Belief Change, Consistency and Argumentation

Florence Dupin de Saint-Cyr

Université Fédérale Toulouse Midi-Pyrénées

Mémoire des travaux et ouvrages

Pour obtenir l'
HABILITATION À DIRIGER DES RECHERCHES

Délivrée par : *Université de Toulouse III Paul Sabatier*

Belief Change, Consistency and Argumentation

November 13, 2015

Florence Dupin de Saint-Cyr – Bannay

Jury :

Salem Benferhat	Professeur, Université d'Artois, Lens	*Examinateur*
Philippe Besnard	Directeur de Recherche CNRS, Université de Toulouse	*Examinateur*
Claudette Cayrol	Professeur, Université de Toulouse	*Examinateur*
Anthony Hunter	Professeur, University College London	*Examinateur*
Gabriele Kern-Isberner	Professeur, Universitaet Dortmund	*Rapporteur*
Jérôme Lang	Directeur de Recherche CNRS, Univ. Paris Dauphine	*Directeur*
Pierre Marquis	Professeur, Université d'Artois, Lens	*Rapporteur*
Nicolas Maudet	Professeur, Université Pierre et Marie Curie, Paris 6	*Rapporteur*
Henri Prade	Directeur de Recherche CNRS, Université de Toulouse	*Examinateur*

École doctorale : *Mathématiques Informatique Télécommunications (MITT)*
Discipline ou spécialité : *Informatique – Intelligence Artificielle*
Unité de recherche : *Institut de Recherche en Informatique de Toulouse (IRIT)*

Contents

Introduction

Intelligence is hard to define[1] but it is closely related to understanding, which seems to be a simpler concept. Understanding or more precisely comprehending means to catch, to take within oneself, to make one's own. Hence artificial intelligence looks like a weird concept, since it seems to require to the computer to *comprehend* something. How could a computer integrate an idea coming from human beings, given that its internal structure has nothing in common with a human being? It is easier to imagine a computer understanding another computer. Thus, I do not see artificial intelligence as a way to make computers able to understand people. My current approach is rather about using computers in order to help people better understand each other and better understand their environment. Hence, I focus on modeling what an ideal human being would obtain by reasoning: a computer is used to derive/predict the inference results which should then be integrated by a human being. To sum up, my view of Artificial Intelligence consists of building a collaboration between human beings and computers in order to combine their different abilities for understanding the world. My work is about defining tools for this collaboration.

Let me focus on the "rational agent basic control loop", described in Algorithm 1 (slightly adapted from the basic algorithm[2] given in [201]). This loop encodes an agent behavior within a "BDI" (for Beliefs, Desires and Intentions) approach.

A rational agent has (lines **1-2**) initial sets of beliefs and intentions (that could be seen as preferred states of the world which the agent aims at achieving). This agent is able to perceive information about the world (line **4**); then it integrates this piece of information (line **5**). After that, it deliberates about what to do next (line **6-7**): it computes the "Desires" (line **6**), *i.e.*, the possible states that are achievable and their associated preferences, which it then filters (line **7**) in order to choose new "Intentions" before executing a plan that achieves them.

My main domain of research relates to Step **5** of this algorithm, *i.e.*, the field of belief change. Indeed understanding the world means being able to integrate knowledge about it (this can be a non-monotonic process) and it means also being able to integrate its dynamics. These two processes require to allow for belief changes. Studying belief

[1] Yoggi, a taxi driver in Washington, gave me this poetic and pro-active definition: "Intelligence is the ability to see the beauty".

[2] In Conclusion, we give Wooldridge's elaborated version of this algorithm.

```
1  B := B_0 ;                                          /* initial beliefs */
2  I := I_0 ;                  /* initial intentions (options to achieve) */
3  while true do
4  |  get_next_percept(p);
5  |  B := belief_integrate(B, p) ;                    /* current beliefs */
6  |  D := options(B, I) ;                          /* options generation */
7  |  I := filter(B, D, I) ;               /* choice of options to achieve */
8  |  π :=plan(B, I) ; /* compute a sequence of actions for achieving I */
9  |  execute(π);
10 end
```

Algorithm 1: Rational Agent Basic Control Loop

change covers not only the analysis of the consequences of a belief change, but also the study of how to produce some changes in order to achieve some goals (see Step **8**).

Moreover, what is summarized in the single step **5** of this algorithm is a more complex task than it appears, since when changes occur, the agent should be able to keep on thinking. Hence the problem of consistency handling is crucial, in two respects: the agent has to adapt its knowledge, goals and behaviors in order to try to remain consistent whenever changes occur, or to be able to cope with inconsistency when it is there.

I have also studied Step **1** of this algorithm *i.e.*, the problem of representing the initial beliefs. Indeed, handling correctly information is also recognized as an "intelligent" ability. There are several kinds of information: in particular we can make a distinction between generic and factual information. A factual piece of information may be viewed as a snapshot observation at a given time, such a piece of information will be called "fact". Generic information can be viewed as the gathering of factual pieces of information into a compact form. We will call this kind of information: "rules". The use of rules is very common in artificial intelligence, since they can help to predict new conclusions in a given context described by a set of factual pieces of information. The first part of this document explores the different frameworks that I used in order to handle generic information, with a particular focus on inconsistency management and change.

During the last ten years, I have been involved in the research on argumentation, it is linked with belief change, since arguing aims at changing someone's opinion (which relates also to Step **8** of this algorithm). It is also linked to consistency handling since, in order to persuade someone, arguments should be consistent or at least have this appearance. Moreover arguments are acceptable only if they can be integrated consistently in the mental state of an agent (again a problem embedded in Step **5** of the algorithm).

Rules, consistency and dynamicity are also key concepts in the argumentation context. Indeed, an argument is often viewed as a pair (support, conclusion) where the support is something that justifies the conclusion. This justification is often due to a generic information that can be present explicitly in the support (in the case of logic-based arguments). It is often the case, especially in natural argumentation, that the

generic information necessary to understand the causal link between the support and the conclusion of an argument is implicit (the argument is then called an "enthymeme"). Some arguments and their conflict relations can also be used to represent generic information: this is one claim of abstract argumentation (in which generic information is not in the content of the argument but in the structure of the relation between arguments).

Argumentation and persuasion are very important in everyday life, hence studying systems able to analyze the persuasion process is a useful challenge which may contribute to rationalize these two activities. This is the subject of the second part of this report.

The third part will be concerned with my work in progress and the directions of research that I would like to follow and that are related to some steps of the algorithm on which I have not given much attention yet in my research, such as Steps **6**, **7** and **9**, which concern preferences and goals.

Notations

In this report, we are often going to refer to a propositional language, it is denoted by $\mathscr{L}$, the vocabulary (set of symbols) on which it is written is denoted by $\mathscr{V}$. The Boolean constants $\top$ and $\bot$ represent respectively the tautology and the contradiction. We use the symbols $\wedge$, $\vee$, $\rightarrow$, $\neg$, $\leftrightarrow$ for the usual connectives "and", "or", "material implication", "not", "material equivalence". Classical inference is denoted by $\vdash$. The set of interpretations on $\mathscr{L}$ is denoted by Ω. Lower Greek letters are used to represent propositional formulas (*e.g.* $\varphi, \psi, \ldots$) and interpretations (*e.g.* ω). The models of a formula φ are represented by $[\varphi]$. $\models$ denotes the semantic inference and the satisfaction. $\equiv$ denotes logical equivalence.

We denote by Φ the *characteristic formula function* of $\mathscr{L}$, it is a function that associates to each interpretation a formula which has this interpretation as unique model: $\forall \omega \in \Omega, \Phi(\omega) \in \mathscr{L}$ and $[\Phi(\omega)] = \{\omega\}$.

For instance, Φ can be defined as follows:

$$\Phi(\omega) = \bigwedge_{v \in \mathscr{V}, \omega \models v} v \wedge \bigwedge_{v \in \mathscr{V}, \omega \models \neg v} \neg v$$

For representing *default rules*, we will use the notation $\alpha \rightsquigarrow \beta$. In this expression α and β are propositional formulas and $\rightsquigarrow$ is a new symbol, it is interpreted as if α then generally β.

For representing *production rules*, we will use the symbol $\mapsto$, $\alpha \mapsto \beta$ is interpreted as if α is proven then β is proven.

In this document, N is a positive integer representing the number of time points considered. $\forall v \in \mathscr{V}$ and $t \in [\![1, N]\!]$, $v_{(t)}$ is a *time-stamped propositional symbol* s.t. $v_{(t)}$ holds in an interpretation ω if and only if v holds at time t in ω. Let $\mathscr{V}_{(N)} = \{v_{(t)} | v \in \mathscr{V}, 1 \leq t \leq N\}$ denotes the set of all time-stamped propositional symbols.

$\mathscr{L}_{(N)}$ is the language built on the vocabulary $\mathscr{V}_{(N)}$, the time-stamped Boolean constants $\top_{(t)}$ and $\bot_{(t)}$ $(1 \leq t \leq N)$ and the usual connectives.

A formula of $\mathscr{L}_{(N)}$ is called a *temporal formula*. Temporal formulas are denoted by capital Greek letters (Ψ, etc.).

For any formula φ of $\mathscr{L}$ and any time point $t \in [\![1, N]\!]$, $\varphi_{(t)}$ is called a t-formula, it is obtained by replacing each symbol v appearing in φ by $v_{(t)}$.

Part I

Beliefs, Consistency and Change

In knowledge representation, conflicting information is difficult to handle. Indeed in this domain, a knowledge base (KB) is a set of logical formulas which is used for describing a system and for deducing new information about it. The difficulty is to reason with an inconsistent KB since classical inference is no more reliable in that case. Inconsistency may come from a change of beliefs when a new piece of information arrives which contradicts what was believed before. Or it may come from the use of contradicting defeasible rules. There are three ways to handle potentially dangerous belief changes: either modify the system in order to prevent inconsistency to occur, or let it occur and manage it, or embed the possibility to change in the system itself, *i.e.*, use a representation of an evolving system.

The first chapter shows several ways to deal with inconsistency: we show how to check consistency thanks to a solver: it may help to detect mistakes or redundancies, then we show how to prevent inconsistency to occur: in fusion of spatial information we define methods to keep the maximum set of consistent information in case of conflicting beliefs, in a priori revision we proposed to forbid some new facts to be added, and to "put shields" on generic formulas. Another way to prevent consistency is to use defeasible rules (enabling to cope with inconsistent generic information), we have developed several methods for reasoning in presence of defeasible rules and we have extended them to handle also uncertainty.

In the second chapter we deal with evolving systems, we first explain the principle of extrapolation and then we extend it in order to make causal ascription which amounts to update a system by a counter-factual. In a third section we describe how we have extended classical belief update theory in order to cope with transition constraints while updating.

Chapter 1

Handling inconsistency

In the following we are going to explore several ways to handle generic knowledge encoding "if a then b" where a and b are factual properties. Each of them is associated to a different way to deal with inconsistency. Here are the different kinds of generic rules that we have used:

- rules that correspond to *material implications* *i.e.*, in this case the rule "if a then b" is equivalent to $\neg a \vee b$, it means that the rule can be used either in forward chaining *i.e.*, from a to deduce b and then possibly use b to trigger another rule, but they can also be used for backward induction *i.e.*, from $\neg b$ to deduce $\neg a$. This expression is used when the rules are assumed to be strong (*i.e.*, do not admit exceptions). In these rules the symbols a and b are representing the same kind of information, *i.e.*, the rule does not deserve any particular status to the symbols according to the fact that they are antecedent or consequent (since the rule can be contraposed).

- rules that correspond to one way implications (denoted by $a \mapsto b$) and are only used with Modus Ponens. In some applications it is useful to have this kind of triggering rules (see Section 1.c) those rules use symbols of the same nature, rules can be used in forward chaining but are not reversible. This kind of rules is classical in AI, and was used in rule-based systems, in Prolog, in ASP. The intuitive meaning is based on a proof notion: in order to trigger a rule it's antecedent has to be established (either by another rule, or by a direct observation). Defeasible rules (see Section 1.d) belong to this kind of rules with the particularity to allow for exceptions, it means that they ase used only forward but even if the antecedent is true the rule is not necessarily triggered.

- generic information may relate pieces of information that are not of the same nature since it is sometimes very important to be able to distinguish the objects for reasoning separately. There are two kinds of such generic information:

  - an association (*e.g.* under the form of a reified formula) that relates two properties of distinct nature, for instance in *spatial or temporal reasoning* where some properties are attached to a region or to a time-point. In that

case the inference about some properties that hold somewhere or sometimes should not be mixed with inferences about the partition of the land or the precedence relations between time points.

- special production rules, where knowing that a given property holds enable to deduce something of a *different* nature. This kind of rule is useful when we want to block the forward chaining at one step inference. This is the case for instance in causal reasoning[1] where an action enables us to deduce facts. It is also the case in decision theory where it is crucial to reason independently on goals and facts (in order to avoid wishful thinking).

The three previous kinds of generic information can all be extended in order to express uncertainty. Indeed the categorization done above is a way to present a structural link between two pieces of information, the strength of this link can be expressed by a certainty level. The more certain we are that a generic information holds the more certainly we can follow the link when the premises of the generic information hold (*i.e.*, when the rule/association is applicable). We have specifically studied such an extension for attributive formulas in the spatial context (see Section 1.b.1) and for defeasible rules (see Section 1.d.1).

In this chapter we give an overview of several formalisms that we have proposed and studied in order to represent different kinds of generic information. We describe four approaches in which we manage to avoid the perturbations that can come from inconsistencies.

1.a Two examples based on material implication

Material implication is the most simple way to represent a generic information, it is a classical logic-based operator hence it is well handled by tools based on classical logic, like SAT or SMT solvers. This implication has been used from the beginning of AI research, for representing dynamic and static knowledge, *e.g.* actions effects (under the form of causal rules see [96]), inheritance properties,... In this section we show how we have used this implication in two different contexts first for encoding industrial texts in order to detect inconsistencies and second for encoding ontologies in a simplified manner. However, material implication is not always the best choice for representing generic information, in the next sections we describe some more expressive ways to do it: enabling to block a backward reasoning, allowing for exceptions, or simply linking two concepts (with no implication constraint).

[1] For a more detailed account on causal rules the user can refer to the chapter written with Andreas Herzig, Jérôme Lang and Pierre Marquis of the French book "Panorama de l'intelligence artificielle", in which we analyze the use of causal rules in the domain of reasoning about change and actions [96].

1.a.1 Encoding industrial requirements

[30] F. Bannay, M.-C. Lagasquie-Schiex, W. Raynaut, and P. Saint-Dizier. Using a SMT solver for risk analysis: detecting logical mistakes in texts. In *International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 867–874. IEEE, novembre 2014
[16] L. Amgoud, F. Dupin de Saint-Cyr, M.-C. Lagasquie-Schiex, and P. Saint-Dizier. Improving risk analysis in procedures via text analysis and reasoning: a road-map. In *International Forum on Industrial Safety (IFIS)*, juillet 2010

In the Lelie project (described in [167]), we have used satisfiability checking for detecting *inconsistencies*, *redundancy* and *incompleteness* in procedural texts and we have implemented a tool (for more details the reader can refer to the articles [30, 16] written with my colleagues Marie-Christine Lagasquie and Patrick Saint-Dizier on the tool developed by the student William Raynaut).

The main generic information in this domain is given in the "requirements" that are associated to an industrial procedure. Requirements are official texts coming from public laws or specific industrial conventions, they are rules governing the actions and facts that are allowed, forbidden or mandatory. It seems very natural to translate these rules by classical material implications (in this work we consider only strong requirements, defeasible advises are not taken into account) of the form $a \rightarrow b$ where $\rightarrow$ is the material implication (*i.e.*, $a \rightarrow b \equiv \neg a \vee b \equiv \neg b \rightarrow \neg a$) hence, contraposition is allowed.

In this project, we have chosen to use the "Z3" solver [154] that respects the formalism that is issued from the Satisfiability Modulo Theory (SMT) area [34], the efficiency of this kind of tools has been stimulated by decades of research and competitions (among SAT-solvers [130] and among SMT-solvers [33]). The SMT language is a variant of first-order logic that enables to use numerical values which are frequent in industrial domains, the availability of quantifiers and function symbols was also one reason for our choice even if we do not use them in the current version of the tool.

The translation is based on a system, called TextCoop [168], which is dedicated to language analysis in particular discourses (taking account of long-distance dependencies). TextCoop identifies (and annotates in XML) some structures which are of interest for our purpose: *e.g.* instructions, requirement statements, prerequisites, warnings, advices, themes, strengths (for requirements), the main verb and its complements, particular instruments (with equipment or product names) or adjuncts such as amounts which are numerical values (PH, Volts, weights, etc.) and temporal complements.

The content analysis is done on two kinds of input data:

- requirements: information describing the context and the precautions with which a given action (included in a procedure) must be carried out. Here is an example of requirement tagged by TextCoop .
  ```
  <requirement> in case of <theme> work at a height </theme> <predicate>
  do
  not use </predicate> <object> ropes </object> </requirement>
  ```

- procedures: ordered sequences of instructions. Here is an example of instruction tagged by TextCoop .

```
<procedure> <predicate> use </predicate> <object> a rope </object> to
tie the harness.  </procedure>
```

Then a list of synonyms is used in order to restrict the vocabulary and manage the term matching aspects between requirements and the related procedures.

In the requirement example given above, the verb is translated into a predicate that takes as arguments a subject (here the agent called `op` for "operator") and a complement/adjunct (here "the rope") : $\neg$`use(op,rope)`. It is given with a theme: `is(theme,work_at_a_height)` and they are linked by a material implication: `is(theme,work_at_a_height)` $\rightarrow \neg$ `use(op,rope)`.

The translation proposed in this work was a classical use of material implication, it was used for translating requirements that are often of the form: "if you are in this situation then you should [not] do this". This kind of information is deontic, but since our goal was to check if procedures were consistent with requirements, the translation with a material implication was sufficient for our purpose.

Our tool is able to detect three kinds of mistakes:

- Inconsistency inside the requirements, or between requirements and instructions. Moreover, thanks to the implementation of an ATMS [74] in Z3 , it has been possible to identify the origin of the inconsistency.
- Redundancy of a new instruction wrt a set of existing instruction: detection is based on checking if the instruction is already inferred by the previous instructions.
- Incompleteness: checking if the requirements and instructions allows the inference of new formulas that are not inferred by the instructions alone (*e.g.* some instructions are missing in order to mention required security principles).

1.a.2 Encoding hierarchical relations

[76] F. Dupin de Saint-Cyr and H. Prade. Logical handling of uncertain, ontology-based, spatial information. *Fuzzy Sets and Systems, Advances in Intelligent Databases and Information Systems*, 159(12):1515–1534, juin 2008

Although Description Logics (see the book of Baader et al. [24]) have been developed as tractable fragments of first order logics, for encoding ontologies, in my work with Henri Prade [76], we have provided a simplified propositional encoding of the intuitive notion of ontology, which is sufficient for handling fusion problems of symbolic information expressed by means of category labels[2]. Hence, we use the term ontology here, in the weak technical sense of a graph structure between concepts, where the arrows encode specializing/subsuming relations. The simple type of ontology that we consider has three characteristics: a label may have sub-labels that represent its sub-categories, a label is

[2]In this section and Section 1.b we describe a work done about fusing uncertain spatial information that took place in the framework of an Inter-Regional Action Project "GEOFUSE: Fusion d'informations géographiques incertaines", partially supported by Midi-Pyrénées and Provence-Alpes-Côte d'Azur regions.

considered to be the reunion of its sub-labels, labels referring to the most specific classes are pairwise mutually exclusive. For instance, Figure 1.1 provides an example of (a part of) an ontology about land cover (where arrows refer to generalization relations).

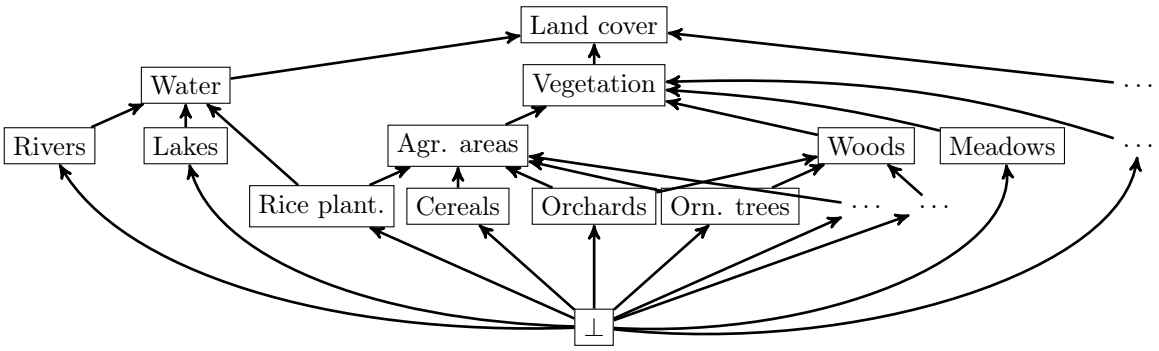


Figure 1.1: *Example of a fragment of an ontology of features*

In the partition of a territory, particular subsets of parcels may have names. The inclusion relation between sets of parcels can be also represented by an ontology, as for features. Note that even if several spatial ontologies may coexist (due to the existence of several points of view in the way the geographical space can be portioned in a meaningful way), all these ontologies share the same set of elementary parcels. See Figure 1.2 that exhibits two spatial ontologies G_s and $G_{s'}$, where in each case, the leaves, *i.e.*, the elementary parcels, are $p_1, \ldots, p_6$. A similar assumption seems more difficult to do for "property" ontologies, used by different sources.

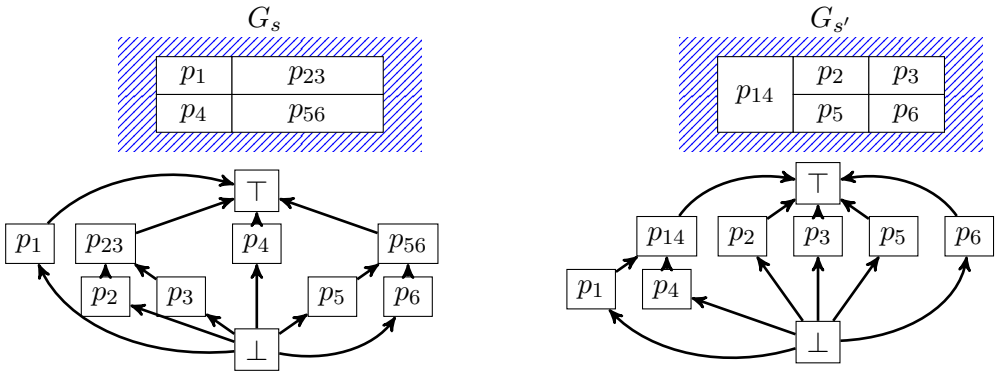


Figure 1.2: *Two space ontologies*

We have defined an *ontology* as a graph $G = (X, U)$ where $X \subseteq \mathscr{L}$ is a set of formulas and U is a set of arcs. Each arc $(\varphi, \psi) \in U$ represents the fact that φ subsumes ψ i.e., the set of objects satisfying φ is a subset of the objects satisfying ψ. An ontology is connected, without circuit and it admits a unique source equal to $\bot$ and a unique sink (called $Sink$). Moreover all the vertices that have $\bot$ as predecessor, called *leaves*,

are pairwise mutually exclusive properties. Each feature which is not a leave nor $\bot$ is equivalent to the disjunction of all its predecessors, i.e., the set of objects satisfying it is the union of the sets of objects satisfying all its predecessors.

The *levels* of the ontology are defined inductively as: Level 0 (L_0) is the set of source vertices, Level i is the set of source vertices in $G \backslash (L_0 \cup \ldots L_{i-1})$ and so on.

Example 1 *The ontology of Fig.1.1 has four levels:*
L_0: $\bot$
L_1: *Rivers; Lakes; Rice plantations; Cereals; Orchards; Orn. trees;Meadows*
L_2: *Water; Agricultural areas; Woods*
L_3: *Vegetation*
L_4: *Land cover*

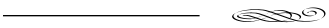
It is not difficult to give a logical encoding to an ontology, as we did in the following definition:

Definition 1 (logical encoding of an ontology) *Any graph $G = (X, U)$ representing an ontology can be associated to a set L_G of formulas such that:*

1.	$\forall (\varphi, \psi) \in U$,	*it holds that*	$\varphi \rightarrow \psi$.
2.	$\forall \varphi \in X \backslash \{L_1 \cup L_0\}$,	*it holds that*	$\varphi \rightarrow \bigvee_{\varphi_i \in U^-(\varphi)} \varphi_i$.
3.	$\forall \varphi, \psi \in L_1$,	*it holds that*	$\varphi \wedge \psi \rightarrow \bot$.
4.	$\forall (\varphi, \psi) \in X \times X$, *s.t.* $\varphi \vdash \psi$,	*it exists a directed path from* φ *to* ψ *in* G.	

Example 2 *Using the logical encoding L_G of the ontology given in Figure 1.1, we get e.g. that* $\{Water\} \cup L_G \vdash Rivers \vee Lake \vee RicePlantations$ *and* $\{Water \wedge Ornamental\_trees\} \cup L_G \vdash \bot$. *We can also establish that* $Water$ *and* $Vegetation$ *have a common subclass since* $\{Water \wedge Vegetation\} \cup L_G \vdash Rice\_plantation$. *Hence,* $L_G \vdash Water \wedge Vegetation \leftrightarrow Rice\_plantation$.

We have only illustrated our view of an ontology for property labels. However, Definition 1 is supposed to apply as well for spatial ontologies.



In the previous frameworks (Lelie project and Ontology representation), we have encoded information as if it was perfect: *i.e.*, complete and certain. In the first example it means that we only consider requirements that should always be fulfilled, and in the second example it means that the inclusion link between properties is perfectly known and assumed to be correct. If we want to manage defeasible requirements or imperfect information in general then this amounts to consider that some formulas are not mandatory, handling this kind of information can be done in a possibilistic setting (as we will see in Section 1.b.1) or with penalties (like in Section 2.b in the context of causal ascription), or probabilities.

1.b Spatial Information

[76] F. Dupin de Saint-Cyr and H. Prade. Logical handling of uncertain, ontology-based, spatial information. *Fuzzy Sets and Systems, Advances in Intelligent Databases and Information Systems*, 159(12):1515–1534, juin 2008
[98] F. Dupin de Saint-Cyr, R. Jeansoulin, and H. Prade. Spatial information fusion: Coping with uncertainty in conceptual structures. In *International Conference on Conceptual Structures (ICCS)*, volume Supplementary Proceedings, pages 66–74. Springer, 2008
[97] F. Dupin de Saint-Cyr, R. Jeansoulin, and H. Prade. Fusing uncertain structured spatial information. In Sergio Greco and Thomas Lukasiewicz, editors, *International Conference on Scalable Uncertainty Management (SUM)*, number 5291 in LNAI, pages 174–188. Springer, octobre 2008
[105] F. Dupin de Saint-Cyr and H. Prade. Multiple-source data fusion problems in spatial information systems. In *(Proc. of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'06)*, pages 2189–2196, Paris, France, july 2006

Generic information is not always implication but concerns a link between two things. Associations may relate information of the same kind (*e.g.* for encoding hierarchical relations) or they may concern two different kinds of information that cannot be mixed, namely geographic parcels and land cover properties. When information is not of the same kind, it is convenient to be able to separate the properties (*e.g.* a Boolean formula) from the object to which they are attached (*e.g.* the spatial location or the time point), this can be done with reified formulas.

1.b.1 Spatially reified formulas

A specific aspect of spatial information is that information is associated to *parcels* that are geographically identified (the parcels are usually defined through partitions of the territory). This means that information is of the type "property-object", where the object is a parcel or a set of parcels. In the literature, a piece of spatial information is often represented

- either in a relational data base style (see *e.g.* Laurini and Thompson work [143]): each parcel is described in terms of attributes, and is thus associated with a set of attribute values.
- or in the formal concept analysis domain (see [199, 116]), a relation specifies the links between parcels and properties. In this domain a *concept* is a pair (extension, intention) whose components are referring to each other in a bi-univocal way. However, in spatial information, the vocabulary is often insufficient for describing any subset of parcels in a non-ambiguous way, or conversely given a set of properties there may be no proper set of parcels that satisfy them and only them.

While the two previous representations could be translated in first order logic, where constants can either represent parcels or properties, very few authors use a logical approach for handling geographical information, up to some noticeable exceptions such as the work of Wurbel, Papini and Jeansoulin [203].

Henri Prade and I have developed a logical framework [76] in which "attributive formulas" are linking a property statement to a set of parcels where it applies. More

precisely, (φ, p) expresses that φ is true for *each* elementary parcel (*i.e.*, set of leaves of a given spatial ontology) satisfying p. Hence, our representational language is built on ordered pairs of formulas of two distinct vocabularies (one for the parcels, say $\mathscr{V}_s$, and one for the properties, say $\mathscr{V}_i$), here denoted (φ, p) and called *attributive formula*. Another understanding would view (φ, p) as the material implication $\neg p \vee \varphi$ in the language based on the union of the two vocabularies. Note that the properties and the parcels may themselves be symbolic labels taken from a *vocabulary* organized in an *ontology* as the ones seen in Section 1.a.2.

Observe that there exist propositional formulas built on the vocabulary $\mathscr{V}_i \cup \mathscr{V}_s$ which cannot be put under the attributive form, *e.g.*, $a \wedge p_1$ where a is a literal of $\mathscr{V}_i$ and p_1 a literal of $\mathscr{V}_s$. However, the introduction of classical connectives ($\wedge$, $\vee$ and $\neg$) between attributive formulas does make sense, since any pair (φ, p) is a classical formula. Thus, formulas such as $\neg(\varphi, p)$ or $(\varphi_1, p_1) \vee (\varphi_2, p_2)$ are logical formulas (but not necessarily attributive ones).

The attributive formulas considered above are similar to the ones encountered in a multi-agent extension [88] of possibilistic logic where they are used to express that any agent in a subset has some belief φ. As in the multi-agent case, this formalism can be extended in order to handle uncertainty of formulas, as well as existential quantification on subsets of parcels.

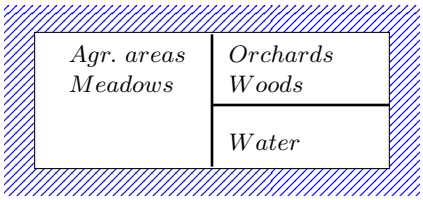


Figure 1.3: *Information about a region*

Example 3 *Let us consider the situation represented on Figure 1.3. Presented as such, the example is ambiguous, because when there are two labels on a parcel, we do not know if they are connected by a conjunction (meaning that several labels apply simultaneously to the parcel) or by a disjunction (meaning that one does not know what is the right label, but one of them applies to the parcel).*

Similarly as the above example, when a label applies to a union of elementary parcels, one may wonder if the label applies to each elementary parcel or, maybe to some of them without knowing which of them. Clearly, this leads to four logical readings of two labels a and b associated with an area covered by two elementary parcels p_1 and p_2:

i.. $(a \wedge b,\ p_1 \vee p_2)$: both a and b apply to each of p_1 and p_2.

ii.. $(a \wedge b,\ p_1) \vee (a \wedge b,\ p_2)$: both a and b apply to p_1 or they both apply to p_2.

iii.. $(a \vee b,\ p_1 \vee p_2)$: a applies to each of p_1 and p_2 or b applies to each of p_1 and p_2.

iv.. $(a \vee b,\, p_1) \vee (a \vee b,\, p_2)$: this is the most imprecise case where we do not know what of a or b applies to what of p_1 or p_2. This case may be particularized by a mutual exclusiveness: $\neg(a, p_1 \vee p_2) \wedge \neg(b, p_1 \vee p_2)$ specifying that each label cannot apply to both parcels.

Note that there is another latent ambiguity, when a label is attached to a parcel, regarding the localization of how the label applies to the parcel: it may apply *everywhere* or only *somewhere* in the parcel. Observe that in the first above reading if a and b are mutually exclusive (*e.g.* "Meadows" and "Agr. areas") the everywhere understanding is impossible.

A third kind of ambiguity is about the implicit use of the "closed world assumption" (CWA). For instance, if a source says that a parcel contains "Meadows" and "Cereals" does it exclude that the parcel would also include "Woods"? "Woods" would be indeed excluded by applying CWA. Note that the application of the "closed world assumption" may help to induce "everywhere" information from "somewhere" information.

Another need when dealing with spatial information is to be able to take into account *any* spatial information even if it is imprecise or pervaded with uncertainty. For instance, one may know that a parcel is covered either by "cereals" or by "meadows", where the two categories are mutually exclusive. Such a disjunctive value is not allowed in standard relational databases, or in standard formal concept analysis, while it raises no problem in a logical representation. Moreover, in case of imprecise information, it might be useful to represent that some alternatives are more likely than others. Disjunction may apply also to non mutually exclusive categories, as in the expression "cereals or vegetation" whose intended meaning could be "it is likely to be cereals, although any other vegetation might be possible". It leads us to introduce uncertainty on properties. This extension is natural in a possibilistic setting [84] since a possibilistic formula (φ, α) can be viewed at the meta level as being only true or false, since either $N(\varphi) \geq \alpha$ or $N(\varphi) < \alpha$ (where N represents a necessity measure):

Definition 2 (uncertain attributive formula) *An* uncertain attributive formula *is a pair* $((\varphi, \alpha), p)$ *meaning that for all elementary parcels satisfying* p, φ *is certain at least at level* α.

As we have seen handling spatial information may be a complex representational task, but once the representation setting is established some other tasks may be tackled: we have defined new inference rules based on classical inference adapted to spatially-reified formulas. Our second study concerns fusion, indeed spatial information is often distributed over different sources using different sensors.

1.b.2 Inference

In the case of spatially-reified information, we have proposed an adaptation of classical inference rules (see [98, 97, 76, 105]). In our view of attributive formulas, it holds that (φ, p) is equivalent to $\neg p \vee \varphi$.

We have shown that the inference rules of possibilistic logic [84] straightforwardly extend to uncertain attributive formulas, *e.g.* the inference rule $(\neg\varphi \vee \varphi', p), (\varphi \vee \varphi'', p') \vdash (\varphi' \vee \varphi'', p \wedge p')$ becomes $((\neg\varphi \vee \varphi', \alpha), p), ((\varphi \vee \varphi'', \beta), p') \vdash ((\varphi' \vee \varphi'', \min(\alpha, \beta)), p \wedge p')$.

The attributive formulas associated with a somewhere reading (where a parcel p is viewed as a collection of more elementary objects o, and (φ, p, s) means that $\exists o \in p, \varphi(o)$) have a different behavior w.r.t. inference. Indeed, the inference rule $(\varphi, p), (\psi, p) \vdash (\varphi \wedge \psi, p)$ is no longer compatible with this reading since $\exists o \in p, \varphi(o)$ and $\exists o' \in p, \psi(o')$ does not entail $\exists o'' \in p, \varphi(o'') \wedge \psi(o'')$.

1.b.3 Fusion

It is often the case that spatial information is provided by different sources, however to exploit them it is important to be able to gather all the pieces of information in only one consistent base. We assume in the following that the sources are equally reliable. When the sources are consistent, fusion amounts to do an union, if they are not, the idea is to discard less information as possible. More formally, suppose we have two pieces of information φ and ψ provided by two different sources. If φ and ψ are consistent, the fusion is straightforward and yields the conjunction $\varphi \wedge \psi$. If $\varphi \wedge \psi \equiv \bot$ then at least one of the two sources should be wrong. But, if we do not want to throw away all pieces of information, we may still assume that one is right, this yielding the disjunction $\varphi \vee \psi$ as a result of the fusion.

In standard possibilistic logic, conflict becomes a matter of degree. When fusionning two possibilistic knowledge bases K_1 and K_2 the idea is to consider the possibility distributions [84] π_1 and π_2 that are semantically respectively associated to them and then to combine those distributions. This yields so-called adaptive fusion operators [89] which are parameterized with the level of inconsistency Inc of the knowledge base $K_1 \cup K_2$. These operators are such that if $Inc = 0$ the conjunctive min-based combination of the distributions is retrieved, and if $Inc = 1$ the disjunctive max-based combination of the distributions is obtained.

Possibilistic information fusion easily extends to attributive formulas expressing spatial information. Indeed, each formula (φ, p) provided by a source is equivalent to the conjunction of formulas (φ, p_i), where the p_i's correspond to the leaves of the spatial ontology used by this source such that $p_i \models p$. The sources may or may not use the same ontology of properties and the same spatial ontology this may lead to inconsistencies. In Section 1.a.2, we have advocated the idea that different spatial ontologies should at least share the same leaves (called elementary parcels).

Then, for each elementary parcel p_i possibilistic information fusion takes place. However, if the two sources do not use the same property ontology, an additional source of knowledge laying bare the links between the two vocabularies is necessary in order to determine if for a given elementary parcel the sources are conflicting or not. This is specially important if the fusion principle depends on the existence and the extent of inconsistency between the sources.

In case of inconsistency, instead of building a disjunction, another possibility would be to weaken the "everywhere" reading leading to inconsistency into a "somewhere" reading.

In the particular case of an inconsistency due to a mutual exclusiveness constraint coming from an ontology, still another way of getting rid of this inconsistency is to introduce new labels in the ontology which would be compatible with the apparently conflicting labels. For instance, two mutually exclusive labels such as "Vegetation" and "Lakes" might be attached to the same area because they are intimately mixed there, hence there is no inconsistency if a new label is introduced for this situation, namely "swamp". This latter option has been proposed by Doukari and Jeansoulin [80].

1.c Production Rules

[95] F. Dupin de Saint-Cyr, B. Duval, and S. Loiseau. A priori revision. In *Lecture Notes in Artificial Intelligence (Proc. of ECSQARU-01)*, pages 488–497, Toulouse, France, September 2001
[103] F. Dupin de Saint-Cyr and S. Loiseau. Validation and refinement versus revision. In *Symposium on verification and validation of knowledge based systems and components (EUROVAV'99)*, pages 163–176, Oslo, Norway, June 1999. Kluwer Academic Publishers

1.c.1 Framework

This section is in the frame of the rule-based systems domain (initially developed under the name "expert system" (*e.g.* MYCIN [179]). A rule-base system is composed of an inference engine and a set of rules called knowledge base (abbreviated "KB") that are "production" rules. Production rules are different from material implications since they can only be used in forward chaining by the inference engine. They are used to derive new pieces of knowledge, starting from some factual information. This framework has had famous applications in the eighties, the two main tasks required to use such rule-based systems were first knowledge acquisition and second knowledge validation. Acquisition techniques were based on interviews of human experts and case studies, nowadays knowledge acquisition is more related to machine learning techniques applied to information that can be found on the web. However the second task, first explored by Sowa et al. [182], is still very important (and maybe more crucial to achieve with the phenomenal rise of mass information to deal with): it consists in measuring the quality of an existing rule base. Note that at that time inconsistency measures on knowledge bases (see *e.g.* Hunter and Konieczny [128]) had not yet been developed. The reader can refer to the articles done with my colleagues Béatrice Duval and Stephane Loiseau [95, 103] for more details about my contribution to this domain.

In my particular study, the rule-based system contains only three kinds of *propositional* formulas that are facts, rules and constraints ; we distinguish between *input symbols* (denoted by uppercase letters), which can compose new factual pieces of information, and other symbols (intern symbols).

We consider the following restrictions: the possible new information is a conjunction of input literals (input symbol or its negation) and the *knowledge base* is a set of Horn clause formulas called *rules* denoted by: $\alpha \mapsto \beta$, where α is a conjunction of literals and β is a non input literal, interpreted as if α is proven then β is proven. This framework allows us to consider Modus Ponens (denoted by $\vdash_{MP}$) as the unique inference relation.

Example 4 *Let us consider the following knowledge base KB_1: Quakers (Qua) are Pacifists (pac), Republicans (Rep) are non Pacifists, Quakers are American (am), Americans like Baseball (bball), and Quakers don't like Baseball.*

$$KB_1 = \left\{ \begin{array}{ll} r1 : Qua \mapsto pac, & r2 : Rep \mapsto \neg pac, \\ r3 : Qua \mapsto am, & r4 : am \mapsto bball, \\ r5 : Qua \mapsto \neg bball & \end{array} \right\}$$

The set of input symbol is $\{Qua, Rep\}$.

With this knowledge base KB_1, if a new piece of information arrives and states that Nixon is both a Quaker and a Republican, it is possible to deduce that Nixon is both pacifist and not pacifist, a contradiction that we want to avoid.

As we noticed at the end of the previous example, when generic information is expressed by production rules, some contradiction may arise if several rules that have a contradicting conclusion are triggered together. In the next section we will see a proposal to avoid this problem.

1.c.2 A priori revision

As we will see in more details in Chapter 2, the problem of belief change [1, 200, 136] is to find what can be inferred from a knowledge base when a new formula φ has been added to it. It is called belief revision when φ is a new piece of information completing what is currently known about the system. Note that classical revision takes place *after* the arrival of a new piece of information φ, so this revision can be called *a posteriori revision.*

The aim of the work I did with Stephane Loiseau [103] and Béatrice Duval [95] was to propose a way to make *a priori revision.* In a priori revision, we have provided a way to "armor" the KB by suppressing some rules and by forbidding to accept some new information such that adding any allowed formula φ to the revised KB will not bring inconsistency. Consequently, in the revised KB, classical monotonic inference relation will always be usable. Given an initial KB, we were able to provide a set of armored KB such that each one will be consistent with any conjunction φ of allowed input literals. In order to do that, we compute a set of diagnoses that can explain a potential inconsistency of the KB. A diagnosis is a pair composed by a set of integrity constraints which define a set of input facts (conjunction of input literals) that are forbidden and by a set of formulas that must be removed from the KB. Applying a diagnosis to a KB is called "armoring" it.

Example 4 (continued): *For Nixon example, here are some diagnoses: $D_0 = \langle \varnothing, \{r_1, r_2, r_3, r_4, r_5\}\rangle$, $D_1 = \langle \varnothing, \{r_1, r_3\}\rangle$, and $D_9 = \langle \{\{Rep, Qua\}\}, \{r_4\}\rangle$. If we consider D_9, the new information "Nixon is both pacifist and Quaker" represented by $Rep \wedge Qua$ is forbidden. Moreover, for any input fact φ that is not forbidden, we can guarantee that $\varphi \cup \{r_1, r_2, r_3, r_5\}$ is consistent.*

One difficulty is that it can exist many such diagnoses. In order to chose a diagnosis, we first discard non minimal ones. Second, we use a penalty approach that provides

criteria to prefer the diagnoses that reject or make useless the less important formulas of KB.

A *minimal diagnosis* is a diagnosis that leads to minimal change in the corresponding armored KB (*i.e.*, which suppresses the smallest number of rules and forbids the smallest number of input facts). If we consider the three diagnoses: $D_{11} = \langle\{\{a, b\}, \{a, c\}\}, \{r1\}\rangle$, $D_{12} = \langle\{\{a, b\}\}, \{r1\}\rangle$ and $D_{13} = \langle\{\{a\}\}, \{r1\}\rangle$. D_{11} is not minimal because it is not necessary to forbid the conjunction of the literals a and c to have a diagnosis; D_{13} is not minimal because D_{12} shows that it is not necessary to forbid all the interpretations satisfying a, it is sufficient to forbid the interpretations having a and b (D_{12} is minimal).

Note that the minimality principle that we used is based on the syntax. For instance, between two diagnoses $D_a = \langle\{\{a, b\}, \{a, \neg b\}\}, \{r1\}\rangle$ and $D_b = \langle\{\{a\}\}, \{r1\}\rangle$, that are equivalent semantically, the minimality criterion leads to prefer D_a. This definition is in accordance with the interpretation of the production rules (that are not equivalent to classical material implication).

A diagnosis explicitly excludes some rules from the knowledge base. It may also happen that some rules become useless in the revised knowledge base because they can never be fired. A rule cannot be fired if its conditions correspond to an impossible conjunction of input literals or if its premise can not be derived from any input.

We have proven [95]] that *If a diagnosis contains a useless rule then it is not minimal.* It means that minimality and uselessness are complementary notions to evaluate diagnoses. We can use them in order to refine the preference relation on diagnoses in a way that we can prefer diagnoses that reject or make useless the less formulas of the initial KB. For any rule r_i of the KB, we assume an associated penalty $\alpha(r_i)$ that represents a degree of confidence in r_i, it must be understood as the cost that the user must pay in order to discard the rule r_i. The diagnoses are compared according to the cost of the rules that are discarded or that become useless.

Example 4 (continued): *We associate a penalty to each rule.*

$r1 : Qua \mapsto pac$ $\quad \alpha_1 = 5$ $\quad r2 : Rep \mapsto \neg pac$ $\quad \alpha_2 = 5$
$r3 : Qua \mapsto am$ $\quad \alpha_3 = 100$ $\quad r4 : am \mapsto bball$ $\quad \alpha_4 = 5$
$r5 : Qua \mapsto \neg bball$ $\quad \alpha_5 = 7$

The penalty associated to $r3$ *means that this rule is very important. The penalty-preferred a priori revised KB corresponds to the diagnosis* D_9.

A difficulty with this approach is to obtain the penalties for the rules. They can be given by an expert. If no penalty is given, each formula can be associated with a penalty equal to 1, this approach is equivalent to count the number of formulas. If the KB represents a default behavior of some components, penalties can be proportional to probabilities associated to a faulty component, as done by de-Kleer and Williams in [75].

Adding a fact to an a priori revised knowledge base is very easy: it only amounts to check if it is allowed. Now, adding a rule $p \mapsto c$ can be done if it does not exist a set of possible inputs which can lead both to the premise p and also to a conclusion incompatible with c. So if a rule is "allowed" then no input fact used with it will bring inconsistencies, this is why when a rule is added there is no new input fact to forbid.

An interesting point is that a priori revision being done once for all, it means that

incrementing can be done iteratively. However some increments are not allowed. The computation of diagnoses as well as the incrementing process is based on a ATMS and described in [95].

The idea of pre-processing a knowledge base was not new, Liberatore [147] had investigated the feasibility of reducing the on-line complexity of some AI problems (diagnosis, planning, reasoning about actions, and belief revision). In particular, he has shown how to use a pre-compilation for iterated revision: first compute a formula K' which can represent the result of the classical iterated revision of K by a sequence of formulas $\varphi_1, \ldots, \varphi_n$. This computation is made once for all given a knowledge base K and the formulas $\varphi_1, \ldots, \varphi_n$. The on-line remaining steps consists only in computing if for a given formula ψ, $K' \models \psi$. The use of pre-processing for revision had also been done by Coste-Marquis and Marquis in [70] for stratified knowledge bases, the aim of the authors was to decide, in polynomial time w.r.t. the size of the KB, if a formula is a logical consequence of the pre-compiled revised knowledge base. Our, approach is linked with these two works, but the main difference is that we place ourselves before the arrival of the pieces of information $\varphi_1, \ldots, \varphi_n$. So we can view a priori revision as a pre-compilation of the pre-compilation done by [147] and [70].

In that view, our approach has not much links with all the approaches proposed for handling iterated-revision, first introduced by Darwiche and Pearl [73], that are based on rankings on interpretations that can evolve with the arrival of new information. The main difference is on the priority given to the new information: in a-priori revision the initial knowledge is prioritary and the idea is to protect it against inconsistent evolution, while in iterated-revision the priority is given to the new pieces of information (see *e.g.* the postulates proposed by Delgrande, Dubois and Lang [77] relating prioritized merging and iterated revision).

Notice that contrarily to classical revision, in our work, after the pre-compilation of the a priori revised KB, addition and iterated addition of new pieces of information is very simple, and can be done in a polynomial time.

The use of constraints and rules has already been done in the field of non-monotonic reasoning (see for instance [41]). Indeed they are very expressive, this is why we thought that our framework should express this two kinds of rules. Notice that some frameworks provide more precise distinctions by ranking the rules in the knowledge base according to their importance. This more precise distinction is done in our proposal by using penalties (as in [102]) associated to the formulas. A refinement of a priori revision could allow to modify the rules of the initial knowledge base. Indeed, in our current work the pre-compilation consists only in removing rules and forbidding input facts, a possible extension should be to attenuate some rules by precising its premise. This kind of attenuation has been done for repairing inconsistencies, see for instance the recent proposal of Doder and Vesic in [79].

1.d Default rules

[106] F. Dupin de Saint-Cyr and H. Prade. Possibilistic Handling of Uncertain Default Rules with Applications to Persistence Modeling and Fuzzy Default Reasoning. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 440–450. AAAI Press, 2006

1.d.1 Default rules and uncertain default rules

A default rule is a typical generic information relating some premises a to a conclusion b. $a \rightsquigarrow b$ translates, in the possibility theory framework, into the constraint $\Pi(a \wedge b) > \Pi(a \wedge \neg b)$ which expresses that having b true is strictly more possible than having it false when a is true [38]. Default rules can not be represented by a material implication (*i.e.*, they are not equivalent to their contraposed formula). This is due to the fact that the premises and the conclusion play a completely distinct role: namely the premise is related to the context *i.e.*, the applicability of the rule while the conclusion is a result that is obtained only when the rule is fired. This is why this kind of rule can only be handled in forward reasoning, like a production rule.

Defeasibility is crucial in order to be able to reason with incomplete information, indeed in many situations it is more convenient to express general behavior concisely without referring to exceptional cases. More precisely, using default rules has three advantages :

- it allows to describe roughly a piece of knowledge without entering into details, hence it is more concise
- it allows to handle default rules together with their exceptions, hence the reasoning with default rules is non-monotonic
- knowing the precise context is not necessary since normality is assumed.

In the following, we consider a set Δ of default rules (as defined in Section 1.d), together with a propositional factual base FC describing all the available information about the context. In my work done with Henri Prade [106], we have proposed a new method for handling default rules which consists in rewriting the rules by expliciting their exceptions in the context. The idea is to generate automatically from Δ a set of non-defeasible rules D in which the condition parts explicitly state that we are not in an exceptional context to which other default rules refer. In the same time, strict rules called "completion rules" stating that we are not in an exceptional situation are added to a new set CR. The use of these completion rules is motivated by the need of reasoning in presence of incomplete information: the completion allows us to still be able to apply the modified rules which now have a more precise condition part. Note that the rules in CR will only be used if they are consistent with the context described in FC (taking D into account). Hence, it only requires to do a consistency test each time the context FC is changed.

We have proposed an algorithm for the rewriting based on System Z[3]. We showed that this algorithm terminates and that the set D of strict rules given by this algorithm is consistent. Note that each rule of the initial default knowledge base is present, modified or not, in the resulting rule base. So, there is no loss of information as with System Z inference. We also showed that the entailment defined on the rewriting method verifies *Reflexivity*, *Left logical equivalence*, *Right weakening*, *Or*, *Cautious monotony*, *Cut* and *Rational monotony.*

Example 5 *Considering* $\Delta = \{b \rightsquigarrow f, b \rightsquigarrow l, b \wedge w \rightsquigarrow \neg f\}$ *where b, f, l and w stand respectively for bird, fly, have legs and wounded, we can rewrite these rules by describing explicitly their exceptions starting from the last stratum. It gives the following knowledge base* $D = \{b \wedge w \rightarrow \neg f, b \wedge \neg w \rightarrow f, b \rightarrow l\}$. *There is only one completion rule:* $CR = \{b \rightarrow \neg w\}$, *hence, in the context* $FC = \{b\}$, *the completion rule is consistent, so it allows us to deduce* $f \wedge l$. *In the context* $FC = \{b \wedge w\}$ *we cannot add the completion rule since it is inconsistent with* FC *so we can conclude* $\neg f \wedge l$.

The non-monotonic inference relation called "Rewriting entailment" is a "rational closure" entailment, and allows us to deduce more conclusions than "System Z" [161] entailment (or its equivalent "best-out" entailment [37]). There has been other proposals for "rational closure" inference from defaults, among them, the "lexicographic entailment" [37, 144] is an approach that is recognized to give good results, in particular, it avoids "blocking of inheritance problems". Meanwhile it has a drawback, it is sensitive to direct or indirect redundancy since it is based on a counting of the rules.

Reasoning under incomplete information by means of rules having exceptions, and reasoning under uncertainty are two important types of reasoning that artificial intelligence has studied at length and formalized in different ways in order to design inference systems able to draw conclusions from available information as it is. They indeed address two distinct problems, in general using symbolic and numerical approaches respectively. However, the joint handling of exceptions and uncertainty has received little attention in non-monotonic reasoning, up to few noticeable exceptions (see the approaches proposed independently by Goldsmidt and Pearl [120], or by Lukasiewicz [149], or by Nicolas, Garcia and Stephan [157]), or even conditional probabilities that do exhibit a kind[4] of non-monotonic behavior when its context part is modified). We have addressed this topic in [76].

In order to be triggered, default rules only require "general" information, which agrees with situations of incomplete information. However, conclusions that we want to privilege in a given context may themselves be pervaded with uncertainty. Indeed, when a rule of the form "if a then generally b" is stated, no estimate of the certainty of having b true in context a is provided, even roughly. In the following definition we introduce the notion

[3]Pearl's System Z [161] gives a stratification of a set of default rules s.t. the first stratum contains the most specific rules, *i.e.*, which do not admit exceptions (at least, expressed in the considered default base), the second stratum has exceptions only in the first stratum and so on.

[4]Indeed translating the default "if a then b generally" by a constraint of the form $Prob(b \mid a) \geq \alpha$ violates System P postulates of non-monotonic reasoning [160, 86].

of uncertain default rule, in Section 1.d.3 we will discuss about the problem of handling this kind of rule that involves a joint processing of defeasibility and uncertainty.

Definition 3 *An* uncertain default rule *is a pair* $(a \rightsquigarrow b, \alpha)$ *where* a *and* b *are propositional formulas of* $\mathscr{L}$*, and* α *is the certainty level of the rule, the symbol* $\rightsquigarrow$ *is a non classical connective encoding a non-monotonic consequence relation between* a *and* b*.*

In the following, for simplicity, we use for certainty levels the real interval scale $[0, 1]$. However a qualitative scale could be used, since only the complete preorder between the levels is meaningful. The intuitive meaning of $(a \rightsquigarrow b, \alpha)$ is "by default" if a is true then b has a certainty level at least equal to α. Note that, in general, there is no relation between the certainty level associated with a default rule and the certainty level associated with a more specific rule. In particular, it would be wrong to assume that the more specific rule always provides a more certain conclusion. The status of being a default rule, is just a proviso for possible exceptional situations to which other rules in the knowledge base may refer.

For instance, the rule "birds with large wings fly" is more certain than "birds fly", while one may consider that the rule "Antarctic birds fly" is less certain than "birds fly", assuming that in Antarctic there are many penguins (that do not fly) together with some more sea birds that fly.

But, even if it is less certain, the specific rule that fits the particular context of incomplete information at hand, is the right one to use.

Moreover, handling uncertainty, at least qualitatively, in a given incomplete information context is crucial in various situations. For example, high level descriptions of dynamical systems often requires both the use of default rules that express persistence (persistence rules are typical defeasible rules) and the processing of uncertainty due to the limitation of the available information. Another illustration concerns fuzzy defaults such as "young birds cannot fly" understood as "the younger the bird, the more certain it cannot fly" (it has been published in [106]).

1.d.2 Particular Applications

Persistence modeling

The "frame" and "qualification" problems [151] are classical representation problems about dynamic systems. The "frame problem" comes from the quasi-impossibility to enumerate every fluent[5] that is not changed by an action. The "qualification problem" refers to the difficulty to exactly define all the preconditions of an action. An idea common to many proposals for solving the frame problem is to use default comportment descriptions for expressing persistence. Stating default transitions may be also useful for coping with the qualification problem. Besides, the available knowledge about the way a real system under study can evolve may be incomplete. This is why uncertainty should also be represented, at least in a qualitative way.

[5] A fluent is a literal representing a property of the world/system which may evolve over time.

In this section, let $\mathscr{A}$ be the set of action symbols. We consider that the symbols set $\mathscr{V}$ contains in addition to the symbols representing facts all the symbols $do(a)$ where $a \in \mathscr{A}$, representing action occurrences. When there is ambiguity, symbols may be indexed by a number representing the time point in which it is considered (see Notations). The evolution of the world is described by uncertain default rules of the form $(\varphi_{(t)} \rightsquigarrow \psi_{(t+k)}, \alpha)$ with $k \geq 1$, meaning that if φ is true at time t then ψ is generally true at time $t+k$ with a certainty level of α.

In order to handle the frame problem, we have chosen to define a frame axiom. Among all the kinds of fluents, we can distinguish persistent fluents (for which a change of value is surprising), from non persistent ones (which are also called dynamic by Sandewall [171]). Here, we assume that a set of non persistent literals NP is defined. Note that occurrences of actions are clearly non persistent fluents: $\{do(a)|a \in \mathscr{A}\} \subseteq NP$.

Definition 4 (frame axiom) *$\forall f \in \mathscr{V}$, if $f \notin NP$ then $(f_{(t)} \rightsquigarrow f_{(t+1)}, p(f))$ and if $\neg f \notin NP$ then $(\neg f_{(t)} \rightsquigarrow \neg f_{(t+1)}, p(\neg f))$ where $p(f)$ is the persistence degree of f.*

The persistence degree depends on the nature of the fluent, for instance, the fluent *asleep* is persistent but it is less persistent than $deaf$.

Given the description of an evolving system composed of a set of uncertain default transition rules Δ describing its behavior (Δ contains laws describing fluents evolutions when times goes by or actions/events occur, and default persistence rules (coming from the frame axiom)) and a possibilistic knowledge base $FC_{(t)}$ that describes the initial state of the world, we have studied the problem of predicting the next state $FC_{(t+1)}$ of the world.

Fuzzy default rules

Let us outline another application for uncertain default where the certainty levels may vary. This setting enables us to handle fuzzy default rules of the form 'the more b, the more it is certain that a implies c is true" (this kind of rules were first introduced by Benferhat, Dubois and Prade in [40]). For instance, "the younger a bird is, the more certain it cannot fly". This kind of rule can be encoded by $(b \rightsquigarrow \neg f, \mu_y)$ where μ_y is a certainty level depending on how young is the bird. For instance, if *tweety* is a bird of known age then the plausible consequence $(\neg fly(tweety), \mu_y(age(tweety)))$ can be obtained.

Decreasing Persistence

The possibility to affect variable levels to a rule may be also useful in order to express decreasing persistence (see *e.g.* the study done with Jérôme Lang in [99]): the more the time goes by the less it is certain that a fluent keeps its value. A decreasing persistence rule is generally of the form $(m_{(t)} \rightsquigarrow m_{(t+d)}, f(m, d))$ where the level attached to the rule depends on the fluent quality (highly persistent or dynamic) and of the length d of the time interval.

1.d.3 Inference with uncertain default rules

The core of the treatment of uncertain default rules that we proposed with Henri Prade [106], is based on the idea of translating them into a set of uncertain (non defeasible) rules. We had chosen to model uncertainty in the qualitative setting of possibility theory [85, 87]. Indeed, this agrees with the qualitative nature of default rules.

Let $U\Delta$ be a set of uncertain default rules of the form $(a \rightsquigarrow b, \alpha)$, while Δ continues to represent a set of default rules without certainty levels. In this approach, two types of levels are involved: namely levels encoding specificity and levels of certainty. Although it is possible to handle specificity by possibilistic logic in the same manner as the certainty levels will be processed in this section, the two types of levels should not be confused and the inference process uses the two scales separately.

We compared three methods for handling default rules, the first one in the possibilistic counterpart [39] of System Z, the second is contextual entailment [41] and the third one is the rewriting method presented in Section 1.d.1. In these three methods, specificity is used to determine which rules are appropriate in the current context. We denote by D the set of strict rules obtained from Δ by applying one of the three methods, and we denote by UD the corresponding set of strict rules associated with their certainty levels. Then, in the resulting base UD, the certainty levels are taken into account in agreement with possibility theory in order to draw plausible conclusions with their certainty levels.

The first method suffers from a drawback called the "drawning effect" in particular when modelizing a dynamic system: all the persistence rules are drowned. One noticeable advantage of the third method is that the deduction can be iterated without recompilation of the default base (whereas it would be necessary with the second method).

1.d.4 Related works

There has been very few works handling both defeasibility and uncertainty, up to the noticeable exception of system Z^+ (defined by Goldsmith and Pearl [120]). In system Z^+, a default rule $(a \rightsquigarrow b)$ is extended with a parameter representing the degree of strength or firmness of the rule and denoted by $(a \rightarrow^{\delta} b)$. In Z^+, the ranking of defaults is obtained by comparing sums of strength degrees, somewhat mixing the ideas of specificity and strength. Separate scales for specificity and certainty are not used is this approach, this may not always yield the expected conclusion.

Nicolas *et al.* [157, 156] also present an approach that deals with defeasibility and uncertainty in a possibilistic framework. But, they combine possibilistic logic with Answer Set Programming rather than using the same setting for default and uncertainty handling. Certainty levels are used in order to help to restore consistency of a logic program by removing rules that are below a level of inconsistency. As our first method, this approach does not avoid the drowning problem, while our two other methods do.

Using an uncertain framework in order to describe an evolving system has been done by many authors, for instance in a probabilistic setting. But reasoning in this setting implies to dispose of many a priori probabilities, this is why using defeasibility may help to reduce the size of information for representing the system. Besides, it is a common idea

to define a frame axiom in terms of default rules (see Lang et al. [142] for an overview or my more recent production with Andreas Herzig, Jerome Lang and Pierre Marquis in the French book "Panorama de l'intelligence artificielle" [96]). But, as far as I know, frame rules are either considered as default rules (see Giunchiglia et al. [119], or Baral and Lobo [31] for instance), or are associated with low priority levels (see Kakas et al. proposal in [132]), but do not involve both default and uncertainty features.

In this chapter we have explored several representations of generic information and have proposed inference mechanisms able to deal with inconsistency problems. The first section concern classical inference and inconsistency checking. In the second section, we have explained how inference can be done with spatially reified formulas. The particularity of this inference is due to the fact that some formulas apply everywhere on a parcel and other only somewhere. Then we showed how spatial information fusion can be handled in a possibilistic setting. The third section describes "a priori revision", this work concerns KB that contains production rules, and the aim is to provide a way to protect them against incoming information that could bring inconsistency. In the fourth section we have presented an approach aiming at reasoning with default rules. The aim of this work is very closed to the aim of the work I did in "a priori revision" since the idea is to transform rules into material implication but it differs on three points: 1) in the rewriting method when the rules may bring inconsistency they are modified and not deleted; 2) the algorithm that we provided for the rewriting method is not based on an ATMS but rather on the stratification given by System Z; 3) the idea of the rewriting method is to enable us to draw defeasible conclusions hence formulas are added for this purpose while in a priori revision the added formulas are constraints forbidding some inputs facts. Note that "rewriting" defaults by mentioning explicit exceptions is reminiscent of techniques used in circumscription-based approaches. Lastly, we have proposed inference mechanisms that allow us to handle rules which are both uncertain and defeasible. The inference method has two steps: first building a set of non defeasible rules that can be used in the current context, and then processing the uncertainty of the identified rules in the setting of possibility theory.

As seen for a priori revision or in the applications of defeasibility, inconsistency handling and representation problems are strongly related to possible changes. Since in order to chose a representation setting one must foresee what will be the possible evolutions of the system hence prepare oneself to manage forthcoming inconsistencies...

Chapter 2

Reasoning about a dynamic system

Intelligence is often related to the ability to adapt oneself to any new environment, hence handling change has been an important research domain in AI. As it is recalled in [96], reasoning about actions and change is maybe the most classical AI topic: in particular it is the subject of the founding article of McCarthy and Hayes [151]. Research in this area had been very productive until the end of the nineties. It has led to propose solutions to the different problems linked to action representation and to elaborate a typology of the different formalisms on the basis of their expressive power [171] and to progress towards automatic reasoning about actions and change.

There are various reasons why an agent may desire to act on the current state of a dynamic system. It is either to modify it in order to obtain a better situation for its own interest, or to maintain a given state, or to ensure that the successive states are not digressing too much from a normal trajectory, or the agent may desire to acquire a better knowledge of it. Such reasons imply some concepts (state, action, observation, etc.) and some process relating them (planning, prediction, explanation, etc.). We focus here on the problem of handling the arrival of new pieces of information that may be contradictory w.r.t. what was previously believed.

The way change is taken into account has an influence on the methods for handling generic information. There is a classical dichotomy (first discovered by Winslett [200] and developed by Katsuno and Mendelzon [136]) between two kinds of evolutions of the beliefs: the first one, called "revision" concerns changes in the beliefs about a system which itself has not changed while the second one called "update" is defined by an evolution of the system itself. Rather than the usual presentation of revision by Alchourrón, Gärdenfors and Makinson [1], that views revision as mapping a closed logical theory K and an input formula α to a closed logical theory $K \star \alpha$. In this report, we choose the syntactical presentation of Katsuno and Mendelzon [136], which views revision as mapping a propositional formula φ (representing the initial belief state) and another propositional formula α (the "input") to a propositional formula $\varphi \star \alpha$ (representing the belief state after revision by α). This simplification is without loss of generality whenever the propositional language is generated by a *finite* set of propositional symbols (which is the case here).

In this chapter, we first explain the process of extrapolation (which amounts to reason about incomplete temporal information and try to complete it), second we present our work about causal ascription (which amounts to find what is the cause of a given temporal fact) and lastly we present our work about the axiomatisation of update operators satisfying transition constraints.

2.a Extrapolation

[101] F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). *Artificial Intelligence*, 175:760–790, janvier 2011
[100] F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). In *International Conference, Principles of Knowledge Representation and Reasoning (KR)*, pages 497–508. Morgan Kaufmann Publishers, avril 2002

The process of extrapolation [101, 100] starts from a set of temporal observations, encoded by a temporal formula Ψ, and tries to complete these observations in order to infer new information assuming that changes are exceptional. Such a process is a specific case of *chronicle completion* [171]. The rationale of belief extrapolation is that as long as nothing tells the contrary, fluents do not change. More precisely, the basic assumptions for extrapolation operators are the following:

1. the agent observes some properties of the world at different time points, but does not have the ability of performing actions; furthermore, if exogenous events occur, they are perceived by the agent through the observations of their effects only (for instance, the occurrence of the event that it rained last night is perceived by actual rain last night, or by seeing the wet ground this morning). This is why changes can be qualified as unpredicted.

2. the system is inertial: by default, it remains in a static state. This assumption justifies the use of a change minimization policy.

We had introduced particular temporal formulas, called scenarios, for speaking about series of observations at different (but precisely located) time points.

Definition 5 (scenarios) *A* scenario Σ *is a conjunction of t-formulas for* $t \in [\![1, N]\!]$*, i.e., a temporal formula of the form* $\varphi^1_{(1)} \wedge ... \wedge \varphi^N_{(N)}$*, where* $\varphi^1, \ldots, \varphi^N$ *are formulas of* $\mathscr{L}$*. To simplify notations, a scenario is written under the form:* $\Sigma = \langle \varphi^1, ..., \varphi^N \rangle$*, where* $\Sigma(i) = \varphi^i$*.* $\mathscr{S}_{(N)}$ *denotes the set of all scenarios*[1]*.*

As observed with spatial attributive formulas, a temporal formula is not necessarily expressible by a scenario. Hence, considering scenarios *only* is not sufficient if we want to express implications between fluents at different time points, or observations whose temporal location is imprecise.

[1]Formally, we should write $\mathscr{S}_{\mathscr{V},N}$, since the set of scenarios has been defined for a given set of propositional symbols $\mathscr{V}$ and a fixed N. However, for sake of lightness, we omit the subscripts when there is no ambiguity.

Example 6 *Let $\mathcal{V} = \{a, b\}$ and $N = 3$. $a_{(1)} \wedge (a_{(2)} \vee \neg b_{(2)}) \wedge \top_{(3)}$ is a scenario, more simply denoted by $\langle a, a \vee \neg b, \top \rangle$. $a_{(1)} \vee b_{(2)}$ is a temporal formula (but not a scenario).*

2.a.1 Extrapolation Semantic

Definition 6 (trajectories) *$TRAJ_{(N)} = 2^{\mathcal{V}_{(N)}}$ denotes the set of all interpretations for $\mathcal{V}_{(N)}$, called* trajectories. *For shortness, a trajectory is represented by a sequence $\tau = \langle \tau(1), \ldots, \tau(N) \rangle$ of interpretations in Ω. The satisfaction of a temporal formula $\Psi \in \mathscr{L}_{(N)}$ by a trajectory $\tau \in TRAJ_{(N)}$ is defined as in standard propositional logic and is denoted by $\tau \models \Psi$. $Traj(\Psi) = \{\tau \in TRAJ_{(N)} | \tau \models \Psi\}$ is the set of trajectories satisfying Ψ. A temporal formula Ψ is consistent iff $Traj(\Psi) \neq \varnothing$. A trajectory τ is* static *iff $\tau(1) = ... = \tau(N)$.*

If τ is a trajectory and $v \in \mathcal{V}$, we denote by $\tau(t)(v)$ the truth value of v at time t in τ, that is, $\tau(t)(v) = true$ if $\tau(t) \models v$ and $\tau(t)(v) = false$ if $\tau(t) \models \neg v$.

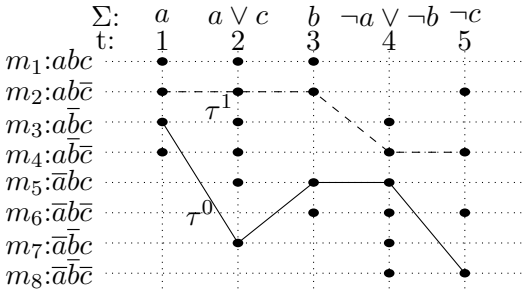


Figure 2.1: *Two trajectories satisfying $\Sigma = \langle a, a \vee c, b, \neg a \vee \neg b, \neg c \rangle$*

Example 7 *Let us consider the scenario $\Sigma = \langle a, a \vee c, b, \neg a \vee \neg b, \neg c \rangle$. The trajectories satisfying Σ are all those connecting the big dots of Figure 2.1. Two of them are represented: $\tau^0 = \langle\ m3, m7, m5, m5, m8\ \rangle$ and $\tau^1 = \langle\ m2, m2, m2, m4, m4\ \rangle$.*

We define the *change set* of a trajectory as the set of all pairs consisting of a literal and a time point such that the literal becomes true at this time point. Note that a trajectory τ can be unambiguously defined by one of its states (for instance, its initial state $\tau(1)$) and its change set.

Definition 7 (change set)
The change set *$Ch(\tau)$ of a trajectory τ is defined by:*

$$Ch(\tau) = \left\{ \langle l, t \rangle \;\middle|\; \begin{array}{l} l \in LIT,\ t \in [\![2, N]\!], \\ \tau(t-1) \models \neg l \text{ and } \tau(t) \models l \end{array} \right\}$$

Notation: $Ch(\tau)(t) = \{\{l \mid \langle l, t \rangle \in Ch(\tau)\}$ is the set of literals changing to true at time point t

For instance, in the previous example, $Ch(\tau^0) = \{\langle \neg a, 2\rangle, \langle b, 3\rangle, \langle \neg b, 5\rangle, \langle \neg c, 5\rangle\}$ and $Ch(\tau^1) = \{\langle \neg b, 4\rangle\}$. $Ch(\tau^0)(5) = \{\neg b, \neg c\}$.

Semantically, extrapolation consists in finding the preferred trajectories satisfying Ψ, with respect to some given preference relation between trajectories (this is similar to many approaches to non-monotonic reasoning, where we select preferred models among those that satisfy a formula). A *preference relation on trajectory* $\preceq$ is a reflexive and transitive relation on $TRAJ_{(N)}$ (not necessarily connected). $\tau \preceq \tau'$ means that τ is at least as preferred[2] as τ'.

In this section, most preference relations are inertial (*i.e.*, static trajectories are always preferred to non-static ones) and change-monotonic (*i.e.*, they can be defined from a comparison of their change sets). We are now in position to formally define an extrapolation operator.

Definition 8 (extrapolation operator)
Let $\preceq$ be an inertial preference relation on $TRAJ_{(N)}$. The extrapolation operator *induced by $\preceq$ maps every temporal formula Ψ to another temporal formula $E_{\preceq}(\Psi)$, unique up to logical equivalence, which is satisfied exactly by the* preferred *trajectories among those that satisfy Ψ. More formally, $E_{\preceq}$ is a mapping from $\mathscr{L}_{(N)}$ to $\mathscr{L}_{(N)}$ such that*

$$Traj(E_{\preceq}(\Psi)) = \min(\preceq, Traj(\Psi))$$

If $\preceq$ is complete then $E_{\preceq}$ is said to be a complete extrapolation operator.

Note that we have required $\preceq$ to be inertial. This condition will be droped for extended extrapolation operators (see Section 2.b).

As said before, we give special attention to the case where the input is a *scenario* and we define the operator that associates with each initial scenario an *extrapolated scenario* defined as the scenario projection of its extrapolation. Formally: Ex is the mapping from $\mathscr{S}_{(N)}$ to $\mathscr{S}_{(N)}$ defined by $Ex(\Sigma) = S(E(\Sigma))$ where $S(\varphi)$ is the scenario approximation of the formula φ s.t. $Mod(S(\varphi)_{(t)}) = \{\tau(t) | \tau \in Traj(\varphi)\}$. Scenario-scenario extrapolation is semantically characterized by:

$$Traj(Ex(\Sigma))(t) = \{\tau(t) \mid \tau \in \min(\preceq, Traj(\Sigma))\}$$

Example 8 *Let $\mathscr{V} = \{a\}$ and $N = 2$. The four possible trajectories are $\tau^1 = \langle a, a\rangle$, $\tau^2 = \langle a, \neg a\rangle$, $\tau^3 = \langle \neg a, a\rangle$ and $\tau^4 = \langle \neg a, \neg a\rangle$. Consider the preference relation $\preceq$ defined by $\tau^1 \sim \tau^4 \prec \tau^2 \sim \tau^3$. Let $\Sigma = \langle \top, \top\rangle$. We have $Traj(\Sigma) = \{\tau^1, \tau^2, \tau^3, \tau^4\}$ and $min(\preceq, Traj(\Sigma)) = \{\tau^1, \tau^4\}$, from which we get $E_{\preceq}(\Sigma) = (a_{(1)} \wedge a_{(2)}) \vee (\neg a_{(1)} \wedge \neg a_{(2)})$ and $Ex_{\preceq}(\Sigma) = Ex_{\preceq}(\top_{(1)} \wedge \top_{(2)}) = \langle \top, \top\rangle$.*

This example shows that unlike formula-formula extrapolation E, scenario-scenario extrapolation Ex generally leads to a loss of information. This is due to the fact that for a given set X of trajectories there is generally no scenario Σ such that $Traj(\Sigma) = X$.

[2]Here, preference has to be interpreted in terms of plausibility, and not in decision-theoretic terms: $\tau \preceq \tau'$ means that τ is *at least as plausible* as τ'.

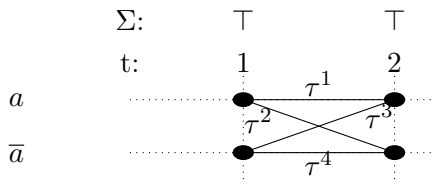


Figure 2.2: *The four trajectories satisfying* $\Sigma = \langle \top, \top \rangle$

2.a.2 Some extrapolation operators

We have studied several examples of typical, inertial and change-based preference relations together with their associated extrapolation operators. Although there are as many change-based preference relations as preference relations on the set of all possible change sets, there is a reasonable number of prototypical change-based relations obtained by making some natural neutrality assumptions on fluents and on time points.

For instance $\preceq_{nct}$ preference relation (*nct* standing for *number of change time points*) prefers trajectories in which changes occur at fewer time points. Now, *nct* does not distinguish trajectories having time points with a single change from the one that has several changes per time points. $\preceq_{nc}$ prefers trajectories with a minimal *number of changes*, and $\preceq_{csi}$ prefers a trajectory to another if there is a *change set inclusion* between their respective change sets.

Example 7 (continued): *The trajectory* τ^1 *is preferred to* τ^0 *wrt nct relation and wrt nc. However,* $\tau^2 = \langle m3, m3, m6, m6, m6 \rangle$ *which has three changes at time point 3 will be equivalent to* τ^1 *with nct while be less preferred by nc.*

In all, we have obtained 15 "natural" preference relations, assuming neutrality between time points and fluents. These relations have been refined by giving three ways of relaxing the neutrality assumption:

chronologically by giving priority to later or earlier time points in change sets. For instance, $\preceq_{csi}$ can be refine into a preference relation that prefers trajectories where changes occurs as late as possible. This policy is called *chronological minimization* in [178] and is extensively discussed in [171]. It is defined by:

$$\tau \preceq_{chr} \tau' \text{ if } Ch(\tau) = Ch(\tau') \text{ or } \exists k \in [\![1, N]\!] \text{ s.t. } \begin{cases} Ch(\tau)(k) \subset Ch(\tau')(k) \\ \forall t < k, Ch(\tau)(t) = Ch(\tau')(t) \end{cases}$$

fluent penalties by giving priority to some changes over others according to the nature of the fluents: this was done by associating penalties to fluents.

event penalties by considering events as sets of dependent changes rather than atomic changes with a event-based penalty preference relation (see Section 2.b for a more detailed study of this kind of relation).

2.a.3 Extrapolation and belief change

Extrapolation and revision

Belief revision is often thought of as dealing with *static worlds*, therefore with formulas pertaining to the same time point. However, as remarked by Friedman and Halpern in [115], "what is important for revision is not that the world is static, but that the propositions used to describe the world are static". That is, nothing prevents us from considering revision operators in a language generated from a set of time-stamped propositional symbols. We have shown in [101], that *belief extrapolation is a specific instance of belief revision (on a time-stamped language)*: extrapolation amounts to *revising the prior belief that all fluents persist throughout time by the observations*. More precisely, the fact that all fluents persist throughout time is represented by a set of persistence rules defined by $PERS = \bigwedge_{v \in \mathscr{V}} \bigwedge_{t=1}^{N-1} v_{(t)} \leftrightarrow v_{(t+1)}$ it is the temporal formula having for models the set of static trajectories ($PERS$ is true if and only if no change occurs between time 1 and time N). This result in turn allowed us for deriving easily a representation theorem for extrapolation.

Theorem 1 *$E : \mathscr{L}_{(N)} \rightarrow \mathscr{L}_{(N)}$ is a complete extrapolation operator iff it satisfies Ex1 to Ex6.*

Ex1 $E(\Psi) \models \Psi$

Ex2 *If $PERS \wedge \Psi$ is satisfiable then $E(\Psi) \equiv PERS \wedge \Psi$.*

Ex3 *If Ψ is satisfiable then $E(\Psi)$ is satisfiable.*

Ex4 *If $\Psi \equiv \Psi'$ then $E(\Psi) \equiv E(\Psi')$.*

Ex5 $E(\Psi) \wedge \Psi' \models E(\Psi \wedge \Psi')$

Ex6 *If $E(\Psi) \wedge \Psi'$ is consistent then $E(\Psi \wedge \Psi') \models E(\Psi) \wedge \Psi'$.*

These postulates Ex1-Ex6 correspond to the postulates R1-R6 of Katsuno and Mendelzon [137][3] . Note that when Ψ is a scenario Σ, (Ex2) comes down to

Ex2S If $\bigwedge_{t=1}^{N} \Sigma(t)$ is satisfiable then for every $t \leq N$, $E(\Sigma)(t) \equiv \bigwedge_{t=1}^{N} \Sigma(t)$.

[3]Katsuno and Mendelzon's formulation of the AGM postulates for belief revision are:

- R1: $\varphi \circ \mu$ implies μ
- R2: If $\varphi \wedge \mu$ is satisfiable then $\varphi \circ \mu \equiv \varphi \wedge \mu$.
- R3: If μ is satisfiable then $\varphi \circ \mu$ is also satisfiable.
- R4: If $\varphi_1 \equiv \varphi_2$ and $\mu_1 \equiv \mu_2$ then $\varphi_1 \circ \mu_1 \equiv \varphi_2 \circ \mu_2$.
- R5: $(\varphi \circ \mu_1) \wedge \mu_2$ implies $\varphi \circ (\mu_1 \wedge \mu_2)$.
- R6: If $(\varphi \circ \mu_1) \wedge \mu_2$ is satisfiable then $\varphi \circ (\mu_1 \wedge \mu_2)$ implies $(\varphi \circ \mu_1) \wedge \mu_2$.

where $\varphi, \varphi_1, \varphi_2, \mu, \mu_1, \mu_2$ are propositional formulas and $\circ$ is the revision operator that associates with two formulas φ and μ a propositional formula denoted $\varphi \circ \mu$ resulting from revising φ by μ.

An immediate – but important – property deriving from Ex1, Ex4 and Ex5 is that extrapolation is idempotent: $E(E(\Psi)) \equiv E(\Psi)$.

The previous representation theorem characterizes extrapolation operators that are defined on complete inertial orderings, but intuitively, it is natural to expect such orderings to be incomplete, allowing two trajectories to be incomparable. We have also shown that incomplete extrapolation operators can also be characterized by the set of postulates Ex1-Ex5, E7, E8 with

E7 If $E(\Psi) \models \Psi'$ and $E(\Psi') \models \Psi$ then $E(\Psi) \equiv E(\Psi')$.

E8 $E(\Psi) \wedge E(\Psi') \models E(\Psi \wedge \Psi')$

Hence technically speaking, an extrapolation operator is a revision operator on a time-stamped language where the initial belief state is *fixed* (to $PERS$). *Is it nothing more than that?* The answer is both positive and negative. Positive, because indeed, any extrapolation operator can be seen as a revision operator. Negative, because the temporal structure makes extrapolation a specific class of belief change operators with its specific properties.

We end this section by briefly discussing the possible connections between extrapolation and Lehmann's *iterated belief revision* [144]. An iterated revision function maps any sequence of formulas $\sigma = \langle \varphi_1, \ldots, \varphi_n \rangle$ to a belief state $[\sigma]$ resulting from the sequence σ of individual revisions (only the final result is considered). The difference between extrapolation and iterated revision is clear when considering the following example: let $\Sigma = \langle a \rightarrow b, a, \neg a \rangle$; any "reasonable" extrapolation operator satisfies $Ex(\Sigma) = \langle a \wedge b, a \wedge b, \neg a \wedge b \rangle$ (the change from a to $\neg a$ between 2 and 3 being certain, the preferred trajectory is the one containing no other changes). Now, Lehmann's iterated revision, and also most iterated revision operators defined on epistemic states (*e.g.*, [42, 73]) give $[a \rightarrow b, a, \neg a] = \neg a$. The reason for this difference is that iterated revision is concerned with pieces of information concerning a *static world*; what evolves is the agent's belief state, not the state of the world. Therefore, once the new information $\neg a$ has "cancelled" the preceding one, the reasons to believe in b have disappeared. This strong "directivity" of time in iterated revision contrasts with extrapolation, where past and future can often be interchanged (as soon as *Reversibility*[4] holds).

Extrapolation and update

The key property of belief update [136, 200] is Katsuno and Mendelzon's postulate **U8** which tells that models of K are updated independently:

$$\mathbf{U8}\ (K1 \vee K2) \diamond \varphi \leftrightarrow (K1 \diamond \varphi) \vee (K2 \diamond \varphi)$$

Belief update only considers two time points and takes as input a pair (K, α) of formulas referring respectively to $t = 1$ and $t = 2$. Rephrasing the framework of belief update in terms of extrapolation will amount to consider a scenario $\Sigma = \langle K, \alpha \rangle$ and

[4] see "Temporal properties"

compute a completed belief set at time 2. At first glance, this extrapolation may look similar to belief update. However, we have proven that this is not the case. The main reason for this result is that as soon as the language contains at least two propositional symbols, the AGM postulates are inconsistent with U8 (see for instance [124]).

Extrapolation and update complete each other and, in order to be able to reason both with implicit and explicit change, we can integrate both. This has been developed in my work about causal ascription [91] (see Section 2.b) under the name event-based extrapolation.

Temporal properties of extrapolation

We have defined and characterized the following *temporal* properties (*i.e.*, they explicitly refer to the flow of time) for extrapolation operators:

- *Inertia* is, by definition, satisfied by all extrapolation operators.
- *Reversibility* expresses that "forward" persistence (inferring beliefs from the past to the future) and "backward" persistence (inferring beliefs from the future to the past) are symmetric. Its formal definition is based on the reverse of a temporal formula which is a formula obtained by replacing each occurrence of any variable $x_{(t)}$ by $x_{(N-t+1)}$; more precisely, an extrapolation operator satisfies Reversibility if and only if $\forall \Psi \in \mathscr{L}_{(N)}$, we have $Reverse(E(\Psi)) = E(Reverse(\Psi))$.

  We have provided a semantic characterization (based on the preference relation on trajectories that should be indifferent to the reversal of trajectories). We have shown that Reversibility holds for many of the "typical" extrapolation operators, which sheds more light on how extrapolation departs from (iterated) revision and update, which obviously don't satisfy it. Note that Reversibility is strongly linked to the equal plausibility of a change from a to $\neg a$.
- The *Markovianity* property says that as soon as there is a *complete* observation at a given time point k, an extrapolation problem can be decomposed into two independent sub-problems: the first one up to k and the second one from k on. For this we had to work with families of extrapolation operators, parameterized by N, rather than for a fixed N, similarly, the length of scenarios could also vary. We were able to provide a sufficient condition for Markovianity based on the fact that the preference relation should be decomposable (*i.e.*, comparing two trajectories may be done by decomposing those trajectories into two parts and comparing the combination of each two parts): if the preference relation is decomposable then Ex satisfies Markovianity.

  We have shown that most operators satisfy Markovianity, but that the property fails typically for operators that use a global minimization like Ex_{icl} (inclusion of changing literal which compares by inclusion the set of literals that have changed).
- *Independence from empty observations* (IEO) expresses that adding an empty observation between two observations should not change anything to the way obser-

vations are extrapolated — or, in other words, the choice of the time unit has no influence on extrapolated beliefs.

We were able to provide a *sufficient* (but not necessary) condition for an extrapolation operator to comply with (IEO) and we showed that all the typical operators that we had defined satisfy it.

2.a.4 Computational issues

We have obtained complexity results for several preference relation $\preceq_x$ concerning the two decision problems:

- EXTRA$_x$: given two temporal formulas Ψ_1, Ψ_2, decide whether $E_x(\Psi_1) \models \Psi_2$;
- EXTRA-SC$_x$: given two scenarios Σ and Σ', decide whether $Ex_x(\Sigma) \models \Sigma'$.

For instance we have showed that EXTRA$_x$ and EXTRA-SC$_x$ are Π_2^p-complete for every $x \in \{csi, icl, chr\}$ while EXTRA$_{nc}$ and EXTRA-SC$_{nc}$ are $\Delta_2^{\mathsf{P}}(O(\log n))$-complete.

Moreover we have provided an algorithm for the practical computation of extrapolated beliefs. A bad news is that techniques for computing extrapolation are generally sensitive to the choice of the preference relation, which means that, to a large extent, the study has to be done independently for each preference relation. Considering all preference relations we had studied would have been far too long. Rather, we had focused on the "prototypical" preference relation $\preceq_{csi}$, and on scenario-scenario extrapolation.

We have shown that the computation of the preferred trajectories of a given scenario Σ w.r.t. $\preceq_{csi}$ (corresponding to the set of what we call *minimal explanations for this scenario*) could be done in four steps :

1. for each variable, compute its *relevant time points* for Σ (*i.e.*, the times points where there are some observations that are "relevant" to v. This notion of relevance was expressed similarly as in Lang et al. [142]), a formula φ is relevant for a variable v if any formula equivalent to φ contains v.
2. generate the *persistence formulas* (of the form $v_{(t)} \leftrightarrow v_{(t')}$) associated to each variable w.r.t. its relevant time points (computed at step 1), in order to express that v persists between two consecutive relevant observations;
3. select the *maximal subsets of persistence formulas consistent with* Σ;
4. compute the *minimal change sets* corresponding to the maximal persistence sets (obtained at step 3).

Given a scenario Σ (of length N), the number of minimal explanations (*i.e.*, minimal change sets that are consistent with Σ) can be very large, even when the number of changes is small:

Example 9 $\Sigma^0 = \langle a \wedge b \wedge c \wedge d,\ \top,\ \neg b,\ d,\ \neg a,\ \top,\ b,\ \neg c \vee \neg d \rangle$ *has* 352 *minimal explanations (and as many minimal trajectories). Now, these 352 minimal explanations can be expressed compactly:* γ *is a minimal explanation for* Σ^0 *iff it contains exactly the following changes:* $\langle \neg a, t \rangle$ *for exactly one* $t \in \{2..5\}$*;*$\langle \neg b, t \rangle$ *for exactly one* $t \in \{2, 3\}$*;* $\langle b, t \rangle$ *for exactly one* $t \in \{4..7\}$*; either* $\langle \neg c, t \rangle$ *for one* $t \in \{2..8\}$ *or* $\langle \neg d, t \rangle$ *for one* $t \in \{5..8\}$*.*

These explanations can be represented more succinctly in what we have called *compact change sets*, which are sets of change "contiguous" sets. Intuitively, a compact change set consists of the disjunction of all the change sets it covers, *e.g.*, two maximally compact change sets are covering the previous example: $\{\langle \neg a, 2, 5 \rangle, \langle \neg b, 2, 3 \rangle, \langle b, 4, 7 \rangle, \langle \neg c, 2, 8 \rangle\}$ and $\{\langle \neg a, 2, 5 \rangle, \langle \neg b, 2, 3 \rangle, \langle b, 4, 7 \rangle, \langle \neg d, 5, 8 \rangle\}$.

We have shown that finding a covering set of compact explanations can be seen as a logic-based abduction problem (where abducibles correspond to elementary changes), and can be computed using dedicated algorithms, hence complexity results and tractable classes that were obtained by Eiter and Gotlob [107] can be used to find some tractable sub-classes of belief extrapolation.

2.b Causal ascription

[101] F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). *Artificial Intelligence*, 175:760–790, janvier 2011
[91] F. Dupin de Saint-Cyr. Scenario Update Applied to Causal Reasoning. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 188–197. AAAI Press, 2008

In [91], we have proposed to define a kind of hypothetical reasoning using both update and extrapolation, which amounts to compute what could have happened if something had been different in a given story. Updating scenarios allows us to define formally the counter-factual aspect of causation: to check if an event is a cause in a given scenario amounts to update this scenario by the non-occurrence of this event. In many situations, this question is fundamental, since it may help to assign responsibilities, it may clarify causal relations, and enable people to distinguish variables which have a determining impact on the future from those which finally have no influence. This approach was developed in the context of the ANR project MICRAC (Computational and cognitive models of causal reasoning) 2005-2008.

2.b.1 Event-Based Extrapolation

In this section, we show how we had introduced explicitly events in basic extrapolation, we had started to work on it in [100] and developed it in [91] in order to encode causality. An event is an operation which induces a change in the normal course of the evolution. In the literature, events are often described by their effects and their preconditions. Here, we assume that the system evolution is described in a similar way as in the work of Boutilier [61], where the plausibility of events, as well as their dynamics, are modeled by ordinal conditional functions.

Definition 9 (Event encoding) *Given a set of event symbols Ev, the two following functions are available:*
- the function ke measures the surprise degree associated with the simultaneous occurrence of a set of events in a situation: $ke : \Omega \times 2^{Ev} \rightarrow \mathbb{N} \cup \{\infty\}$
- the function km that allows to obtain a penalty distribution over the possible situations (representing the surprise degree associated with the different situations after the simultaneous occurrences of a set of events in a situation): $km : \Omega \times 2^{Ev} \times \Omega \rightarrow \mathbb{N} \cup \{\infty\}$.

An infinite surprise degree means that the occurrence of the event (resp. the transition) is impossible. The two available characteristic functions ke and km are supposed to be coherent with a given action theory. ke generalizes the function "Precond" (for preconditions) of STRIPS [110] which defines necessary conditions for the occurrence of one event. It is supposed to be defined for any set of events (events which took place simultaneously). Here, the function km does not impose to have deterministic events as in the case of the function `Result` of the situations calculus [151].

Note that this definition is under the strong assumption of a Markovian behavior of the system, *i.e.*, the evolution of the system does not depend on its history but only on its current state. Taking into account non Markovian fluents would need a more heavy formalism and was left outside the scope of scenario update.

An "inert" system (in the classification of Sandewall [171]) is a system where no fluent is temporary. Thus, if no event occurs then the state of the world does not change and the occurrence of any event is surprising.

Definition 10 (Inert System) *The system is* inert *iff* $\forall m, m' \in \Omega$, $\forall ev \subseteq Ev$,

1. $km(m, \varnothing, m') = 0 \Leftrightarrow m = m'$ *and*
2. $ke(m, ev) = 0 \Leftrightarrow ev = \varnothing$

Thanks to temporal formulas it is possible to represent scenarios containing at the same time observations of facts and observations of event occurrences.

Definition 11 (Mixed Temporal Formula) *Let Ev be a set of event symbols (denoted $\varepsilon^1, \ldots, \varepsilon^P$)*[5]*. Let $\mathscr{V}' = \mathscr{V} \cup Ev$ and ${\mathscr{V}'}_{(N)} = \{v_{(t)} | (v \in \mathscr{V}$ and $t \in [\![1, N]\!])$ or $(v \in Ev$, and $t \in [\![1, N-1]\!])\}$ its set of associated time-stamped variables. We denote by LIT' the set of literals built from $\mathscr{V}'$. A* mixed temporal formula *is built on the variables of ${\mathscr{V}'}_{(N)}$ with the usual connectors and constants. Let ${\mathscr{L}'}_{(N)}$ be the set of these formulas.*

The formula: $d_{(1)} \wedge \varepsilon^{cd}_{(1)} \wedge \neg d_{(2)}$ is an example of mixed temporal formula expressing that the door was open (d was true) at time point 1 and the event ε^{cd} has occurred (meaning for instance that "somebody has closed the door") at time point 1 and at time point 2 the door was closed.

[5] More precisely, the symbol ε^i does not denote the event itself but its occurrence.

Definition 12 (Mixed Trajectory) *A* mixed situation *s is an interpretation of $\mathscr{V}'$. A* mixed trajectory *corresponds to a truth value assignment to the variables of ${\mathscr{V}'}_{(N)}$. Every mixed trajectory τ can be represented by a sequence $\tau = \langle \tau(1), \ldots \tau(N) \rangle$ of interpretations of $\mathscr{V}'$. Let $TRAJ'_{(N)}$ denote the set of mixed trajectories.*
A mixed trajectory is called static *if*
$\begin{cases} \forall t \in [\![1, (N-1)]\!], & e(\tau(t)) = \varnothing \text{ and} \\ \forall t \in [\![1, N]\!], & f(\tau(1)) = ... = f(\tau(N)). \end{cases}$
where $f(s)$ denote the facts that hold in the situation s (interpretations of $\mathscr{V}$) and $e(s)$ the events that occur in s (interpretations of Ev).

The cost of a trajectory corresponds to the sum for each time point of the surprise degree associated to the occurrence of the events at this time point added to the surprise degree to reach the following situation being given the occurrence of these events at the previous time point.

Definition 13 (Cost of Mixed Trajectory) *The* cost *of a trajectory τ is:*

$$k(\tau) = \sum_{t=1}^{N-1} ke(f(\tau(t)), e(\tau(t)) + km(f(\tau(t)), e(\tau(t)), f(\tau(t+1)))$$

where $f(s)$ denote the facts that hold in the situation s and $e(s)$ the events that occur in s.

We consider the cost of the events which occurred between time point 1 and N-1 without taking account of those which occurred at the last moment since they have no impact. Let us note that, here, the surprise degree associated with the occurrence of an event or the surprise degree associated with obtaining a given situation are quantified in terms of penalties. Besides, the use of penalties to characterize surprise degrees has been proposed initially by Sandewall in [170]. This choice is justified by the intrinsically additive and compensatory character of surprises, but the use of other measures like probabilities or possibilities is completely possible. One can refer to [102] for a study of the links between these various measures.

Belief extrapolation extended to mixed formulas is defined by means of a preference relation on trajectories which minimizes their cost, *i.e.*, which minimizes the surprise degree associated with the events of these trajectories.

Definition 14 *Given a mixed temporal formula Ψ, the extended extrapolation of Ψ is defined by: $EE : {\mathscr{L}'}_{(N)} \to {\mathscr{L}'}_{(N)}$ such that: $Mod(EE(\Psi)) = \{\tau \mid \tau \in \min(k, Mod(\Psi))\}$*

Example 10 *Two agents a and b share their life between Toulouse and London. I received a postcard from London but I could not read the signature, it was one of them but I do not know who exactly. Hence a or b was in London at time point 1: $a_{(1)} \vee b_{(1)}$ where $a_{(t)}$ (resp. $b_{(t)}$) denotes the fact that a (resp. b) is in London at time point t. I learn that the day after they have exchanged secret documents, hence either they were together*

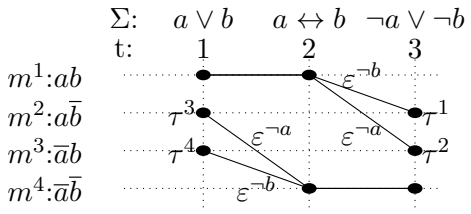


Figure 2.3: *Spy story*

either in London or in Toulouse that day: $a_{(2)} \leftrightarrow b_{(2)}$. *I know that one of them was seen in Toulouse two days after:* $\neg a_{(3)} \vee \neg b_{(3)}$. *I know that they prefer not to travel together in order to avoid suspicion. Hence, I can extrapolate four possibilities (considering only the less surprising ones, see Figure 2.3): either they were both in London, they met there and then one of them left (it makes two possibilities* τ^1 *and* τ^2 *according to the identity of the agent who left), either only one of them was in London, he left London and they met in Toulouse (it also gives two possibilities* τ^3 *and* τ^4*).* $Traj(EE(\Sigma)) = \{\tau^1, \tau^2, \tau^3, \tau^4\}$. *If we encode the event of living London to Toulouse by* $\varepsilon^{\neg a}$ *for the first agent (and* $\varepsilon^{\neg b}$*) for the second agent, we obtain* $EE(\Sigma) \models (\varepsilon^{\neg a}_{(1)} \oplus \varepsilon^{\neg b}_{(1)}) \oplus (\varepsilon^{\neg a}_{(2)} \oplus \varepsilon^{\neg b}_{(2)})$ *where* $\oplus$ *is the exclusive or.*

We have shown that an extended extrapolation operator EE satisfies the extended postulates of belief extrapolation $Ex1, Ex2' \ldots Ex6$ where $Ex2'$ replace the postulate $Ex2$ since $PERS'$ involve event occurrence whereas the initial postulate did not, more precisely,
Ex2' : if $PERS' \wedge \Psi$ is consistent then $E(\Psi) \equiv PERS' \wedge \Psi$
with $PERS' = \bigwedge_{t \in [\![1, N-1]\!]} (\bigwedge_{v \in \mathcal{V}} v_{(t)} \leftrightarrow v_{(t+1)} \wedge \bigwedge_{\varepsilon \in Ev} \neg \varepsilon_{(t)})$.

We can rephrase extended extrapolation in terms of revision since it amounts to revise the formulas describing the natural evolution of the system (static and dynamic laws) by the formula to extrapolate. This extrapolation operator is based on a preference relation on trajectories which minimizes at the same time the occurrences of surprising events and the surprising transitions. Note that in this version of Ex2', the system is supposed to be inert (see Definition 10).

2.b.2 The Question of "What would have occurred if..."

The question tackled in this paragraph is: being given a factual story and a knowledge of the dynamics of the system, what would occur if someone imposes a change in this story? Within our formal framework, the story is a sequence of observations (events or facts), *i.e.*, a mixed scenario or more generally, a mixed temporal formula. The knowledge of the dynamics of the system corresponds to laws of evolution of the world summarized by the cost functions ke and km. Let us examine again the spy-story example:

Example 10 (continued): *Now, there are two interesting questions:*

1. *What can I conclude if I learn that Gatwick airport in London was closed (because of a strike of the controllers) at time point 1, preventing any flight between London and Toulouse between time point 1 and 2:*$\neg\varepsilon^{\neg a}_{(1)} \wedge \neg\varepsilon^{\neg b}_{(1)}$.

2. *What would have happened if the police had set up a security check in Gatwick airport? (and would have been able to arrest any suspect agent, hence forbidding them to fly):* $\neg\varepsilon^{\neg a}_{(1)} \wedge \neg\varepsilon^{\neg b}_{(1)}$

While the two questions are expressed by the same formula they do not imply the same reasoning. The first question implies that I should take into account a new piece of information about this story, I should complete my knowledge, it means that among all possible trajectories, I am able to be more selective and pick the ones that are in agreement with the new piece of information, or if no trajectory is in agreement, I should correct minimally my beliefs in order to obtain a possible trajectory. In our case, I will conclude that they were both in London and they left after their meeting.

The second question implies that I should make an hypothetical reasoning about what the story becomes if something change, hence I should consider every possibility and see how it may evolve. Hence the four initial trajectories (see Figure 2.3)should be reviewed, for τ^1 *and* τ^2 *the hypothesis does not change anything, if the two agents where in London and met there then they were not at the airport when the police check point could have been set up. While for* τ^3 *and* τ^4 *corresponding to the trajectories in which an agent should leave London at time point 1, the consequences of such a supposition could be more dramatic, since the agent could have been arrested in London and then he could not have exchanged information with its colleague and so on...*

In this toy example, we can see that the first reasoning is a belief revision while in the second case the reasoning is an update. A main claim of this study is that the question of "what would have happened if ..." is an *update operation*. Note that the update of this example has been chosen with a formula containing only occurrences of events but we could have chosen to update by a fact, for instance we could have wondered what would have occurred if the first agent had been in Toulouse at time 2: $\neg a_{(2)}$.

To define in practice the update of mixed temporal formulas by an instantaneous formula, we need to define a preference relation between trajectories w.r.t. any initial given trajectory.

Many preorderings can be proposed to compare two trajectories with respect to a same third (we have adapted every change-based preorderings defined for classical extrapolation in order to obtain preorderings w.r.t. a given trajectory). For our specific purpose of hypothetical reasoning, we have provided a new preference relation called *chronological closeness* such that the trajectories that are identical to the initial trajectory w.r.t. events occurrences should be preferred. Concerning facts they also should be identical before the time point of the instantaneous formula. Indeed in the spy story example, we can notice that the facts that were believed to hold after time point 2 are no longer necessarily true. In summary, we chose to minimize chronologically the event distances on the entire trajectory and minimize chronologically the fact distances only until the change time point. This relation implies that any trajectory is closer to itself

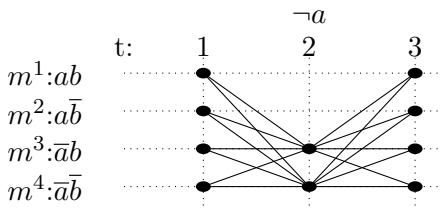


Figure 2.4: *Trajectories satisfying* $\neg a_{(2)}$

than to other trajectories. Note that this is not a strict preference, *i.e.*, the relation suggested is not strictly faithful in the sense of Katsuno and Mendelzon[6] some trajectories may have the same sequence of events and differ only in the sequence of situations after time point p but be as preferred as the reference trajectory. This idea is an extension of the idea of branching time concerning the future and linear time concerning the past [66].

Intuitively, the update of a mixed temporal formula Ψ by an instantaneous mixed formula $\varphi_{(t)}$, denoted $\Psi \diamond_t \varphi_{(t)}$, consists in calculating for each preferred (*i.e.*, less surprising) trajectory τ satisfying Ψ, the possible trajectories satisfying $\varphi_{(t)}$ that are closest to τ. The result is the union for each initial trajectory of the closest trajectories obtained.

Example 10 (continued): *If we want to know what would have occurred if the first agent had been in Toulouse at time point 2 then it is necessary to calculate* $\Sigma \diamond_2 \neg a_{(2)}$. *It requires to compute the possible trajectories in which* $\neg a_{(2)}$ *holds; then select in this set, the trajectories closest to each of the 4 initial trajectories drawn on Figure 2.3.*

In this example, we suppose that the definitions of the surprise degrees associated with the transitions of the system allow to obtain the 32 possible trajectories satisfying $\neg a_{(2)}$*: 16 trajectories passing by* $\neg a_{(2)} b_{(2)}$ *and 16 by* $\neg a_{(2)} \neg b_{(2)}$ *see Figure 2.4.*

Note that τ^3 *and* τ^4 *belong to this set.* τ^3 *is strictly closer to itself than the 31 other trajectories. It is the same for* τ^4. *In* τ^1 *there was only one event occurring at time 2, namely* $\varepsilon^{\neg b}$. *We can found one trajectory having the same curses of events, namely:* $\tau^5 = \langle (m^3, \varnothing), (m^3, \{\varepsilon^{\neg b}\}), (m^4, \varnothing)\rangle$. *This trajectory is the closest to* τ^1 *w.r.t. chronological closeness among all the 32 possible trajectories. Concerning* τ^2*, there is no trajectory identical for event occurrences, but two trajectories have only one difference with it: namely the two static trajectories* $\tau^6 = \langle (m^3, \varnothing), (m^3, \varnothing), (m^3, \varnothing)\rangle$ *and* $\tau^7 = \langle (m^4, \varnothing), (m^4, \varnothing), (m^4, \varnothing)\rangle$. *But* τ^6 *is closer since its situation at time point 1 less differs from* τ^2 *than* τ^7. *Finally, we obtain exactly 4 trajectories satisfying* $\Sigma \diamond_{(2)} \neg a_{(2)}$*:* τ^3, τ^4,τ^5 *and* τ^6. *In each trajectory the first agent stays in Toulouse at time 3. In this hypothetical reasoning, one cannot conclude on the fact that the two agents could meet or not.*

We have shown that the syntactic operator corresponding to the *chronological closeness preference* relation is an update operator in the sense of the definition we gave in [81]. More precisely, in the proposal that we made in [81], the postulate U1 is not nec-

[6] A *faithful assignment* is a function that associates with each $\omega \in \Omega$ a complete preorder $\preceq_\omega$ such that $\forall \omega_1 \in \Omega$, $\omega \prec_\omega \omega_1$ where $\prec$ is defined classically from $\preceq$ as follows: $x \prec y$ iff ($x \preceq y$ and $y \not\preceq x$)

essary since it can be deduce from U3bis, U5 and U10. The set U3bis, U4, U5, U8, U9 and U10 is minimal and complete with respect to the representation theorem[7] where
U1 $(K \diamond \varphi)$ implies φ.
U3bis $K \diamond \varphi$ consistent implies $K \diamond (\varphi \vee \psi)$ consistent.
U4 If $K1 \leftrightarrow K2$ and $\varphi^1 \leftrightarrow \varphi^2$ then $(K1 \diamond \varphi^1) \leftrightarrow (K2 \diamond \varphi^2)$.
U5 $(K \diamond \varphi) \wedge \psi$ implies $(K \diamond (\varphi \wedge \psi))$.
U8 $(K1 \vee K2) \diamond \varphi \leftrightarrow (K1 \diamond \varphi) \vee (K2 \diamond \varphi)$.
U9 If K is deductively closed and $(K \diamond \varphi^1) \wedge \varphi^2$ is consistent then $(K \diamond (\varphi^1 \wedge \varphi^2))$ implies $((K \diamond \varphi^1) \wedge \varphi^2))$.
U10 $\exists \varphi \in \mathscr{L}$ s.t. $K \diamond \varphi$ is inconsistent.
Note that we have chosen to propose a more general update operator than in Katsuno and Mendelzon framework since we wanted to enable impossible updates and also we wanted to allow not strict faithfulness. This is why in [81] we added U10 and removed the postulates U2 and U3 since U3 imposes that updates are always possible[8] and U2 which has been often discussed in the literature is well known to impose inertia[9]. We will come back on a more precise explanation of the postulates in Section 2.c. Strict faithfulness, in this context, is not always desirable (see Example 11) except in the particular case of a deterministic system. Let us note however that it is completely possible to particularize the preference relation suggested in order to make it faithful.

Example 11 *Let us consider a scenario Σ in which a die has been rolled on a green carpet at time point 1 and its face is a six at time point 2. Σ does imply that the "carpet is green" but if we update Σ by the fact (that we already know) that "the carpet is green" denoted by cg, we do not necessarily obtain the same result: $\Sigma \diamond_2 cg_{(2)}$ is not equivalent to Σ, since, if the encoding of ke and km represent the transition of a fair die (hence non deterministic), it gives 6 possible trajectories.*

2.b.3 Causality and Scenario Update

The numerous works in the field of causality lead to several definitions of causality, we describe only two notions that we have used in our model, namely counter-factuality and manipulability. Indeed, one of the first definitions was given by Lewis [145], which introduces "counter-factuality" in 1973. This definition is based on the existence of possible worlds and on a similarity distance between worlds. A causes B in a counter-factual way when one can affirm that if A had not occurred in a world as close as possible of the current world then B would not have occurred. Another type of definition uses manipulability theory, in particular Von Wright [192] formalizes the idea that causality is related

[7] The representation theorem for update of Katsuno and Mendelzon is: There is an operator $\diamond : \mathscr{L} \times \mathscr{L} \to \mathscr{L}$ satisfying U1, U2, U3, U4, U5, U8, U9 *if and only if* there is a *faithful* assignment that associates with each $\omega \in \Omega$ a complete preorder denoted $\preceq_\omega$ such that $\forall \varphi, \alpha \in \mathscr{L}$, $[\varphi \diamond \alpha] = \bigcup_{\omega \in [\varphi]} \{\omega' \in [\alpha]$ such that $\forall \omega'' \in [\alpha], \omega' \preceq_\omega \omega''\}$. (If postulates U6 and U7 are considered instead of U9 then the theorem relates the update operator to a family of partial preorders).

[8] **U3** If K and φ are consistent then $(K \diamond \varphi)$ is consistent

[9] **U2** If K implies φ then $(K \diamond \varphi)$ is equivalent to K.

to the concept of intervention. A causes B if while forcing A to be true, one forces B to be true and by doing nothing to change the value of A then it does not change the value of B. In their structural model approach, Halpern and Pearl [123] have also proposed a counter-factual definition of causality (encoded by causal graphs with endogenous and exogenous variables).

We were interested in searching for a concrete cause of a particular fact or event. In our model, this cause could either be an event that has occurred or was coming from some facts that hold in the initial situation. The counter-factual nature of the cause was used in the following way: A causes B if A and B are true in the initial mixed temporal formula, and if B is false after the update of this initial formula by $\neg A$. Formally:

Definition 15 (cause)
Given a mixed temporal formula $\Psi \in \mathscr{L}'_{(N)}$,

$$A_{(t)} \text{ causes } B_{(N)} \text{ iff } \begin{cases} t = 1 \text{ or } A \in \mathscr{L}_{Ev} \\ EE(\Psi) \models A_{(t)} \wedge B_{(N)} \\ \Psi \diamond_t \neg A_{(t)} \models \neg B_{(N)} \end{cases}$$

where $\mathscr{L}_{Ev}$ *is the set of formulas built on* Ev.

Several authors also use chronicles to study causality, for instance, one can refer to [36] or [138]. These approaches use an interventionist modeling of causality. Indeed, in [138], the authors define the concept of voluntary cause which implies a deliberated choice of the agent among its possible actions. In the approach of Belnap *et al.*, the authors are interested in the representation of the fact that the agent "could have act differently". Only actions of agents are regarded as "true" causes. This definition is not ours because we were not interested in the problem of perceived causality but in the event causation problem (looking for the particular causes of a fact in a given scenario). We are however in agreement with the fact that causes should correspond to actions or events (to reflect that causality is related to manipulability as preached by Von Wright). The works of [138] and of [36] use modal logic for the definition of the possible evolutions of the worlds. In our work, we use similar concepts since we also define a preference relation on trajectories, but we use a less complex formalism based simply on propositional logic.

In this approach, we made the assumption that the reported observations were *reliable*, it would be interesting to consider the possible existence of bad perceptions of the world, it would require to utilize both revision and update operators, these two operators being applied to temporal formulas.

The distance between trajectories which we use in this work is rather simple, an interesting prospect would be to use the DNA sequences alignment techniques used in bio-data processing in order to calculate events sequence alignments. Thus, as in the algorithm of [155], we could associate a cost with the addition, the withdrawal, the substitution or the shift of an event in a sequence. Then we could define the distance between a trajectory and another by the cost of the best alignment between these two trajectories.

2.c Belief update with transition constraints

[50] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Enforcement in Argumentation is a kind of Update. In *International Conference on Scalable Uncertainty Management (SUM)*, number 8078 in LNAI, pages 30–43. Springer-Verlag, 2013
[94] F. Dupin de Saint-Cyr, P. Bisquert, C. Cayrol, and M.-C. Lagasquie-Schiex. Argumentation Update in YALLA (Yet Another Logic Language for Argumentation). *under submission to IJAR*, 2015

In the following, we present a recent work about update [50] in which our aim was to restrain the possible changes that can be done on a state of the world. It amounts to have a different definition of update than Katsuno and Mendelzon [136] in which we can take into account some constraints about the possible transitions. I had already developped this kind of idea in [81] by introducing the set of postulates described in the previous section, in which some transitions are not possible. This new definition of postulates is more refined because it allows us to precise what are the possible transitions. This result was required in order to be able to take into account the notion of authorized operation in the argumentation domain where some operations are not allowed according to the user knowledge or to the target system (see Section 4.c).

In [50], we have defined an update operator based on a set of authorized transitions as follows:

Definition 16 (Update operator related to a set of authorized transitions)

- *Let* $\mathcal{T} \subseteq \Omega \times \Omega$ *be the set of all the* authorized transitions *between states of the world.*

- $\forall \varphi, \psi \in \mathscr{L}$*, a transition from* φ *to* ψ *satisfies* $\mathcal{T}$*, denoted* $(\varphi, \psi) \models \mathcal{T}$*, iff* $([\varphi] \neq \varnothing$ *and* $\forall \omega \in [\varphi]$, $\exists \omega' \in [\psi]$, $(\omega, \omega') \in \mathcal{T})$.

- *An* update operator $\diamondsuit$ *is a mapping relative to a set of authorized transitions* $\mathcal{T} \subseteq \Omega \times \Omega$ *from* $\mathscr{L} \times \mathscr{L} \rightarrow \mathscr{L}$ *which associates with*

  - *any formula* φ *giving information about an initial state of the world,*
  - *and any formula* α*,*

  a formula, denoted $\varphi \diamondsuit \alpha$*, characterizing the states of the world in which* α *holds, that can be obtained from states satisfying* φ *by a change belonging to* $\mathcal{T}$*.*

We have been able to define a set of rational postulates for $\diamondsuit$. These postulates are constraints that aim at translating the idea of update under authorized transitions. Some postulates coming from standard update are suitable, namely U1, since it ensures that after an update the constraints imposed by α are true. U2 postulate is optional, it imposes that if α already holds in a state of the world then updating α means no change. This postulate imposes inertia as a preferred change, this may not be desirable in all situations. U3 imposes that if a formula holds for some states of the world and if the update piece of information also holds for some state then the result of update

should give a non empty set of states. Here, we did not want to impose that any update is always possible since some state of the world may be unreachable from others. So we had proposed to replace U3 by a postulate called $E3$ based on the set of authorized transitions $\mathcal{T}$: $\forall \varphi, \psi, \alpha, \beta \in \mathscr{L}$

E3: $[\varphi \Diamond \alpha] \neq \varnothing$ if and only if $(\varphi, \alpha) \models \mathcal{T}$.

Due to the definition of $(\varphi, \alpha) \models \mathcal{T}$, E3 handles two cases of update impossibility: no possible transition and no world (*i.e.,* no state of the world where φ holds or no state where α holds). U4 is suitable in our setting since update operators are defined semantically. U5 is also suitable for update since it says that states of the world updated by α in which β already holds are states in which the constraints α and β are updated. Due to the fact that we want to allow for update failure, this postulate had been restricted to "complete" formulas[10]

E5: if $card([\varphi]) = 1$ then $(\varphi \Diamond \alpha) \wedge \beta \models \varphi \Diamond (\alpha \wedge \beta)$.

U8 captures the decomposability of update with respect to a set of possible input states of the world. We slightly change this postulate in order to take into account the possibility of failure, namely if updating something is impossible then updating it on a larger set of states is also impossible, else the update can be decomposable:

E8: if ($[\varphi] \neq \varnothing$ and $[\varphi \Diamond \alpha] = \varnothing$) or ($[\psi] \neq \varnothing$ and $[\psi \Diamond \alpha] = \varnothing$)
then $[(\varphi \vee \psi) \Diamond \alpha] = \varnothing$
else $[(\varphi \vee \psi) \Diamond \alpha] = [(\varphi \Diamond \alpha) \vee (\psi \Diamond \alpha)]$.

Postulate U9 is a kind of converse of U5 but restricted to a "complete" formula φ *i.e.,* such that, $card([\varphi]) = 1$, this restriction is required in the proof of Katsuno and Mendelzon's Theorem (recalled in footnote 7) as well as in Theorem 2.

As already noticed for Katsuno and Mendelzon's postulate, the presence of U1 is not necessary, it is still the case in the new setting since U1 can be derived from E3, E5 and E8. We have shown that E3, U4, E5, E8 and U9 constitute a minimal set. We have obtained the following representation theorem, saying that an update operator satisfying these postulates can be defined by means of the definition of a family of preorders on states of the world.

Definition 17 *Given a set $\mathcal{T} \subseteq \Omega \times \Omega$ of authorized transitions, an* assignment respecting $\mathcal{T}$ *is a function that associates with each $\omega \in \Omega$ a complete preorder $\preceq_\omega$ such that $\forall \omega_1, \omega_2 \in \Omega$, if $(\omega, \omega_1) \in \mathcal{T}$ and $(\omega, \omega_2) \notin \mathcal{T}$ then $\omega_1 \prec_\omega \omega_2$.*

Theorem 2 *Given a set $\mathcal{T} \subseteq \Omega \times \Omega$ of authorized transitions, there is an operator $\diamond : \mathscr{L} \times \mathscr{L} \to \mathscr{L}$ satisfying E3, U4, E5, E8, U9 if and only if there is an assignment respecting $\mathcal{T}$ such that $\forall \omega \in \Omega$, $\forall \varphi, \alpha \in \mathscr{L}$,*

[10] Note that $card([\varphi]) = 1$ if and only if $\exists \omega \in \Omega$ such that $[\varphi] = [\omega]$.

(1) $[\varphi \diamond \alpha] = \varnothing$ *if* $\exists \omega \in [\varphi]$ *such that* $[\Phi(\omega) \diamond \alpha] = \varnothing$

(2) $[\varphi \diamond \alpha] = \bigcup_{\omega \in [\varphi]} [\Phi(\omega) \diamond \alpha]$ *otherwise*

$$\textit{(3)}\ [\Phi(\omega) \diamond \alpha] = \left\{ \omega_1 \in \Omega \;\middle|\; \begin{array}{l} \omega_1 \in [\alpha] \textit{ and} \\ (\omega, \omega_1) \in \mathcal{T} \textit{ and} \\ (\forall \omega_2 \in [\alpha] \textit{ such that } (\omega, \omega_2) \in \mathcal{T}, \omega_1 \preceq_\omega \omega_2) \end{array} \right\}$$

In other words (1) and (2) allow us to define the update of a formula φ wrt the update of its individual models, with (1) stating that if one of them can not be updated the whole update fails, and (2) stating that otherwise the whole update corresponds to the union of the individual updates defined by (3).

This result is a significant headway, but as usual for a representation theorem, it gives only a link between the existence of an assignment of preorders and the fact that an update operator satisfies the postulates. It does not give any clue about how to assign these preorders *i.e.*,, how to design precisely an update operator. However, let us illustrate this setting in an example where the assignment and the authorized transitions are given.

Example 12 *Let us consider three variables xx, xy and t meaning respectively "Mrs. X is alive", "Mr. X is alive", "Mr. and Mrs. X are together in the same room". Suppose that we know that at a given time point Mr. X is alive (xy), we do not know whether Mrs. X is alive and if they are together. However we know that Mrs. X cannot be alive if they are together. It means that among the eight worlds: $w_1 = (xx, xy, t)$, $w_2 = (xx, xy, \overline{t})$, $w_3 = (xx, \overline{xy}, t)$, $w_4 = (xx, \overline{xy}, \overline{t})$, $w_5 = (\overline{xx}, xy, t)$, $w_6 = (\overline{xx}, xy, \overline{t})$, $w_7 = (\overline{xx}, \overline{xy}, t)$, $w_8 = (\overline{xx}, \overline{xy}, \overline{t})$ there are three possible worlds representing the situation: w_2, w_5 and w_6. We know that some transitions are not possible from this time point to the next time point: it is impossible that Mrs. X (respectively Mr. X) rises from the dead, i.e., every transition from $(\overline{xx}, .., ..)$ to $(xx, .., ..)$ (respectively $(.., \overline{xy}, ..)$ to $(.., xy, ..)$) does not belong to $\mathcal{T}$.*

A gunshot has been heard and "Mrs. X was found dead". It means that the world has evolved in such a way that xx is false. Let us consider a particular assignment satisfying $\mathcal{T}$ with the following preference relations on transitions:

- $\forall i \neq 6, w_6 \prec_{w_5} w_i$*: if Mrs. X is dead and Mr. X is alive in the same room, then it is likely that at the next instant Mr. X has left since he does not like to stay with dead people,*
- $\forall i \neq 2, w_2 \prec_{w_2} w_i$*: if Mr. and Mrs. X are alive and not together then it is more plausible that at the next instant it is still the case, otherwise it is more plausible that they met and stay alive than that one of them dies, which in turn is more plausible than both of them died separately and so on*[11]*:* $w_1 \prec_{w_2} \{w_4, w_6\} \prec_{w_2}$

[11] w_5 is also considered as less plausible since it would require two steps for passing from w_2 to w_5 namely killing Mrs. X and gathering Mr. X and Mrs. X, moreover we know that Mr. X does not like to be in such an equivocal situation.

$\{w_3, w_5, w_7, w_8\}$

- $\forall i \neq 6, w_6 \prec_{w_6} w_i$: *if Mrs. X is dead and Mr. X is alive but elsewhere, then it is more plausible that it is still the case at the next instant.*

In that case $[xy \wedge (t \rightarrow \neg xx)\Diamond_{\mathcal{T}}\neg xx] = \{w_6\}$ *since for every* $w \in \{w_2, w_5, w_6\}$, $w_6 \prec_w w'$ $\forall w'$ *s.t.* $w' \models \neg xx$ *and* $(w, w') \in \mathcal{T}$.[12]

From Theorem 2 we have deduce two simple cases of impossibility: if the initial situation or the goal are impossible then update is impossible (this result is a kind of converse of U3). Note that there are some cases where U2 does not hold together with E3, U4, E5, E8 and U9. If U2 is imposed then the update operator is associated with a preorder in which a given state is always closer to itself than to any other state of the world. This is why it imposes to have a faithful assignment (see Footnote 6). In that case, the relation represented by $\mathcal{T}$ should be reflexive. We have also shown that if we remove the constraint about authorized transitions (by setting $\mathcal{T} = \Omega \times \Omega$) then we recover Katsuno and Mendelzon theorem.

2.d Related works

Probably the most related approach to belief extrapolation is the generic class of belief change operators proposed by Berger et al. [43]. This class of belief change operators is actually a subclass of the set of extrapolation operators (although the authors use the terminology "iterated updates" – which we think is not adequate, as discussed before, and is probably the first approach, chronologically speaking, on belief extrapolation. Shapiro and Pagnucco [177] introduced a framework which is a situation calculus version of a specific extrapolation operator (minimizing the number of exogenous events) in which ontic actions (and thus updates) are possible. Booth and Nittka [60] take a subjective view of belief revision that could be seen as a subjective version of extrapolation: given an observer and an observed agent, the observer tries to make inferences about what the agent believed (or will believe) at a given moment, based on an observation of how the agent has responded to some sequence of previous belief revision inputs over time. Assuming a framework for iterated belief revision which is based on sequences, they construct a model of the agent that "best explains" the observation. The comparison between [60] and extrapolation suggests a promising issue for further research: "subjective extrapolation" would consist in starting with a scenario describing what we know of the agent's beliefs at different time points (possibly using formulas from doxastic logic), and then find the most plausible events that occurred and that explain the changes in the agent's beliefs.

[12] Note that the reasoning process would have been different if the coroner had discovered that Mrs. X was already dead before the gunshot. This process is a *revision* and would have amount to complete the initial knowledge hence to deduce that at the initial time, only two worlds were possible w_5 and w_6, denoted $[xy \wedge (t \rightarrow \neg xx) \star \neg xx] = \{w_5, w_6\}$.

The integration of unexpected change (via belief extrapolation) and ontic actions (via belief update) has been investigated in a few proposal. Generalized update of Boutilier [61] consists in finding out which events (from a given set of events E) most likely occurred between two time points t_1 and t_2, and use the knowledge about the dynamics of these events to reason about what was true at t_1 and what is true at t_2. We have shown that for a fixed formula α, there exists a non-inertial extrapolation operator simulating an update by α. The framework developed by Hunter and Delgrande in [127] allows not only for unexpected changes, but also fallible observations, and actions (without exceptional effects). Their approach is somehow similar to extrapolation, in the sense that it is based on the selection of preferred trajectories. However, they commit to a specific choice of a preference relation, that makes use of integer-valued ranking functions that can be seen as associating "surprise degrees" both to unexpected events and incorrect observations. This restriction on preference relations is the main reason for their Proposition 7, that states that some extrapolation operators are not representable in their framework. The approach of Liberatore and Schaerf [148] is also a fairly general system (BReLS) aiming at integrating revision, update and merging. It deals with time-stamped observations and consider two semantics: using the "trajectory" semantics and assuming that there is no more than one observation at each time point, we obtain our extrapolation operator induced by the penalty-induced relation $\preceq_k$; the other semantics ("pointwise") yields iterated update (but is incompatible with extrapolation because of U8 which underlies this semantics). Lastly, Friedman and Halpern [115] have defined a very general framework for belief change, of which revision and update are two specific instances. Now, for all classes of preference relations studied in this article, scenario extrapolation is also an instance of a belief change system satisfying the assumption called *prior* by Friedman and Halpern.

Belief extrapolation addresses a problem similar to dynamic diagnosis, where the goal is to identify a complete evolution of the world which best fits a given (maybe incomplete) history that contains some abnormal behaviors. Dynamic diagnosis approaches (see *e.g.* [62] for an early survey of definitions) can be handled differently according to the logical language used for modeling the system, the actions and the observations (*e.g.*, action languages in [185] and [32], logic programs in [26]), the nature of the available actions, the type of diagnosis (abductive or consistency-based), and the selection principle for finding plausible diagnoses. Consistency-based dynamic diagnosis consists in finding a series of failures across time which is *consistent* with the observations and the system description (while an abductive diagnosis together with the system description should allow to *deduce* the observations). Therefore, extrapolation can be seen both as a simplification (because extrapolation assumes inertia and that no action is available to the agent) and a generalization (because diagnosis makes use of a specific selection criterion compare to the rich set of selection operators available for extrapolation) of consistency-based dynamic diagnosis. For instance, the criterion in Thielscher [185] is chronological and minimizes abnormalities in the initial state while taking account of the prior likelihood of failure of the components; Baral et al. [32] minimize the set of faulty components and exogenous actions and propose a diagnosis-plan that aims at selecting dynamically a diagnosis by

giving the sequence of tests to be done in order to discriminate between the unobserved possible faulty components; this kind of checking plan is also used by Balduccini and Gelfond [26]. Thus, extrapolation offers a much more systematic and principled study of such operators (through an axiomatisation and a representation result), our results contributes to bridging dynamic diagnosis and belief change.

We chose to model beliefs across time using explicit time points (thus making use of a propositional logic of reified time). Since our results, and the preference relations we have focused on, do not exploit the metric nature of time, they would still hold in a similar way if we had chosen to use a purely symbolic representation of time such as in modal logics for belief change (*e.g.*, [176, 189]) or epistemic or doxastic extensions of the situation calculus (*e.g.*, [78, 172]). Finally, using a modal language would open the door to new opportunities; in particular, having a specific belief modality for each time point (or situation) allows to express mutual inter-temporal beliefs such as "at time 3 I believed that φ held at time 2 while I now believe that φ did not hold at time 2".

Extrapolation is adequate for reasoning about time-stamped observations on a changing world while belief update is not. The key point is postulate U8 which, by requiring that all models of the initial belief set be updated separately, forbids us inferring new beliefs about the past from later observations. In belief update, the input α should rather be interpreted as the projection of the expected effects of some "explicit change" , or more precisely, the *expected* (not the *observed*) effect of the action (or event) "make α true", see [140] for further discussion. We see that the crucial issues are *observability* (what do we observe about the world at what time?) and *predictability of change*. Belief extrapolation deals with observation and unexpected change, while belief update is suitable for expected change without observations. In Sandewall's taxonomy [171], extrapolation is adequate for the action-free subclass of K-**IS** (correct knowledge, inertia and surprises) while update is adequate for the class K_{p}-**IA** (no observations after the initial time, inertia and alternative results of actions). Event-based extrapolation is rather in K-**AS**.

In the last section we have presented a part of our last study published in [50] where we came back again to the idea of finding more convenient update postulates than those of Katsuno and Mendelzon [136]. Indeed, this set of postulates has already been broadly discussed and criticized in the literature. Most critics were concerning the point that Postulate U2 imposes inertia (which is not always suitable), Herzig [125] proposes to restrict possible updates by taking into account integrity constraints, *i.e.,*, formulas that should hold before and after update, our transition constraints generalizes this proposal. On the same point, in [81], we had already worked on a proposal to not impose inertia and to allow for update failure even if the formulas are consistent (see Section 2.b). One benefit of the new model is that it does not require to invent an artificial inaccessible state (as in our old postulate U10). This idea to restrict transitions in update was first introduced by Cordier and Siegel [68]. Their proposal goes beyond our idea since they allow for a greater expressive power by using prioritized transition constraints. However, they do not provide postulates nor representation theorem associated with their update operator.

Our main contributions in this chapter are first a thorough investigation of the extrapolation process, its use for causal reasoning and finally the investigation on a more expressive update operator. These contributions are both foundational and computational. On the foundational side, we have defined a very general and structured class of extrapolation operators. We have established a precise connection between extrapolation and belief revision, showing that extrapolation can be seen as a particular case of revision over a time-stamped language. This led us to give an axiomatic framework for extrapolation including representation theorems inherited from representation theorems for belief revision and sufficient (and sometimes necessary) conditions for some specific temporal properties (reversibility, Markovianity, independence of empty observations) to be satisfied. We also gave an impossibility result showing that an extrapolation operator cannot be an update operator. Lastly we have provided new axioms for update operators that respect transition constraints. On the computational side, we have identified the computational complexity of belief extrapolation and provided a method for computing extrapolated beliefs, for a specific (yet representative enough) operator. In the next chapters we will see how these notions are of interest in the argumentation framework.

Part II

Belief Change and Argumentation

Argumentation is a reasoning model based on the construction and the comparison of arguments. Arguments are reasons for believing in statements, or for performing actions. Hence it can be used for reasoning or in a decision context, it may involve only one or several agents. Argumentation has three expressive powers: it offers the opportunity to express pros and cons (through the conflict relation between arguments), to justify some claims (an argument can be viewed as a justification of a claim), to explain some logical conclusion or some decisions (by being able to produce the set of arguments that support those decisions/conclusions). In the representation field of artificial intelligence we are often more interested in designing reasoning principles that can be explained in a high-level language rather than black-box methods that give conclusions without being able to explain the reasons that leads to them (for instance methods based on neural-network or statistical learning approaches...). Since the principle of explanation is its main building block, argumentation theory falls perfectly within the scope of Artificial Intelligence. It is studied for modeling agent's internal reasoning, namely for handling inconsistent, incomplete or uncertain information (*e.g.*, [45, 117]) and for making decisions (*e.g.*, [21, 114]).

There is also an extensive literature devoted to modeling agent's interactions, namely dialogs in which agents may exchange arguments with each other, like persuasion (*e.g.*, [18, 164]) and negotiation (*e.g.*, [134, 159]). Indeed, the most natural application of argumentation is for persuasion purposes. The ability to persuade other people is commonly viewed as an intelligent feature hence it is another reason to justify its study in AI. For instance, many persuasion debates have marked human history: Herodotus debate on the three government types, Valiadolid debate, the Bohr-Einstein debate about quantum mechanics, presidential TV debates... The "winner" is often considered as very clever and skilled. Incidentally, oratory featured in the original Olympics and there exist teaching lessons for being a good orator (*e.g.* the book of Schopenhauer [173]). A good orator is someone who is able to make his point of view adopted by the public whatever the truth is and whatever his adversary may say. This skill and cleverness is a big challenge for human being in everyday life as well as in History since debates are both very common and very influential. This is why it is important that artificial intelligence focuses on this field of research. This implies to develop at least two features: representing and handling persuasion dialogs, designing good artificial orators (able to find strategies to win a debate) in order to be able to analyze persuasion process for controlling them and for increasing the awareness of the citizens.

Chapter 3

Arguments

In this chapter, we will describe two representations of arguments, the first one has been introduced by Dung [90], it is an abstract framework in which arguments are abstract entities about which only one thing is known (and represented): namely their binary conflict relations. Hence, the origin of arguments is supposed to be unknown. The second representation of argument uses a structured pair (support, conclusion). We present the classical representation of so called logic-based arguments where the support is a set of formulas that entails the conclusion. However this kind of ideal argument is not easy to build in practice and we propose to use a relaxed version of it, where some information can be missing either in the support part or in the conclusion. These kind of arguments are called enthymemes.

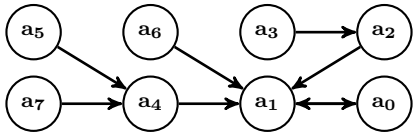
3.a Abstract Arguments

[47] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in argumentation systems: exploring the interest of removing an argument. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 275–288. Springer-Verlag, octobre 2011
[94] F. Dupin de Saint-Cyr, P. Bisquert, C. Cayrol, and M.-C. Lagasquie-Schiex. Argumentation Update in YALLA (Yet Another Logic Language for Argumentation). *under submission to IJAR*, 2015

In this section, we describe a model of abstract argumentation. Due to the modeling of change (see Chapter 4), we have realized that it is important to be able to consider several argumentation systems, hence we had extended Dung's definitions in order to be able to deal with an argumentation universe in which several argumentation systems can be defined. It is an extension of the classical approach proposed by Dung [90]: we consider a set $\mathcal{A}_U$ of symbols (denoted by lower case letters) representing a set of arguments and an attack relation $\mathcal{R}_U$ on $\mathcal{A}_U \times \mathcal{A}_U$. The pair $(\mathcal{A}_U, \mathcal{R}_U)$, called *universe*, allows us to represent the set of possible arguments together with their interactions. $\mathcal{A}_U$ may be infinite, maybe a set of logic-based arguments built from a knowledge base, or may be explicitly provided as in the following example created by Pierre Bisquert in [47]:

Example 13 *During a trial concerning a defendant (Mr. X), several arguments can be involved to determine his guilt. The set of arguments $\mathcal{A}_U$ and the relation $\mathcal{R}_U$ are given below.*

a_0	*Mr. X is not guilty of premeditated murder of Mrs. X, his wife.*
a_1	*Mr. X is guilty of premeditated murder of Mrs. X.*
a_2	*The defendant has an alibi, his business associate has solemnly sworn that he met him at the time of the murder.*
a_3	*The close working business relationships between Mr. X and his associate induce suspicions about his testimony.*
a_4	*Mr. X loves his wife so deeply that he asked her to marry him twice. A man who loves his wife cannot be her killer.*
a_5	*Mr. X has a reputation for being promiscuous.*
a_6	*The defendant had no interest to kill his wife, since he was not the beneficiary of the huge life insurance she contracted.*
a_7	*The defendant is a man known to be venal and his "love" for a very rich woman could be only lure of profit.*



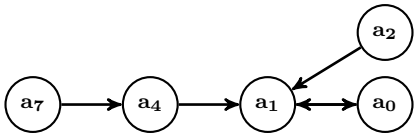
Given a universe $(\mathcal{A}_U, \mathcal{R}_U)$, an argumentation system is defined as follows.

Definition 18 *An* argumentation graph $\mathcal{G}$ *on* $(\mathcal{A}_U, \mathcal{R}_U)$ *is a pair* $(\mathcal{A}, \mathcal{R})$ *where*

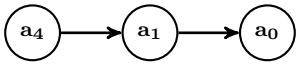
- $\mathcal{A} \subseteq \mathcal{A}_U$ *is the* finite *set of vertices of* $\mathcal{G}$ *called "arguments" and*
- $\mathcal{R} \subseteq \mathcal{R}_U \cap (\mathcal{A} \times \mathcal{A})$ *is its set of edges, called "attacks".*

The set of argumentation graphs that may be built on the universe $(\mathcal{A}_U, \mathcal{R}_U)$ *is denoted by* $\mathcal{G}_U$*.*

Example 14 *The prosecutor is trying to make accepted the guilt of Mr. X and her knowledge is summarized in her argumentation system ($\mathcal{G}_{pros}$) inside the universe described in Example13:*



Moreover, the prosecutor knows the content of the jury's argumentation system ($\mathcal{G}_{jury}$). Indeed, she had given the argument a_1 against the argument a_0 and knows that in the jury's mind Mr. X is more plausibly guilty than innocent inducing a preference on the attack from a_1 to a_0 over the attack from a_0 to a_1 (which is thus neglected), the lawyer had answered by uttering a_4 attacking this suspicion of guiltiness:



The acceptable sets of arguments ("extensions") inside an argumentation system $(\mathcal{A}, \mathcal{R})$ are computed using "semantics" that are based on the following notions defined for any argument $\alpha \in \mathcal{A}$ and any set of arguments $S \subseteq \mathcal{A}$:

- S *attacks* α if and only if $\exists \beta \in S$ such that $\beta \mathcal{R} \alpha$.
- S is *conflict free* if and only if $\nexists \alpha, \beta \in S$ such that $\alpha \mathcal{R} \beta$.
- S *defends* an argument α if and only if S attacks any argument attacking α. The set of the arguments defended by S is denoted by $\mathcal{F}(S)$. More generally, S *indirectly defends* α if and only if $\alpha \in \bigcup_{i \geq 1} \mathcal{F}^i(S)$.
- S is an *admissible set* if and only if it is both conflict free and defends all its elements.

For abstract argumentation, we only consider the semantics proposed by Dung [90]):

Definition 19 (Dung's semantics) *Given $(\mathcal{A}, \mathcal{R})$ an argumentation graph with $\mathcal{E} \subseteq \mathcal{A}$*

- *$\mathcal{E}$ is a* complete extension *of $(\mathcal{A}, \mathcal{R})$ if and only if $\mathcal{E}$ is an admissible set and every argument which is defended by $\mathcal{E}$ belongs to $\mathcal{E}$.*
- *$\mathcal{E}$ is a* preferred extension *of $(\mathcal{A}, \mathcal{R})$ if and only if $\mathcal{E}$ is a maximal (with respect to set-inclusion $\subseteq$) admissible set.*
- *$\mathcal{E}$ is* the only grounded extension *of $(\mathcal{A}, \mathcal{R})$ if and only if $\mathcal{E}$ is a minimal (with respect to $\subseteq$) complete extension.*
- *$\mathcal{E}$ is a* stable extension *of $(\mathcal{A}, \mathcal{R})$ if and only if $\mathcal{E}$ is conflict-free and attacks any argument not belonging to $\mathcal{E}$.*

The status of an argument is determined by its presence in the extensions of the selected semantics. For example, an argument can be "skeptically accepted" (resp. "credulously") if it belongs to all the extensions (resp. at least to one extension) and be "rejected" if it does not belong to any extension.

In the PhD thesis of Pierre Bisquert and furthermore in [94], our purpose was to build a first-order logical theory able to describe properties of sets of vertices in a graph, hence able to manipulate extensions. This logical formalism called YALLA$_U$ was introduced in order to make a bridge between argumentation theory and belief change theory, this precise link will be explained in Section 4.c. The signature Σ_U includes as many symbols as elements in $2^{\mathcal{A}_U}$.

Definition 20 (Signature) *$\Sigma_U = (V_{const}, V_f, V_P)$ where $V_{const} = \{c_\perp, c_1, \ldots, c_p\}$ with $p = 2^{|\mathcal{A}_U|} - 1$, $V_f = \{union^2\}$ and $V_P = \{on^1, \rhd^2, \subseteq^2\}$.*

The semantics of YALLA$_U$ is defined thanks to a structure over Σ_U. Such a structure is associated with an argumentation system $(\mathcal{A}, \mathcal{R})$ built on the universe $\mathcal{A}_U$, viewed as an attack relation between sets of arguments. We have $\mathcal{A} \subseteq \mathcal{A}_U$ and $\mathcal{R} \subseteq \mathcal{R}_U \cap (\mathcal{A} \times \mathcal{A})$.

Definition 21 (Structure) *A* structure $\mathcal{M}$ *of signature* Σ_U*, associated with* $(\mathcal{A}, \mathcal{R})$*, is a pair* $(\mathcal{D}, \mathcal{I})$ *with* $\mathcal{D} = 2^{\mathcal{A}_U}$ *the* domain *and* $\mathcal{I}$ *an* interpretation function *associating:*

- *a unique element of* $\mathcal{D}$ *to each constant symbol* c_i *(in particular* $\varnothing$ *is associated with* $c_\perp$*),*
- *the* binary set union *operator (function from* $\mathcal{D}^2$ *to* $\mathcal{D}$*) to the function symbol union,*
- *the characterization of the* subsets *of* $\mathcal{A}$ *to the symbol on: on*(S) *if and only if* $S \subseteq \mathcal{A}$
- *the* binary set inclusion *relation (binary relation on* $\mathcal{D}^2$*) to the predicate symbol* $\subseteq$*,*
- *the* attack relation *between sets of arguments induced by* $\mathcal{R}$*, and defined by* $S_1 \mathcal{R} S_2$ *if and only if* $S_1 \subseteq \mathcal{A}, S_2 \subseteq \mathcal{A}$ *and* $\exists x_1 \in S_1, \exists x_2 \in S_2, (x_1 \mathcal{R} x_2)$*, to the predicate symbol* $\rhd$*.*

We have proposed specific axioms called AX_U for the predicates $\subseteq$, *union*, $\rhd$ and *on* in order to conform their behavior to set inclusion, union and attacks between sets, and to translate the meaning that a set of arguments belongs to an argumentation system respectively. The axiomatisation is sound and complete (since for any formula φ of YALLA$_U$: $AX_U \models \varphi$ if and only if $AX_U \vdash_{Sys} \varphi$, where $\vdash_{Sys}$ is the inference consequence based on any sound and complete Axiomatic System Sys for predicate calculus).

We have presented several examples illustrating the expressive power of YALLA$_U$. In particular, we have shown that we can precisely describe an argumentation system by its characteristic formula[1], this formula has the given argumentation system for unique model.

Definition 22 (Characteristic formula of an Argumentation System) *The function* Φ_U *associated with YALLA*$_U$ *is defined by:* $\Phi_U : \mathcal{G}_U \rightarrow$ *YALLA*$_U$*, s.t.*

$$\Phi_U((\mathcal{A}, \mathcal{R})) = on(\mathcal{A}) \wedge \bigwedge_{x \in \mathcal{A}_U \setminus \mathcal{A}} \neg(on(\{x\}) \wedge \bigwedge_{(x,y) \in \mathcal{R}} (\{x\} \rhd \{y\}) \wedge \bigwedge_{(x,y) \in \mathcal{R}_U \setminus \mathcal{R}} \neg(\{x\} \rhd \{y\}).$$

$\Phi_U(\mathcal{A}, \mathcal{R})$ *is called the* characteristic formula *of* $(\mathcal{A}, \mathcal{R})$.

Example 15 *The argumentation systems* $\mathcal{G}_{jury}$ *of Example 13 can be expressed by:* $\Phi_U(\mathcal{G}_{jury}) = on(\{a_0, a_1, a_4\}) \wedge (\{a_4\} \rhd \{a_1\}) \wedge (\{a_1\} \rhd \{a_0\}) \wedge \bigwedge_{a \in \{a_2, a_3, a_5, a_6, a_7\}} \neg on(\{a\}) \wedge \neg(\{a_0\} \rhd \{a_1\})$

[1]The constant symbols of the language YALLA$_U$ are abusively denoted by the elements of $2^{\mathcal{A}_U}$.

YALLA$_U$ allows us to express incomplete knowledge by using disjunctions. Moreover, thanks to the idea to interpret a term by a set of arguments, we have been able to provide formulas that express the criteria about sets of arguments underlying the traditional argumentation semantics (such as conflict-freeness, defense, admissibility etc.). For instance a term t represents a conflict-free set in $(\mathcal{A}, \mathcal{R})$ if and only if $(\mathcal{A}, \mathcal{R}) \models on(t) \wedge (\neg(t \rhd t))$.

In the literature several logical formalisms have already been proposed for encoding argumentation systems. For instance, Villata et al. [191] have proposed a logical formalism for representing (and reasoning about) the extensions of traditional semantics. This work follows the work of Besnard and Doutre [44] where the arguments are denoted by symbols of the language enabling the user to write formulas whose models are sets of arguments. The purpose of this work is to characterize the extensions. A language with a similar expressive power has been proposed by Coste-Marquis et al. [69], with another purpose. The idea was to generalize Dung's formal framework [90] by taking into account additional constraints (expressed in a logical form) about the admissible sets of arguments. A logical language was also proposed by Wooldridge et al. [202] in which it is possible to express acceptability, conflicts and defense notions. However, this formalism is devoted to logical arguments (see next Section).

Those works are related to our YALLA$_U$ proposal since these languages enable also to describe and reason about argumentation systems, however none of them enables the user to express structural properties of an abstract argumentation system together with its semantical properties (which is the main purpose of YALLA$_U$). In Section 4.c, the language YALLA$_U$ is used in order to apply belief update concepts to argumentation.

3.b (Support, Claim) Arguments

[93] F. Dupin de Saint-Cyr. Handling enthymemes in time-limited persuasion dialogs. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 149–162. Springer-Verlag, 2011
[92] F. Dupin de Saint-Cyr. A first attempt to allow enthymemes in persuasion dialogs. In *DEXA International Workshop: Data, Logic and Inconsistency (DALI)*, pages 332–336. IEEE Computer Society - Conference Publishing Services, 2011

In this section we are going to explore another representation framework that uses structured arguments.

It is generally admitted that a logic-based argument is composed of two parts: a support and a claim, such that the support is a logical minimal proof of the claim [126]. In everyday life, there is nearly no "logical argument", we often give an argument without mentioning implicit common knowledge. Otherwise an argument would be very long to express and boring to listen (it could even be infinite when each part of the support of a claim should in turn be completely explained). Shortly speaking a logical argument is not into line with Gricean maxims[2]. Approximate arguments, called enthymeme by

[2]In [122], Grice describes four categories of cooperative principles: Quantity ("make your contribution as informative as is required (for the current purposes of the exchange)" and "do not make your contribution more informative than required" (Grice conceded that this last part is not mandatory), Quality

Aristotle, is a syllogism keeping at least one of the premises or conclusion unsaid.

Handling enthymeme has two advantages: first it allows to deal with more concrete cases where agents want to shorten their arguments. Note that the problem of implicit knowledge was one of the motivation for non-monotonic reasoning which aims at reasoning despite a lack of information, indeed argumentation is clearly linked with default reasoning (Dung's semantics were inspired from the ASP domain). Second it may involve a strategic matter, namely droping a premise may remove a possible attack or may enable to cheat by pretending that implicit knowledge can help to prove a claim while it is not the case...

Enthymemes have already been studied in the literature (by Walton and Reed and Macagno and also by Black and Hunter in [198, 193, 55, 196]), in [93, 92], we have adopted and extended the definitions proposed by those authors:

Definition 23 (logical arguments)
A logical argument *is a pair* $\langle S, \varphi \rangle$ *such that:* $\begin{cases} (1) & S \subseteq \mathscr{L}, \varphi \in \mathscr{L} \\ (2) & S \nvdash \bot, \\ (3) & S \vdash \varphi, \\ (4) & \nexists S' \subset S \text{ s.t. } S' \vdash \varphi \end{cases}$

The set of logical-arguments that can be built on a set $E \subseteq \mathscr{L}$ *of formulas is denoted by* $\texttt{Args}(E)$. *S is called the support (*$\texttt{Supp}$*) and c is the conclusion (*$\texttt{Conc}$*).*

More generally, the definition of argument can be grounded on Tarski's logics [183] as in [4]: *i.e.*, those logics are defined by pairs $(\mathscr{L}, \mathrm{CN})$ where $\mathscr{L}$ is a set of well-formed *formulas* and CN is a *consequence operator* that satisfies the following basic properties:

- *Expansion:* $X \subseteq \mathrm{CN}(X)$
- *Idempotence:* $\mathrm{CN}(\mathrm{CN}(X)) = \mathrm{CN}(X)$
- *Absurdity:* $\mathrm{CN}(\{x\}) = \mathscr{L}$ for some $x \in \mathscr{L}$

The notion of *consistency* is then defined as follows: A set $X \subseteq \mathscr{L}$ is *consistent* w.r.t. a logic $(\mathscr{L}, \mathrm{CN})$ iff $\mathrm{CN}(X) \neq \mathscr{L}$. It is *inconsistent* otherwise.

In such a setting an argument that is built from a *knowledge base* $\Sigma \subseteq \mathscr{L}$ is a pair (X, x) s.t. $X \subseteq \Sigma$, X is consistent $x \in \mathrm{CN}(X)$ and $\nexists X' \subset X$ such that $x \in \mathrm{CN}(X')$. An argument (X, y) is *atomic* iff $X = \{x\}$ and $\mathrm{CN}(\{x\}) = \mathrm{CN}(\{y\})$. An argument (X', x') is a *sub-argument* of an argument (X, x) iff $X' \subseteq X$.

In [6], different conflict relations between logic-based arguments have been studied. For instance the relation "Undercut" defined as followed will be used in Section 5.a.

Definition 24 (Undercut) *Let* $A_1 = (X_1, x_1)$, $A_2 = (X_2, x_2)$ *be two logical arguments,* A_1 undercuts A_2 *if* $\exists\ h_2 \in X_2$ *such that* $x_1 \equiv \neg\ h_2$.

("Try to make your contribution one that is true: Do not say what you believe to be false, Do not say that for which you lack adequate evidence"), Relation ("Be relevant"), Manner ("related to HOW" to say things: "Avoid obscurity of expression, Avoid ambiguity, Be brief (avoid unnecessary prolixity), Be orderly").

Works by linguists [165, 169] have emphasized the main forms of counter-argumentation that may take place in every day life dialogs. The first one concerns undermining the conclusion of another argument. It is known in AI as "rebuttal" [108]. The second common form of attacks, known in AI as "assumption attack" [108], consists of undermining a premise in the support of another argument.

Definition 25 (Rebuttal and Assumption attack)
- An argument α *rebuts* an argument β iff the set $\{\texttt{Conc}(\alpha), \texttt{Conc}(\beta)\}$ is inconsistent.
- An argument α *assumption-attacks* an argument β iff $\exists x \in \texttt{Supp}(\beta)$ s.t. the set $\{\texttt{Conc}(\alpha), x\}$ is inconsistent.

In a context of reasoning, Dung's extensions can be used in order to define the plausible conclusions denoted $\texttt{Output}(\mathcal{A}, \mathcal{R})$ to be drawn from a knowledge base Σ. The idea is to infer a formula x from Σ iff x is the conclusion of an argument that is *skeptically accepted*. However Amgoud and Besnard [5] have shown that the conclusions that can be obtained from a knowledge base by using Dung's extensions with logic-based arguments are not always rational according to the attack relation and the semantic chosen. This drawback of Dung's proposal for structured arguments was a motivation for me to develop new approaches for reasoning about these arguments (see Sections 5.b and 6.b).

The notion of approximate argument is independent of the logic used:

Definition 26 (approximate arguments [45, 126])
An approximate argument *is a pair* $\langle S, \varphi \rangle$ *where* $S \subseteq \mathscr{L}$ *and* $\varphi \in \mathscr{L}$.

In other words, an approximate argument is simply a pair (support,claim) and when the support is a minimal proof of the claim this argument is called a logical argument. Note that an approximate argument does not need to have a consistent support S and it is not required that its conclusion φ is a logical consequence of S. In order to be able to deal with arguments that have incomplete support or incompletely developed conclusion we first defined an incomplete argument and then extend the enthymeme formalization proposed by Black and Hunter in [55].

Definition 27 (incomplete argument) *An* incomplete argument *is a pair* $\langle S, \varphi \rangle$ *where* $S \subseteq \mathscr{L}$ *and* $\varphi \in \mathscr{L}$ *(i.e.,* $\langle S, \varphi \rangle$ *is an approximate argument) such that:*

$$\begin{cases} (1) & S \nvdash \varphi \\ (2) & \exists \psi \in \mathscr{L} \text{ s.t. } \langle S \cup \{\psi\}, \varphi \rangle \text{ is a logical argument} \end{cases}$$

In this definition, the first condition expresses the fact that the argument is strictly incomplete, *i.e.,* the support is not sufficient to infer the conclusion. The second one imposes that it is possible to complete it in order to obtain a logical argument. Logical or incomplete arguments are particular distinct cases of approximate arguments. Note that the support of an incomplete argument should be consistent or else adding any formula to it would still give an inconsistent support (hence violate condition (2) for logical arguments). Moreover S should be consistent with φ. Our definition is a slight variation of Hunter's concept of *precursor*, which he defines as an approximate argument

$\langle S, \varphi\rangle$ such that $S \nvdash \varphi$ and $S \nvdash \neg\varphi$. Hence an "incomplete argument" is a "precursor" but the converse is false. The small difference lays in the fact that a completed precursor may not be minimal, for instance $\langle\{a, b, a \wedge b\}, c\rangle$ is a "precursor" and not an "incomplete argument" since any completion would have a non minimal support (*i.e.*, in Definition 26, (4) will not hold).

Example 16 *In [173], Schopenhauer gives the following example (in order to explain the extension stratagem*[3]*) "I asserted that the English were supreme in drama. My opponent attempted to give an instance to the contrary, and replied that it was a well-known fact that in music, and consequently in opera, they could do nothing at all". The argument of Schopenhauer's opponent is an incomplete argument. Indeed, "in music (*m*), and consequently in opera (*o*), English are not supreme (*$\neg s$*)" maybe transcribed into the following approximate argument:* $a = \langle\{m \rightarrow \neg s\}, o \rightarrow \neg s\rangle$. *And by adding the formula* $o \rightarrow m$ *to its support we obtain the following logical argument:* $b = \langle\{m \rightarrow \neg s, o \rightarrow m\}, o \rightarrow \neg s\rangle$.

There are two ways to "complete" an argument: either by adding premises, then the support should be strictly included in the completed support or by specifying the conclusion, then it should be inferred by the union of the completed conclusion and support but should differ from the previous conclusion.

Definition 28 (enthymeme) *Let* $\alpha = \langle S, \varphi\rangle$ *and* $\alpha' = \langle S', \varphi'\rangle$ *being approximate arguments,* $\langle S', \varphi'\rangle$ completes $\langle S, \varphi\rangle$ *iff*
$\begin{cases} (1) & S \subset S' \text{ and } \varphi = \varphi' \text{ or} \\ (2) & S \subseteq S' \text{ and } \{\varphi'\} \cup S' \vdash \varphi \text{ and } \varphi \neq \varphi' \end{cases}$

α *is an* enthymeme *for* α' *iff* α' *is a logical argument and* α' *completes* α.

Our definition extends the definition of [55] in the sense that it allows to cover arguments whose conclusion is an implicit claim requiring implicit support (the following example would not be considered as an enthymeme by [55]).

Example 16 (continued): *We may build an infinity of logical arguments decoding an incomplete argument. For instance,* α *is an enthymeme for the logical argument* β *but also for the logical argument:* $\gamma = \langle\{m \rightarrow \neg s, o \rightarrow m, o, o \rightarrow d\}, \neg(d \rightarrow s)\rangle$.

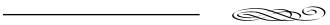
The following function gives the set of logical arguments that can be built from a knowledge base Σ and that are enthymemes for a given argument.

Definition 29 (Decode) *Let* $\Sigma \subseteq \mathscr{L}$ *and* $\langle S, \varphi\rangle \in$ AArg, $\mathsf{Decode}_\Sigma(\langle S, \varphi\rangle) = \{\langle S', \varphi'\rangle \in$ Arg *such that* $S' \backslash S \subseteq \Sigma$, $\varphi' \in \Sigma$ *and* $\langle S, \varphi\rangle$ *is an enthymeme for* $\langle S', \varphi'\rangle\}$.

In the previous example, it holds that $\beta, \gamma \in \mathsf{Decode}_{\mathscr{L}}(\alpha)$.

[3] This stratagem consists in extending what the adversary has said in order to invalidate the generalization obtained.

Logical, approximate, and enthymeme arguments are usually called logic-based arguments, and are often used for a reasoning task, in Section 5.b we have proposed a persuasion dialog protocol in which the agents exchange this kind of arguments.



In this chapter we have presented two ways to model arguments. The second one is using a structured description of arguments. This structure is used to evaluate them with respect to the current knowledge, the precise content of an argument is described under a logical formalism which facilitates the evaluation.

In contrary, in an abstract argumentation system the evaluation depends only on the attack relation which is supposed to be given. This implies to use this model carefully in applications, since the attacks between arguments should be in accordance with the definitions of acceptable sets of arguments (called extensions): for instance an argument acceptable with regard to the definition of the "stable semantics" should belong to a set of arguments that are attacking every other argument.

Moreover it has been shown by Amgoud and Besnard in [5] that the combination of logic-based arguments with Dung's semantics may give counter-intuitive results. Furthermore, when dealing with logic-based argumentation the attack notion is not necessary since arguments can be evaluated by themselves, this is why the second approach did not integrate this attack notion at all.

However the attack relation between arguments may enable a graphic representation of conflicts under the form of a digraph which is in general a visual user friendly representation. The different tasks that can be achieved within these formalisms will be described in the two next Chapters.

Chapter 4

Change in abstract argumentation systems

[64] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Revision of an Argumentation System. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 124–134. AAAI Press, 2008
[65] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in Abstract Argumentation Frameworks: Adding an Argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010
[47] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in argumentation systems: exploring the interest of removing an argument. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 275–288. Springer-Verlag, octobre 2011
[48] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Duality between Addition and Removal: a Tool for Studying Change in Argumentation. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, volume 297 of *Communications in Computer and Information Science*, pages 219–229. Springer, juillet 2012
[49] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Characterizing change in abstract argumentation systems. In *Trends in Belief Revision and Argumentation Dynamics*, volume 48 of *Studies in Logic*, pages 75–102. College Publications, 2013
[50] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Enforcement in Argumentation is a kind of Update. In *International Conference on Scalable Uncertainty Management (SUM)*, number 8078 in LNAI, pages 30–43. Springer-Verlag, 2013
[94] F. Dupin de Saint-Cyr, P. Bisquert, C. Cayrol, and M.-C. Lagasquie-Schiex. Argumentation Update in YALLA (Yet Another Logic Language for Argumentation). *under submission to IJAR*, 2015

During my PhD and later on, I had acquired a scientific culture in belief-change theory, meanwhile Claudette Cayrol and Marie-Christine Lagasquie had developed studies on the topic of abstract argumentation. We decided to focus on what those two domains could bring to each other, it appeared to be a good idea, since this domain now called "dynamic of argumentation" is very flourishing nowadays... At the beginning, we wanted to know how the extensions of an argumentation system evolve with the arrival of a new information. The natural idea was to take into account the arrival of a new argument. But in fact, at start, this idea didn't seduce the community since argumentation was exactly made to be a non-monotonic principle hence adding an argument was simply viewed as a part of an arguing process. However, we had the idea to show some properties that could hold for the extensions after an argument addition, without being obliged

to recompute precisely those extensions. This was the subject of Aurore Miquel's Master 2 internship [152] and the subject of Pierre Bisquert PhD [46], it gave birth to several developments:

- we first proposed 4 kinds of minimal operations that can be done to an argumentation system: adding one argument with its interactions, removing an argument, adding one attack between existing arguments, removing one attack.
- we proposed a typology of the possible kinds of changes that could be produced by such operations
- we tried to characterize the changes of this typology w.r.t. some conditions about the initial argumentation system and about the operation (this was done only for addition and removal of one argument)
- we made a parallel between performing operations on an argumentation system and update
- we have provided a tool for computing the necessary operations to obtain a given property in an argumentation system.

At start we had only considered the addition of an argument, and we had called this change "a revision" in [64], however we had received several critics against this name, because revision is related with consistency while in argumentation this notion does not exist. However, now this term has been adopted by many people it is a pity since we have discovered later that the changes we were speaking of is more related to update than revision. In [65], we studied the properties that hold on a system after adding an argument w.r.t. to some conditions about that argument.

Due to the way natural argumentation is done, examples of additions were very easy to find. It was not so obvious for removal, we were the first to study it, and during the PhD of Pierre Bisquert we were able to find some examples of suppression, and to study what are the properties of removal (see [47]). A typical example of the need to remove an argument is when an objection is accepted during a trial, since the objected argument should be removed from the reports about the trial, hence from its associated argumentation system.

In the following we detail the typology of the change properties, the characterizations that were obtained, the link with update and the tool, all these results can be found in Pierre Bisquert's PhD [46].

4.a A typology of change properties

When we started to work on argumentation dynamics we first tried to define a typology of the possible changes in [65]. But since it was done with the point of view of change generated by adding one argument, we have generalized it in [47].

Given an elementary operation among adding/removing one argument together with its set of interactions, adding/removing one attack, we may face different changes concerning the extensions of the initial argumentation system. Note that we consider that the "semantics" does not change.

This typology aims at distinguishing the possible impacts of the changes on three levels by comparing things "before" (*i.e.*, in the initial system whose set of extensions is called E and one of them[1] is called $\mathcal{E}$) and "after" the change (*i.e.*, in the resulting system, its set of extensions is denoted by E' and one of them is denoted $\mathcal{E}'$):

- the set of extensions: the number of extensions may increase (the change is called *extensive*, denoted e), may decrease (it is a *restrictive* change, denoted r) or may remain *constant* (denoted c). We refine these criteria by taking into account the particular cases where there is no extension (denoted 0), only one empty (denoted $1v$), only one non-empty extension (denoted $1nv$), several extensions (denoted k or j with j representing a number lower than k). For the constant case, we refine the classification as follows: a *c-expansive* (respectively *c-narrowing*) change concern a set of extensions in which each extension strictly increases (resp. decreases). If the set of extensions remains the same the change is called *c-conservative* (c-cons for short) , finally all other constant changes are called *c-altering*. This first classification is described in Table 4.1.

Initial System	**Final System** $E' = \varnothing$	$E' = \{\varnothing\}$	$E' = \{\mathcal{E}'\}, \mathcal{E}' \neq \varnothing$	$\lvert E'\rvert > 1$
$E = \varnothing$	c-cons	$\times$	$e_{\varnothing-1ne}$	$e_{\varnothing-k}$
$E = \{\varnothing\}$	$\times$	c-cons	c_{1e-1ne}	e_{1e-k}
$E = \{\mathcal{E}\}$, $\mathcal{E} \neq \varnothing$	$r_{1ne-\varnothing}$	r_{1ne-1e}	c-cons, if $E = E'$ c-expansive, if $\mathcal{E} \subset \mathcal{E}'$ c-limitative, if $\mathcal{E}' \subset \mathcal{E}$ c-altering, else	e_{1ne-k}
$\lvert E\rvert > 1$	$r_{k-\varnothing}$	r_{k-1e}	r_{k-1ne}	e_{j-k}, if $\lvert E\rvert < \lvert E'\rvert$ c-cons, if $E = E'$ c-expansive, if $\lvert E\rvert = \lvert E'\rvert$ and (1) c-limitative, if $\lvert E\rvert = \lvert E'\rvert$ and (2) c-altering, if $\lvert E\rvert = \lvert E'\rvert$ and (3) r_{k-j}, if $\lvert E\rvert > \lvert E'\rvert$

$\times$: impossible case.
(1): $\forall \mathcal{E} \in E, \exists \mathcal{E}' \in E'$ s.t $\varnothing \neq \mathcal{E} \subset \mathcal{E}'$ and $\forall \mathcal{E}' \in E', \exists \mathcal{E} \in E$ s.t $\varnothing \neq \mathcal{E} \subset \mathcal{E}'$
(2): $\forall \mathcal{E} \in E, \exists \mathcal{E}' \in E'$ s.t $\varnothing \mathcal{E}' \subset \mathcal{E}$ and $\forall \mathcal{E}' \in E', \exists \mathcal{E} \in E$ s.t $\varnothing \neq \mathcal{E}' \subset \mathcal{E}$
(3): $E \neq E'$ and not (1) and not (2)

Table 4.1: *Classification of change wrt extensions set*

- the extensions themselves: we have defined several change properties capturing the monotony of the acceptability of sets of arguments, they are described in Table 4.2.

[1] when there is one

For instance *expansive monotony* amounts to have a change such that every set of arguments that were conjointly accepted before change are still accepted conjointly after.

	Expansive Monotony	Restrictive Monotony
Simple	$\forall \mathcal{E} \in E, \exists \mathcal{E}' \in E'$ s.t $\mathcal{E} \subseteq \mathcal{E}'$	$\forall \mathcal{E}' \in E', \exists \mathcal{E} \in E$ s.t $\mathcal{E}' \subset \mathcal{E}$
Credulous	$\bigcup_{\mathcal{E} \in E} \mathcal{E} \subseteq \bigcup_{\mathcal{E}' \in E'} \mathcal{E}'$	$\bigcup_{\mathcal{E}' \in E'} \mathcal{E}' \subseteq \bigcup_{\mathcal{E} \in E} \mathcal{E}$
Skeptical	$\bigcap_{\mathcal{E} \in E} \mathcal{E} \subseteq \bigcap_{\mathcal{E}' \in E'} \mathcal{E}'$	$\bigcap_{\mathcal{E}' \in E'} \mathcal{E}' \subseteq \bigcap_{\mathcal{E} \in E} \mathcal{E}$

Table 4.2: *Classification of changes wrt extensions monotonicity*

- the arguments: we have defined several changes that can be considered wrt the status of one particular argument, say x. They are described in Table 4.3 where E_x (resp. E'_x) is the set of the extensions that contain x before (resp. after) change. For instance, *Total Acceptability Establishment* consists in making x belong to all extensions after the change while it was not accepted before.

Initial System	**Final System**		
	$E'_x = \varnothing$	$0 \subset E'_x \subset E'$	$\varnothing \subset E'_x = E$
$E_x = \varnothing$	Reject Conservation	Partial Acceptability Establishment	Total Acceptability Establishment
$\varnothing \subset E_x \subset E$	Removal of Credulous Acceptability	Credulous Acceptability Conservation	Global Establishment of Acceptability
$\varnothing \subset E_x = E$	Skeptical Acceptability Removal	Acceptability Reduction	Skeptical Acceptability Conservation

Table 4.3: *Classification of changes wrt the argument x acceptability*

The previous typology is a kind of catalog of the possible change properties in argumentation. Some changes may be considered as useful according to the role of the user, *e.g.* a debate moderator may be interested in focusing or enlarging the dialog depending on the remaining time, while an orator may have dialog strategies and may want to focus on particular arguments. Let us review some of these properties:

- A "decisive" change (*e.g.* c_{1e-1ne}) is useful to lower ignorance since after this change one and only one extension remains. It can be used by a moderator for concluding the debate.
- An "expansive" change (*e.g.* c-expansive) increases the accepted arguments while conserving those already accepted, it can also be used by a moderator or by an orator in order to convince a larger audience about the current view of the debate.
- A "conservative" change (*e.g.* c-conservative) may be a more neutral attitude that can be adopted by a moderator or an orator that does not want to deliver new information but wants to participate (very useful political waffle).

- "Monotony" allows us to focus on some particular arguments and may be used strategically by an orator.
- "Questioning" change (*e.g.* e_{j-k}) and "destructive" change (*e.g.* r_{k-1e}) are increasing ignorance either by augmenting the possible views or by destroying any coherent view, they may be used desperately by a strategical orator/manager that wants to forbid any decision to be made.
- An "altering" change (*e.g.* c-altering) allows to completely change the point of view, it may also be done to reverse the course of the debate.

4.b Characterizations

Once we have defined the framework in which change properties can be classified, we can address the important issue of providing characterizations for these properties: *i.e.*, conditions on the argumentation system and on the change operation that are necessary or sufficient to guarantee that the properties are satisfied.

We have obtained characterization results either by a direct proof [48] or by using an indirect proof based on other results about the dual operation (addition being the dual of removal) [49]. The following proposition is an example of characterization:

Characterization 15 of [49] : *Let $\mathcal{G} = (\mathcal{A}, \mathcal{R})$ be an argumentation system, and $\mathcal{E}$ its grounded extension. For any operation o of the form $\langle \oplus, z, \mathcal{R}_z \rangle$ executable by an agent on $\mathcal{G}$, for any argument $x \in \mathcal{A} \cup \{z\}$, if $\nexists y \in \mathcal{A}$ such that $(y, z) \in \mathcal{R}$ and $\{z\}$ indirectly defends x and $x \notin \mathcal{E}$ then $x \in \mathcal{E}'$ where $\mathcal{E}'$ denotes the grounded extension of $\mathcal{G}' = o(\mathcal{G})$.*

This proposition established in [49] is a characterization of a change that *sets up the acceptability* of an argument. Indeed it concerns the goal of "enforcement" of the argument x (as we will see in Section 4.c). Thanks to this characterization, we know (without requiring a new computation of the extensions) that if an operation adds an argument z, such that z is not attacked and indirectly defends another argument x which was not accepted under the grounded semantics, then x will become accepted.

Our characterizations can be considered as a guide for selecting the change operation to perform in order to obtain a desired property on an argumentation system and they may also be used as a tool for predicting the result of a change operation in a given context.

4.c Axioms (belief change)

In the literature, the operation to perform on an argumentation system in order to ensure that a given set of arguments is accepted given a set of authorized changes is called "enforcement" [35]. This enforcement may be done more or less easily, since it may involve more or less changes (costs to add/remove arguments may be introduced).

Example 17 *For instance, suppose that a lawyer knows the current state of knowledge of the jury (under the form of an argumentation system) and suppose she wants to make*

the audience to accept a set of arguments. In order to achieve this goal, she has to make a change to the audience argumentation system, either by adding an argument or by making an objection about an argument that was uttered before (in order to remove it). The aim of the speaker will be to find the least expensive change to perform.

This example is a particular case of a more general enforcement operator. Since we could consider cases where the agent does not know exactly the argumentation system on which she must make a change but knows only some information about it (*e.g.* some arguments that are accepted or that are present in the system). In this more general case, the idea is to ensure that the argumentation system after change satisfies a given property (that maybe different from the acceptability of a precise argument) whatever the initial system was.

The key idea that we have developed in [50] is the parallel between belief update theory [200, 136] and enforcement in argumentation. Enforcement consists in searching for the argumentation systems that are closest to some given starting argumentation systems, in a set of argumentation systems in which some target arguments are accepted. This gives us the parallel with preorders on worlds in belief update. Hence worlds correspond to argumentation systems while formulas should represent knowledge about these argumentation systems. In classical enforcement this knowledge is expressed in terms of a description of an initial argumentation system and a set of arguments that one wants to see accepted. This is why we have proposed to introduce a propositional language in which this kind of information may be expressed. This language called YALLA$_U$ has been described in Section 3.a and had enabled us to generalize enforcement with a broader expressiveness.

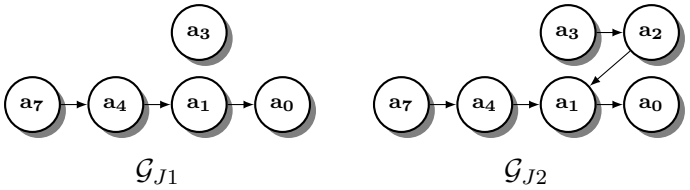
Note that we rather face an update problem, since an agent wants to change an argumentation system in order to satisfy a particular goal. A revision approach would apply to situations in which the agent learns some information about the initial argumentation system and wants to correct its knowledge about it. This would mean that the argumentation system has not changed but the awareness of the agent has evolved.

We have established the representation theorem for a generalized update operator (*i.e.*, operator with a set of authorized transitions), in order to capture generalized enforcement operators. We have proven that a generalized enforcement is a kind of update with transition constraints that capture "authorized operations" on argumentation systems. Indeed some operations may be allowed or not according to the knowledge (encoded by an argumentation system) of the user and according to the target argumentation system on which he wants to act.

The axioms that are verified by an enforcement operator are the postulates U1, E3, U4, E5, E8, U9. Those postulates are not specific of argumentation systems they only characterized an update operator satisfying transition constraints. This is why we needed to refine our representation theorem with properties that are adapted to argumentation system, hence that take into account the attack relations and the definitions of extensions. A result of our research, presented in [94], was the idea to use the characterizations as specific update postulates for argumentation, by translating them into a formula of YALLA$_U$ with an update operator inside. More precisely we have provided two proposi-

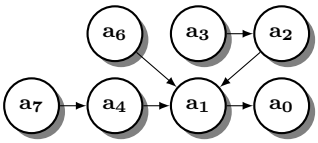
tions (one for the addition and one for the removal of an argument) which are kinds of argumentation postulate builder, allowing to build a generalized enforcement postulate corresponding to every result already established by a characterization.

Example 18 *Let us suppose that the lawyer thinks that the jury's knowledge may be represented by two argumentation systems $\mathcal{G}_{J1}$ and $\mathcal{G}_{J2}$ that are equally possible. Translated in YALLA, they are the two models of the formula $\varphi = \Phi_U(\mathcal{G}_{J1}) \vee \Phi_U(\mathcal{G}_{J2})$.*



$\mathcal{G}_{J1}$ $\mathcal{G}_{J2}$

Given the knowledge of the lawyer:



She wants to have a_0 accepted under the grounded semantics. Let us suppose that she is only authorized to perform an elementary change (because the judge left her only one word to add) and we can also assume that she is not allowed to object against arguments that are not present and not able to add arguments that she does not know: let us denote $\mathcal{T}_l$ this set of authorized transitions. Among this set of elementary changes $\mathcal{T}_l$, she prefers addition to removal. Then it means that she should find if the following formula[2] has some models:

$$[\varphi \diamond_{\mathcal{T}_l} (\exists p, G(p) \wedge (\{a_0\} \subseteq p))]$$

Here are some operations for the lawyer, the first level is the preferred ones, the second is less preferred, the third level contains operation that are not authorized.

$(\oplus, a_2, \{(a_2, a_1)\})$	$(\oplus, a_2, \{(a_2, a_1), (a_3, a_2)\})$
$(\oplus, a_6, \{(a_6, a_1)\})$	
$(\ominus, a_0, \varnothing)$	$(\ominus, a_1, \varnothing)$
$(\ominus, a_3, \varnothing)$	$(\ominus, a_4, \varnothing)$
$(\ominus, a_7, \varnothing)$	
$(\ominus, a_2, \varnothing)$	$(\oplus, a_5, \{(a_5, a_1)\})$
...	

Hence she can operates the two changes: $(\oplus, a_2, \{(a_2, a_1)\})$ and $(\oplus, a_6, \{(a_6, a_1)\})$.

Note that $(\oplus, a_6, \{(a_6, a_1)\})$ could have been obtained directly by using Characterization 15 of [49] (see Section 4.b) where a_6 play the role of z.

[2]In this formula, $G(p)$ is a shortcut for the formula in YALLA_U that expresses that p is the grounded extension.

4.d Tool

Several software tools have already been designed for helping a user to reason with arguments. They are meant to be cognitive assistants (see for instance [190, 184]) or they focus on a particular domain (*e.g.* [22, 2] in the domain of case-based legal reasoning). Our proposal presented in [51] was more at a strategical level since it aimed at helping the user to persuade an audience, hence our tool was made to provide means to do it (arguments that could be added, or objection that could be done). More precisely, the tool developed by Pierre Bisquert[51] is able to produce as output the list of actions executable by the agent in order to achieve its goal. This software uses the characterizations presented in Section 4.b. The software can handle three semantics, the grounded, the preferred and stable semantics. In addition, at the current stage, the software handles only some types of change (addition and removal of arguments only).

The tool is organized around two modules: an argumentation system (AS for short) handler, and an inference engine (for computing the change operations), see Figure 4.1. The outputs of Module 1 are inputs for Module 2.

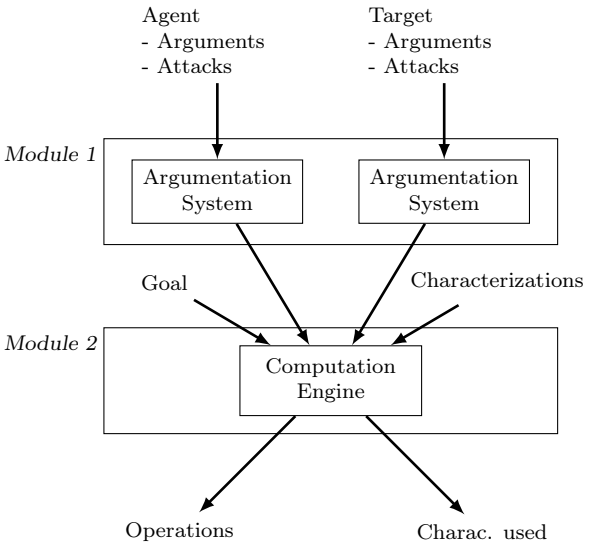


Figure 4.1: *Architecture of the tool.*

The first module encoded in Python may handle the creation of various AS, defined by giving a set of arguments and a set of attacks, hence in accordance with our theoretical framework, the tool makes it possible to create one AS for the agent and one for its target. Moreover, the semantics used in the target system is also specified (it will be used to check the achievement of the goals of the agent). The two AS as well as the extensions of the target, are then transmitted to the second module.

The second module encoded in Prolog computes the change operations. More precisely, it allows to answer to the question "What are the operations executable by the agent on the target system that can achieve her goal?". This module requires a goal and a

set of characterizations. The characterizations translate naturally into logical rules (this is what gave us the idea to develop the logical language YALLA), and the mechanism of unification allows us to generate and easily filter the operations wrt the AS and the goal of the agent.

We have presented a theoretical framework and a tool able to find a change operation which achieves a goal given a target AS and given arguments and attacks from a "source" AS (representing the knowledge of an agent). An important issue about knowledge dynamics is the establishment of a set of axioms that characterize rational changes. Hence it was important to situate our framework in the field of belief change theory. We first have generalized classical update postulates in order to take into account a set of authorized transitions (see Section 2.c). Since the update approach is based on classical propositional logic, we have defined a logical language (called YALLA$_U$ in Section 3.a) for representing argumentation systems. Then we have shown that the change operations on argumentation systems are update operations. Moreover due to previous works about dynamics in argumentation we have been in position to provide a set of new postulates (based on change characterizations) that are specific of argumentation update.

The tool that we have provided could be used as a cognitive assistant for human agents and may also help to locate gaps of characterization. We have studied the behavior of this tool by means of two experimental protocols used for generating benchmarks made of random AS and random goals.

We can see the characterizations as guides to decide about the action to perform on a target graph. This decision should be done according to the nature of the desired change (given by our typology), according to the difficulty to check if the condition of the characterization holds, and according to the typicality of the property to show (which relates to an estimation of the risk associated to the action).

Since computing the extensions after change is very expensive (in spite of the progress made in [146]), one perspective is to show that our approach is less expensive. This will require to determine the computational complexity of the tests required to check the applicability of some characterizations. Indeed, some characterizations use complex concepts: for example the indirect defense of an argument by a set.

It seems necessary to extend our trial example in order to allow for a real interaction between the prosecutor and the lawyer. In a more general way, this implies to study the changes operated by an agent on its own system when another agent carries out a modification of the target AS. In other words it means to study the *revision* of argumentation systems.

The main difficulty in this work appeared to be the abstract argumentation theory itself, since in order to study the evolution of extensions, the first step was to find examples where extensions are natural, then make the system evolve. I have been stroken by the fact that it is very difficult to find examples where Dung's extensions are solutions

of a problem (see more details in Section 6.a). This could be seen as a drawback since it is important to show that an approach is applicable for real problems. However, although all our characterizations were done for Dung's semantics, our results are usable with any kind of characterization *e.g.* that could rely on other definitions of accepted argument. This is due to our definition of the logical language YALLA$_U$ which allows us to go beyond Dung's classical semantics, since it may capture any definition of an acceptable set of arguments. Nevertheless, this difficulty incited me to design new models for argumentation especially in the context of persuasion dialogs and group decision (see Sections 5.b and 6.b).

Chapter 5

Dialogs

Argumentation plays a central role in the domain of dialog systems. Indeed natural argumentation involves a discussion between at least two agents (it may be restricted to only one agent when she is deliberating with herself, but in that case she is playing two distinct roles). A dialog system is built on three main components: i) a *communication language* specifying the locutions that will be used by agents during a dialog, ii) a *protocol* specifying the set of rules governing the dialog such as who is allowed to say what and when? and iii) agents' strategies for selecting their moves at each step in a dialog.

We first describe a general Agent Communication Language (ACL) based on commitments and penalties, in which everything that is said commits either the speaker or the hearer. Then we focus on *persuasion* dialogs. Persuasion concerns two (or more) agents who disagree on a state of affairs, they engage in the dialog in order to persuade the others to change their minds either in *public*, *i.e.*, in presence of an audience (*e.g.* [58, 63, 139]) or in *private* (*e.g.* [18, 164, 204]). Private persuasion is more concerned by the evolution of the argumentation systems when adding arguments received from the other party: an agent becomes persuaded of a claim if this claim becomes supported by its argumentation system. In general, I have been more interested in public persuasion like *e.g.* a political debate where both candidates are trying to rather convince the voters than their adversary.

At the time of our study, while there were numerous works on dialog protocols (e.g, whether a dialog terminates, or whether turn shifts equally between agents etc), no work had been done on criteria for evaluating the dialogs generated. In real life, two people listening to the same political debate may disagree on the "winner" and have different feelings about the dialog itself. Hence, it seems important to be able to compare objectively the *quality* of different dialogs. Such a comparison may help to design protocols that enforce agents to produce better dialogs w.r.t. chosen criteria. This is the object of the second section of this chapter.

Enabling agents to use approximate arguments in persuasion dialogs makes more room for strategies and generalizes the dialog setting. Our third contribution concerns the definition of an ACL enabling enthymemes in which we have characterized the notion of common knowledge and designed a protocol enforcing agents to converge towards more

agreements.

Our last contribution in this domain is an axiomatisation of logic-based argumentative persuasion dialog systems which is very different from the one defined for logic-based argumentative reasoning systems. It led us to claim that the two domains should be dealt with different mechanisms.

5.a Commitment and penalties

[12] L. Amgoud and F. Dupin de Saint-Cyr. Measures for persuasion dialogs: A preliminary investigation. In *Computational models of argument (COMMA)*, pages 13–24. IOS Press, 2008
[11] L. Amgoud and F. Dupin de Saint-Cyr. A new semantics for ACL based on commitments and penalties. *International Journal of Intelligent Systems*, 23(3):286–312, 2008
[10] L. Amgoud and F. Dupin de Saint-Cyr. Towards ACL semantics based on commitments and penalties. In *European Conference on Artificial Intellige*, pages 235–239. IOS Press, 2006
[9] L. Amgoud and F. Dupin de Saint-Cyr. A Semantics for Agent Communication Languages based on commitments and penalties. In *International Workshop on Computational Logic in Multi-Agent Systems (CLIMA)*, pages 28–39. Springer, 2005

In complex multi agent systems, the agents may be heterogeneous and possibly designed by different programmers. Thus, the importance of defining a standard framework for ACL with a *clear semantics* has been widely recognized. The definition of an ACL from a syntactic point of view amounts to list the different *speech acts* [23, 175] that agents can perform, the semantic definition defines the conditions under which a given speech act can be played. It should be *verifiable*, *i.e.*, it should be possible to check whether a system conforms to a particular ACL or not, *clear* and *practical* [201]. Although a number of significant agent communication languages have been developed, at the time of our study, obtaining a suitable formal semantics for ACLs which satisfies the above objectives was remaining one of the greatest challenges of multi-agent theory.

Indeed, most classical proposals fail to meet these objectives. For instance, *mentalistic semantics* (*e.g.* KQML [111] and FIPA [112]) based on the mental states (beliefs and intentions) of the interacting agents are not verifiable as shown in [201] since they assume, more or less explicitly, that agents are "sincere" and "cooperative". The most popular category of semantics is the *social* one. In this kind of approach, as developed in [67, 180, 181], primacy is given to the interactions among the agents. The semantics is based on social *commitments*. A commitment is an engagement taken by an agent towards a set of agents. Commitments are induced by uttering speech acts. For example, by affirming a data, the agent commits on the truth of that data. After a promise, the agent is committed to carrying it out. While this approach had overcame the limitation of the mentalistic approach by being verifiable, at the time of our study, it was still suffering from some weak points, in particular, the concept of commitment was *ambiguous*.

In our proposal developed in [12, 11, 10, 9], we defined a formal semantics which is *social* in nature. Our main contribution was to give a translation of each possible speech act in terms of a commitment for the sender or for the receiver (without referring to the mental state of the agent) translated in terms of *penalty* to be paid by the agent who has not fulfill its commitment. Indeed when a question is uttered, a commitment

for giving an answer is created, and the debtor is the hearer. Note that this does not mean that the hearer should necessarily give an answer. A dialogue protocol may impose such a condition, but the problem of dealing with protocols was beyond the scope of our research, our aim was to give a clear and verifiable *meaning* to each speech acts..

Formally, let $\mathcal{A} = \{a_1, \ldots, a_n\}$ be a set of variables denoting *agents* identifiers. Each agent is assumed to have a *role* allowing it to have the control over a subset of formulas in $\mathscr{L}$. By having a control over a formula, we mean that the agent is allowed to alter the truth value of that formula, formally: $\texttt{Role} : \mathcal{A} \longmapsto 2^{\mathscr{L}}$. The roles are supposed to be visible to all the agents. Thus, each agent is aware about the formulas that it can control, and about the formulas under the control of the other agents. Let $\mathcal{S}$ denote the set of speech acts. A *move* m is a tuple (s, R, act, x) where s and R are agents and set of agents (the sender and the receiver(s) respectively), *act* is a speech act (denoted by $\texttt{Act}(m)$) and x, denoted by $\texttt{Content}(m)$, is either a *consistent formula* of $\mathscr{L}$ or a *logical argument* of $Arg(\mathscr{L})$ (see Definition 23). When the sender ($\texttt{Sender}(m)$) and receiver are not important or implicit we denote the move by $act{:}x$. Let $\mathscr{M}$ denote the set of all the possible moves based on $\mathcal{S}$. An example of a move is $(a_1, \{a_2\}, \texttt{Question}{:}\varphi)$ where φ encodes "the sky is blue". This move means that a_1 asks to a_2 whether the sky is blue or not. Here `Question` is a speech act and φ is a propositional formula.

In the proposal developped with Leila Amgoud [11], we used the following set $\mathcal{S}$ of basic speech acts based on the proposal of Searle [174] that are commonly used in the literature (for instance [18, 19, 121, 159, 163, 204]) for modling the different types of dialogues identified by Walton and Krabbe [195]:

$$\mathcal{S} = \{\texttt{Assert}, \texttt{Argue}, \texttt{Declare}, \texttt{Question}, \texttt{Request}, \texttt{Challenge}, \texttt{Promise}\}.$$

Our purpose was to show that we were able to represent each of them in terms of commitments. Indeed, for each of these speech acts, Table 5.1 show its syntax, its meaning, the induced commitment and an example of content.

In addition to the above speech acts, we have considered another act called `Retract` which does not belong to the different categories of speech acts defined by Searle [174]. It can be seen as a meta-level act allowing agents to *withdraw* commitments already made. Allowing such a move makes it possible for the agents to have a kind of non-monotonic behavior (*i.e.*, to change their points of view, to revise their beliefs, etc.) without being sanctioned. Syntactically, $\texttt{Retract}{:}m$ is a meta-move with m being itself a move (*i.e.*, $m \in \mathscr{M}$).

We have proposed to store the various moves uttered during a dialogue in *commitment stores* (as in [150]) which are visible to all agents. Hence, contrarily to the mental states of an agent that are private the commitments of an agent are visible to all the agents. For instance if an agent a_i makes a request r to another agent a_j, the request (r) is stored in the commitment store of a_j. Hence, a_j is said *committed* to answer to it. Formally, a *commitment store* CS_i associated with a_i is a pair $CS_i = \langle A_i, O_i \rangle$ with: $A_i \subseteq \{m \in \mathscr{M} | \texttt{Act}(m) \in \{\texttt{Assert}, \texttt{Argue}, \texttt{Declare}, \texttt{Promise}\}\}$: it contains the commitments that the agent has willingly taken, and $O_i \subseteq \{m \in \mathscr{M} \mid \texttt{Act}(m) \in \{\texttt{Question}, \texttt{Request}, \texttt{Challenge}\}\}$ which contains the commitments required to her by the others.

Syntax	Meaning	Commitment	Example
`Assert`:x, $x \in \mathscr{L}$	inform that x holds	sender should defend x against opposite argument	"this article can be published"
`Argue`:a, $a \in Arg(\mathscr{L})$	support a claim by an argument	sender should defend it against attacks	"articles revealing private information cannot be published, hence this article cannot be published."
`Declare`:x, $x \in \mathscr{L}$	make x hold	sender should have the right to do it (*i.e.*, the right "role")	"John and Mary are husband and wife"
`Question`:x, $x \in \mathscr{L}$	ask if x holds	receiver should answer (give an argument for x or for $\neg x$)	"John and Mary are married" (?)
`Request`:x, $x \in \mathscr{L}$	ask to make x hold	receiver should act to make x hold	"a_2 is paid" (when a_2 asks for being paid)
`Challenge`:x, $x \in \mathscr{L}$	ask for an explanation (argument) for x	receiver should present an argument for x	"this article can be published" (why?)
`Promise`:x, $x \in \mathscr{L}$	commit oneself to make x true in the future	sender should do it (one day)	"a_2 is paid" (for saying that a_2 will be paid)

Table 5.1: *The speech acts and their commitments*

A commitment store is supposed to be *empty* at the beginning of a dialogue. Then, each move uttered during a dialogue is stored in a commitment store except the move retract. Indeed, this last does not commit neither its sender nor its receiver to anything.

We have introduced a function `PROP` that computes the set of formulas representing the state of the world according to what has been uttered during the dialogue. Note that Questions, Challenges and Requests are not considered in the definition of this function since they don't describe the state of the world while formulas that appear in assertions and arguments are directly taken into account. However, things are different with the formulas related to a move `Declare`. Indeed, by definition, after `Declare`:x the world evolves in such a way that x becomes true. Consequently, one has to *update* the whole set of propositions previously uttered (in this work we did not precise the update operator to be used this had been left outside the scope of this framework).

Now, it is natural to associate with each commitment a penalty that sanctions agents when the commitment is violated. The need of *cumulating* sanctions when several violations have occurred is a reason for using a penalty based framework which is built on *additivity*. For this purpose we have adapted the penalty logic framework, that we had proposed in [102] for handling inconsistency in knowledge bases: the penalties associated to the violated commitments in the CS were assumed to depend only on the corresponding speech act. Each speech act in $\mathcal{S}$ is supposed to have a *cost* which is a strictly positive integer or the infinity: $\texttt{Cost} : \mathcal{S} \longmapsto \mathbb{N}^\star \cup \{+\infty\}$. This captures the idea that some speech acts are more important than others. For instance, violating a promise

may be more costly than not answering a question. Since a commitment store is empty at the beginning of a dialogue, its initial penalty is equal to 0. In [11], we have described the conditions of violations of the different speech acts, together with the place where they are stored and their associated penalty.

For instance, an `Assert` move is violated if it is possible to build an argument whose conclusion is opposed to it from the set of propositions uttered by the agent, *i.e.*, if the agent is self-contradictory. When, a commitment is fulfilled or withdrawn the penalty of the commitment store decreases.

Example 19 *Let us consider the following dialog (assuming that the moves are allowed by a given protocol), we give the evolution of* CS_1*:*
Give me a nail please: $(a_2,\{a_1\},\textit{Request},a2n)$

A_1	O_1
$\varnothing$	`Request`:$a2n$

$c(CS_1) = \texttt{Cost}(\textit{Request})$

(where $a2n$ *stands for "*a_2 *can have the nail")*
No. : $(a_1,\{a_2\},\textit{Assert},\neg a2n)$

A_1	O_1
`Assert`:$\neg a2n$	`Request`:$a2n$

$c(CS_1) = 0$

Why not?: $(a_2,\{a_1\},\textit{Challenge},\neg a2n)$

A_1	O_1
`Assert`:$\neg a2n$	`Request`:$a2n$ `Challenge`:$\neg a2n$

$c(CS_1) = \texttt{Cost}(\texttt{Challenge})$

Because I want to hang a mirror (hm) and thus I need this nail (nn). I cannot give you a nail if I need it.: $(a_1,\{a_2\},\texttt{Argue},(\{hm,\ hm \rightarrow nn,\ nn \rightarrow \neg a2n\},\ \neg a2n))$

A_1	O_1
`Assert`:$\neg a2n$ `Argue`:$(\{hm,\ hm \rightarrow nn,\ nn \rightarrow \neg a2n\},\ \neg a2n)$	`Request`:$a2n$ `Challenge`:$\neg a2n$

$c(CS_1) = 0$

In this dialogue the agent a_1 has an exemplary behavior since after each move, the penalties associated with its commitment store are canceled. It means that a_1 does not contradict itself (regarding the properties she has used in assertions and arguments), and that a_1 has answered to all the requests (negatively but she did it) and to the challenge made by a_2.

We have shown that the proposed semantics satisfies some desirable properties. Namely, the semantics sanctions only bad behaviors of agents, and any bad behavior is sanctioned *i.e.*, if the commitment store has a strictly positive cost then it means that there is a violated move in the commitments, and conversely if there is a violated move then the commitment store will have a strictly positive cost. An important result is the fact that if the total penalty of part A_i is null then all the stated information is consistent.

Based on the formalization of the notion *independence* of propositional formulas[1] by Lang et al. in [141] we have defined a notion of Independence between two moves

[1]Let φ, ϕ be two propositional formulas, and Σ a set of formulas, φ is *new* for ϕ w.r.t. Σ iff:

- $\exists\ (S,\phi) \in \texttt{Arg}(\Sigma \cup \{\varphi\})$ and $(S,\phi) \notin \texttt{Arg}(\Sigma)$, or

given a dialogue. Roughly speaking, two moves are independent if the formulas in the content of the first move are independent from the ones of the second move given all the propositions already uttered. This notion allows us to capture the fact that sometimes the content of a move maybe of no interest with respect to the current state of knowledge when this speech act is answering a question or a challenge. It also allowed us to show that the violation status does not change when an independent move is uttered (provided that this move is not a `Declare` since in that case the update operator has to be, what we called, "Independence compatible"). We have also shown that if two formulas are independent w.r.t. the formulas of a commitment store, then the penalty of two moves conveying these formulas is decomposable.

The contribution of this approach can be summarized as follows: we have clarified the origin of each commitment induced from a speech act. We have proposed a new semantics in terms of commitments associated to violation penalties. All the violation criteria are based on what has been exchanged (and not on the knowledge bases of agents). This makes the semantics *verifiable*. Contrarily to existing social semantics that focus only on speech acts isolated from the context of the dialog, our semantics is defined on the basis of moves uttered during a dialog which ensures that it is *practical*. Note that in order to add a new speech act, one needs simply to define a new violation criterion and a penalty associated with it.

With our ACL, one does not need to specify the different moves allowed after each move in the protocol itself. Agents only need to minimize the penalty to pay at the end of the dialog. This give birth to very flexible protocols and consequently, the agent's strategies become very rich. Besides, the notion of penalty may play a key role in defining agent's *reputation* and *trust* degrees. It is clear that an agent that pays a lot of penalties during dialogues may lose its credibility, and will no longer be trusted. Examining more deeply penalties can help to figure out agents profiles: cooperative agent, consistent agent, thoughtful agent (*i.e.*, agent that respects its promises)...

5.b Persuasion dialogs with Enthymemes and limited time

[93] F. Dupin de Saint-Cyr. Handling enthymemes in time-limited persuasion dialogs. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 149–162. Springer-Verlag, 2011
[92] F. Dupin de Saint-Cyr. A first attempt to allow enthymemes in persuasion dialogs. In *DEXA International Workshop: Data, Logic and Inconsistency (DALI)*, pages 332–336. IEEE Computer Society - Conference Publishing Services, 2011
[25] J. Balax, F. Dupin de Saint-Cyr, and D. Villard. DebateWEL: An interface for Debating With Enthymemes and Logical formulas. In *European Conference on Logics in Artificial Intelligence (JELIA)*, volume 7519 of *Lecture Notes in Computer Science*, pages 476–479. Springer, 2012

As said in introduction of this chapter, persuasion dialog models have already been widely developed in the literature but as far as I know the dialog persuasion systems that

- $\exists\ (S, \neg\phi) \in \texttt{Arg}(\Sigma \cup \{\varphi\})$ and $(S, \neg\phi) \notin \texttt{Arg}(\Sigma)$

φ is said to be *independent* from ϕ w.r.t. Σ otherwise.

have been developed either did not define what is an argument or were always assuming that an argument is a "perfect" minimal proof of a formula, no formal persuasion dialog system able to handle enthymemes had yet been defined at the time of my work on this subject. This is why in [93, 92, 25] my purpose was to develop a dialog system in which it is possible to use "approximate argument" (as defined in Section 3.b) hence to take into account implicit information. Indeed in enthymeme handling, it may be interesting to focus on what is missing. We have seen in Example 16 that some stratagems given by Schopenhauer for taking victory in a dialog are based on the use of enthymemes.

The ACL that I have defined in this study, is specific for handling enthymemes in a *persuasion* dialog involving *two agents*, denoted a and $\overline{a}$. The specificity comes from the facts that imperfect arguments should be dealt differently: the hearer can agree: it means that he recognizes the link, or can ask for a completion in order to see the link and so on. Hence the speech acts that were allowed differ from the one of the previous section. I have considered a set of eleven speech acts (they are described precisely with their three associated effects (locutionary, illocutionary and perlucotionary [23]) in [93]). Although some speech acts are "assertive" (according to Searle [174]) namely Assert and Argue, as in the previous Section we claim that they are "commissives" in the sense that they commit the utterer to avoid to contradict them. While Close is clearly a "declarative" speech act (for ending the dialog), it is less obvious for Retract, Dismantle (that retract an argument) since they are not only "declarative" but also "assertive" (because the retracted formulas or dismantled arguments correspond to assertion of the form "I assert neither φ nor $\neg\varphi$" "I assert neither that S is a valid proof for φ nor that it is not") and "commissive" (since they are assertive).

The commitments induced by the different speech acts during a dialog are transcribed into a commitment store (CS) as previously. This structure is different since we had to store (approximate) arguments separately from formulas since, because of their imperfection, approximate arguments are not expressible in terms of formulas. Moreover, I have added a common knowledge store that contains all the mutually agreed contents of moves (see an example on Figure 5.1). It gives a tuple $(F_a, A_a, R_a, F^\circ, A^\circ, F_{\overline{a}}, A_{\overline{a}}, R_{\overline{a}})$ representing respectively, formulas asserted by a, arguments said by a, requests towards a, common formulas, common arguments, formulas asserted by $\overline{a}$, arguments said by $\overline{a}$ and requests toward $\overline{a}$.

In [93], I have described the effects and preconditions of each move done by agent a towards the other agent denoted $\overline{a}$. For instance, the move Argue is used with a content which is an approximate argument: $\langle S, \varphi\rangle$. The agent a has the right to do it if the following preconditions hold: $\langle S, \varphi\rangle \notin A^\circ \cup A_a \cup A_{\overline{a}}$ and $S \cup \{\varphi\} \cup F_a \cup \mathsf{form}(A_a) \cup F^\circ$ consistent. It means that the argument should not have already been uttered either by a or $\overline{a}$ (hence it should not belong to common knowledge arguments nor to the set of not yet agreed arguments uttered by a or $\overline{a}$). Moreover the formulas of this argument should be consistent with the formulas that have been used by a or that are in the common knowledge.

The post-conditions of this move are: $\langle S, \varphi\rangle \in A_a$ and (Challenge φ) $\notin R_a$ and (Agree $\langle S, \varphi\rangle$) $\in R_{\overline{a}}$. It means that the effects of doing this Argue move are commitments to a

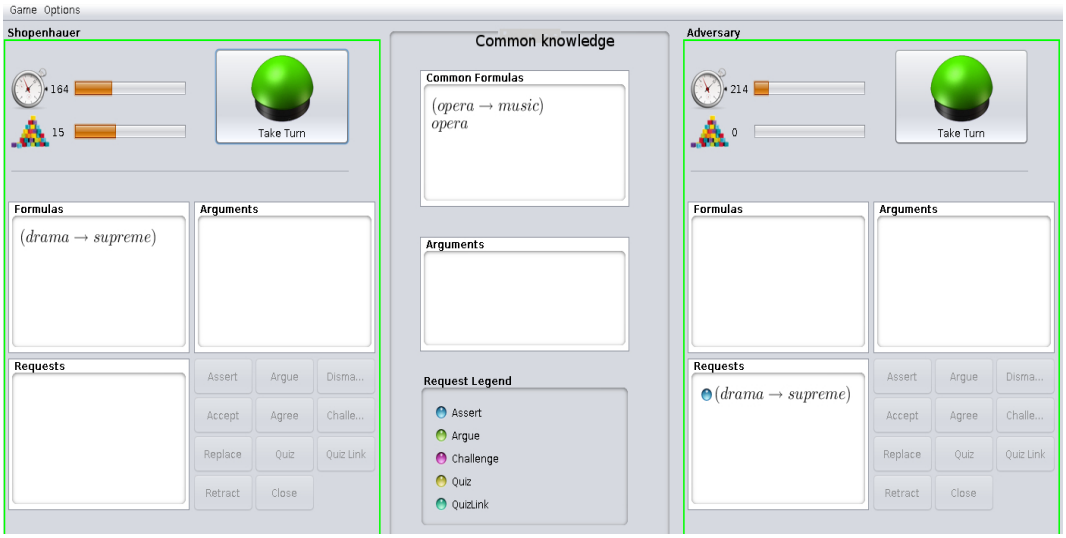


Figure 5.1: *Commitment store (screen shot of the tool described in [25])*

and $\overline{a}$, namely this argument is added to the set of arguments of a (a is then committed to not contradict it in the future unless a dismantles it). If there was a current Challenge done by $\overline{a}$ on the formula φ then a is no more committed to answer to it. $\overline{a}$ is now committed to Agree with this argument before the end of the dialog, unless it manages to make a dismantle it.

Another example is the Quiz move: Quiz(S, φ) can be done by an agent a only if there is no logical argument completing (S, φ) that can be built from the common knowledge and the formulas already asserted by a. In other words, the agent cannot understand the argument a (at least in what she has said, nothing shows that the agent can do it).

After defining the ACL we have defined a protocol which is a Boolean function that checks if a move is acceptable at a given stage of the dialog. We have proposed to define this function on the basis on the content of the commitment store. The protocol defines what is a persuasion dialog wrt an initial common knowledge (F, A) (where F and A are possibly empty sets of formulas and approximate arguments assumed to be consistent). In our definition, it is a sequence of moves such that there is a sequence of states of the CS that have good properties wrt to this sequence of moves. For instance at start $CS_1 = (\varnothing, \varnothing, \varnothing, F, A, \varnothing, \varnothing, \varnothing)$, and at each step, the preconditions of the move m_i should hold in CS_i and the post-conditions are applied to CS_i in order to obtain a new state CS_{i+1} except if the move is Close (which should be allowed in this stage, *i.e.*, it requires that all the commitments of the agent are fulfilled) and that the other agent has already closed his participation to the dialog. If these "ending conditions" are not possible then the dialog has no end.

Example 16 (continued): *Let us consider the following persuasion sub-dialog*

$$D = \left(\begin{array}{lll} (Schopenhauer, & \mathsf{Assert}, & d \rightarrow s &), \\ (Adversary, & \mathsf{Argue}, & \alpha_1 = \langle \{m \rightarrow \neg s\}, o \rightarrow \neg s \rangle &), \\ (Schopenhauer, & \mathsf{Argue}, & \alpha_2 = \langle \{d \leftrightarrow t \vee s\}, m \rightarrow \neg d \rangle &), \\ (Adversary, & \mathsf{Agree}, & \alpha_2 &) \end{array} \right)$$

Suppose that common knowledge is the following: $F^{\circ} = \{o \rightarrow m, o\}$ *meaning that "opera is music" and that "opera exists". Table 5.2 describes the commitment stores of each participants.*

After these moves the dialog is not finished since two requests are not yet answered. Schopenhauer has to options either (1) to agree with α_1 *(since it is consistent with common knowledge) then he would have no more commitments and his adversary will be obliged either to accept the first claim or to provide another argument against it or (2) he may ask his adversary to precise the link that argument* α_1 *has with the formulas already asserted. In that case the adversary would not be able to* Replace *his argument since the logical argument that completes* α_1 *and related to the subject is* c *(*$\langle \{m \rightarrow \neg s, o \rightarrow m, o, o \rightarrow d\}, \neg(d \rightarrow s) \rangle$*) whose support is now inconsistent with the common knowledge (see Table 5.2).*

Since a persuasion dialog may be infinite, in [93], we have introduced a particular persuasion dialog where the speaking time is restricted. This notion has required to define the duration of the moves (all moves were associated to a duration of 1 except for Assert, Argue, Replace in which the size of their content was taken into account). Now, the time-limited persuasion dialog wrt to an initial common knowledge (F, A) and a total duration T, is a variant of a persuasion dialog where the CS are equipped with two integers for counting the remaining time of each agent, the starting condition is then: $CS_1 = (\varnothing, \varnothing, \varnothing, T, F, A, \varnothing, \varnothing, \varnothing, T)$ and at each step the duration is taken into account by checking in the precondition if the duration of the moves does not overlap the remaining time of the agent, the CS after a move is updated in a way that the remaining time of the agent is decreased by the duration of the move he has just uttered. The termination condition for the dialog has changed because it can stop because the agents have no more speaking time.

We have shown that a time-limited persuasion dialog is finite. This last property is important due to enthymemes, the requests for completion of arguments can sometimes be infinite depending on common knowledge... We have shown that when a dialog is closed by the two participants then they have fulfilled all their commitments. We have also shown that common knowledge is increasing after the persuasion dialog and can be used as initial common knowledge for future dialogs.

Our proposal was a first attempt to handle enthymemes in persuasion dialogs. The ambition was to handle incomplete information both in the premises and in the claim of an argument. The latter is more difficult to handle and has required to introduce a new speech act Quizlink allowing to ask for an insight about what is hiding behind the claim. In some cases, one may agree with an argument that is not related with the subject but when he understands the underlying implication he wants to reject it.

After the third move

Schopenhauer			Common knowledge		Adversary		
Form.	Args	Requests	Form. (F°)	Args (A°)	Form.	Args	Requests
$d \rightarrow s$	α_2	(Agree α_1)	$o \rightarrow m$ o			α_1	(Accept $d \rightarrow s$) (Agree α_2)

After the fourth move

$d \rightarrow s$		(Agree α_1)	$o \rightarrow m$ o $d \leftrightarrow t \vee s$ $m \rightarrow \neg d$	α_2		α_1	(Accept $d \rightarrow s$)

If the move $(Schopenhauer, \mathsf{Quizlink}, \alpha_1)$ is done
Then the move $(Adversary, \mathsf{Dismantle}, \alpha_1)$ should be done, leading to:

$d \rightarrow s$			$o \rightarrow m$ o $d \leftrightarrow t \vee s$ $m \rightarrow \neg d$	α_2			(Accept $d \rightarrow s$)

If the Adversary has no other argument linked with the subject, then he is forced to do the move $(Adversary, \mathsf{Accept}, d \rightarrow s)$ in order to be authorized to close the dialog:

			$o \rightarrow m$ o $d \leftrightarrow t \vee s$ $m \rightarrow \neg d$ $d \rightarrow s$	α_2			
Schopenhauer			Common knowledge		Adversary		

Table 5.2: *Commitments stores of Schopenhauer and his Adversary*

In our two proposals of ACL we have only represented what is publicly uttered, since we have considered that we do not have access to the agent's mind. This way to apprehend the public statements is also done for instance by [118], a public utterance is called "grounded" in their framework. Their approach allows to deal with inconsistent assertions (which is not allowed in our second framework)considering that it is up to the other agent to detect and denounce inconsistency by asking to its adversary to "resolve" it. Dealing with possible inconsistent assertions is a challenge for further developments of our second approach, however we could argue that what is public should be consistent in order to be civilized and respectful of the audience and of the debate quality.

During the dialogs the public utterances are stored and may evolve when arguments are retracted or replaced, in my opinion it is a more natural view of argumentation than the classical "static view of argument" introduced by Dung (see Section 6.a), in which an argument is not allowed to be changed by its utterer when it is attacked since all arguments are kept together with the attacks among them.

A very appealing development of this framework concerns the strategical part, we

plan to translate our protocol rules into the Game player project language GDL2 [186], indeed in GDL2 it is possible to handle games with imperfect information. After this translations strategies coming from game theory and strategies dedicated to dialog games (*e.g.* [17]) could be compared. Moreover, an important perspective would be to take into account the duration of the moves for choosing the strategy to achieve the persuasion goal.

5.c Persuasion Dialog quality

[14] L. Amgoud and F. Dupin de Saint-Cyr. On the quality of persuasion dialogs. *Studies in Logic, Grammar and Rhetoric, Argument and Computation*, 23(36):69–98, 2011

[13] L. Amgoud and F. Dupin de Saint-Cyr. Extracting the core of a persuasion dialog to evaluate its quality. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, volume LNAI 5590, pages 59–70. Springer-Verlag, 2009

[11] L. Amgoud and F. Dupin de Saint-Cyr. A new semantics for ACL based on commitments and penalties. *International Journal of Intelligent Systems*, 23(3):286–312, 2008

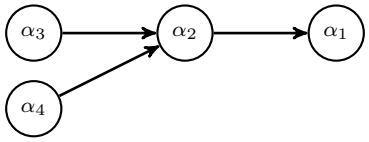
In [14, 13, 11] we have investigated objective criteria for analyzing *already generated logic-based argumentative persuasion dialogs* whatever the protocol and the strategies that are used. We place ourselves in the role of an external observer who tries to evaluate a dialog, we have proposed three points of view: measures that evaluate the quality of exchanged *arguments*, measures that analyze the *behavior* of each participating agent, measures of the *global properties of the dialog* itself.

In what follows, a persuasion dialog is considered as an exchange of logic-based *arguments*[2] between two or more agents. We had assumed that each agent involved in a dialog recognizes any logic-based argument of $Arg(\mathscr{L})$ and any attack in $\mathcal{R}_{\mathscr{L}}$ (an unspecified binary relation s.t. $\mathcal{R}_{\mathscr{L}} \subseteq Arg(\mathscr{L}) \times Arg(\mathscr{L})$). This assumption does not mean that each agent is aware of all the arguments. But, it means that *agents use the same logical language and the same definitions of argument and attack relation.* The *subject* of such a dialog is an argument and its *aim* is to determine the status of that argument. Since only arguments are exchanged, it means that the speech act is systematically an Argue, hence the moves are reduced to triple $\langle s, R, \alpha \rangle$ where s and R are respectively the sender and the Receiver as in Section 5.a, and α is a logic-based argument ($\alpha \in Arg(\mathscr{L})$) referred as the content of the move ($\texttt{Content}(m)$). Formally, the kind of persuasion dialog D studied here is a finite[3] sequence of n moves: $\langle m_1, \ldots, m_n \rangle$. built under a given protocol. A sub-dialog of D is a sub-sequence $\langle m_1, \ldots, m_i \rangle$, $i \leq n$. In the framework studied, an argumentation system ($\texttt{AS}_D$) is associated to D in order to evaluate the status of its subject (which is the first argument uttered: $\texttt{Subject}(D) = \texttt{Content}(m_1)$) under the grounded extension. $\texttt{AS}_D$ is a pair $(\texttt{Args}(D), \mathcal{R}(D))$ where $\texttt{Args}(D)$ is the set of arguments uttered and $\mathcal{R}(D)$ the attacks between them according to $\mathcal{R}_{\mathscr{L}}$.

[2]Note that in [18], other kinds of moves (like questions, assertions) may be exchanged in a persuasion dialog it is also the case in our proposal with enthymemes (see Section 5.b).

[3]We assume that the dialog D is finite, this assumption is not too strong since a main property of any protocol is the termination of the dialogs [187].

Example 20 *Let D_1 be a dialog between two agents a_1 and a_2 with $D_1 = \langle\langle a_1, \{a_2\}, \alpha_1\rangle$, $\langle a_2, \{a_1\}, \alpha_2\rangle$, $\langle a_1, \{a_2\}, \alpha_3\rangle$, $\langle a_1, \{a_2\}, \alpha_4\rangle$, $\langle a_2, \{a_1\}, \alpha_1\rangle\rangle$. The subject of D_1 is the argument α_1. Let us assume the following attacks among some of these arguments.*



Thus, $\texttt{Args}(D_1) = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ *and* $\mathcal{R}(D_1) = \{(\alpha_2, \alpha_1), (\alpha_3, \alpha_2), (\alpha_4, \alpha_2)\}$.

The *output* of a dialog is the status of the argument under discussion (*i.e.*, the subject). In example 20, the grounded extension of $\texttt{AS}_{D_1}$ is the set $\{\alpha_1, \alpha_3, \alpha_4\}$. Thus, the output is an acceptation.

During a dialog, agents utter arguments that may have different *weights*. A weight may highlight the *quality of information involved in the argument* in terms, for instance, of certainty degree. It may also be related to the cost of revealing an information. In [6], several definitions of arguments' weights have been proposed, and their use for comparing arguments has been studied. It is worth noticing that the same argument may not have the same weight from one agent to another. In what follows, a weight in terms of a numerical value is associated to each argument. The greater this value is, the better the argument.

$$\texttt{weight} : Arg(\mathscr{L}) \longrightarrow \mathbb{N}^*$$

On the basis of arguments' weights, it is possible to compute the weight of a dialog D as follows:$\texttt{Weight}(D) = \sum_{\alpha \in \texttt{Args}(D)} \texttt{weight}(\alpha)$ and the contribution of an agent to a dialog: $\texttt{Contr}(\texttt{a}_\texttt{i}, \texttt{D})$ of an agent a_i to the dialog D is the weight of what she has said over the global weight of the dialog.

It is clear that the `Weight` measure is (non strictly) monotonic wrt dialog increasing[4] contrarily to `Contr`. However the contribution of the agent who will present the next move will never decrease.

Example 20 (continued): $D_1^{a_1} = \{\alpha_1, \alpha_3, \alpha_4\}$ *and* $D_1^{a_2} = \{\alpha_2\}$. *Suppose that an external agent who wants to analyze this dialog assigns the following weights to arguments:* $\texttt{weight}(\alpha_1) = 1$, $\texttt{weight}(\alpha_2) = 4$, $\texttt{weight}(\alpha_3) = 2$ *and* $\texttt{weight}(\alpha_4) = 3$. *The contributions of the two agents are respectively* $\texttt{Contr}(a_1, D_1) = 6/10$ *and* $\texttt{Contr}(a_2, D_1) = 4/10$.

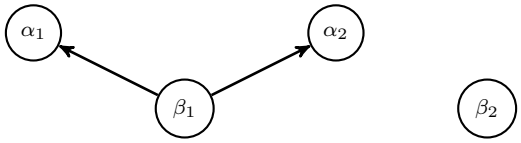
The *behavior of an agent* in a given persuasion dialog may be analyzed on the basis of three main criteria: i) its degree of *aggressiveness* in the dialog, ii) the *source of its arguments*, *i.e.*, whether it builds arguments using its own formulas, or rather the ones revealed by other agents, and finally iii) its degree of *coherence* in the dialog.

The first criterion, *i.e.*, the aggressiveness $\texttt{Agr}(a_i, a_j, D)$ of an agent a_i in a dialog D against agent a_j, measures to what extent an agent was attacking arguments sent by

[4]Note that due to the definition of the weight of a dialog, if an agent says several times the same argument it is only counted once.

the other agent(s), *i.e.*, the number of arguments attacking the ones of the other agents wrt to its total number of arguments uttered. An aggressive agent prefers to destroy arguments presented by other parties rather than presenting arguments supporting her own independent point of view.

Example 21 *Let D_2 be a persuasion dialog between the agents a_1 and a_2. Assume that* $\texttt{Args}(D_2) = \{\alpha_1, \alpha_2, \beta_1, \beta_2\}$, $D_2^{a_1} = \{\alpha_1, \alpha_2\}$, $D_2^{a_2} = \{\beta_1, \beta_2\}$ *and the conflicts are depicted in the figure below.*



The aggressiveness degrees are $\texttt{Agr}(a_1, a_2, D_2) = 0$ *and* $\texttt{Agr}(a_2, a_1, D_2) = 1/2$.

The second criterion concerns the way arguments are built either from the agent's own knowledge base, or by using formulas revealed by other agents. In [17], Leila Amgoud and Nicolas Maudet have argued that it is interesting to turn out an agent's argument against itself in order to weaken its position, it minimizes the risk of being attacked subsequently. The *degree of loan* $\texttt{Loan}(a_i, a_j, D)$ of an agent a_i wrt agent a_j is tho ratio of the formula owned by a_j that have been used by a_i over all the formulas it has used, where a formula is owned by an agent if it is revealed *for the first time* by that agent.

The third criterion concerns the coherence of an agent. There are two kinds of self contradiction: *explicit* contradiction when an agent presents an argument and a counter-argument in the same dialog, and an *implicit* contradiction appearing in a "complete" version of the agent a_i's argumentation system denoted $\mathsf{CAS}(D^{a_i})$. This complete argumentation system takes into account not only the set of arguments which are explicitly expressed in a dialog by an agent, *i.e.*, $\texttt{Args}(D^{a_i})$, but also all the arguments that may be built from the set of formulas involved in the arguments of $\texttt{Args}(D^{a_i})$. Due to the monotonic construction of arguments, for any set A of arguments, $A \subseteq \text{Arg}(\texttt{Formulas}(A))$ but the reverse is not necessarily true. We have defined a *measure of incoherence* of an agent in a dialog as the ratio of the number of effective attacks inside its arguments in her CAS over the Cartesian product of these arguments.

We have shown that if an agent is aggressive towards itself, then it is incoherent but the converse is not always true (since aggressiveness is not computed on the complete argumentation system). Similarly, we showed that if agent a_i is aggressive towards agent a_j and if all the formulas of a_i are borrowed from a_j, then a_j is for sure incoherent. Note that incoherence is not necessarily a bad behavior, it depends on the aim of the participants: the goal may either be to win the debate whatever the other says or to discuss and take into account new information. In the last case, changing its opinion is a self-contradiction but may be a constructive attitude.

It is very common that a dialog contains redundancies or useless moves since agents may deviate from the subject of the dialog. Thus, only some arguments may be useful

for computing the output of the dialog. We have characterized the useful moves in a dialog by identifying the *ideal* version of a dialog. Formally, if there exists a path from the argument presented by the agent towards the argument representing the subject in the graph of the argumentation system associated to the dialog, in that case the move is said *relevant*. If the path is a directed one the move is said *useful*: useful moves are those that have a direct influence on the status of the subject.

Example 21 (continued): *Assume that* $\texttt{Subject}(D_2) = \alpha_1$. *It is clear that* α_2, β_1 *are relevant while* β_2 *is not and* β_1 *is useful while* α_2 *is not.*

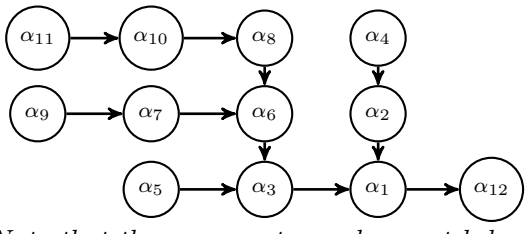
We defined a measure, called $\texttt{Relevance}(D)$, that computes the percentage of moves that are relevant in a dialog D. In Example 21, $\texttt{Relevance}(D) = 3/4$. It is clear that the greater this degree is, the better the dialog. When the relevance degree of a dialog is equal to 1, this means that agents did not deviate from the subject.

Inspired by works on proof procedures [7] that were proposed in the argumentation theory in order to check whether an argument is accepted or not, we have computed and characterized a sub-dialog, called *ideal*, of the original one that is concise. The closer a dialog is to its ideal sub-dialog, the better is its quality. In order to compute an ideal sub-dialog, we build a tree, called Dialog tree denoted by D^t, which is a finite tree with the subject of the persuasion dialog as root and the branches are all the possible dialog branches that can be built from D. A dialog branch is a kind of partial sub-graph of $\texttt{AS}_D$ in which the nodes contain arguments and the arcs represent inverted attacks. Note that arguments that appear at even levels are not allowed to be repeated. Moreover, these even levels arguments should attack (without being attacked by) the preceding argument. Such a branch should be maximal.

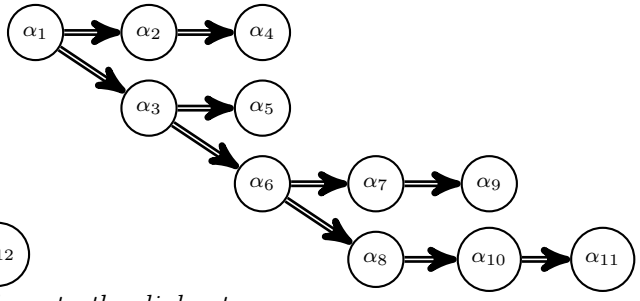
We have shown that each persuasion dialog has exactly one corresponding dialog tree and that the status of the subject of the original persuasion dialog D is exactly the same in both argumentation systems $\texttt{AS}_D$ and $\texttt{AS}_{D^t}$ (where $\texttt{AS}_{D^t}$ is the argumentation system whose arguments are all the arguments that appear in the dialog tree D^t and whose attacks are obtained by inverting the arcs between those arguments in D^t).

Example 22

Let us consider D_3 *whose subject is* α_1 *and whose graph is the following:*



The dialog tree associated to this dialog is:



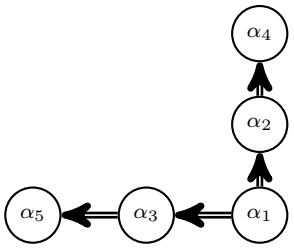
Note that the argument α_{12} *does not belong to the dialog tree.*

In order to compute the status of the subject of a dialog, we can consider the dialog tree as an And/Or tree. This distinction between nodes is due to the fact that under

the grounded extension, an argument is accepted if it can be defended against all its attackers. A dialog tree can be decomposed into one or several trees called *canonical trees* that are sub-trees of D^t with root $\texttt{Subject}(D)$ and which contains all the arcs starting from an even node and exactly one arc starting from an odd node.

We have shown that a canonical tree which branches are all of even-length is sufficient to reach the same outcome as the original dialog in case the subject is accepted, the smallest of these canonical trees (in terms of number of nodes) is called *ideal*. When the subject is rejected, the whole dialog tree is necessary to ensure the outcome, this tree is the *ideal* tree in that case.

Example 22 (continued):
The subject α_1 of dialog D_3 is accepted since there is a canonical tree whose branches are of even length (it is the canonical tree on the right in the next Example). It can also be checked that α_1 is in the grounded extension $\{\alpha_1, \alpha_4, \alpha_5, \alpha_8, \alpha_9, \alpha_{11}\}$ of $\texttt{AS}_{D_3}$.
The dialog D_3 has the canonical tree shown on the right which is ideal.



Note that the ideal dialog exists but is not always unique (in Example 20 two ideal dialogs can be defined either with arguments $\{\alpha_3, \alpha_2, \alpha_1\}$ or $\{\alpha_4, \alpha_2, \alpha_1\}$). It is clear that the closer (in terms of set-inclusion of the exchanged arguments) the dialog from its ideal version, the better the dialog.

Generally speaking, protocols are the high level rules that govern a dialog, these rules should ensure "correct" dialogs, *i.e.*, dialogs that terminate and reach their goals. However, they do not say anything on the quality of the dialogs. In our papers [14, 13, 11], we have argued that there are criteria for measuring this quality. Indeed, under the same protocol, different dialogs on the same subject may be generated, and some of them may be judged better than others. We have given three kinds of reasons for such a judgement (arguments used, agent behaviors and conciseness of the dialog) each one has been translated into a quality measure.

From our results, it seems natural that a protocol penalizes irrelevant and not useful moves (until there is a set of arguments that relate them to the subject). Note that in our proposal, the only thing that matters in order to obtain a conclusion is the final set of interactions between the exchanged arguments. But the criteria of being relevant to the previous move or at least to a move not too far in the dialog sequence could be taken into account for analyzing dialog quality.

Furthermore, it may be the case that from the set of formulas involved in a set of arguments, new arguments may be built. This gives birth to a new set of arguments and to a new set of attack relations called complete argumentation system CAS associated with a dialog. Hence, it could be interesting to define dialog trees on the basis of the CAS. However, some arguments of the CAS may require formulas owned by other agents, it would mean that an ideal but practicable dialog would require the agents to utter their arguments in an efficient sequence (each agent should be able to build each argument at

each step). This address a kind of collaborative planning problem.

5.d Axioms for persuasion dialogs compared with reasoning

[15] L. Amgoud and F. Dupin de Saint-Cyr. An axiomatic approach for persuasion dialogs. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 618–625. IEEE Computer Society, novembre 2013

An increasing number of systems supporting *public persuasion dialogs*, including some that we have developed, are using (abstract or structured) argumentation theories that were initially developed for non-monotonic reasoning. In our study published in [15], we have considered logic-based instantiations of Dung's framework [90], namely those based on deductive logics [4][5] and we have proposed some basic postulates for persuasion dialogs. We have shown that these postulates are incompatible with the ones proposed for non-monotonic reasoning [3].

In this section, we refer to a dialog system by $\mathcal{D}S$. For the purpose of our study, and without loss of generality, we had focused on persuasion dialogs between *two* agents P and C. Each of them is equipped with a knowledge base Σ_k (with $k \in \{P, C\}$) and an argumentation system $\mathcal{T}_k = (\mathcal{A}_k, \mathcal{R}_k)$. In order to stay in a general setting we consider arguments grounded on a Tarskian logic (based on a pair $(\mathscr{L}, \text{CN})$ where CN is a consequence relation) which is the language shared by the two agents. In the considered argumentation systems, arguments are evaluated using any Dung's semantics [90] (see Section 3.a Definition 19), the set of extensions of an argumentation system $\mathcal{T}$ is denoted $\texttt{Ext}(\mathcal{T})$.

A persuasion dialog is a valid sequence of moves based on a set of speech acts $\mathscr{S}$, *i.e.*, a sequence that satisfies the rules of a given protocol (unspecified in this study). The only restriction on $\mathscr{S}$ is that it should contain at least two kinds of speech acts, namely "Argue" whose content is an argument and "Assert" whose content is a formula of $\mathscr{L}$. Apart from the agents argumentation systems, a third argumentation system is associated to each persuasion dialog in order to evaluate the exchanged arguments (with Argue moves). Defining the corresponding attack relation is more tricky since the agents may use different relations (for instance, P may use the "undercut relation" [162] whereas C "assumption attack" [109]). They may also choose distinct semantics for the evaluation of arguments. In what follows, we assume the existence of a third relation denoted $\mathcal{R}$ which results from a merging of the two relations $\mathcal{R}_P$ and $\mathcal{R}_C$ using an operator $\otimes$ not specified in this paper. Thus, $\mathcal{R} = \mathcal{R}_P \otimes \mathcal{R}_C$. More precisely, the *argumentation system associated* with a dialog $\mathcal{D}$ contains the arguments $\texttt{Args}(\mathcal{D})$ that were either uttered by Argue moves, or that are arguments $\{(\{x\}, x)$ built from $Assert : x$ moves. The attacks $\mathcal{R}(\mathcal{D})$ are defined by $\mathcal{R}$ restricted to $\texttt{Args}(\mathcal{D})$. The outcome of a persuasion dialog $\mathcal{D}$ is the status of its subject wrt $\texttt{AS}_\mathcal{D} = (\texttt{Args}(\mathcal{D}), \mathcal{R}(\mathcal{D}))$ *i.e.*, $\texttt{Output}(\texttt{AS}_\mathcal{D})$.

We propose a minimum number of postulates that any persuasion dialog system should satisfy. The first postulate concerns the finiteness of the generated dialogs. This

[5] Note that the results of our study hold also in case of rule-based systems.

requirement is already known in the literature. In [131], protocols should ensure termination. Here, we require finiteness not only for the number of moves but also for the content of each move. For instance, it is not allowed to assert $x \wedge x \wedge \ldots$.

Finiteness: For all persuasion dialog $\mathcal{D}$ generated by a dialog system $\mathcal{DS}$, $size(\mathcal{D}) \in \mathbb{N}$ where $size(\mathcal{D}) = \sum_{m \in \mathcal{D}} sizemove(m)$ with $sizemove(m)$ is the number of atoms used in the content of m plus 1.

The second important postulate concerns the formalism that is used for computing the outcomes of dialogs. In our context, Dung's system should ensure sound results. Namely, extensions (under any semantics) represent various positions in a dialog. Thus, they should be coherent. This leads to a consistency postulate.

Consistency: For all persuasion dialog $\mathcal{D}$ generated by a dialog system $\mathcal{DS}$, for all $\mathcal{E} \in \texttt{Ext}(\texttt{AS}_{\mathcal{D}})$, $\{\texttt{Conc}(a) \mid a \in \mathcal{E}\}$ is consistent.

In persuasion dialogs, agents try to convince other parties to accept some assertion by putting forward arguments. These latter are intended to justify the assertion. Thus, it is unacceptable to justify an assertion by itself. Consequently atomic arguments are forbidden. Similarly, tautologies are not allowed in dialogs.

Non triviality: For all persuasion dialog $\mathcal{D}$ generated by a dialog system $\mathcal{DS}$, for all $\alpha \in \{\texttt{Content}(m) \mid m \in D, \texttt{Act}(m) = \text{Argue}\}$, α is not atomic and $\texttt{Conc}(\alpha)$ is not a tautology (*i.e.*, $\texttt{Conc}(\alpha) \notin \text{CN}(\varnothing)$).

The aim behind building systems for persuasion dialogs is to automate such dialogs and to conduct efficient ones. These systems should capture as much as possible natural dialogs, it is thus important for a dialog system to capture the two classical forms of attacks, namely "rebuttal" and "assumption attacks" (see Definition 25). The following postulate ensures this by constraining the attack relation $\mathcal{R}$.

Expressivity: For all persuasion dialog $\mathcal{D}$ generated by a dialog system $\mathcal{DS}$, for all $\alpha, \beta \in \texttt{Args}(\mathcal{D})$, if α rebuts β then $(\alpha, \beta) \in \mathcal{R}$, and if α assumption-attacks β then $(\alpha, \beta) \in \mathcal{R}$.

The next postulate is also about expressive power since for any non-trivial subject, it constrains the dialog system to be able to generate at least one dialog in which this subject is accepted and one dialog in which it is not.

Non-determinism: For all formula $x \in \mathscr{L}$, s.t. $x \notin \text{CN}(\varnothing)$ and $\text{CN}(\{x\}) \neq \mathscr{L}$, there exist at least two dialogs $\mathcal{D}_1$ and $\mathcal{D}_2$ generated by a dialog system $\mathcal{DS}$, such that $\texttt{Subject}(\mathcal{D}_1) = \texttt{Subject}(\mathcal{D}_2) = x$ and $\texttt{Output}(\mathcal{D}_1) \neq \texttt{Output}(\mathcal{D}_2)$.

Note that if the set of non-trivial formulas of $\mathscr{L}$ (*i.e.*, without considering tautologies and contradictions) is infinite then any dialog system satisfying non-determinism can generate an infinite number of dialogs. Another requirement that seems important for a dialog system is to allow dissimulation, indeed each agent should be able to dissimulate information.

Dissimulation: For any agent k, such that Σ_k contains at least two non trivial distinct formulas, then for any $x \in \Sigma_k$ s.t. $x \notin \text{CN}(\varnothing)$, there exists a dialog $\mathcal{D}$ generated by a dialog system $\mathcal{DS}$, such that $x \notin \texttt{Output}(\texttt{AS}(\mathcal{D}))$.

We have shown that the persuasion dialog postulates can be satisfied all together by a dialog system with an exception: consistency is not compatible with expressivity. But this exception is due to Dung's framework which cannot be instantiated by symmetric attack relations [4]. Thus, the non compatibility of consistency and expressivity does not mean that the two postulates are not required for dialog systems.

We have investigated whether the set of postulates defined for reasoning systems are suitable for dialog systems and vice versa. Indeed, a set of postulates for argumentative reasoning systems was proposed in [3]. The postulates can be satisfied all together. The first one ensures that each extension supports consistent conclusions. The second postulate concerns the closure of its output under the consequence operator CN. The third postulate concerns *sub-arguments* (see Section 3.b). It ensures that the acceptance of an argument should imply also the acceptance of all its sub-parts.

Consistency : Let $\mathcal{T} = (\mathcal{A}, \mathcal{R})$ be an AS over a base Σ. For all $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$, $\bigcup_{\alpha \in \mathcal{E}} \{\texttt{Conc}(\alpha)\}$ is consistent.

Closure under CN : Let $\mathcal{T} = (\mathcal{A}, \mathcal{R})$ be an AS over a base Σ. For all $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$, $\bigcup_{\alpha \in \mathcal{E}} \{\texttt{Conc}(\alpha)\} = \text{CN}(\bigcup_{\alpha \in \mathcal{E}} \{\texttt{Conc}(\alpha)\})$.

Closure under sub-arguments : Let $\mathcal{T} = (\mathcal{A}, \mathcal{R})$ be an AS over a base Σ. For all $\mathcal{E} \in \texttt{Ext}(\mathcal{T})$, if $\alpha \in \mathcal{E}$, then $\texttt{Sub}(\alpha) \subseteq \mathcal{E}$.

We have shown that the three postulates of the reasoning system cannot be satisfied by a dialog system since in this latter the set of exchanged arguments is not complete (due to the finiteness of dialogs and also to the fact that in dialogs, some arguments are considered as trivial and thus do not need to be exchanged). Similarly, we have shown that four postulates of the dialog system cannot be satisfied by the argumentation system. Moreover, we have established that the outcome of a dialog system can be different from the outcome that should be obtained by a reasoning system that would use a knowledge base containing all the formulas exchanged during the dialog. Note that our study holds for any other logic-based instantiation of the abstract framework of Dung [90], like ASPIC system [71].

Since early nineties, there is an increasing number of works trying to formalize dialogs in which agents may exchange arguments. Persuasion and negotiation dialogs have received particular attention from AI community. Several systems were developed for each of them. In those systems arguments are exchanged in order to support *claims* in persuasion dialogs and *offers* in a negotiation context. The arguments are then evaluated using "classic" argumentation systems that were originally developed for non-monotonic reasoning or for reasoning about inconsistent information.

Our study has revealed that a dialog system needs particular argumentation systems for evaluating its outcomes. Those systems should obey the nature of dialog. This work

can be extended in different ways. The first one consists of defining argumentation systems that are more suitable for public persuasion dialogs and that ensure the postulates discussed in this paper. Another future work consists of defining new postulates for dialogs, namely for capturing manipulation in dialogs.

In this chapter we have seen several aspect of dialog handling, the language, the protocol, the quality of the dialog, the axioms that are required for dialog systems. In all these works it appears that Dung's framework is not well adapted. In the following chapter we have proposed to use a new kind of argumentation system (already presented in Section 6.b) in a framework of collective decision making. This system could be seen as a very particular dialog where the agents are only allowed to speak once and at the same time, this is not a dialog but a kind of vote in which we wanted to lower the risk of manipulation...

Part III

Work in progress and projects

In this part, I am exposing my current subjects of work and also the long term perspectives that seems appealing for me. They are gathered in the next chapter.

I start by explaining why I don't see much interest in going on working on Dung's abstract argumentation theory.

Then I focus on a new structure called BLA on which I am currently working with Romain Guillaume. This structure is designed for argumentative group decision making, it offers many perspectives that I describe briefly as for instance group preferences elicitation.

Third, I describe a new approach that belongs to my long term project to go beyond classical rationality and incorporate more human-being practical abilities for reasoning and deciding in our systems. This approach developped with Pierre Bisquert and Madalina Croitoru is an attempt to encode S1-S2 reasoning of Tversky and Kahneman.

The notion of association that is used in S1-reasoning deserve a deep study in order to be used for aprehending the reasoning about analogical arguments.

Chapter 6

Perspectives

6.a Why Dung's framework does not suit me ?

After several years of work on abstract argumentation, I am more and more convinced that practical intuition is lacking behind the framework introduced by Dung. Indeed all Dung's theory about how to reason with arguments [90] was introduced in order to encompass both non-monotonic reasoning and logic programming in a general theory, and the two applications provided by Dung were far from argumentation namely n-person games imputations and the stable marriage problem. The theory is based exclusively on a graph structure that represents directed attacks between arguments. The fact that graphs are structures that computer scientists appreciate a lot is not sufficient to justify the use of Dung's theory as the unique way to see argumentation. Indeed, in my opinion, Dung's theory is not related *intuitively* to everyday life argumentation. Here is a more precise list of critics:

- There is no definition of what is an argument, an abstract argument is simply a vertex in a graph. Attacks between arguments are neither constrained. The lack of constraints on these definitions is problematic since some information is missing to understand and validate arguments and their relationship. It seems important to be able to say that some entities are *not* arguments and that the attacks should be related to the nature of the arguments involved in it. In Dung's framework, any graph is acceptable for representing relations between arguments (while symmetric attacks, cycles, self-attacked arguments may translate very different situations that should be dealt with care).

- Moreover the idea that arguing does not modify the arguments previously uttered and the related attacks is far from real argumentation where after a counter-argument, the attacked argument is often either amended/precised by its utterer or removed (these completions or retractions were the subject of my proposal of persuasion dialog protocol with enthymemes see Section 5.b). Hence in real debates no argument remains attacked: either it disappears or it is corrected and the attack disappears. In that sense, Dung's framework could be seen as a "static view

of argumentation", not in the sense that nothing might be added, but in the sense that flawed arguments stay in the system and cannot be improved.

- Inference in abstract argumentation is defined by the selection of a set of vertices in the graph, called extensions, that together have good properties. This notion of group of arguments as well as the defense notion (which allows to consider as accepted an argument that has been attacked but such that each of its attackers are attacked) is related to the "static view of argumentation". The defense notion is debatable since if the argument has been attacked it means that something is wrong with it, unless the attack does not hold. The only case considered is that the attack may not hold (since there is no other information about the argument). But the attack being unrelated to other information, what is assumed in this framework is that the attack does not hold if the attacker is not acceptable (leading to a recursive definition of defense).

- Dung's framework has already been criticized in the particular case of logic-based arguments with attack relations built from the logic-based information inside the arguments. Leila Amgoud and Philippe Besnard have shown [5] that "stable, semi-stable and preferred semantics either lead to counter-intuitive results or provide no added value w.r.t. naive semantics" (where only non-attacked arguments are considered as acceptable). They have also established that "ideal and grounded semantics either coincide and generalize the free consequence relation developed by Benferhat, Dubois and Prade in 1997, or return arbitrary results. Consequently, Dung's framework seems problematic when applied over deductive logical formalisms".

- More generally, I think that Dung's framework is not only badly adapted for logic-based arguments but also very difficult to implement in general, *i.e.*, it is very difficult to give a meaning to the vertices of the graph and to the arrows in terms of natural argumentation (with both the ideas of conflict and preference) in order that the set of accepted arguments according to Dung mean something intuitive (except with the naive semantics). During Pierre Bisquert thesis we have even tried to find an implementation outside the scope of argumentation, *e.g.* arguments are allocation pairs and attacks are preferences between conflicting allocations. In this implementation of Dung's framework, the intuitive meaning of an extension is not very salient but may nevertheless be considered. Indeed, the grounded extension corresponds to a strict envy-free part of a possible affectation, the preferred extensions are partial envy-free affectations and the stable extensions are partial affectation such that any modification of an allocation in this affectation is strictly worst. So, a direction of research could be to find useful implementations of Dung's framework outside argumentation, following the ideas of Dung himself in his paper.

All these drawbacks (including the result stating that using Dung's framework in Dialog is not suitable see Section 5.d) led me to work on different proposals aiming at capturing natural argumentation. I have already made a first proposal for dealing with persuasion dialogs with enthymemes (which was outside the scope of Dung's framework

see Section 5.b). I am currently working with Romain Guillaume on formalizing precisely the notion of argument, attack, validity of an argument in the context of decision making. Moreover, when speaking of natural argumentation, it comes immediately to mind that the syntactic aspect of what is said is not the only thing that impacts the argument acceptation. This is why together with Pierre Bisquert and Madalina Croitoru, we have proposed (see Section 6.d.1) to take extra-information into account such as *e.g.* the source, the cognitive engagement of the receiver, etc.

In those proposals, the idea to consider more precise definitions of argument enables us to evaluate the arguments individually, bypassing the concept of "defense".

6.b Arguments for decision making

[28] F. Bannay and R. Guillaume. Towards a transparent deliberation protocol inspired from supply chain collaborative planning. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*. Springer, juillet 2014
[29] F. Bannay and R. Guillaume. Qualitative deliberation based on bipolar leveled sets of arguments under incomplete distributed knowledge. *under submission to JAIR*, 2015

The fact that an argumentation system is a structure that can encode generic information is used here with a completely distinct point of view. Indeed in this new setting, generic information is encoded under the form of an argument viewed as an association between a reason and the goal that is achieved when this reason holds. And in this model, factual information is a way to validate the arguments that are applicable in some given context.

The structure that gathers these arguments is called a BLA (Bipolar Leveled set of Arguments) and is dedicated to decision making, this framework was proposed with Romain Guillaume in [28, 29]. We have been able to give a clear and well-founded semantics of a *decision argument*. Indeed as seen in the introduction of this part, a first aim is to be able to give a clear meaning to arguments and attacks and to integrate these definitions in the process to decide if the argument is acceptable or not.

Given a set $\mathscr{C}$ of candidates[1] about which some knowledge is available, an argument is viewed as *a reason for believing that, by default, a given goal can be achieved* by selecting a candidate c. This relation between beliefs and preferences (in terms of goals) comes from the fact that in decision problem, arguments should encode a kind of expected utility notion. More precisely, we have considered two distinct languages $\mathscr{L}_F$ (a propositional language based on a vocabulary $\mathscr{V}_F$) representing information about some features that are believed to hold for a candidate and $\mathscr{L}_G$ (based on a distinct vocabulary $\mathscr{V}_G$) representing information about the achievement of some goals when a candidate is selected. The idea to have distinct languages is both to differentiate beliefs and desires for sake of clarity but also for the purpose of avoiding manipulation since one language will be used by the voters while the other will not be accessible during the vote. A *decision argument* α is a pair (φ, g) where $\varphi \in \mathscr{L}_F$ and $g \in LIT_G$.

Example 23 *If the candidates are people applying for a job then the argument (unmotivated, not efficient for the job) could be understood as "if the candidate is unmotivated*

[1] Candidates are also called alternatives in the decision literature.

then a priori the goal to have an efficient person for the job is failed".

A BLA gathers the set of possible arguments which may be used in order to evaluate the admissibility of a given candidate. Moreover, the BLA is structured in order to represent some key notions that are involved in argumentative decision making, namely beliefs and goals (which are inside the arguments), levels of importance of the arguments, polarities wrt the goal and attacks.

- the level $l(\alpha)$ of each argument α depends both on the importance of the goal and on the "credibility" of the argument, This *level is supposed to be given* by the decision makers. Note that the precise value of the level of an argument is not meaningful, *only the rank ordering of the levels* is taken into account.

- the polarity $pol(\alpha)$ depends only on the goal ($\texttt{Conc}(\alpha)$), if the fact that the achievement of the goal is either wished, then $pol(\alpha) = \oplus$, or dreaded, denoted $pol(\alpha) = \ominus$. It is also *supposed to be given* for each goal in LIT_G.

- the attack relation $\mathcal{R}$ is defined between two *conflicting* arguments (*i.e.*, arguments with opposite goals) of the same level. The attack is directed from the argument whose conclusion holds when both reasons are present. This *direction is supposed to be given* by the decision makers. Indeed it requires extra-information: either one argument α attacks the other one β (when agents have agreed that when $K \vdash \texttt{Supp}(\alpha) \wedge \texttt{Supp}(\beta)$ the goal $\texttt{Conc}(\alpha)$ is achieved where K will represent the common knowledge about features that hold) or there is a symmetric attack (when the agents have agreed that when $K \vdash \texttt{Supp}(\alpha) \wedge \texttt{Supp}(\beta)$ we don't know which goal among $\{\texttt{Conc}(\alpha), \texttt{Conc}(\beta)\}$ is achieved. Note that the latter is a case where the arguments destroy each other.

Example 24 *Figure 6.1 illustrates the BLA corresponding to a recruitment example where the features are: eb (educational background), gp (good personality), i (introverted candidate), jhop (job hopper), lpe (long professional experience), spe (professional experience in the specialty of the job) and u (unmotivated candidate). The goals are: ap (anti-social personality), ej: (efficient for the job), et (easy to train), st (a stable person).*

We can explain the attack from $(u, \neg ej)$ to (eb, ej) and (pe, ej) by the fact that when a candidate is unmotivated, even if she has a good professional experience or a good educational background, the goal "efficient for the job" will not be achieved by hiring this candidate.

Specificity may be used to justify the attack between $(jhop \wedge \neg spe \wedge lpe, et)$ and $(jhop \wedge lpe, \neg et)$ since a job hopper that has a long professional experience is generally not easy to train but if this experience was not in the specialty of the job, it means that this person has a good adaptability hence will be easy to train.

The realization of the goal of an argument (which could be viewed as similar to the "admissibility of an argument" in abstract argumentation theory) is possible only if *this argument is not attacked* (it corresponds to the so-called "naive" semantics).

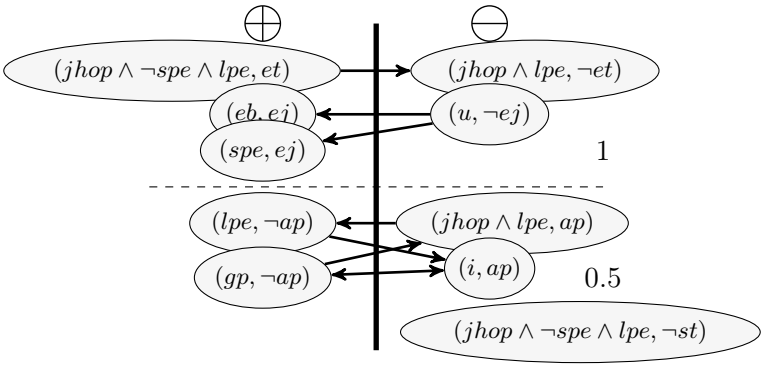


Figure 6.1: *Recruitment BLA*

The BLA defines all possible information required to make a decision. Unfortunately, in an uncertain context, only little information maybe available. In this section we propose a method for analyzing the acceptability of a candidate w.r.t. a BLA. First, we present the available information and the notion of instantiated BLA, called valid BLA. Then, we define thresholds for acceptability and study their relations with classical decision rules of qualitative bipolar decision making.

The BLA structure is meant to be a common language for enabling several agents to make a consensual decision about a candidate to choose, it is used as a kind of qualitative collective utility function for group decision making. Once this BLA is established the decision makers can use it in order to check if a candidate is acceptable. The main interest of the structure is that it differentiates clearly the goals to achieve from the practical facts concerning a candidate. Indeed a decision maker is only allowed to use the language of facts hence to describe features that the candidate possesses but he is not allowed to express her opinion about this candidate. The common opinion should be derived from the BLA.

More precisely, given a candidate $c \in \mathscr{C}$, and a *consistent* knowledge base K_c, $K_c \subseteq \mathscr{L}_F$, representing the knowledge of a decision maker about the candidate c, and given a formula φ describing a configuration of features ($\varphi \in \mathscr{L}_F$), the decision maker has three possibilities, φ holds for candidate c (denoted $K_c \vdash \varphi$), or not (denoted $K_c \vdash (\neg\varphi)$ or the feature φ is unknown for c (denoted $K_c \nvdash \varphi$ and $K_c \nvdash \neg\varphi$).

6.b.1 Validity of arguments, admissibility of candidates

From a knowledge bate K we have defined the validity of an argument, simply by the fact that this argument can be used in the context (since its premise holds): an argument $a = (\varphi, g)$ is valid according to K if $K \vdash \varphi$. A valid BLA consists of a restriction of a BLA to its valid arguments.

Example 25 *Let us consider that we know that a candidate has the features eb, lpe and jhop (see Example 24), the Valid BLA is represented in Figure 6.2.*

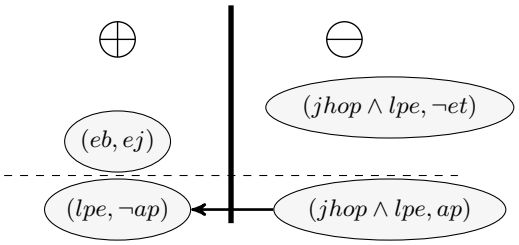


Figure 6.2: *A Valid BLA*

We have simply defined the notion of *realized goal* as the conclusion of a valid argument not attacked by any other valid argument. This means that the only thing to check is the existence of attacks on arguments: the defense notion is not useful for our purpose.

Example 26 *According to the valid BLA of Example 25, the goal ej is realized since there is only one argument, namely (eb, ej), concerning this goal in the Valid BLA. Similarly, the goal $\neg et$ is realized, hence et is failed. The goal ap is also realized since the argument $(jhop \wedge lpe, ap)$ attacks the argument $(lpe, \neg ap)$. The goal ph is not realized since there is no argument concerning this goal that is valid (i.e., whose reasons are known) in our context.*

From the status of a goal, we have defined an admissibility status for a candidate $c \in \mathscr{C}$ given a Valid BLA $\langle \mathcal{A}, l, pol, \mathcal{R} \rangle$ for this candidate. For instance, a *necessary admissible* candidate has positive arguments with goals of maximum importance that are realized (*i.e.*, unattacked) and all the negative goals of the same importance are failed. A *possibly admissible* candidate has at least one unattacked positive argument of maximum importance. An *indifferent candidate* has no unattacked arguments (positive or negative), while a *controversial candidate* is both supported and criticized by unattacked arguments of maximum importance.

Example 27 *The candidate described by the arguments given in the valid BLA of Example 25 is controversial, since at the most important level we have both a realized positive goal, namely ej and a realized negative goal: $\neg et$.*

We have shown that these admissibility status are rational wrt the decision rules defined by [57], since an inadmissible candidate cannot be preferred to an admissible candidate wrt these decision rules.

6.b.2 Possibilistic reading of a BLA

We have proposed a *possibilistic reading* of a BLA that justifies this structure in the qualitative decision theory by using a *possibility measure* [87], Π on the beliefs language $\mathscr{L}_F$ and a *degree* $\mu(x) \in [0, 1]$ (where 1 is the most important level) associated to the goals. These two measures have allowed us to define formally the arguments, the levels and the attacks. Indeed an argument is viewed as a kind of default rule saying that by

default when the reason holds the goal is realized, (*i.e.*, an argument (φ, g) is encoded in a possibilistic setting [84] by: $\Pi(\varphi \wedge g) \geq \Pi(\varphi \wedge \neg g)$). Moreover the level of an argument depends on the possibility of observing φ and g that hold together (denoted $\Pi(\varphi \wedge g)$) and on the importance of the goal (denoted $\mu(g)$). More precisely, it can be interpreted as a risk/opportunity level since it is the aggregation of a chance measure (here the possibility measure) and a utility measure (here a qualitative utility degree μ).

The attack definition depends on extra-information which can also be based on a possibility measure. Intuitively α attacks β if they are conflicting and at the same level, and when the reasons of α and β hold together, the goal of α is more plausibly realized than the one of β. More precisely the attack $\alpha R \beta$ with $\alpha = (\varphi_\alpha, g)$ and $\beta = (\varphi_\beta, \neg g)$ when $l(\alpha) = l(\beta)$, can be interpreted as: $\Pi(\varphi_\alpha \wedge \varphi_\beta \wedge g) \geq \Pi(\varphi_\alpha \wedge \varphi_\beta \wedge \neg g)$. We also interpret a path in the BLA as a constraint linking all the reasons present in the arguments of the path with the realization of the goal of the first argument.

Note that the users are not at all obliged to precise possibility degrees or utilities, but once the BLA is given (levels of arguments and attack relations) it is possible to estimate these measures, hence a BLA formalizes how human beings integrates a belief measure with a preference degree in a given context.

6.b.3 Group decision with a BLA

We have shown that some multi-criteria decision situations (such as the existence of veto argument, or of compensation between several arguments compared to only one) can be captured by specific BLAs.

Finally, we have shown that in a context of multi-agent decision the BLA is well protected against manipulation. This was done by proposing two basic strategies in which the agents do not necessarily reveal all the information they know depending on their private opinion about the candidate. It appeared that the use of strategies cannot lead to make accepted a candidate that would have been rejected if all the information had been revealed *i.e.*, decision makers cannot betrayed the consensual properties and goals expressed by the BLA.

6.b.4 Related work

Our proposal of decision argument is not a logic-based argument since there is no explicit deductive link between the reason and the goal. Those arguments could be viewed as enthymemes (see Section 3.b) but with the precision that the support and the conclusion are of different nature, namely based respectively on beliefs and preferences. Moreover the decision argument is itself a more or less objective link between the reason and the goal realized: this gives more freedom to the decision makers for building the arguments, hence we do not impose a complex definition based on a value-based transition system like in the work of Black and Atkinson [54] (in which they impose a deductive link between conditions, actions and goals).

In the domain of multi-criteria decision, Dubois and Fargier [82, 83] have proposed a qualitative bipolar approach in which a candidate is associated with two distinct sets

of positive "arguments" (pros) and negative "arguments" (cons). The arguments are not structured they are just labelled as positive or negative w.r.t. the decision goal without attack relation between them. A function π assesses the level of importance of each argument for the decision maker. In the same domain, Amgoud and Prade [20] have proposed a bipolar argumentation-based approach for decision and distinguish two types of arguments: the epistemic arguments and the practical arguments. The epistemic arguments are used to deal with inconsistent knowledge and the practical arguments [166] are in favor or against a decision. Practical argumentation [201, 8, 194] consists in answering the question "what is the right thing to do in a given situation". This question is clearly related to a decision problem and several works are using argumentative approaches to tackle it: for instance Bonet and Geffner [56] have a very similar view of what we call arguments, since they use defeasible rules in favor or against a given action. However, in this kind of approach there is no attack relation defined between two practical arguments, while we only consider these attacks in our framework and are not interested in dealing with inconsistency problems.

Our need to have arguments for making a decision on the ground of factual reasons could be related to the notion of argument for reasoning about actions and values of Fox and Parsons [113]. In their approach an argument is a triplet of the form (claim, ground, value) the claim can be either a sentence or an action and the value can correspond to a confidence degree or an expected utility. The authors define inference mechanism on this kind of triplets according to the nature of the claim and value. While this process seems very interesting to study, our aim is less ambitious, namely we do not handle inconsistency, nor actions, but we focus on interactions between conflicting decisions about the same goal and on how to decide if a goal will be achieved when some facts are true.

6.c Elicitation of Preferences and Beliefs of a Group

To sum up, a BLA is a new framework for decision making under incomplete and distributed knowledge. The BLA is established between voters before the vote and it describes the priorities among goals, the importance level of arguments and the contradictions among them. The voters will give the features corresponding to the current candidate, then the BLA will be instantiated and it will automatically lead to an admissibility decision determined by the instantiated BLA (by checking the non attacked arguments in favor or against the candidate).

This framework can be used by only one human agent in order to decide whether a candidate is admissible. This agent can use the BLA to clarify and express its criteria and then to decide accordingly. But the BLA demonstrates the full extent of its usefulness in the case where knowledge is distributed over several agents who have personal preferences but who want to collaborate in order to make a good decision for the group. The group of human agents can vote by giving the features that concern the candidate, by a simultaneous vote.

A first benefit of the BLA is its visual aspect allowing to be easy to read and to

create, a second benefit is that it provides a neutral process to compute a group decision.

In the following Sections we present several directions of future work.

6.c.1 Zoom/Unzoom on Multi-layered BLA

[104] F. Dupin de Saint-Cyr and S. Loiseau. Alignement cognitif de symboles. In *Journées Francophones Modèles Formels de l'Interaction (MFI)*, pages 249–254. Cépaduès Editions, mai 2003

The zoom/unzoom reasoning is grounded on the fact that visualization is very important: the ability to access to a global and focused view is an interesting feature that I had study in a preliminary work about cognitive alignment of symbols [104, 72]. This study could be extended in order to build efficient and interactive tools for visualizing and manipulating arguments.

It would be interesting to consider a more complex structure, namely a multi-layered BLA, on which it would be possible to zoom/unzoom on the features (thanks to the use of an ontology) or on the links (due to defeasibility properties for instance) or on the goals (a filtering procedure could show only the goals without their reasons, and their conflicts relations and levels, it could also be possible to integrate a decomposition into sub-goals).

Another visualization of a BLA could be done by projecting it according to the different domains to which the goals or reasons are referring to, it may underline independent parts and may help to decompose the decision according to separate independent fields. Moreover we could introduce a filtering operation which could delete some arguments concerning either information that is not allowed or goals that are not relevant. This kind of operation should be studied in order to determine whether deleting some argument may induce changes upon the levels or the attacks: this means to be able to update/revise BLAs according to different points of view.

6.c.2 Interactive Elicitation

As said with the possibilistic reading of a BLA, knowing how people rank order decision arguments and how they solve the conflicts between those arguments may enable us to compute associated utility and plausibility functions. My project is to provide tools to allow people to play with the levels and attacks of a BLA in order to show them what are the hided meaning of their choices and conversely if they want to impose some utility values to some features, the tool could show how it impacts the decisions.

This is why I would like to work on the development of softwares for handling the creation/modification of a BLA, the vote about a candidate, and the decision process. We could start from tools already existing in the argumentative decision support technology. Indeed, in this domain, many interactive argumentation platforms have been developed : Walton and Reed's platform [197] called "Araucaria", Baldwin and Price's "Debategraph" [27], van Gelder's [190] "Rationale" are examples of those tools. Karacapidilis and Pappis in [135] have proposed a platform called "Hermes" that allows a group of users to converge towards a common specification of the solution of the decision problem in terms of criteria

and ideal solution. Then the system is able to apply similarity measures in order to provide the candidate that is closer to the ideal solution. Introne and Iandoli [129] have used the PENDO platform to demonstrate that an argumentative formulation is more user-friendly than mathematical formulation and that the decision-making performance is enhanced by the use of an argumentative tool.

Hence, the studies in decision support technology domain are well in accordance with our intuition, since a BLA is a visual representation of arguments. However our definitions are done in a restricted formal language with a clear semantics, this is not the case in those works that allow for natural language arguments that maybe related by several kinds of links (more or less easy to establish or validate). Nevertheless, the techniques that may help a group to communicate and converge towards a common general knowledge about the decision to make, could be very useful for constructing the initial BLA.

6.c.3 From individual beliefs and preferences to collective BLAs

Building a BLA with a group of people is a complex task since it involves a collaborative aspect. I would like to study protocols that would help people to express their own preferences, together with their preference as group members (as in the study of social ties done by Frédéric Moisan [153]).

Moreover, two directions maybe explored considering the aggregation of individual beliefs and preferences in order to create a collective BLA:

- either aggregate beliefs and preferences at an individual level in order to create an individual BLA and then merge them in a common BLA
- or merge the beliefs and aggregate the preferences independently then create the common BLA.

The fusion and aggregation mechanisms are to be defined. Moreover, in order to design those mechanisms, we could have a look in the decision support system area, where the fact that the knowledge is incomplete and distributed is well apprehended. A recent proposal by Ouerdane & al. [158] could also help us to build a collective BLA, indeed a BLA may correspond to their defeasible "cognitive artifact" (*i.e.*, the formulation of the decision problem and the evaluation model). This approach can be viewed as a protocol to build a BLA, it is an iterative non-monotonic process encoded in a logic-based argumentation framework based on the proposal of Kakas and Moraitis [133]. This framework is based on several argumentation schemes (among which the one that says that a claim holds when enough supportive reasons can be provided and no exceptionally strong negative reason is known, this kind of idea is related to our definition of "realized" goal). Hence the collective decision model is built by exchanging arguments (that are not of the same nature as our decisive arguments), this process may be a starting point for my own project.

Moreover, since the BLA is a kind of common knowledge, an idea of protocol could involve two-person persuasion debates in which the result is the common agreed knowledge

as in Section 5.b. It would be interesting to study how to organize a series of two-person debates inside a group of people. The way those debates are organize *e.g.* the structure of the graph of debates (balanced tree, or "comb" structure...) may completely change the final resulting common knowledge. This study could take into account the credulous or skeptical profile of the agents, *i.e.*, the will to accept argument. This kind of research is related to the social choice theory domain.

How does this collective building relates to fairness division ? I would like to study this aspect with researchers in social choice theory and also with psychologists in order to build practical experiments on an argumentation support software, this could provide benchmarks that will be helpful for studying strategies, manipulation properties, and fairness feelings.

Note that in our initial vision of a BLA (coming from the collaborative spirit of supply chain management domain), the BLA should represent the aggregation of the individual preferences of the voters when they forget their individual interests *i.e.*, it should result in aggregating the *individual preferences as group member of each voter*. However non-cooperative attitudes are interesting to study: for instance how to take into account the fact that some individual preferences or some beliefs may not be expressible since they are taboo or confidential. Another aspect is the importance of arguments, we could take inspiration from the work of Bonzon et al. [59]. They propose to consider what they call "coalition of arguments" which are set of arguments that can be valid together. They provide some preferences over arguments and attacks and they compute a Shapley value according to the potential set in which the argument can be. This measure represents the potential impact of the argument. It could be interesting to integrate this kind of impact measure in our work, either for building a BLA or for choosing to reveal or not some information that can activate an argument.

6.d Decrypting persuasion

6.d.1 Appreciative Argument Evaluation

[53] P. Bisquert, M. Croitoru, and F. Dupin de Saint-Cyr. Towards a dual process cognitive model for argument evaluation. In *SUM*, 2015. under submission
[52] P. Bisquert, M. Croitoru, and F. Dupin de Saint-Cyr. Four ways to evaluate arguments according to agent's engagement. In *Brain Intelligence*, 2015. under submission

A natural claim is that the syntactic content of arguments are not the only thing to take into account in order to understand human persuasion. The intended meaning, the personality of the speaker, her eloquence etc. have an impact. Moreover in order to capture the fact that some syntactic information are implicit it is important to see the syntactic content of an argument as a kind of default rule (which can also implicitly refer to other default information).

I have started to work on this subject with Pierre Bisquert and Madalina Croitoru [53, 52]. More precisely we have proposed to formalize the evaluation process of an argument by a human agent. We have focused on appreciative arguments: it is an argument that

contains the expression of an agent opinion about something. An appreciation is a pair (formula, flag) where the flag is in $\{+, -, ?\}$ meaning favorable, negative and neutral respectively. An appreciative argument is a quadruplet $(s, h, w, (c, f))$ representing the source s of the argument, its support h that is divided into beliefs and appreciations, an optional warrant w (which is a kind of rule that allows to infer new opinions from old opinions and beliefs, it is called an a-rule), and a conclusion which is an appreciation (c, f).

We have proposed two systems for evaluating arguments, the system is chosen according to the cognitive availability of the hearer with respect to the subject of the argument. Our two systems are inspired by the highly influential work of Tversky and Kahnemann [188], the first system (called S1) deals with quick and instinctive thoughts and is based on associations such as cause-effect, resemblance, valence, etc. The second system (called S2) is used as little as possible and is a slow and conscious process that deals with what we commonly call reason.

We have used a hash-table in order to encode the S1 system, the entries are formulas and the cells contains an appreciative flag and a stack of formulas (such that the top element corresponds to what comes immediately to mind). Reasoning with such a hash-table simply consists in following the associative links until a frank opinion is obtained about the formula.

We have formalized the rational reasoning by using a default logic (contextual entailment [41]) for the belief part and a unification method for the opinion part. More precisely, the agent has a knowledge base which contains on the one hand a set of default rules and a set of facts for the beliefs, and on the other hand a set of appreciation pairs for representing its opinions, together with a set of a-rules. Evaluating rationally an argument amounts to check if the knowledge expressed in the support is believable (thanks to a default reasoning), and if the appreciations expressed in the argument are compatible with what can be obtained by the a-rules of the knowledge base, and finally if the conclusion of the argument is well related to the premise part thanks to a unification with the (maybe implicit) warrant.

We have proposed four level of engagement: unconcerned, quiescent, engaged, enthusiastic: the more the agent is engaged in the evaluation of the argument the more the agent uses a rational reasoning instead of an associative reasoning. We plan to integrate more graduality hence to provide a mechanism that integrates the associative reasoning S1 into the rational system S2. Furthermore a more general extension could gradually integrate an even more rational system, that could provide meta-inferences about the inference system itself, generate new heuristics or even new inference mechanisms...

The logic that we have proposed for encoding the S2 system has to be further studied, it is related both to decision theory and practical reasoning and more research is required in order to see if some proposal already exist for handling our "a-rules".

6.d.2 Analogical arguments

I plan to work on analogical arguments, I would like to build a system for decoding analogical arguments. My idea is to use the notion of substitution to check if by substituting

analogous objects no inconsistency occur in the knowledge base. More precisely in the analogy $a : b :: c : d$, expressing that a is to b what c is to d, in order to check if this analogy holds it would be interesting to see what happens when substituting a to c and b to d in the knowledge base, the stronger the inconsistency it triggers the weaker the analogy.

Moreover there is a kind of enthymemes in the field of analogical arguments, since sometimes they are incomplete. For instance the comparison may be done purposely on something that is commonly rejected, hence the implicit information is that the argument is about rejecting something.

Some incomplete analogies are called metaphors, like "Federer is the Mozart of tennis", they imply two computations: first to complete the analogy then to check if it is correct. Moreover some analogies are more accurate than others, it seems that there is a graduality in the fit of analogies, I would like to explore how this kind of imprecision of analogies can be captured in a defeasible setting.

Incomplete analogies when there are not poetic maybe used as satire or more generally as jokes. Hence studying analogical arguments may be a first step towards understanding and modeling laughter. Laughter is another intelligent ability since it requires inference mechanisms and also implies the laugher to adopt an appropriate distance wrt the subject of the joke.

Conclusion

Let us come back to the control loop of a rational agent proposed by Wooldridge and already evoked in Introduction. This time we give Wooldridge's elaborated version [201] (see Algorithm 2). In this version, after the determination of a plan to achieve her goals, the agent executes its actions sequentially and after each of them, she evaluates their impact and can deliberate again about her intentions (depending on a function called `reconsider` which evaluates roughly the necessity to reconsider the goals). This algorithm is a good overview of the tasks to develop in order to have a better representation of human reasoning, which is important for understanding and predicting human behavior and thus for designing artificial rational agents that could help human beings to reason and decide.

```
1  B := B_0 ;                                          /* initial beliefs */
2  I := I_0 ;                  /* initial intentions (options to achieve) */
3  while true do
4  |  get_next_percept(p);
5  |  B := belief_integrate(B, p) ;                    /* current beliefs */
6  |  D := options(B, I) ;                           /* options generation */
7  |  I := filter(B, D, I) ;                /* choice of options to achieve */
8  |  π :=plan(B, I) ; /* compute a sequence of actions for achieving I */
9  |  while not (empty (π) or succeeded (I, B) or impossible (I, B)) do
10 |  |  a := pop(π) ;                     /* unstack the first action of π */
11 |  |  execute(a);
12 |  |  get_next_percept(p);
13 |  |  B := belief_integrate(B, p);
14 |  |  if reconsider(I, B) then D := options (B, I); I := filter (B, D, I);
15 |  |  if not sound(π, I, B) then π := plan (B, I);
16 |  end
17 end
```

Algorithm 2: Rational Agent Elaborated Control Loop

The first line concerns the initialisation of beliefs. I have explored this aspect when dealing with common knowledge and generic information: it is the subject of Chapter 1. The second line is about intentions (that are viewed as preferences on states of the

world). I am not an expert of preference representation but I have started recently to work on it, namely for decision making (in which preferences are integrated to beliefs see Section 6.b) and when dealing with appreciative arguments (in Section 6.d.1).

I have already worked a lot on the function `belief_integrate`, which relates to belief revision and update (see Chapter 2 and Chapter 4 in the framework of argumentation). As alluded in footnote 1, intelligence can be seen as an ability to "perceive well" the world, *i.e.*, to direct our perception in order to obtain the most relevant information for our purpose. This is why the function `get_next_percept` is worth mentioning, because it can be directed towards a better "comprehension" of the world. It can be done by taking into account more parameters, which is what I intend to do for instance concerning argumentation, in which I would like to incorporate perlocutionary aspects.

The two functions `options` and `filter` are crucial for rational decision making. Computing the options relates to the predictive aspect and is linked to the ideas of laws and rules (plausible beliefs), the filtering aspect is also linked to beliefs but also to social conventions and preferences. Indeed the first one gives an extent of the possibilities that the agent may consider and the second one enables the agent to choose the best goals according to its preferences. It seems important for a rational agent to be able to generate as many possible options in order to be able to find the best ones and be aware of the other ones. It is also important that the `filter` function should make a good balance between idealism and pragmatism, *i.e.*, should find a rational threshold between freedom and respect of the laws (social or natural laws); in other words, the classical exploration-exploitation dilemma.

Even if this algorithm is sequential, it seems that often, for efficiency matters, people intertwine the five steps **4**, **5**, **6**, **7**, **8** *i.e.*, perception, integration, computation of the opportunities, filtering and planning. The ability to do all these steps together when it is well done is very interesting and could be called clear-sightedness or foresight. For this purpose, an interesting function to study is `reconsider`: it translates the tendency/ability to adapt someone's intentions in regard of new opportunities, or new failures. In the view proposed by Wooldridge, what is very appealing is that this function should be very efficient, since it is only a glimpse about the necessity of a precise re-computation. This kind of problem relates to my zoom/unzoom reasoning project (see Section 6.c.1), and to the aim to finding the good focal distance in order to be able to reason and decide correctly.

Lastly, there is no explicit account of the social aspect of rationality in this algorithm, but it should be integrated in most of the functions (see my proposals in Chapter 5), and for sure the desires and intentions are not only individual but also social (see *e.g.* my project about the BLA Section 6.c).

Bibliography

[1] C. 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.

[2] V. Aleven. Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. *Artificial Intelligence*, 150(1-2):183–237, 2003.

[3] L. Amgoud. Postulates for logic-based argumentation systems. *International Journal of Approximate Reasoning*, 2013.

[4] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *International Conference on Scalable Uncertainty Management (SUM)*, pages 42–55. Springer-Verlag, 2010.

[5] L. Amgoud and P. Besnard. Logical limits of abstract argumentation frameworks. *Journal of Applied Non-Classical Logics*, 23(3):229–267, 2013.

[6] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning*, 29(2):125–169, 2002.

[7] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197 – 216, 2002.

[8] L. Amgoud, C. Devred, and M.-C. Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In *Argumentation in Multi-Agent Systems*, pages 37–56. Springer, 2009.

[9] L. Amgoud and F. Dupin de Saint-Cyr. A Semantics for Agent Communication Languages based on commitments and penalties. In *International Workshop on Computational Logic in Multi-Agent Systems (CLIMA)*, pages 28–39. Springer, 2005.

[10] L. Amgoud and F. Dupin de Saint-Cyr. Towards ACL semantics based on commitments and penalties. In *European Conference on Artificial Intellige*, pages 235–239. IOS Press, 2006.

[11] L. Amgoud and F. Dupin de Saint-Cyr. A new semantics for ACL based on commitments and penalties. *International Journal of Intelligent Systems*, 23(3):286–312, 2008.

[12] L. Amgoud and F. Dupin de Saint-Cyr. Measures for persuasion dialogs: A preliminary investigation. In *Computational models of argument (COMMA)*, pages 13–24. IOS Press, 2008.

[13] L. Amgoud and F. Dupin de Saint-Cyr. Extracting the core of a persuasion dialog to evaluate its quality. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, volume LNAI 5590, pages 59–70. Springer-Verlag, 2009.

[14] L. Amgoud and F. Dupin de Saint-Cyr. On the quality of persuasion dialogs. *Studies in Logic, Grammar and Rhetoric, Argument and Computation*, 23(36):69–98, 2011.

[15] L. Amgoud and F. Dupin de Saint-Cyr. An axiomatic approach for persuasion dialogs. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 618–625. IEEE Computer Society, novembre 2013.

[16] L. Amgoud, F. Dupin de Saint-Cyr, M.-C. Lagasquie-Schiex, and P. Saint-Dizier. Improving risk analysis in procedures via text analysis and reasoning: a road-map. In *International Forum on Industrial Safety (IFIS)*, juillet 2010.

[17] L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning (NMR)*, pages 409–417, 2002. Special session on Argument, Dialogue, Decision.

[18] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proc. of the International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, 2000.

[19] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the European Conference on Artificial Intelligence, ECAI'00*, pages 338–342, Berlin, Germany, August 2000. IOS Press.

[20] L. Amgoud and H. Prade. Comparing decisions on the basis of a bipolar typology of arguments. In Giacomo Della Riccia, Didier Dubois, Rudolf Kruse, and Hans Joachim Lenz, editors, *Preferences and similarities*, pages 249–264. Springer, april 2008.

[21] L. Amgoud and H. Prade. Using Arguments for Making and Explaining Decisions. *Artificial Intelligence*, 173:413–436, february 2009.

[22] K.D. Ashley. An AI model of case-based legal argument from a jurisprudential viewpoint. *Artificial Intelligence Law*, 10:163–218, 2002.

[23] J. Austin. *How to Do Things With Words.* Cambridge (Mass.), 1962. Paperback: Harvard University Press, 2nd edition, 2005.

[24] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider (eds). *The Description Logic Handbook.* Cambridge University Press, 2002.

[25] J. Balax, F. Dupin de Saint-Cyr, and D. Villard. DebateWEL: An interface for Debating With Enthymemes and Logical formulas. In *European Conference on Logics in Artificial Intelligence (JELIA)*, volume 7519 of *Lecture Notes in Computer Science*, pages 476–479. Springer, 2012.

[26] M. Balduccini and M. Gelfond. Diagnostic reasoning with a-prolog. *Theory and Practice of Logic Programming*, 3(4+ 5):425–461, 2003.

[27] P. Baldwin and D. Price. Debategraph. `http:debategraph.org`, 2006.

[28] F. Bannay and R. Guillaume. Towards a transparent deliberation protocol inspired from supply chain collaborative planning. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*. Springer, juillet 2014.

[29] F. Bannay and R. Guillaume. Qualitative deliberation based on bipolar leveled sets of arguments under incomplete distributed knowledge. *under submission to JAIR*, 2015.

[30] F. Bannay, M.-C. Lagasquie-Schiex, W. Raynaut, and P. Saint-Dizier. Using a SMT solver for risk analysis: detecting logical mistakes in texts. In *International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 867–874. IEEE, novembre 2014.

[31] C. Baral and J. Lobo. Defeasible specifications in action theories. In *Proc. of the 15^{th} IJCAI*, pages 1441–1446, 1997.

[32] C. Baral, S. McIlraith, and T. Cao Son. Formulating diagnostic problem solving using an action language with narratives and sensing. In *Proc. of the 7^{th} KR*, pages 311–322, Breckenridge, Colorado (USA), 2000.

[33] C. Barrett, L. de Moura, and A. Stump. Smt-comp: Satisfiability modulo theories competition. In *Computer Aided Verification*, pages 20–23. Springer, 2005.

[34] C. Barrett, A. Stump, and C. Tinelli. The SMT-LIB Standard: Version 2.0. In A. Gupta and D. Kroening, editors, *Proceedings of the 8th International Workshop on Satisfiability Modulo Theories (Edinburgh, UK)*, 2010.

[35] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In *Proceeding of the 2010 conference on Computational Models of Argument: Proceedings of COMMA 2010*, pages 75–86, Amsterdam, The Netherlands, The Netherlands, 2010. IOS Press.

[36] N. Belnap, M. Perloff, M. Xu, P. Bartha, M. Green, and J. Horty. *Facing the future: agents and choices in our indeterminist world.* Oxford University Press Oxford, 2001.

[37] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax- based entailment. In Ruzena Bajcsy, editor, *Proc. of the 13th IJCAI*, pages 640–645, Chambery, France, 1993. Morgan-Kaufmann.

[38] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In W. Swartout B. Nebel, C. Rich, editor, *Proc. of the 3rd KR*, pages 673–684, Cambridge, MA, October 1992.

[39] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1):259–276, 1997.

[40] S. Benferhat, D. Dubois, and H. Prade. Towards fuzzy default reasoning. In T. Sudkamp R. Davé, editor, *18th International Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, pages 23–27, New York, USA, June 1999.

[41] S. Benferhat and F. Dupin de Saint-Cyr. Contextual handling of conditional knowledge. In *Proc. of the 6th Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, volume 3, pages 1369–1374, Granada, Spain, july 1996.

[42] S. Benferhat, S. Konieczny, O. Papini, and R. Pino-Pérez. Iterated revision by epistemic states: axioms, semantics and syntax. In *Proceedings of ECAI-2000*, pages 13–17, 2000.

[43] S. Berger, D. Lehmann, and K. Schlechta. Preferred history semantics for iterated updates. *Journal of Logic and Computatation*, 9(6):817–833, 1999.

[44] P. Besnard and S. Doutre. Checking the acceptability of a set of arguments. In J. Delgrande and T. Schaub, editors, *10th International Workshop on Non-Monotonic Reasoning (NMR 2004), Whistler, Canada, June 6-8, 2004, Proceedings*, pages 59–64, 2004.

[45] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203 – 235, 2001.

[46] P. Bisquert. *Étude du changement en argumentation. De la théorie à la pratique.* Thèse de doctorat, Université Paul Sabatier, Toulouse, France, Décembre 2013.

[47] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in argumentation systems: exploring the interest of removing an argument. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 275–288. Springer-Verlag, octobre 2011.

[48] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Duality between Addition and Removal: a Tool for Studying Change in Argumentation. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, volume 297 of *Communications in Computer and Information Science*, pages 219–229. Springer, juillet 2012.

[49] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Characterizing change in abstract argumentation systems. In *Trends in Belief Revision and Argumentation Dynamics*, volume 48 of *Studies in Logic*, pages 75–102. College Publications, 2013.

[50] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Enforcement in Argumentation is a kind of Update. In *International Conference on Scalable Uncertainty Management (SUM)*, number 8078 in LNAI, pages 30–43. Springer-Verlag, 2013.

[51] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Goal-driven Changes in Argumentation: A theoretical framework and a tool. In *International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 610–617. IEEE Computer Society, 2013.

[52] P. Bisquert, M. Croitoru, and F. Dupin de Saint-Cyr. Four ways to evaluate arguments according to agent's engagement. In *Brain Intelligence*, 2015. under submission.

[53] P. Bisquert, M. Croitoru, and F. Dupin de Saint-Cyr. Towards a dual process cognitive model for argument evaluation. In *SUM*, 2015. under submission.

[54] E. Black and K. Atkinson. Choosing persuasive arguments for action. In *10th Int. Conf. on Autonomous Agents and Multi-Agent Systems*, 2011.

[55] E. Black and A. Hunter. Using enthymemes in an inquiry dialogue system. In *Proc of the 7th Int. Conf. on Autonomous Agents and Multiag. Syst. (AAMAS 2008)*, pages 437–444, 2008.

[56] B. Bonet and H. Geffner. Arguing for decisions: A qualitative model of decision making. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 98–105. Morgan Kaufmann Publishers Inc., 1996.

[57] JF. Bonnefon, D. Dubois, and H. Fargier. An overview of bipolar qualitative decision rules. In G. Della Riccia, D. Dubois, R. Kruse, and H-J. Lenz, editors, *Preferences and Similarities*, volume 504 of *CISM Courses and Lectures*, pages 47–73. Springer, 2008.

[58] E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *10th International Conference on Autonomous Agents and Multiagent Systems*, pages 47–54, 2011.

[59] E. Bonzon, N. Maudet, and S. Moretti. Coalitional games for abstract argumentation. In *Proceedings International Conference on Computational Models of Argument (COMMA'14)*, 2014.

[60] R. Booth and A. Nittka. Reconstructing an agent's epistemic state from observations. In *IJCAI*, pages 394–399, 2005.

[61] C. Boutilier. A unified model of qualitative belief change: a dynamical systems perspective. *Artificial Intelligence*, 1-2:281–316, 1998.

[62] V. Brusoni, L. Console, P. Terenziani, and D. Dupré. A spectrum of definitions for temporal model-based diagnosis. *Artificial Intelligence*, 102(1):39–79, 1998.

[63] E. Cabrio and S. Villata. Combining textual entailment and argumentation theory for supporting online debates interactions. In *50th Annual Meeting of the Association for Computational Linguistics*, pages 208–212, 2012.

[64] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Revision of an Argumentation System. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 124–134. AAAI Press, 2008.

[65] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in Abstract Argumentation Frameworks: Adding an Argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010.

[66] E. Clarke and E. Emerson. Design and synthesis of synchronization skeletons using branching time temporal logic. In Dexter Kozen, editor, *Logics of Programs*, volume 131 of *Lecture Notes in Computer Science*, pages 52–71. Springer Berlin Heidelberg, 1982.

[67] M. Colombetti. A commitment-based approach to agent speech acts and conversations. In *Proceedings of the Workshop on Agent Languages and Conversation Policies. I4th International Conference on Autonomous Agents*, pages 21–29, 2000.

[68] M.-O. Cordier and P. Siegel. Prioritized transitions for updates. In *ECSQARU*, pages 142–150, 1995.

[69] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Proc. of KR*, pages 112–122, Lake District, 2006.

[70] S. Coste-Marquis and P. Marquis. Compiling stratified belief bases. In *Proceedings of the 14th European Conference on Artificial Intelligence*, pages 23–27, Berlin, 2000.

[71] D2.2. Towards a consensual formal model: inference part. *Deliverable of ASPIC project*, 2004.

[72] C. Daligny. Validation de connaissances et affichage cognitif d'informations redondantes. Technical report, Laboratoire d'Etudes et de Recherche en Informatique d'Angers, France, 1999.

[73] A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.

[74] J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[75] J. de Kleer and B. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97 – 130, 1987.

[76] F. Dupin de Saint-Cyr and H. Prade. Logical handling of uncertain, ontology-based, spatial information. *Fuzzy Sets and Systems, Advances in Intelligent Databases and Information Systems*, 159(12):1515–1534, juin 2008.

[77] J. Delgrande, D. Dubois, and J. Lang. Iterated revision as prioritized merging. In *International Conference on Principles of Knowledge Representation and Reasoning (KR), Lake District (UK), 02/06/2006-05/06/2006*, pages 210–220, http://www.aaai.org/Press/press.php, 2006. AAAI Press.

[78] R. Demolombe and M. Parra del Pilar Pozos. A simple and tractable extension of situation calculus to epistemic logic. In *Foundations of Intelligent Systems*, pages 515–524. Springer, 2010.

[79] D. Doder and S. Vesic. How to decrease and resolve inconsistency of a knowledge base? In *7th International Conference on Agents and Artificial Intelligence (ICAART'15)*, pages 27–37, 2015.

[80] O. Doukari and R. Jeansoulin. Space contained conflict revision to allow consistency checking of spatial decision support. In *AGILE 2007 Conference*, Aalborg, DK, May 2007.

[81] D. Dubois, F. Dupin de Saint-Cyr, and H. Prade. Update postulates without inertia (regular paper). In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Selected papers of the Europ. Conf. ECSQARU'95), Fribourg, Switzerland,* , number 946 in LNAI, pages 162–170, http://www.springerlink.com/, juillet 1995. Springer-Verlag.

[82] D. Dubois and H. Fargier. Qualitative decision making with bipolar information. *KR*, 6:175–186, 2006.

[83] D. Dubois and H. Fargier. Qualitative bipolar decision rules: Toward more expressive settings. In *Preferences and Decisions*, pages 139–158. Springer, 2010.

[84] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of logic in Artificial Intelligence and logic programming*, volume 3, pages 439–513. Clarendon Press - Oxford, 1994.

[85] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, 1988.

[86] D. Dubois and H. Prade. Conditional objects and non-monotonic reasoning. In *Proc. of the 2nd KR*, pages 175–185, Cambridge, MA, avril 1991. Morgan Kaufmann.

[87] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified Representation of Uncertainty and Imprecision*, pages 169–226. Kluwer Academic Publishers, 1998.

[88] D. Dubois and H. Prade. Towards a multiple-agent extension of possibilistic logic. In *Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2007)*, London, UK, 23-26 July 2007.

[89] D. Dubois, H. Prade, and R. Yager. Merging fuzzy information. In H. Prade J. Bezdek, D. Dubois, editor, *Fuzzy Sets in Approximate Reasoning and Information Systems*, pages 335–401. Kluwer, Boston, the handbooks of fuzzy sets series edition, 1999.

[90] P. M. 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.

[91] F. Dupin de Saint-Cyr. Scenario Update Applied to Causal Reasoning. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 188–197. AAAI Press, 2008.

[92] F. Dupin de Saint-Cyr. A first attempt to allow enthymemes in persuasion dialogs. In *DEXA International Workshop: Data, Logic and Inconsistency (DALI)*, pages 332–336. IEEE Computer Society - Conference Publishing Services, 2011.

[93] F. Dupin de Saint-Cyr. Handling enthymemes in time-limited persuasion dialogs. In *International Conference on Scalable Uncertainty Management (SUM)*, number 6929 in LNAI, pages 149–162. Springer-Verlag, 2011.

[94] F. Dupin de Saint-Cyr, P. Bisquert, C. Cayrol, and M.-C. Lagasquie-Schiex. Argumentation Update in YALLA (Yet Another Logic Language for Argumentation). *under submission to IJAR*, 2015.

[95] F. Dupin de Saint-Cyr, B. Duval, and S. Loiseau. A priori revision. In *Lecture Notes in Artificial Intelligence (Proc. of ECSQARU-01)*, pages 488–497, Toulouse, France, September 2001.

[96] F. Dupin de Saint-Cyr, A. Herzig, J. Lang, and P. Marquis. Raisonnement sur l'action et le changement. In P. Marquis, O. Papini, and H. Prade, editors, *Panorama de l'intelligence artificielle. Ses bases méthodologiques, ses développements*, volume 1, chapter 12, pages 255–282. Cépaduès, http://www.cepadues.com, 2014.

[97] F. Dupin de Saint-Cyr, R. Jeansoulin, and H. Prade. Fusing uncertain structured spatial information. In Sergio Greco and Thomas Lukasiewicz, editors, *International Conference on Scalable Uncertainty Management (SUM)*, number 5291 in LNAI, pages 174–188. Springer, octobre 2008.

[98] F. Dupin de Saint-Cyr, R. Jeansoulin, and H. Prade. Spatial information fusion: Coping with uncertainty in conceptual structures. In *International Conference on Conceptual Structures (ICCS)*, volume Supplementary Proceedings, pages 66–74. Springer, 2008.

[99] F. Dupin de Saint-Cyr and J. Lang. Reasoning about unpredicted change and explicit time. In *Inter. Joint Conf. on Qualitative and Quantitative Practical Reasoning (ECSQARU/FAPR'97), Bad Honnef, Germany,* , pages 223–236, http://www.springerlink.com/, juin 1997. Springer-Verlag.

[100] F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). In *International Conference, Principles of Knowledge Representation and Reasoning (KR)*, pages 497–508. Morgan Kaufmann Publishers, avril 2002.

[101] F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). *Artificial Intelligence*, 175:760–790, janvier 2011.

[102] F. Dupin de Saint-Cyr, J. Lang, and T. Schiex. Penalty logic and its link with Dempster-Shafer theory. In *Proc. of the* 10^{th} *Conf. on Uncertainty in Artificial Intelligence*, pages 204–211. Morgan Kaufmann, July 1994.

[103] F. Dupin de Saint-Cyr and S. Loiseau. Validation and refinement versus revision. In *Symposium on verification and validation of knowledge based systems and components (EUROVAV'99)*, pages 163–176, Oslo, Norway, June 1999. Kluwer Academic Publishers.

[104] F. Dupin de Saint-Cyr and S. Loiseau. Alignement cognitif de symboles. In *Journées Francophones Modèles Formels de l'Interaction (MFI)*, pages 249–254. Cépaduès Editions, mai 2003.

[105] F. Dupin de Saint-Cyr and H. Prade. Multiple-source data fusion problems in spatial information systems. In *(Proc. of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'06)*, pages 2189–2196, Paris, France, july 2006.

[106] F. Dupin de Saint-Cyr and H. Prade. Possibilistic Handling of Uncertain Default Rules with Applications to Persistence Modeling and Fuzzy Default Reasoning. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 440–450. AAAI Press, 2006.

[107] T. Eiter and G. Gottlob. The complexity of logic-based abduction. *Journal of the Association for Computing Machinery*, 42(1):3–42, 1995.

[108] M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as logical uncertainty. In *ECSQARU'93*, pages 85–90, 1993.

[109] M. Elvang-Gøransson, P. Krause, and J. Fox. Acceptability of arguments as "logical uncertainty". In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 85–90. Springer, 1993.

[110] R. Fikes and N. Nilsson. Strips : A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[111] T. Finin, Y. Labrou, and J. Mayfield. KQML as an agent communication language. In *J. Bradshaw, ed. Softaware agents, MIT Press, Cambridge.*, 1995.

[112] FIPA. ACL message structure specification. In *FIPA 02. Foundation for Intelligent Physical Agents*, 2002.

[113] J. Fox and S. Parsons. On using arguments for reasoning about actions and values. In *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning, Stanford*, pages 55–63, 1997.

[114] J. Fox and S. Parsons. Arguing about beliefs and actions. In *Applications of Uncertainty Formalisms*, pages 266–302, 1998.

[115] N. Friedman and J.Y. Halpern. Modeling beliefs in dynamic systems. part ii: revision and update. *JAIR*, 10:117–167, 1999.

[116] B. Ganter and R. Wille. *Formal Concept Analysis, Mathematical Foundations*. Springer-Verlag, 1999.

[117] A. García and G. Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(2):95–138, January 2004.

[118] B. Gaudou, A. Herzig, and D. Longin. A Logical Framework for Grounding-based Dialogue Analysis. *Electronic Notes in Theoretical Computer Science*, 157(4):117–137, 2006.

[119] E. Giunchiglia, J. Lee, V. Lifschitz, N. McCain, and H. Turner. Nonmonotonic causal theories. *Artificial Intelligence*, 153:49–104, 2004.

[120] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.

[121] T. Gordon. The pleadings game. *Artificial Intelligence and Law*, 2:239–292, 1993.

[122] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, San Diego, CA, 1975.

[123] J. Halpern and J. Pearl. Causes and explanations: A structural-model approach. part i: Causes. *The British journal for the philosophy of science*, 56(4):843–887, 2005.

[124] A. Herzig. Logics for belief base updating. In *Belief Change*, pages 189–231. Springer, 1998.

[125] A. Herzig. On updates with integrity constraints. In *Belief Change in Rational Agents*, 2005. `http://drops.dagstuhl.de/opus/volltexte/2005/334`.

[126] A. Hunter. Real arguments are approximate arguments. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI'07)*, pages 66–71. MIT Press, 2007.

[127] A. Hunter and J. Delgrande. Belief change in the context of fallible actions and observations. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21-1, page 257. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[128] A. Hunter and S. Konieczny. Measuring inconsistency through minimal inconsistent sets. *KR*, 8:358–366, 2008.

[129] J. Introne and L. Iandoli. Improving decision-making performance through argumentation: An argument-based decision support system to compute with evidence. *Decision Support Systems*, 64:79–89, 2014.

[130] M. Järvisalo, D. Le Berre, O. Roussel, and L. Simon. The international SAT solver competitions. *AI Magazine*, 33(1):89–92, 2012.

[131] M. Johnson, P. McBurney, and S. Parsons. When are two protocols the same? In *Communication in Multiagent Systems*, pages 253–268, 2003.

[132] A. Kakas, R. Miller, and F. Toni. E-res: reasoning about actions, events and observations. In *LPNMR-01*, pages 254–266, 2001.

[133] A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 883–890. ACM, 2003.

[134] A. Kakas and P. Moraitis. Adaptive agent negotiation via argumentation. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems (AAMAS '06)*, pages 384–391, New York, NY, USA, 2006. ACM.

[135] N. Karacapilidis and C. Pappis. Computer-supported collaborative argumentation and fuzzy similarity measures in multiple criteria decision making. *Computer and Operations Research*, 27:653–671, 2000.

[136] H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In J. Allen and al., editors, *Proc. of the 2^{nd} Inter. Conf. on Principles of Knowledge Representation and Reasoning*, pages 387–394, Cambridge, MA, 1991.

[137] H. Katsuno and A. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[138] D. Kayser and A. Mokhtari. Time in a causal theory. *Annals of Mathematics and Artificial Intelligence*, 22(1-2):117–138, 1998.

[139] D. Kontarinis, E. Bonzon, N. Maudet, and P. Moraitis. Picking the right expert to make a debate uncontroversial. In *Comp. Models of Argument*, pages 486–497, 2012.

[140] J. Lang. Belief update revisited. In *IJCAI*, volume 7, pages 6–12, 2007.

[141] J. Lang, P. Liberatore, and P. Marquis. Conditional independence in propositional logic. *Artificial Intelligence Journal*, Volume 141(1-2):79–121, 2002.

[142] J. Lang, F. Lin, and P. Marquis. Causal theories of action: A computational core. In *Proceedings of 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, pages 1073–1078, Acapulco, 2003.

[143] R. Laurini and D. Thompson. *Fundamentals of spatial information systems*. 37. Academic press, 1992.

[144] D. Lehmann. Belief revision, revised. In *Proc. of IJCAI'95*, pages 1534–1540, 1995.

[145] D Lewis. *Counterfactuals*. Harvard University Press, 1973.

[146] B. Liao. *Efficient Computation of Argumentation Semantics*. Intelligent Systems. Academic Press, Oxford, UK, 2014.

[147] P. Liberatore. On the compilability of diagnosis, planning, reasoning about actions, belief revision, etc. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 144–155, 1998.

[148] P. Liberatore and M. Schaerf. BReLs: A system for the integration of knowledge bases. In *Proc. of the 7^{th} KR*, pages 145–152, Breckenridge, Colorado (USA), 2000.

[149] T. Lukasiewicz. Weak nonmonotonic probabilistic logics. *Artificial Intelligence*, 168(1-2):119–161, October 2005.

[150] J. MacKenzie. Question-begging in non-cumulative systems. *Journal of philosophical logic*, 8:117–133, 1979.

[151] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Mitchie, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.

[152] A. Miquel. Révision d'un système d'argumentation : une première approche. Rapport de recherche IRIT/RR–2007–25–FR, IRIT, Université Paul Sabatier, Toulouse, septembre 2007.

[153] F. Moisan. *The bonds of society: an interdisciplinary study of social rationality*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2013.

[154] L. Moura and N. Bjørner. Z3: An efficient smt solver. In C.R. Ramakrishnan and J. Rehof, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 4963 of *Lecture Notes in Computer Science*, pages 337–340. Springer Berlin Heidelberg, 2008.

[155] S. Needleman and C. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–53, 1970.

[156] P. Nicolas, L. Garcia, C. Lefevre, and I. Stéphan. Possibilistic uncertainty handling for answer set programming. *Annals of Mathematics and Artificial Intelligence*, 47(1-2):139–181, June 2006.

[157] P. Nicolas, L. Garcia, and I. Stéphan. A possibilistic inconsistency handling in anwer set programming. In *Proc. of ESCQARU'05*, pages 402–414. Springer, 2005.

[158] W. Ouerdane, Y. Dimopoulos, K. Liapis, and P. Moraitis. Towards automating decision aiding through argumentation. *Journal of Multi-Criteria Decision Analysis*, 18(5-6):289–309, 2011.

[159] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[160] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[161] J. Pearl. System Z : a natural ordering of defaults with tractable applications for default reasoning. In M. Vardi, editor, *Proceedings of theoretical aspects of reasoning about knowledge*, pages 121–135, San Mateo, 1990. Morgan Kaufman.

[162] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence Journal*, 57:1–42, 1992.

[163] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *7th European workshop on Logic for Artificial Intelligence, JELIA'00*, Berlin, 2000.

[164] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[165] G. Quiroz, D. Apothéloz, and P. Brandt. How counter-argumentation works. *Argumentation Illuminated*, pages 172–177, 1992.

[166] J. Raz. *Practical reasoning*. Oxford University Press, 1978.

[167] P. Saint-Dizier. Lelie: An intelligent assistant for the analysis and the prevention of risks in industrial processes, 2011-2013.

[168] P. Saint-Dizier. *Challenges of Discourse Processing: the case of technical texts*. Cambridge University Press, http://www.cambridge.org/, février 2014.

[169] C. Salavastru. *Logique, Argumentation, Interprétation*. L'Harmattan, Paris, 2007.

[170] E. Sandewall. The range of appplicability of some non-monotonic logics for strict inertia. *Journal of logic computation*, 4(5):581–615, 1994.

[171] E. Sandewall. *Features and Fluents*. Oxford University Press, 1995.

[172] R. Scherl and H. Levesque. Knowledge, action, and the frame problem. *Artificial Intelligence*, 144(1):1–39, 2003.

[173] A. Schopenhauer. *The Art of Always Being Right: 38 Ways to Win an Argument*. http://en.wikisource.org/wiki/The_Art_of_Always_Being_Right, 1831. Orig. title: *Die Kunst, Recht zu behalten* (Transl. by T. Saunders in 1896).

[174] J. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge U. Press, 1969.

[175] J. Searle and D. Vanderveken. Foundations of illocutionary logic. In *Cambridge University Press*, 1985.

[176] K. Segerberg. Belief revision from the point of view of doxastic logic. *Logic Journal of IGPL*, 3(4):535–553, 1995.

[177] S. Shapiro and M. Pagnucco. Iterated belief change and exogenous actions in the situation calculus. In *ECAI*, volume 16, page 878, 2004.

[178] Y. Shoham. *Reasoning about Change - Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, 1988.

[179] E. Shortliffe. *Computer-based medical consultations*. North-Holland, 1976.

[180] M. Singh. Agent communication languages: Rethinking the principles. In *IEEE Computer*, pages 40–47, 1998.

[181] M. Singh. A social semantics for agent communication languages. In *IJCAI'99 Workshop on Agent Communication Languages*, pages 75–88, 1999.

[182] M. Suwa, A. Scott, and E. Shortliffe. An approach to verifying completeness and consistency in a rule-based expert system. *Ai Magazine*, 3(4):16, 1982.

[183] A. Tarski. *Logic, Semantics, Metamathematics (E. H. Woodger, editor))*, chapter On Some Fundamental Concepts of Metamathematics. Oxford Uni. Press, 1956.

[184] G. Tecuci, D. Marcu, M. Boicu, D. Schum, and Russell K. Computational theory and cognitive assistant for intelligence analysis. In *Proceedings of STIDS*, pages 68–75, 2011.

[185] M. Thielscher. A theory of dynamic diagnosis. *Linkoping Electronic Articles in Computer and Information Science*, 2(11), 1997.

[186] M. Thielscher. A general game description language for incomplete information games. In *Proceedings of AAAI*, pages 994–999, 2010.

[187] P. Torroni. A study on the termination of negotiation dialogues. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 1223–1230. ACM, 2002.

[188] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[189] J. van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.

[190] T. van Gelder. The rationale for Rationale$^{\text{TM}}$. *Law, probability and risk*, 6(1-4):23–42, 2007.

[191] S. Villata, G. Boella, D. Gabbay, L. van der Torre, and J. Hulstijn. A logic of argumentation for specification and verification of abstract argumentation frameworks. *Annals of Mathematics and Artificial Intelligence*, 66(1-4):199–230, 2012.

[192] G. von Wright. Causality and determinism. *The Journal of Philosophy*, 73(8):213–218, 1976.

[193] D. Walton. The three bases for the enthymeme: A dialogical theory. *Journal of Applied Logic*, 6:361–379, 2008.

[194] D. Walton. *Argumentation schemes for presumptive reasoning.* (first edition 1996) Routledge, 2013.

[195] D. Walton and E. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.* State University of New York Press, Albany, NY, 1995.

[196] D. Walton and F. Macagno. Enthymemes, argumentation schemes, and topics. *Logique et Analyse*, 205:39–56, 2009.

[197] D. Walton and C. Reed. Diagramming, argumentation schemes and critical questions. In *Anyone Who Has a View*, pages 195–211. Springer, 2003.

[198] D. Walton and C. Reed. Argumentation schemes and enthymemes. *Synthese*, 145:339–370, 2005.

[199] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered Sets*, pages 445–470. D. Reidel, Dordrecht-Boston, 1982.

[200] M. Winslett. Reasoning about action using a possible models approach. In *Proc. of the 7^{th} National Conference on Artificial Intelligence*, pages 89–93, St. Paul, 1988.

[201] M. Wooldridge. *Reasoning about rational agents.* MIT press, 2000.

[202] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In *4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), July 25-29, 2005, Utrecht, The Netherlands*, pages 560–567. ACM, 2005.

[203] E. Wurbel, O. Papini, and R. Jeansoulin. Revision: an application in the framework of GIS. In *proc. of the 7th International Conference on Principles of Knowledge Representation and Reasoning, KR'2000*, pages 505–516, Breckenridge, Colorado, USA, avril 2000.

[204] S. Zabala, I. Lara, and H. Geffner. Beliefs, reasons and moves in a model for argumentative dialogues. In *Proc. 25th Latino-American Conf. on Computer Science*, 1999.