psychology, psychologists conducted mainly the empirical test of precise, formal and well documented patterns of inference or postulates of rationality. Such empirical tests will naturally be conducted during this project, but in addition to this well known methodology, a new methodology will be introduced that allows to infer new properties or patterns of inference that emerge spontaneously from the activity of human inferential system at work on processes of argumentation. In this context, the role of the psychologists is not to identify the best argumentation systems but to identify some postulates and criteria used in the human process of argumentation, and to validate the complete set of postulates and criteria identified in the project under a psychological point of view.

The realization of software for mapping argumentation and nonmonotonic reasoning A similar research project was carried out in the 1990s in the domain of nonmonotonic reasoning. Because links between argumentation systems and common sense reasoning exist, they can be exploited in a cognitive and psychological perspective, by experiments aiming at comparing theoretical postulates with common sense reasoning principles shared by human beings. A software dedicated to automated argumentation reasoning under various semantics of argumentation will help for this study. Since no such system is available today we plan to develop such a system.

5.3 Organisation

This project could be composed of four steps:

Step 1 A comprehensive study of existing studies, from a theoretical and an experimental perspective.

Step 2 Definition of general principles (like rationality postulates in nonmonotonic reasoning) in order to characterize argumentation systems. With this purpose, we will follow three different lines of research: to find principles by considering only the argumentation point of view; to use the rationality postulates defined in nonmonotonic reasoning context using a translation of an argumentation system into a nonmonotonic reasoning framework (so we need a special tool for carrying out this translation); and to identify new principles by considering the results of some experimental studies.

Step 3 Formalisation of criteria for comparing argumentation systems. Among these criteria, we will find the respect of the principles previously defined, but also the cautiousness of the inference (either at the level of the computed extensions, or at the level of the justified arguments) and other criteria raised at the algorithmic point of view.

Step 4 A theoretical and experimental validation of these principles and criteria.

People involved in this project will be of course computer scientists in AI (with solid knowledge in argumentation and nonmonotonic inference) but also psychologists (in order to review the psychological and philosophical literature concerning the existing results about argumentation postulates, to make an experimental inquiry into new properties of the human inferential system in argumentation, and to validate the sets of criteria and principles identified in this project).

This project has been proposed as ANR white project ("projet ANR blanc") in January 2009. And 11 people belonging to 4 french research laboratories are involved in this project.

My contribution at this project: I have proposed this project so my first contribution would be the project co-ordination. But I am also interested by the theoretical aspect concerning the definition of rationality postulates for argumentation and the formalization of comparison criteria for argumentation systems. To work on these points would be a ideal to further increase my knowledge about nonmonotonic reasoning and argumentation.

Bibliography

[ABCM04] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Justifying practical reasoning. In C. Reed F. Grasso and G. Carenini, editors, *Proceedings of the Fourth Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 87–90, 2004.

[ABP05] Leila Amgoud, Jean-François Bonnefon, and Henri Prade. An argumentation-based approach to multiple criteria decision. In *Proc. of ECSQARU*, pages 269–280, 2005.

[AC98] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In G. F. Cooper and S. Moral, editors, *Proc. of the 14th Uncertainty in Artificial Intelligence*, pages 1–7, Madison, Wisconsin, 1998. Morgan-Kaufmann.

[AC02a] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, 29:125–169, 2002.

[AC02b] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.

[ACLS04a] L. Amgoud, C. Cayrol, and MC. Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In *Proc. of the 10th NMR-UF workshop*, pages 1–9, 2004.

[ACLS04b] Leila Amgoud, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In J. Delgrande and T. Schaub, editors, *Proc. of the 10th NMR workshop (Non Monotonic Reasoning), Uncertainty Framework subworkshop*, pages 1–9, Whistler, BC, Canada, 2004.

[ACLSL08a] Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23:1062–1093, 2008.

[ACLSL08b] Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23:1062–1093, 2008.

[ADLS08] Leila Amgoud, Caroline Devred, and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In Lin Padgham and David Parkes, editors, *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), Estoril - Portugal, 12/05/08-16/05/08*, pages 429–436. ACM, 2008.

[ADLS09] Leila Amgoud, Caroline Devred, and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In Iyad Rahwan and Pavlos Moraïtis, editors, *Argumentation in multi-agent systems*, volume "LNAI 5384", pages 35–54. Springer Verlag, 2009.

[AGM85] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[AK05] Leila Amgoud and Souhila Kaci. On the generation of bipolar goals in argumentation-based negotiation. In Iyad Rahwan, Pavlos Moraitis, and Chris Reed, editors, *Argumentation in Multi-Agent Systems: (Proceedings of the First International Workshop (ArgMAS'04): Expanded and Invited Contributions)*, volume 3366 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin, Germany, 2005.

[ALS07] Leila Amgoud and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. Technical Report IRIT/RR- -2007-14- -FR, IRIT-UPS, June 2007.

[AM02] L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proc. of NMR*, pages 409–417, 2002.

[Amg99] Leila Amgoud. *Contribution à l'intégration des préférences dans le raisonnement argumentatif*. PhD thesis, Université Paul Sabatier, Toulouse, July 1999.

[Amg03] Leila Amgoud. A formal framework for handling conflicting desires. In *Proc. of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2003*, pages 552–563, 2003.

[Amg05] L. Amgoud. Towards a formal model for task allocation via coalition formation. In *Proc. of AAMAS*, pages 1185–1186, 2005.

[AMP00] Leila Amgoud, Nicolas Maudet, and Simon Parsons. Arguments, Dialogue and Negotiation. In W. Horn, editor, *Proc. of the 14^{th} ECAI (European Conference of Artifical Intelligence)*, pages 338–342, Berlin, Germany, August 2000.

[AP02] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.J. Meyer and M. Tambe, editors, *Intellignet Agents VIII: Agent theories, architectures and languages*, pages 190–205, Berlin, Germany, 2002. Springer.

[AP04] L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In *Proc. of the 9^{th} International Conference on the principles of knowledge representation and reasoning (KR)*, 2004.

[AP07] Leila Amgoud and Henri Prade. Practical reasoning as a generalized decision making problem. In *Proc. of MFI'07*, pages 15–24. Annals of LAMSADE, 2007.

[AP08] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 2008.

[BBD$^+$04a] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. CP-nets : A Tool for Representing and Reasoning with Conditional *Ceteris Paribus* Preference Statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

[BBD$^+$04b] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. Preference-Based Constrained Optimization with CP-nets. *Computational Intelligence*, 20(2):137–157, 2004. Special Issue on Preferences.

[BBHP99] Craig Boutilier, Ronen I. Brafman, Holger H. Hoos, and David Poole. Reasoning with Conditional Ceteris Paribus Preference Statements. In *Uncertainty in Artificial Intelligence (UAI'99)*, pages 71–80, 1999.

[BBN04] S. Benferhat, J. F. Bonnefon, and R. M. Da Silva Neves. An experimental analysis of possibilistic default reasoning. In D. Dubois, C. Welty, and M. Williams, editors, *In Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning (KR 2004)*, pages 130–139, 2004.

[BBN05] S. Benferhat, J. F. Bonnefon, and R. M. Da Silva Neves. An overview of possibilistic handling of default reasoning: an experimental study. *Synthese*, pages 53–70, 2005.

[BBPW04] J. Bennaim, S. Benferhat, O. Papini, and E. Würbel. An answer set programming encoding of prioritized removed sets revision: application to GIS. In J. Alferes and J. Leite, editors, *Proceedings of JELIA (LNAI: Logics for AI)*, pages 604–616, Lisbonne, Portugal, September 2004. Springer Verlag.

[BC98] T. Bench-Capon. Specification and implementation of toulmin dialogue game. In *Proceedings of JURIX 98*, pages 5–20, GNI, Nijmegen, 1998.

[BC02] Trevor J.M. Bench-Capon. Value based argumentation frameworks. In Benferhat and Giunchiglia, editors, *Proc. of the 9th International Workshop on Nonmonotonic Reasoning (session on Argument, Dialogue and Decision)*, pages 444–453, Toulouse, France, 2002.

[BCD$^+$93] Salem Benferhat, Claudette Cayrol, Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency management and prioritized syntax-based entailment. In Ruzena Bajcsy, editor, *Proc. of the 13th IJCAI*, pages 640–645, Chambéry, France, 1993. Morgan-Kaufmann.

[BCDD07] T. Bench-Capon, S. Doutre, and P.E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171:42–71, 2007.

[BDKP02] Salem Benferhat, Didier Dubois, Souhila Kaci, and Henri Prade. Bipolar representation and fusion of preferences in the possibilistic logic framework. In *Proc. of KR*, pages 158–169, 2002.

[BDKT97] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[Ber73] C. Berge. *Graphs and Hypergraphs*. North-Holland Mathematical Library, 1973.

[BG98] Salem Benferhat and Laurent Garcia. Traitement de bases de connaissances incohérentes localement stratifiées et application au raisonnement tolérant les exceptions. In LASMEA, editor, *Proc. of the 11th RFIA*, pages 233–242, Clermont Ferrand, France, 1998. AFCET.

[BG07] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10–15):675–700, 2007.

[BH01] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128 (1-2):203–235, 2001.

[BLSL05] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Jeux booléens statiques et représentation compacte de préférences (rapport préliminaire). In Andreas Herzig, Yves Lespérance, and Abdel-Illah Mouaddib, editors, *Proc. of the 3rd MFI (Journées francophones des Modèles formels de l'interaction)*, Caen, France, 2005.

[BLSL06a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Compact preference representation for Boolean games. In *Pacific Rim International Conference on Artificial Intelligence (PRICAI), Guilin, China, 07/08/2006-11/08/2006*, volume 4099, pages 41–50, http://www.springerlink.com/, août 2006. Springer-Verlag.

[BLSL06b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Jeux booléens statiques et représentation compacte de préférences. Rapport de recherche IRIT/RR–2006-13–FR, IRIT, Université Paul Sabatier, Toulouse, juillet 2006.

[BLSL07a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Compact preference representation in Boolean games. KNAW Academy Colloquium, Amsterdam, 05/02/07-07/02/07, 2007.

[BLSL07b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Dependencies between players in Boolean games. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Hammamet, Tunisie, 31/10/2007-02/11/2007*, pages 743–754, http://www.springerlink.com, novembre 2007. Springer.

[BLSL07c] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. Rapport de recherche IRIT/RR–2007-13–FR, IRIT, Université Paul Sabatier, Toulouse, juin 2007.

[BLSL07d] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. Workshop on Logics and Collective Decision making (LCD'07), Lille, 13/03/2007-14/03/2007, 2007.

[BLSL08a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Coalitions efficaces dans les jeux booléens. In *Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA), Amiens, 23/01/08-25/01/08*, page (support électronique). Laboratoire Modélisation, Information et Systèmes (MIS) - Université de Picardie Jules Vernes, janvier 2008.

[BLSL08b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. In Krzysztof Apt and Robert van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 283–297. Amsterdam University Press, 2008.

[BLSLZ06] Elise Bonzon, Marie-Christine Lagasquie-Schiex, Jérôme Lang, and Bruno Zanuttini. Boolean games revisited. In *European Conference on Artificial Intelligence (ECAI), Riva del Garda, Italy, 28/08/2006-01/09/2006*, pages 265–269, http://www.springerlink.com/, 2006. Springer-Verlag.

[BLSLZ09] Elise Bonzon, Marie-Christine Lagasquie-Schiex, Jérôme Lang, and Bruno Zanuttini. Compact preference representation and boolean games. *Journal of Autonomous Agents and Multi-Agent Systems*, 18(1):1–35, feb 2009.

[Boc03] A. Bochman. Collective argumentation and disjunctive programming. *Journal of Logic and Computation*, 13 (3):405–428, 2003.

[Bon07] Elise Bonzon. *Modélisation des interactions entre agents rationnels : les jeux booléens*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France, novembre 2007.

[Bou94] Craig Boutilier. Towards a logic for qualitative decision theory. In *Proc. of KR*, pages 75–86, 1994.

[Bre89] Gerhard Brewka. Preferred subtheories: An extended logical framework for default reasoning. In N.S. Sridharan, editor, *Proc. of the 11^{th} IJCAI*, pages 1043–1048, Detroit, MI, 1989. Morgan-Kaufmann.

[Bre94] Gerhard Brewka. Reasoning about priorities in default logic. In *Proc. of AAAI-94*, pages 940–945, Seattle, WA, 1994.

[CA05] Martin Caminada and Leila Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the 20th National Conference on Artificial Intelligence(AAAI 2005)*, pages 608–613. AAAI Press, 2005.

[CA07] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5–6):286–310, 2007.

[Cam06] Martin Caminada. Semi-stable semantics. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *Proc. of the 1^{st} COMMA (Computational models of arguments)*, pages 121–128. IOS Press, 2006.

[Cas96] Thierry Castell. Computation of prime implicates and prime implicants by a variant of the Davis and Putnam procedure. In *Proceedings of the Eighth International Conference on Tools with Artificial Intelligence*, pages 428–429. IEEE, 1996.

[Cas97] Thierry Castell. *Consistance et déduction en logique propositionnelle*. Thèse, Université Paul Sabatier, Toulouse, France, 1997.

[Cay95] Claudette Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In Chris S. Mellish, editor, *Proc. of the 14th IJCAI*, pages 1443–1448, Montréal, Canada, 1995. Morgan Kaufmann.

[CCCJ05] Guillaume Cabanac, Max Chevalier, Claude Chrisment, and Christine Julien. A social validation of collaborative annotations on digital documents. In *Proc. of IWAC*, pages 31–40, 2005.

[CCCJ07] Guillaume Cabanac, Max Chevalier, Claude Chrisment, and Christine Julien. Collective annotation: Perspectives for information retrieval improvement. In *Proc. of RIAO*, 2007.

[CDLS06a] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Handling controversial arguments in bipolar argumentation systems. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *Proc. of Computational Models of Argument (COMMA)*, pages 261–272. IOS Press, 2006.

[CDLS06b] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Prise en compte des arguments controversés dans des systèmes d'argumentation bipolaires. Technical Report IRIT/RR–2006-01–FR, CRIL-2006003, IRIT, CRIL, 2006.

[CDLS07] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Handling Ignorance in Argumentation: Semantics of Partial Argumentation Frameworks. In Khaled Mellouli, editor, *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Hammamet, Tunisie, 31/10/07-02/11/07*, number 4724 in LNAI, pages 259–270, http://www.springerlink.com, 2007. Springer.

[CDLSM02a] Claudette Cayrol, Sylvie Doutre, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Défense minimale : un raffinement de la sémantique préférée pour les systèmes d'argumentation. Rapport de recherche 2002-03-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, January 2002.

[CDLSM02b] Claudette Cayrol, Sylvie Doutre, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Minimal defence: a refinement of the preferred semantics for argumentation frameworks. In Benferhat and Giunchiglia, editors, *Proc. of the 9th International Workshop on Nonmonotonic Reasoning (session on Argument, Dialogue and Decision)*, pages 408–415, Toulouse, France, 2002.

[CDM03] C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Journal of logic and computation*, 13:377–403, 2003.

[CdSCLS08] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Revision of an argumentation system. In Gerhard Brewka and Jerome Lang, editors, *Proc. of KR 2008*, pages 124–134. AAAI Press, 2008.

[CLS93] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Comparaison de relations d'inférence non-monotone : étude de complexité. Rapport de recherche 93-23R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, September 1993.

[CLS94a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Classification de relations d'inférence non-monotone : la prudence et les propriétés de déduction. Rapport de recherche 94-49R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, November 1994.

[CLS94b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the complexity of non-monotonic entailment in syntax-based approaches. In *Proc. of the 11th ECAI workshop on Algorithms, Complexity and Commonsense Reasoning*, Amsterdam, Nederland, 1994.

[CLS95] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Non-monotonic syntax-based entailment: A classification of consequence relations. In C. Froidevaux and J. Kohlas, editors, *Lecture Notes in Artificial Intelligence 946 (Proc. of ECSQARU-95)*, pages 107–114, Fribourg, Switzerland, 1995. Springer Verlag.

[CLS97] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Quelques réflexions sur des relations d'inférence non-monotone graduelles. Rapport de recherche 97-28R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, June 1997.

[CLS00] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Une approche décisionnelle du traitement de l'incohérence inspirée par le raisonnement non monotone. In *Proc. of the 12th RFIA*, pages 277–285 (vol. 3), Paris, France, 2000.

[CLS01] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Évaluation graduelle d'arguments. Rapport de recherche 2001-18-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, November 2001.

[CLS02a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In *Proc. of the 9th IPMU (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, pages 83–90, Annecy, France, 2002.

[CLS02b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Vers une acceptabilité graduelle des arguments dans les sytèmes d'argumentation. Rapport de recherche 2002-42-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, December 2002.

[CLS03a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Critique et amélioration de l'évaluation graduelle par tuples pour le traitement des circuits. Rapport de recherche 2003-13-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, September 2003.

[CLS03b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual acceptability in argumentation systems. In *Proc. of the 3rd CMNA (International workshop on computational models of natural argument)*, pages 55–58, Acapulco, Mexique, 2003.

[CLS03c] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In B. Bouchon-Meunier, L.Foulloy, and R.R. Yager, editors, *Intelligent Systems for Information Processing: From representation to Applications*, chapter Reasoning, pages 179–190. Elsevier, 2003.

[CLS03d] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In *Intelligent Systems for Information Processing: From representation to Applications*, chapter Reasoning, pages 179–190. Elsevier, 2003.

[CLS03e] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Vers une acceptabilité graduelle des arguments dans les systèmes d'argumentation. In *Proc. of the 2nd MFI (Journées francophones des Modèles formels de l'interaction)*, pages 41–50, Lille, France, 2003.

[CLS04] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolarité en argumentation. Rapport de recherche 2004-07-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, February 2004.

[CLS05a] C. Cayrol and MC. Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In *Proc of the* 8^{th} *ECSQARU*, pages 366–377, 2005.

[CLS05b] C. Cayrol and MC. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proc of the* 8^{th} *ECSQARU*, pages 378–389, 2005.

[CLS05c] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In Luis Godo, editor, *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 366–377, Barcelone, Spain, 2005. Springer-Verlag.

[CLS05d] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[CLS05e] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In Luis Godo, editor, *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 378–389, Barcelone, Spain, 2005. Springer-Verlag.

[CLS07a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments in bipolar argumentation frameworks. In *Proc. of the 7th CMNA (International workshop on Computational Models of Natural argument)*, pages 14–20, Hyderabad, India, 2007.

[CLS07b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments in bipolar argumentation frameworks. In *Proc. of CMNA*, pages 14–20, 2007.

[CLS09] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolar abstract argumentation systems. In Iyad Rahwan and Guillermo Simari, editors, *Argumentation in Artificial Intelligence*. Springer, 2009. À paraître.

[CLSM00] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Une approche méthodologique de la résolution de conflits. In *Proc. of the 12th RFIA*, pages 215–224 (vol. 1), Paris, France, 2000.

[CLSS96a] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Non-monotonic reasoning: from complexity to algorithms. In *Proc. of the 4th International Symposium on Artificial Intelligence and Mathematics*, pages 25–28, Fort Lauderdale, Florida, 1996.

[CLSS96b] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Non-monotonic reasoning: from complexity to algorithms. Rapport de recherche 96-07R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, March 1996.

[CLSS98] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Nonmonotonic reasoning: from complexity to algorithms. *Annals of Mathematics and Artificial Intelligence*, 22(3-4):207–236, 1998.

[CMDK$^+$05] Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, and Pierre Marquis. Argumentation and fusion. In *Proc. of the 20th AAAI*, Pittsburg, Pennsylvania, 2005.

[CMDK$^+$07] Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, and Pierre Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence, Argumentation in Artificial Intelligence*, 171(10-15):730–753, octobre 2007.

[CMDM05] S. Coste-Marquis, C. Devred, and P. Marquis. Inference from contreversial arguments. In *Proceedings of the 12th International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2005)*, pages 606–620, 2–6 december 2005.

[CMDM06] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR' 06)*, pages 112–122, 2006.

[CRS93] Claudette Cayrol, Véronique Royer, and Claire Saurel. Management of preferences in assumption-based reasoning. In R. Yager and B. Bouchon, editors, *Advanced methods in AI. Lecture notes in computer science 682*, pages 13–22. Springer Verlag, 1993. Extended version in Technical Report IRIT-CERT, 92-13R (University Paul Sabatier Toulouse).

[DBC01] Paul E. Dunne and Trevor J.M. Bench-Capon. Coherence in finite argument systems. Technical report 01-006, University of Liverpool, Department of Computer Science (U.L.C.S.), 2001.

[DBC02] Paul E. Dunne and Trevor J.M. Bench-Capon. Coherence in finite argument system. *Artificial Intelligence*, 141(1-2):187–203, 2002.

[DF05] Didier Dubois and Hélène Fargier. On the qualitative comparison of sets of positive and negative affects. In *Proc. of ECSQARU*, pages 305–316, 2005.

[DJ04] V.D. Dang and N.R. Jennings. Generating coalition structures with finite bound from the optimal guarantees. In *Proc. of AAMAS*, pages 564–571, 2004.

[DLP91] Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency in possibilistic knowledge bases - to live or not to live with it. In L.A. Zadeh and J. Kacprzyk, editors, *Fuzzy logic for the Management of Uncertainty*, pages 335–351. Wiley and sons, 1991.

[DM01] S. Doutre and J. Mengin. Preferred Extensions of Argumentation Frameworks: Computation and Query Answering. In A. Leitsch R. Goré and T. Nipkow, editors, *IJCAR 2001*, volume 2083 of *LNAI*, pages 272–288. Springer-Verlag, 2001.

[Dom02] Carmel Domshlak. *Modeling And Reasoning About Preferences With CP-nets*. PhD thesis, Ben-Gurion University of the Negev, Negev, Israel, 2002.

[Dou02] Sylvie Doutre. *Autour de la sémantique préférée des systèmes d'argumentation*. Thèse, Université Paul Sabatier, IRIT, 2002.

[DP06] Didier Dubois and Henri Prade. A bipolar possibilitic representation of knowledge and preferences and its applications. In *Proc. of WILF (LNCS 3849)*, pages 1–10, 2006.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[EGFK93a] M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as logical uncertainty. In *Proc. of ECSQARU'93 – LNCS 747*, pages 85–90, 1993.

[EGFK93b] Morten Elvang-Goransson, John Fox, and Paul Krause. Dialectic reasoning with inconsistent information. In David Heckerman and Abe Mamdani, editors, *Proc. of the 9th UAI*, pages 114–121, Washington, DC, 1993. Morgan-Kaufmann.

[EGH95] M. Elvang-Gøransson and A. Hunter. Argumentative logics: Reasoning with classically inconsistent information. *Data and Knowledge Engineering*, 16:125–145, 1995.

[EGW08] U. Egly, S. A. Gaggl, and S. Woltran. Aspartix: Implementing argumentation frameworks using answer-set programming. In *In Proceedings of ICLP*, 2008.

[FGS04] M. Falappa, A. García, and G. Simari. Belief dynamics and defeasible argumentation in rational agents. In *Proc. of NMR*, pages 164–170, 2004.

[Gab85] D. M. Gabbay. Theoretical foundations for nonmonotonic reasoning in expert systems. *Logics and models of Concurrent Systems*, 1985.

[Gar98] Laurent Garcia. *Raisonnement local dans les bases de connaissances ordonnées : Applications au traitement des incohérences et au raisonnement plausible*. Thèse, Université Paul Sabatier, IRIT, 1998.

[Gef92] H. Geffner. *Default reasoning: Causal and Conditional Theories*. MIT Press, 1992.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-completeness*. W.H. Freeman and Company, New York, 1979.

[GM94] Peter Gärdenfors and David Makinson. Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65:197–245, 1994.

[GNT04] M. Ghallab, D. Nau, and P. Traverso. *Automated planning, theory and practice*. Elsevier, Morgan Kaufmann, 2004.

[Gor95] T. F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[Got92] Georg Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, 1992.

[Har04a] G. Harman. Practical aspects of theoretical reasoning. *The Oxford Handbook of Rationality*, pages 45–56, 2004.

[Har04b] Paul Harrenstein. *Logic in Conflict*. PhD thesis, Utrecht University, 2004.

[HO07] U. Hahn and M. Oaksford. The rationality of informal argumentation: A bayesian approach. *Psychological Review*, 114:704–732, 2007.

[Hun04] A. Hunter. Making argumentation more believable. In *Proc. of AAAI*, pages 269–274, 2004.

[HvdHMW01] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. Boolean Games. In J. van Benthem, editor, *Proceedings of the 8th International Conference on Theoretical Aspects of Rationality and Knowledge (TARK'01)*, volume Theoretical Aspects of Rationality and Knowledge, pages 287–298. San Francisco, Morgan Kaufmann, 2001.

[HvdT03] Joris Hulstijn and Leendert van der Torre. Combining goal generation and planning in an argumentation framework. In Tom Heskes, Peter Lucas, Louis Vuurpijl, and Wim Wiegerinck, editors, *Proceedings of the 15th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2003)*, pages 155–162, Katholieke Universiteit Nijmegen, October 2003.

[HvdT04] Joris Hulstijn and Leendert van der Torre. Combining goal generation and planning in an argumentation framework. In Anthony Hunter and Jerome Lang, editors, *Proceedings of the Workshop on Argument, Dialogue and Decision at the International Workshop on Non-monotonic Reasoning (NMR 2004)*, Whistler, Canada, June 2004.

[JK90] Ulrich Junker and Kurt Konolige. Computing the Extensions of Autoepistemic and Default Logics with a Truth Maintenance System. In *Proc. of the 8th AAAI*, pages 278–283, 1990.

[JV99a] Hadassa Jakobovits and Dick Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law (ICAIL'99)*, pages 53–62, 1999.

[JV99b] Hadassa Jakobovits and Dick Vermeir. Robust semantics for argumentation frameworks. *Journal of logic and computation*, 9(2):215–261, 1999.

[KAEF95] Paul Krause, Simon Ambler, Morten Elvang, and John Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.

[Kle86] Johan De Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[KLM90] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[KM03] 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.

[KP01] Nikos Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information systems*, 26(4):259–277, 2001.

[KP02] S. Konieczny and R. Pino Pérez. Merging information under constraints: a qualitative framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.

[Laf03] Pascal Lafourcade. Application de la résolution de conflits logiques, aux conflits des problèmes d'ordonnancement. Rapport de DEA. DEA RCFR, Université Paul Sabatier Toulouse, 2003.

[Lan94] Jérôme Lang. Syntax-based default reasoning as probabilistic model-based diagnosis. In *Proc. of the 10th UAI*, pages 391–398, Seattle, Washington, 1994.

[Leh92] Daniel Lehmann. Another perspective on default reasoning. Rapport de recherche 92-12, Leibniz Center for Research in Computer Science. Hebrew University of Jerusalem, Israel, July 1992.

[Leh95] Daniel Lehmann. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence*, 15:61–82, 1995.

[Lev92] François Levy. Reason maintenance systems and default theories. Pré-publication 92-2, LIPN, Université Paris Nord, 1992.

[LM92] Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail ? *Artificial Intelligence*, 55:1–60, 1992.

[LS92] Marie-Christine Lagasquie-Schiex. Étude et évaluation de MVL (MultiValued Logics) théorie et outil de M.L.Ginsberg. Rapport de recherche 92-39R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, October 1992.

[LS94] Marie-Christine Lagasquie-Schiex. La complexité des relations d'inférence non-monotone dans les approches syntaxiques. In *Proc. of the 2nd RJCIA*, pages 65–72, Marseille, France, 1994.

[LS95a] Marie-Christine Lagasquie-Schiex. *Contribution à l'étude des relations d'inférence non-monotone combinant inférence classique et préférences*. Thèse, Université Paul Sabatier, IRIT, 1995.

[LS95b] Marie-Christine Lagasquie-Schiex. Implémentation de quelques algorithmes en raisonnement non-monotone (annexe de la thèse numéro 2217). Rapport de recherche 95-22-R, IRIT, France, December 1995.

[LS95c] Marie-Christine Lagasquie-Schiex. Une classification des relations d'inférence non-monotone dans les approches syntaxiques. In *Proc. des 5emes journées nationales du PRC-GDR Intelligence Artificielle*, page 373, Nancy, France, 1995. Présentation effectuée en tant que chercheur invité (meilleur papier jeune chercheur en Intelligence Artificielle).

[LVW02] Jérôme Lang, Leendert Van der Torre, and Emil Weydert. Utilitarian desires. *Journal of Autonomous Agents and Multi-Agents Systems*, 5(3):329–363, 2002.

[Mac79] J. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical logic*, 8:117–133, 1979.

[MCLS05] Matthieu Mardi, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. Bipolarité en argumentation : acceptabilité et algorithmes. Rapport de recherche 2005-20-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, October 2005.

[Men95] Jérôme Mengin. A theorem prover for default logic based on an extended resolution principle and prioritized conflict resolution. In *Lecture Notes in Artificial Intelligence 946 (Proc. of ECSQARU-95)*, 1995.

[Nas50] John Nash. Equilibrium points in n-person games. *Procedings of the National Academy of Sciences*, 36:48–49, 1950.

[NBR02] R. Neves, J. F. Bonnefon, and E. Raufaste. An empirical test of patterns for nonnomotonic reasoning. *Annals of Mathematics and Artificial Intelligence*, 2002.

[Neb91] Bernhard Nebel. Belief revision and default reasoning: Syntax-based approaches. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. of the 2nd KR*, pages 417–428, Cambridge, MA, 1991. Morgan-Kaufmann.

[Nie95] Ilkka Niemelä. Towards Efficient Default Reasoning. In *Proc. of the 14th IJCAI*, pages 312–318, 1995.

[Nie02] J.M. Nielsen. On the number of maximal independent sets in a graph. Technical Report RS 02-15, Center for Basic Research in Computer Science (BRICS), April 2002.

[NOC08] J. C. Nieves, M. Osorio, and U. Cortés. Preferred extensions as stable models. *Theory and Practice of Logic Programming*, 8(4), 2008.

[NP06] S.H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation. In *Proc. of the 3^{rd} WS on Argumentation in multi-agent systems*, 2006.

[OCH08] M. Oaksford, N. Chater, and U. Hahn. Human reasoning and argumentation: The probabilistic approach. In J. E. Adler and L. J. Rips, editors, *Reasoning: Studies of Human Inference and its Foundations*. Cambridge University Pres, 2008.

[ON08] Nir Oren and Timothy J. Norman. Semantics for evidence-based argumentation. In *Proc. of COMMA*, pages 276–284, 2008.

[Pap92] O. Papini. A complete revision function in propositional calculus. In B. Neumann, editor, *Proceeding of the European Conference on Artificial Intelligence (ECAI)*, pages 339–343. John Wiley and sons, Ltd., 1992.

[Par97] Simon Parsons. Normative argumentation and qualitative probability. In *Proc. of ECSQARU*, pages 466–480, 1997.

[PC05] F. Paglieri and C. Castelfranchi. Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in mas. In *Argumentation in Multi-Agent Systems*, pages 78–94. Springer, 2005.

[PG00] J. Pollock and A. Gillies. Belief revision and epistemology. *Synthese*, 122(1-2):69–92, 2000.

[PJ96] S. Parsons and N. R. Jennings. Negotiation through argumentation—a preliminary report. In *Proceedings of the 2nd International Conference on Multi Agent Systems*, pages 267–274, 1996.

[PL92] Gadi Pinkas and Ronald P. Loui. Reasoning from inconsistency: A taxonomy of principles for resolving conflict. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. of the 3rd KR*, pages 709–719, Cambridge, MA, 1992. Morgan-Kaufmann.

[PS97] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.

[RA06] Iyad Rahwan and Leila Amgoud. An Argumentation-based Approach for Practical Reasoning . In Gerhard Weiss and Peter Stone, editors, *5th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2006, Hakodate, Japan*, pages 347–354, New York, USA, 2006. ACM Press.

[Raz78] J. Raz. Practical reasoning. *Oxford, Oxford University Press*, 1978.

[Rei87] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Rip98] L. J. Rips. Reasoning and conversation. *Psychological Review*, 105:411–441, 1998.

[RMF$^+$08] N. D. Rotstein, M. O. Moguillansky, M. A. Falappa, A. J. García, and G. R. Simari. Argument theory change: revision upon warrant. In P. Besnard, S. Doutre, and A. Hunter, editors, *Proc. of the 2^{nd} International Conference of Computational Models of Argument (COMMA)*, pages 336–347. IOS Press, 2008.

[RN95] Stuart Russell and Peter Norvig. *Artificial intelligence. A modern approach*. Prentice Hall, 1995. ISBN : 0-13-103805-2.

[RPRS08] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: a game-theoretical investigation. In *Proc. of COMMA*, pages 324–335, 2008.

[SGC04] Guillermo R. Simari, Alejandro J. Garcia, and Marcela Capobianco. Actions, planning and defeasible reasoning. In *Proceedings of the 10th International Workshop on Non-Monotonic Reasoning, Whistler BC, Canada*, pages 377–384, 2004.

[Sir04] Denis Sireyjol. Prallèles entre théorie de l'argumentation et théorie des jeux. Rapport de DEA. DEA RCFR, Université Paul Sabatier Toulouse, 2004.

[SL92] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[Sme93] Philippe Smets. Probability of deductibility and belief functions. Rapport de recherche 93-5.2, IRIDIA Université libre de Bruxelles, Bruxelles, 1993.

[Som94] Léa Sombé. A glance at revision and updating in knowledge bases. *International journal of intelligent systems*, 9(1):1–27, 1994.

[Tha03] P. Thagard. Why wasn't o. j. convicted: Emotional coherence in legal inference. *Cognition and Emotion*, 17:361–383, 2003.

[Tho00] Richmond H. Thomason. Desires and defaults: A framework for planning with inferred goals. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR 2000: Principles of Knowledge Representation and Reasoning, Proceedings of the Seventh International Conference, Breckenridge, Colorado, USA*, pages 702–713. Morgan Kaufmann Publishers, 2000.

[TKI08] M. Thimm and G. Kern-Isberner. A distributed argumentation framework using defeasible logic programming. In *Proceedings of COMMA 2008*, pages 381–392, 2008.

[Tou58] S. Toulmin. *The Uses of Arguments*. Cambridge University Press, Mass., 1958.

[TP94] S. W. Tan and J. Pearl. Specification and evaluation of preferences under uncertainty. In *Proc. of KR*, pages 530–539, 1994.

[TP05] Yuqing Tang and Simon Parsons. Argumentation-based dialogues for deliberation. In Frank Dignum, Virginia Dignum, Sven Koenig, Sarit Kraus, Munindar P. Singh, and Michael Wooldridge, editors, *Proceedings of the 4th international joint conference on autonomous agents and multi agent systems (AAMAS 2005), Universiteit Utrecht, The Netherlands*, pages 552–559, New York NY, USA, 2005. ACM Press.

[Ver96] Bart Verheij. Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In J.-J.C. Meyer and L.C. van der Gaag, editors, *Proceedings of the 8th International Dutch Conference on Artificial Intelligence*, pages 357–368, 1996.

[Ver99] Bart Verheij. Automated argument assistance for lawyers. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law*, pages 43–52. ACM Press, 1999.

[Ver03] Bart Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic in Computation*, 13:319–346, 2003.

[Vid99] Laurent Vidal. Résolution de conflits guidée par les préférences. Rapport de stage, DESS I.R.R. (Intelligence artificielle, Reconnaissance des formes et Robotique), France, June 1999.

[vNM44] John von Newmann and Oskar Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.

[Vre97] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[Was99] R. Wassermann. Full acceptance through argumentation - a preliminary report. In *Proc. of IJCAI Workshop Practical Reasoning and Rationality*, 1999.

[Win88] M. Winslett. Reasoning about action using a possible models approach. In *Proc. of AAAI*, pages 89–93, 1988.

[WK95] D. Walton and E. Krabbe. *Commitment in dialogue. Basic concepts of interpersonal reasoning*. State University of New York Press, 1995.

[Woo00] Michael J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, 2000.

[WPJ00] E. Würbel, O. Papini, and R. Jeansoulin. Revision: an application in the framework of GIS. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 505–516, Breckenridge, Colorado, USA, April 2000.

[WvdH97] C. Witteveen and W. van der Hoek. A general framework for revising nonmonotonic theories. In J. Dix, U. Furbach, and A. Nerode, editors, *Proc. of the 4^{th} International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR) – LNAI*, volume 1265, pages 258–272. Springer, 1997.

[Xuo92] N.H. Xuong. *Mathématiques discrètes et informatique*. Masson, 1992.

Appendix A

Appendices

A.1 Supervision work

A.1.1 Supervision for Master degree

Student's name	% supervision	period	topic
L. VIDAL	50	1999	2 “Conflict resolution”
L. DUSSEUX	50	2002	2 “Conflict resolution”
P. LAFOURCADE	33	2003	2 “Conflict resolution”
D. GAY	50	2004	3 “Argumentation”
D. SIREYJOL	50	2004	3-4 “Argumentation and Games”
M. MARDI	50	2005	3 “Argumentation”
A. MIQUEL	33	2007	3 “Argumentation”

A.1.2 Supervision for PHD Thesis

PHD student's name	% supervision	period	topic
S. DOUTRE (associate professor at Toulouse I since 2005)	33	1999-2002	3 “Argumentation”
Title of the PHD thesis: “Around the preferred semantics of argumentation systems”			
E. BONZON (associate professor at Paris Descartes since 2008)	50	2004-2007	4 “Games”
Title of the PHD thesis: “Modelization of interactions between rational agents: Boolean games”			

A.2 Organisation of publications

A.2.1 By type of publication

Type of publication	Number	Year of publication
International journals with review committee	5	1998 à 2009
Book parts	3 (+ 1 in press)	2002 à 2009
International conferences with review committee and published proceedings	13	1996 à 2008
International conferences and workshops with review committee but not published proceedings	4	2002 à 2007
National conferences with review committee and published proceedings	6	1994 à 2008
Other conferences	4	1994 à 2008
Technical reports	16	1992 à 2008
PHD Thesis	1	1995
TOTAL	52 (+ 1 in press)	1992 à 2009

A.2.2 By research topic

Research topic	Number	Year of publication
Nonmonotonic inference (Topic 1)	14	1992 à 2000
Conflict resolution (Topic 2)	1	2000 à 2003
Argumentation (Topic 3)	26 (+ 1 in press)	2001 à 2009
Games (Topic 4)	11	2004 à 2008
TOTAL	52 (+ 1 in press)	1992 à 2009

Submitted papers:

- IJCAI (international conference)
- IJIS (international journal)
- JOLC (international journal)

A.3 Project participation and collaborations

I have worked on the following projects:

- an European project in the framework of the 6ème PCRD (*Programme Cadre de Recherche et Développement de la commission européenne on argumentation*), named ASPIC ("Argumentation Service Platform with Integrated Components"), which has started in January 2004 for three years;
- a CNRS national project in the framework of the *Réseau Thématique Pluridisciplinaire* (RTP 11) "Information et Intelligence: Raisonner et Décider" on "the bipolar representation in reasoning and decision" for one year (from October 2003 to October 2004).

Submitted project:

- an "ANR project" (ANR: *Agence Nationale pour la Recherche*) on the argumentation (project AxSA: Axiomatisation of argumentation systems) (see Chapter 5 on page 151).

Collaborations apart from projects:

- in local (RPDMP and LILAC teams at IRIT): Claudette CAYROL, Leila AMGOUD, Sylvie DOUTRE, Florence DUPIN DE SAINT-CYR, Jérôme LANG, Jérôme MENGIN;
- in national: Sylvie COSTE-MARQUIS (CRIL Laboratory at Lens), Sébastien KONIECZNY (CRIL Laboratory at Lens), Pierre MARQUIS (CRIL Laboratory at Lens), Caroline DEVRED (LERIA Laboratory at Angers), Elise BONZON (CRIP5 Laboratory at Paris Descartes), Bruno ZANUTTINI (GREYC Laboratory at Caens).