Actes des 15èmes Journées d'Intelligence Artificielle Fondamentale - JIAF 2021

Zied Bouraoui, Sylvie Doutre

AfIA

Association française
pour l'Intelligence Artificielle

JIAF

Journées d'Intelligence Artificielle Fondamentale

PFIA 2021

Crédit photo : Flicr/xlibber

Table des matières

Éditorial

Journées d'Intelligence Artificielle Fondamentale

Les Journées d'Intelligence Artificielle Fondamentale (JIAF) constituent un rendez-vous annuel de la communauté francophone travaillant sur l'Intelligence Artificielle Fondamentale. Les thématiques de recherche sont relatives aux méthodes et outils fondamentaux de l'Intelligence Artificielle. Les journées sont composées d'exposés de synthèse, permettant à la communauté de découvrir des thématiques connexes au travers d'exposés de spécialistes, et de communications sélectionnées par le comité de programme.

Les thématiques de recherche des JIAF sont relatives aux méthodes et outils fondamentaux de l'Intelligence Artificielle :

- Définition de modèles de représentation des informations (croyances, connaissances, préférences, obligations et permissions, actions, incertitude, confiance, réputation) : langages des logiques classiques ou non classiques, modèles possibilistes, ontologies, langages à base de contraintes, représentations graphiques, etc.
- Définition et automatisation de raisonnements sur ces informations : raisonnement spatio-temporel, dynamique des informations, révision de croyances, fusion d'informations symboliques, raisonnement par argumentation, raisonnement causal, raisonnement abductif, raisonnement à partir de cas, etc.
- Mise au point de méthodes de codage des informations et d'algorithmes de traitement efficaces : compilation de connaissances, SAT, contraintes, ASP, etc.
- Modélisation formelle de l'interaction : entre utilisateurs et systèmes informatiques, entre entités informatiques autonomes (agents), intégration de ces deux aspects dans les divers agents conversationnels, agents de recherche, assistants personnels.
- Choix social, théorie des jeux, algorithmes pour les jeux.
- Pour des objectifs de décision, planification, ordonnancement, diagnostic, apprentissage et dans différents contextes d'application, comme par exemple le Web sémantique.

Ces 15èmes Journées d'Intelligence Artificielle Fondamentale (JIAF 2021) ont eu lieu du 1 au 2 juillet 2021, dans le cadre de la Plate-Forme Intelligence Artificielle (PFIA). Les éditions précédentes se sont déroulées à Angers (en ligne - 2020), Toulouse (2019), Amiens (2018), Caen (2017), Montpellier (2016), Rennes (2015), Angers (2014), Aix-en-Provence (2013), Toulouse (2012), Lyon (2011), Strasbourg (2010), Marseille (2009), Paris (2008) et Grenoble (2007).

Zied Bouraoui, Sylvie Doutre

Comité de programme

Président

— Zied Bouraoui (CRIl, Univ Artois & CNRS)
— Sylvie Doutre (IRIT, Université Toulouse 1 Capitole)

Membres

— Francesco Belardinelli (IBISC, Université d'Évry)
— Elise Bonzon (LIPADE, Université Paris Descartes)
— Tristan Cazenave (LAMSADE, Université Paris Dauphine)
— Nadia Creignou (LIS, Aix-Marseille Université)
— Tiago de Lima (CRIL, Université de Lens)
— Jérôme Euzenat (LIG, INRIA)
— George Katsirelos (MIAT, INRA)
— Sébastien Konieczny (CRIL, CNRS)
— Jean Lieber (LORIA, INRIA)
— Pierre Marquis (CRIL, Université d'Artois)
— Marie-Laure Mugnier (LIRMM, Université de Montpellier)
— Amedeo Napoli (LORIA, CNRS)
— Odile Papini (LIS, Aix-Marseille Université)
— Meltem ÖZTÜRK (LAMSADE, Université Paris Dauphine)
— Laurent Perrussel (IRIT, Université Toulouse 1 Capitole)
— Sophie Pinchinat (IRISA, INRIA)
— Stéphanie Roussel (ONERA)
— Julien Rossit (LIPADE, Université Paris Decartes)
— Serena Villata (I3S, CNRS)
— Bruno Zanuttini (GREYC, UNICAEN)

Actes JIAF 2021

Constrained Incomplete Argumentation Frameworks

Jean-Guy Mailly
LIPADE, Université de Paris, France
jean-guy.mailly@u-paris.fr

Résumé

Des opérations comme le changement ou la fusion de croyances ont été adaptées au formalisme de l'argumentation abstraite. Cependant, ces opérations peuvent nécessiter l'expression d'une certaine forme d'incertitude ou de disjonction dans le résultat, ce qui n'est pas représentable au moyen de systèmes d'argumentation classiques. Pour cette raison, certaines approches de révision ou de fusion de systèmes d'argumentation produisent un ensemble de systèmes d'argumentation ou d'extensions, représentant en quelque sort une disjonction de systèmes/d'extensions. En parallèle, la notion de systèmes d'argumentation incomplets a été proposée. Il s'agit de systèmes d'argumentation dans lesquels l'existence de certains arguments ou attaques peut être incertaine. Un système d'argumentation incomplet peut être associé à un ensemble de completions, c'est-à-dire des systèmes d'argumentation classiques qui correspondent aux différents façons de « résoudre l'incertitude ». Ces systèmes incomplets peuvent sembler être de bons candidats pour une représentation compacte d'une « disjonction » de systèmes d'argumentation, mais nous prouvons que ce modèle n'est pas suffisamment expressif pour représenter n'importe quelle disjonction. Ensuite nous introduisons les systèmes d'argumentation incomplets avec contrainte, qui ajoutent au système une formule propositionnelle qui détermine quelles complétions doivent être utilisées pour raisonner. Nous prouvons que ce modèle est suffisamment expressif pour représenter n'importe quel ensemble de systèmes d'argumentation ou d'extensions, sans pour autant provoquer une augmentation de la complexité du raisonnement par rapport aux systèmes d'argumentation incomplets préexistants. Enfin, nous montrons que notre nouveau formalisme peut être utilisé pour modéliser le forçage d'extensions.

Abstract

Operations like belief change or merging have been adapted to the context of abstract argumentation. However, these operations may require to express some uncertainty or some disjunction in the result, which is not representable in classical AFs. For this reason, some of these works require a set of AFs or a set of extensions as the outcome of the operation, somehow to represent a disjunction of AFs/extensions. In parallel, the notion of Incomplete AFs (IAFs) have been developed recently. It corresponds to AFs where the existence of some arguments or attacks may be uncertain. Each IAF can be associated with a set of classical AFs called completions, that correspond to different ways of "resolving the uncertainty". While these IAFs could be good candidates for a compact representation of a "disjunction" of AFs, we prove that this model is not expressive enough. Then we introduce Constrained IAFs, that include a propositional formula allowing to select the set of completions used for reasoning. We prove that this model is expressive enough for representing any set of AFs, or any set of extensions. Moreover, we show that the complexity of credulous and skeptical reasoning is the same as in the case of IAFs. Finally, we show that CIAFs can be used to model extension enforcement.

1 Introduction

Representing uncertainty and reasoning with uncertain information is of utmost importance in artificial intelligence. Indeed, there are many reasons that may lead an intelligent agent to face uncertainty or impossibility to choose between alternatives. For instance, she can receive information from different sources, which can have different degrees of reliability. This information can be incompatible with her previous knowledge, or with information provided by other sources. This kind of problem can be formalized as belief change operations ("How to incorporate a new piece of information to my knowledge if it is not logically consistent?") [1, 19, 20] or belief merging ("How to give a coherent representation of several agent's knowledge even if they are globally inconsistent?") [21]. In this kind of application, a simple way to deal with the uncertainty of the result is the logical disjunction : if the result of revising an agent's knowledge is "I am not sure whether a is true or b is true.", then it can be expressed with $a \vee b$. However, there are formalisms where this kind of simple representation of undecidedness cannot be done. For instance, in abstract argumentation frameworks (AFs)

[16], either there is certainly an attack between two arguments, or there is certainly no attack between them. But an agent cannot express something like "I am not sure whether a attacks b or not." AFs have been extended in this direction : Partial AFs (PAFs) [8] allow to represent uncertain attacks. Later, Incomplete AFs (IAFs) [6, 5] have been proposed, as a generalization of PAFs where also arguments can be uncertain. Reasoning with a PAF or an IAF is possible thanks to a set of *completions*, that are classical AFs that correspond to the different possible worlds encoded in the uncertain information. While this framework allows to express uncertainty in abstract argumentation in a rich way, there are still situations that cannot be modeled. Consider, *e.g.*, that an agent faces the information "Either a attacks b, or b attacks a, but I am not sure whether is true.". There is no way to represent this information with an IAF. However, this may be necessary in some situations. For instance, several adaptations of belief change [10, 12] or merging [8, 13] to abstract argumentation lead to results that can contain such an uncertainty over the result, impossible to be represented by a single AF. So, these works propose to represent the "disjunction" in the result as a set of AFs, or even as a set of extensions (and it is also known that not every set of extensions can be represented by a single AF [17]).

In this paper, we define a generalization of IAFs, that adds a constraint to it. The constraint in a Constrained IAF (CIAF) is a propositional formula that allows to specify which subset of the completions of the IAF should be used for reasoning. We show that this framework is more expressive than IAFs, in the sense that any set of AFs can be the set of completions of a CIAF. Also, any set of extensions can be obtained from (the completions of) a CIAF. We prove that, despite being more expressive than IAFs, the complexity of credulous and skeptical reasoning does not increase compared to IAFs, under various classical semantics.

Interestingly, we also identify a relation between our CIAFs and extension enforcement [3]. This operation consists in modifying an AF s.t. a given set of arguments becomes part of an extension. Classical enforcement operators are based on expansions, *i.e.* addition of arguments and attacks s.t. the attack relation between former arguments remain unchanged. Theoretical results show under which conditions enforcement is possible under expansions. However, these results may suppose the possibility to perform unnatural expansions, like adding a new argument that attacks all the undesired arguments. In a real dialogue, such an "ultimate attacker", that defeats every unwanted argument, is not likely to exist. We show that completions of a CIAF can be used to model the set of expansions that are available to an agent, and then enforcement is possible iff the desired set of arguments is credulously accepted w.r.t. the CIAF.

The paper is organized as follows. Section 2 describes background notions of abstract argumentation. Our first contributions are presented in Section 3 : the definition of CIAFs, the properties of the framework regarding its expressivity, and finally the computational complexity of credulous and skeptical acceptance. Then in Section 4, we show how CIAFs can be used to model scenarios of extension enforcement. We discuss related work in Section 5, and finally Section 6 concludes the paper and highlights some topics of interest for future research.

2 Background

2.1 Dung's Abstract Argumentation

Abstract argumentation was introduced in [16], where arguments are abstract entities whose origin or internal structure are ignored. The acceptance of arguments is purely defined from the relations between them.

Definition 1 (Abstract AF)**.** *An* abstract argumentation framework *(AF) is a directed graph $\mathcal{F} = \langle A, R\rangle$, where A is a set of* arguments*, and $R \subseteq A \times A$ is an* attack relation*.*

We say that *a attacks b* when $(a, b) \in R$. If $(b, c) \in R$ also holds, then *a defends c* against b. Attack and defense can be adapted to sets of arguments : $S \subseteq A$ attacks (respectively defends) an argument $b \in A$ if $\exists a \in S$ that attacks (respectively defends) b.

Example 1. *Let $\mathcal{F} = \langle A, R\rangle$ be the AF depicted at Figure 1, with $A = \{a, b, c, d, e\}$ and $R = \{(b, a), (c, a), (c, d), (d, b), (d, c), (e, a)\}$. Each arrow represents an attack. d defends a against both b and c, since these are attackers of a that are, in turn, both attacked by d.*

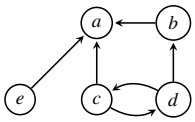


Figure 1 – The AF $\mathcal{F}$

In [16], Dung introduces different semantics to evaluate the acceptability of arguments. They are based on two basic concepts : *conflict-freeness* and *admissibility*.

Definition 2 (Conflict-freeness and Admissibility)**.** *Given $\mathcal{F} = \langle A, R\rangle$, a set of arguments $S \subseteq A$ is :*

- *conflict-free iff $\forall a, b \in S$, $(a, b) \notin R$;*
- *admissible iff it is conflict-free, and defends each $a \in S$ against each of its attackers.*

We use $\mathrm{cf}(\mathcal{F})$ and $\mathrm{ad}(\mathcal{F})$ for denoting the sets of conflit-free and admissible sets of an argumentation framework $\mathcal{F}$.

The intuition behind these principles is that a set of arguments may be accepted only if it is internally consistent (conflict-freeness) and able to defend itself against potential threats (admissibility). The semantics proposed by Dung can defined as follows [16].

Definition 3 (Extension Semantics)**.** *Given $\mathcal{F} = \langle A, R\rangle$, an admissible set $S \subseteq A$ is :*

- *a complete extension iff it contains every argument that it defends ;*
- *a preferred extension iff it is a $\subseteq$-maximal complete extension ;*
- *the unique grounded extension iff it is the $\subseteq$-minimal complete extension ;*
- *a stable extension iff it attacks every argument in $A \setminus S$.*

The sets of extensions of an AF $\mathcal{F}$, for these semantics, are denoted (respectively) $\text{co}(\mathcal{F})$, $\text{pr}(\mathcal{F})$, $\text{gr}(\mathcal{F})$ and $\text{st}(\mathcal{F})$.

Given any semantics σ, we can define the status of any (set of) argument(s), namely *skeptically accepted* (belonging to each σ-extension), *credulously accepted* (belonging to some σ-extension) and *rejected* (belonging to no σ-extension). Given an AF $\mathcal{F}$ and a semantics σ, we use (respectively) $\text{sk}_\sigma(\mathcal{F})$, $\text{cr}_\sigma(\mathcal{F})$ and $\text{rej}_\sigma(\mathcal{F})$ to denote these sets of arguments.

Example 2. *We consider again $\mathcal{F}$ given at Figure 1. Its extensions $\sigma(\mathcal{F})$ for the different semantics $\sigma \in \{\text{co}, \text{pr}, \text{gr}, \text{st}\}$, as well as the sets of accepted arguments, are given at Table 1.*

σ	$\sigma(\mathcal{F})$	$\text{cr}_\sigma(\mathcal{F})$	$\text{sk}_\sigma(\mathcal{F})$
co	$\{e\}, \{d, e\}, \{b, c, e\}$	$\{b, c, d, e\}$	$\{e\}$
pr	$\{d, e\}, \{b, c, e\}$	$\{b, c, d, e\}$	$\{e\}$
gr	$\{e\}$	$\{e\}$	$\{e\}$
st	$\{d, e\}, \{b, c, e\}$	$\{b, c, d, e\}$	$\{e\}$

TABLE 1 – Extensions and Accepted Arguments of $\mathcal{F}$ for $\sigma \in \{\text{co}, \text{pr}, \text{gr}, \text{st}\}$

For more details about argumentation semantics, we refer the interested reader to [16, 2].

2.2 Incomplete AFs

Now, we describe Incomplete Argumentation Frameworks [8, 6, 5].

Definition 4 (Incomplete AF)**.** *An* Incomplete Argumentation Framework *(IAF) is a tuple $\mathcal{I} = \langle A, A^?, R, R^?\rangle$, where A and $A^?$ are disjoint sets of arguments, and $R, R^? \subseteq (A \cup A^?) \times (A \cup A^?)$ are disjoint sets of attacks.*

Elements from A and R are certain arguments and attacks, *i.e.* the agent is sure that they appear in the framework. On the opposite, $A^?$ and $R^?$ represent uncertain arguments and attacks. For each of them, there is a doubt about their actual existence.

Example 3. *Let us consider $\mathcal{I} = \langle A, A^?, R, R^?\rangle$ given at Figure 2. We use plain nodes and arrows to represent certain arguments and attacks,* i.e. $A = \{a, b, c, d, e\}$ *and* $R = \{(b, a), (c, a), (d, b), (d, c)\}$*. Uncertain arguments are represented as dashed square nodes (*i.e. $A^? = \{f\}$*) and uncertain attacks are represented as dotted arrows (*i.e. $R^? = \{(e, a), (f, d)\}$*).*

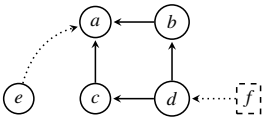


FIGURE 2 – The IAF $\mathcal{I}$

The notion of completion in abstract argumentation was first defined in [8] for Partial AFs (*i.e.* IAFs with $A^? = \emptyset$), and then adapted to IAFs. Intuitively, a completion is a classical AF which describes a situation of the world coherent with the uncertain information encoded in the IAF.

Definition 5 (Completion of an IAF)**.** *Given $\mathcal{I} = \langle A, A^?, R, R^?\rangle$, a completion of $\mathcal{I}$ is $\mathcal{F} = \langle A', R'\rangle$, such that $A \subseteq A' \subseteq A \cup A^?$ and $R_{|A'} \subseteq R' \subseteq R_{|A'} \cup R^?_{|A'}$, where $R_{|A'} = R \cap (A' \times A')$ (and similarly for $R^?_{|A'}$).*

The set of completions of an IAF $\mathcal{I}$ is denoted $\text{comp}(\mathcal{I})$.

Example 4. *We consider again the IAF from Figure 2. Its set of completions is described at Figure 3. Since both a and e are certain arguments, the attack (e, a) appears in half of the completions. On the contrary, the attack (f, d) appears in half of the completions* where f appears.

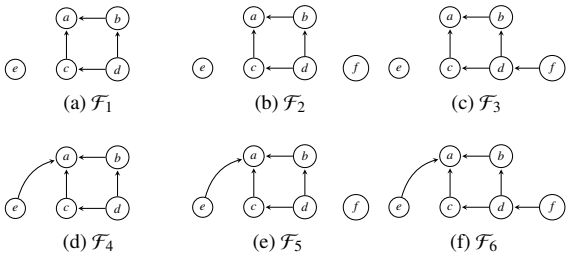


FIGURE 3 – The Completions of $\mathcal{I}$

Concerning the question of compact representation of a set of AFs by the mean of an incomplete AF, the following example proves that some sets of AFs (even simple ones) cannot be represented by an IAF.

Example 5. *Suppose that the result of revising an AF [10] is the set* $\mathfrak{F} = \{\mathcal{F}_1 = \langle\{a,b\},\{(b,a)\}\rangle, \mathcal{F}_2\langle\{a,c\},\{(c,a)\}\rangle\}$. *The question is to determine whether this set can be compactly represented by a single IAF. Reasoning towards a contradiction, we suppose that there is an IAF* $\mathcal{I} = \langle A, A^?, R, R^?\rangle$ *such that* $\text{comp}(\mathcal{I}) = \mathfrak{F}$. *Since* a *belongs to both* $\mathcal{F}_1$ *and* $\mathcal{F}_2$, *it must belong to the certain arguments* A. *On the contrary, the uncertain arguments are* $A^? = \{b,c\}$, *each of them belong to some (but not all) completions. Already we face a contradiction : since* $A = \{a\}$ *and* $A^? = \{b,c\}$, *there should be some completions that only contain* a, *and some completions that contain the three arguments* a, b, c. *This is not the case. So* $\mathcal{I}$ *does not exist.*

3 Constrained IAFs

Now we introduce the Constrained Incomplete Argumentation Frameworks, that generalize IAFs by adding a constraint on the set of possible completions.

3.1 Constraints on Completions

Intuitively, for a given $\mathcal{I}$, a constrained version of it is a pair $\langle\mathcal{I}, C\rangle$ where $C \subseteq \text{comp}(\mathcal{I})$. Then, reasoning on $\langle\mathcal{I}, C\rangle$ requires to use only C instead of the full set of completions of $\mathcal{I}$. For instance, let us consider applications where the uncertainty about the world is encoded as a set of AFs or a set of extensions (for instance, revision [10, 12] or merging of argumentation frameworks [8, 13]). This set of AFs or extensions may not be encodable in a single IAF, while being representable with a single Constrained IAF.

But rather than defining the constraint with a set of completions, we define a logical language to express information on the structure of an AF, *i.e.* a propositional langage such that the models of a formula correspond to AFs, inspired by [9] for selecting extensions. This is a more compact representation of the constraint.

Definition 6 (Constraint)**.** *Given* A *a set of arguments, we define the set of propositional atoms* $Prop_A = Arg_A \cup Att_A$ *where*

- $Arg_A = \{\text{arg}_a \mid a \in A\}$;
- $Att_A = \{\text{att}_{a,b} \mid (a,b) \in A \times A\}$.

Then, $\mathcal{L}_A$ *is the propositional language built from* $Prop_A$ *with classical connectives* $\{\neg, \vee, \wedge\}$. *A constraint over* A *is any formula from* $Prop_A$.

The satisfaction of a constraint by an AF is defined as follows.

Definition 7 (Constraint Satisfaction)**.** *Given* A *a set of arguments, let* $\phi \in \mathcal{L}_A$ *be a formula. The set of models of* ϕ *is denoted* $\text{mod}(\phi)$. *We say that an AF* $\mathcal{F} = \langle A', R\rangle$ *with* $A \subseteq A$ *and* $R \subseteq A' \times A'$ satisfies ϕ *if and only if there is a model* $\omega \in \text{mod}(\phi)$ *such that*

- $A' = \{a \in A \mid \omega(\text{arg}_a) = \top\}$, *and*
- $R = \{(a,b) \in A \times A \mid \omega(\text{att}_{a,b}) = \top\}$.

The intuition behind Definition 7 consists in assigning to a variable arg_a the truth value *true* if and only if the argument a belongs to the AF $\mathcal{F}$, and similarly the variable $\text{att}_{a,b}$ is assigned *true* if and only if there is an attack (a,b) in $\mathcal{F}$. An AF then satisfies the formula if and only if it can be mapped to a propositional interpretation that satisfies the formula.

3.2 Definition and Expressivity of CIAFs

Definition 8 (Constrained IAF)**.** *A* Constrained Incomplete Argumentation Framework *(CIAF) is a tuple* $C = \langle A, A^?, R, R^?, \phi\rangle$, *where* $\langle A, A^?, R, R^?\rangle$ *is an IAF, and* $\phi \in \mathcal{L}_{A\cup A^?}$ *is a constraint.*

The constraint ϕ is used to select a subset of the completions of the IAF $\mathcal{I}_C = \langle A, A^?, R, R^?\rangle$. The completions of a CIAF are then defined as follows.

Definition 9 (Completions of a CIAF)**.** *Given* $C = \langle A, A^?, R, R^?, \phi\rangle$ *a CIAF, we define its set of completions by*

$$\text{comp}(C) = \{c \in \text{comp}(\mathcal{I}_C) \mid c \text{ satisfies } \phi\}$$

where $\mathcal{I}_C = \langle A, A^?, R, R^?\rangle$.

Example 6. *Let* $C = \langle A, A^?, R, R^?, \phi\rangle$ *be a CIAF such that* $\mathcal{I}_C$ *is the IAF from Figure 2, and* $\phi = \text{att}_{e,a} \wedge \text{arg}_f$. *Recall that the completions of* $\mathcal{I}_C$ *are given at Figure 3. Only two of them satisfy* ϕ, *namely* $\mathcal{F}_5$ *(Fig. 3e) and* $\mathcal{F}_6$ *(Fig. 3f). In the other completions of* $\mathcal{I}_C$, *either there is no attack* (e,a), *or there is no argument* f, *thus* ϕ *is no satisfied by these completions. So* $\text{comp}(C) = \{\mathcal{F}_5, \mathcal{F}_6\}$.

Let us mention that, in order to be meaningful, the constraint ϕ must satisfy some conditions. Indeed, there must be at least one model of ϕ such that arg_a is true for each $a \in A$, $\text{att}_{a,b}$ is true for each $(a,b) \in R$, and $\text{att}_{a,b}$ is false for each $(a,b) \in ((A\cup A')\times(A\cup A'))\setminus(R\cup R^?)$. Otherwise, $\text{comp}(C)$ is trivially empty. More generally, a CIAF C is *over-constrained* when $\text{comp}(C) = \emptyset$.

Now, we focus on the expressivity of CIAFs, *i.e.* given a set of AFs (or a set of extensions), is there a CIAF such that its completions (or the extensions of its completions) correspond to the given set ? We show that, in both cases, the answer is yes.

Compact Representation of a Set of AFs First, we define a particular formula, that is only satisfied by one given AF.

Definition 10. *Given* A *a set of arguments, and* $\mathcal{F} = \langle A', R\rangle$ *with* $A' \subseteq A$, *and* $R \subseteq A' \times A'$, *we define* $\psi_\mathcal{F} \in \mathcal{L}_A$ *as*

$$\begin{aligned}\psi_\mathcal{F} \quad = \quad & (\bigwedge_{a\in A'} \text{arg}_a) \wedge (\bigwedge_{a\in A\setminus A'} \neg\,\text{arg}_a) \\ & \wedge(\bigwedge_{(a,b)\in R} \text{att}_{a,b}) \wedge (\bigwedge_{(a,b)\in(A\times A)\setminus R} \neg\,\text{att}_{a,b})\end{aligned}$$

Proposition 1. *Let* $\mathfrak{F} = \{\mathcal{F}_1 = \langle A_1, R_1\rangle, \ldots, \mathcal{F}_n = \langle A_n, R_n\rangle\}$ *be a set of AFs. There is a CIAF* $C = \langle A, A^?, R, R^?, \phi\rangle$ *such that* $\text{comp}(C) = \mathfrak{F}$.

Démonstration. Let us build a CIAF $C = \langle A, A^?, R, R^?, \phi\rangle$ such that $\text{comp}(C) = \mathfrak{F}$. To do that, we first choose $A = \emptyset$ and $A^? = \bigcup_{i=1}^n A_i$, *i.e.* all the arguments that appear in an AF from $\mathfrak{F}$ are uncertain. Similarly, all the attacks are uncertain, *i.e.* $R = \emptyset$ and $R^? = \bigcup_{i=1}^n R_i$. With all these choices, we define an IAF that has all the possible completions on arguments and attacks from $\mathfrak{F}$. In order to restrict the completions to exactly the AFs in $\mathfrak{F}$, we define $\phi = \bigvee_{i=1}^n \psi_{\mathcal{F}_i}$, where $\psi_{\mathcal{F}_i}$ is the formula that is only satisfied by the AF $\mathcal{F}_i$, following Definition 10. The AFs that satisfy ϕ are exactly the ones in $\mathfrak{F}$, so we have $\text{comp}(C) = \mathfrak{F}$. □

Let us exemplify this result.

Example 7. *We continue Example 5. For* $\mathfrak{F} = \{\mathcal{F}_1 = \langle\{a,b\},\{(b,a)\}\rangle, \mathcal{F}_2 = \langle\{a,c\},\{(c,a)\}\rangle\}$, *we define* $C = \langle A, A^?, R, R^?, \phi, \rangle$, *with* $A = \emptyset$, $A^? = \{a,b,c\}$, $R = \emptyset$, $R^? = \{(b,a),(c,a)\}$, $\phi = \psi_{\mathcal{F}_1} \vee \psi_{\mathcal{F}_2}$, *where*

$$\psi_{\mathcal{F}_1} = \text{arg}_a \wedge \text{arg}_b \wedge \neg\, \text{arg}_c \wedge \text{att}_{b,a} \wedge (\bigwedge_{(x,y)\in(\{a,b,c\}\times\{a,b,c\})\setminus\{(b,a)\}} \neg\, \text{att}_{x,y})$$

and

$$\psi_{\mathcal{F}_2} = \text{arg}_a \wedge \neg\, \text{arg}_b \wedge \text{arg}_c \wedge \text{att}_{c,a} \wedge (\bigwedge_{(x,y)\in(\{a,b,c\}\times\{a,b,c\})\setminus\{(c,a)\}} \neg\, \text{att}_{x,y})$$

We have $\text{comp}(C) = \mathfrak{F}$.

Representing a Set of Extensions Now, we focus on the expressibility of a set of extensions with a CIAF.

Proposition 2. *Let* $\mathfrak{E} = \{E_1, \ldots, E_n\}$ *be a set of extensions, and* σ *a semantics. There is a CIAF* $C = \langle A, A^?, R, R^?, \phi\rangle$ *such that* $\bigcup_{c\in\text{comp}(C)} \sigma(c) = \mathfrak{E}$.

Démonstration. First, let us define $A = \bigcup_{i=1}^n E_i$, *i.e.* it is the set of all the arguments that appear in some extension. Then, for each $E_i \in \mathfrak{E}$, we define $\mathcal{F}_i = \langle A, R_i\rangle$ such that $R_i = \{(a,b) \mid a \in E_I, b \in A \setminus E_i\}$, *i.e.* each argument in E_i is unattacked, and it attacks all the arguments that are not in the extension. For any σ defined in this paper, [1] E_i is the only extension of $\mathcal{F}_i$. Thus, $\bigcup_{i=1}^n \sigma(\mathcal{F}_i) = \mathfrak{E}$. From Proposition 1, there is C such that $\text{comp}(C) = \{\mathcal{F}_1, \ldots, \mathcal{F}_n\}$. This concludes the proof. □

3.3 Complexity Issues

In this section, we investigate the complexity of reasoning with CIAFs.

1. And arguably most semantics defined in the literature.

Observation 3. *Verifying whether an AF (or a completion) satisfies a constraint* ϕ *is a polynomial task : the correspondance between an AF and an interpretation* ω *described in Definition 7 can be done polynomially, as well as checking whether* ω *satisfies* ϕ.

Now we study the complexity of credulous and skeptical reasoning. More specifically,we study the following decision problems :

Cred-σ Given $C = \langle A, A^?, R, R^?, \phi\rangle$ a CIAF, $a \in A$, and σ a semantics, is $a \in \bigcup_{\mathcal{F}\in\text{comp}(C)} \bigcup_{S\in\sigma(\mathcal{F})} S$?

Skep-σ Given $C = \langle A, A^?, R, R^?, \phi\rangle$ a CIAF, $a \in A$, and σ a semantics, is $a \in \bigcap_{\mathcal{F}\in\text{comp}(C)} \bigcap_{S\in\sigma(\mathcal{F})} S$?

These problems correspond to *possible credulous acceptance* (PCA) and *necessary skeptical acceptance* (NSA) for IAFs [5]. We prove that complexity does not increase from IAFs to CIAFs.

Proposition 4. *The following hold :*

1. *For* $\sigma \in \{\text{ad}, \text{st}, \text{co}, \text{gr}, \text{pr}\}$, Cred-$\sigma$ *is* NP*-complete.*
2. *For* $\sigma \in \{\text{st}, \text{co}, \text{gr}\}$, Skep-$\sigma$ *is* coNP*-complete.*
3. Skep-pr *is* Π_2^P*-complete.*

Démonstration. Hardness comes from known results for IAFs [5], and the fact that any IAF can be transformed into a CIAF with a tautological constraint. Membership for item *1.* is proved by the following non-deterministic polynomial algorithm. Guess a completion $\mathcal{F} = \langle A', R'\rangle$ of C, and a set of arguments $S \subseteq A'$ s.t. $a \in S$. Checking whether S is a σ-extension of $\mathcal{F}$ is polynomial for $\sigma \in \{\text{ad}, \text{st}, \text{co}, \text{gr}\}$. For $\sigma = \text{pr}$, checking whether S is admissible is enough for proving that a belongs to some preferred extension. Hence the result. The same kind of technique can be used for membership for item *2.* For item *3.*, the argument from [5] still holds for CIAFs : it can be translated in a $\forall X, \exists Y, f(X,Y)$ formula, where $f(X,Y)$ is polynomially decidable. □

Concerning skeptical reasoning, the problem is trivial under $\sigma = \text{ad}$, as usual, since $\emptyset$ is admissible in any AF. So there is no skeptically accepted argument in any completion.

Finally, let us mention that credulous and skeptical acceptance can be generalized to sets of arguments. The generalized versions consists, respectively, in determining whether a given set of arguments S satisfies $S \subseteq \bigcup_{\mathcal{F}\in\text{comp}(C)} \bigcup_{S\in\sigma(\mathcal{F})} S$, or $S \subseteq \bigcap_{\mathcal{F}\in\text{comp}(C)} \bigcap_{S\in\sigma(\mathcal{F})} S$. The generalized versions keep the same complexity.

4 CIAFs and Extension Enforcement

4.1 Expansion-based Enforcement

Let us introduce the concepts of expansion of an AF, and extension enforcement [3].

Definition 11. *Let $\mathcal{F} = \langle A, R\rangle$ be an AF. An* expansion *of $\mathcal{F}$ is an AF $\mathcal{F}' = \langle A \cup A', R \cup R'\rangle$ s.t. $A' \neq \emptyset$ and $A \cap A' = \emptyset$. An expansion is called* normal *if $\forall (a, b) \in R'$, $a \in A'$ or $b \in A'$. Moreover, a normal expansion is* strong *(resp.* weak*) if $\forall (a, b) \in R'$, $a \notin A$ (resp. $b \notin A$).*

In words, an expansion adds some arguments, and possibly attacks. In the case of a normal expansion, the only added attacks concern at least one new arguments, *i.e.* the attacks between the former arguments are not modified. Finally, a normal expansion is strong (resp. weak) if it adds only strong (resp. weak) arguments, *i.e.* arguments that are not attacked (resp. do not attack) the former arguments. The fact that $\mathcal{F}'$ is an expansion of $\mathcal{F}$ is denoted $\mathcal{F} \preceq_E \mathcal{F}'$ (and normal, strong, weak expansions are respectively denoted by $\preceq_N, \preceq_S, \preceq_W$).

Definition 12. *Given an AF $\mathcal{F} = \langle A, R\rangle$, a set of arguments $S \subseteq A$, and a semantics σ, the AF $\mathcal{F}'$ is a normal (resp. strong, weak) σ-enforcement of S in $\mathcal{F}$ iff $\mathcal{F}'$ is a normal (resp. strong, weak) expansion of $\mathcal{F}$, and $\exists E \in \sigma(\mathcal{F}')$ s.t. $S \subseteq E$.*

Definition 12 only considers "non-strict" enforcement, *i.e.* the desired set of arguments must be included in an extension of the new AF. Strict enforcement is defined in a similar manner, but the desired set of arguments must exactly correspond to an extension.

[3] gives some (im)possibility results for this kind of operation. However, some possibility results rely on examples that are not representative of realistic argumentation-based dialogues. The following example is inspired by [3, Theorem 4].

Example 8. *Let $\mathcal{F} = \langle A, R\rangle$ be the AF given at Figure 1. Recall that its stable extensions are* $\text{st}(\mathcal{F}) = \{\{d, e\}, \{b, c, e\}\}$. *Now let $S = \{a, d\}$ be the set of arguments to be enforced. We can define the (strong) expansion $\mathcal{F}' = \langle A \cup \{x\}, R \cup R'\rangle$ where x is a fresh argument, and $R' = \{(x, y) \mid y \in A \setminus S\}$. $\mathcal{F}'$ is shown at Figure 4. $\mathcal{F}'$ has a*

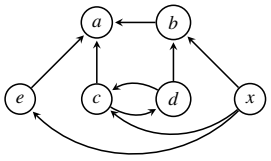


Figure 4 – The Expansion $\mathcal{F}'$ Enforces $S = \{a, d\}$

single stable extension $\{x, a, d\}$, thus it is a strong enforcement of S in $\mathcal{F}$.

Example 8 illustrates the (theoretical) possibility to enforce any (conflict-free) set of arguments if strong (or normal) expansions are permitted. However, in an application context like dialogue (*e.g.* argumentation-based negotiation [15] or persuasion [7]), the existence of an "ultimate" attacker like x, that defeats all the undesired arguments, is unlikely.

4.2 Enforcement as Credulous Acceptability in CIAFs

To handle the problem highlighted by Example 8, we propose to take into account the set of arguments and attacks that an agent has at her disposal for participating to the debate. This means that we parameterize the expansion operation by the set of possible expanded AFs resulting of using some of the available arguments and attacks.

Definition 13. *Given $\mathcal{F} = \langle A, R\rangle$ an AF, $\mathcal{A}$ a set of available arguments s.t. $A \cap \mathcal{A} = \emptyset$, and $\mathcal{R} \subseteq ((A \cup \mathcal{A}) \times (A \cup \mathcal{A})) \setminus (A \times A)$, we say that $\mathcal{F}' = \langle A', R'\rangle$ is a $\mathcal{A}$-$\mathcal{R}$-parameterized expansion of $\mathcal{F}$ (denoted by $\mathcal{F} \preceq^{\mathcal{A},\mathcal{R}} \mathcal{F}'$) iff*

— $\mathcal{F} \preceq_E \mathcal{F}'$,
— $A \subseteq A' \subseteq A \cup \mathcal{A}$,
— $R' = (R \cup \mathcal{R}) \cap (A' \times A')$.

We use $\preceq_N^{\mathcal{A},\mathcal{R}}$ (resp. $\preceq_S^{\mathcal{A},\mathcal{R}}$, $\preceq_W^{\mathcal{A},\mathcal{R}}$) to denote $\mathcal{A}$-$\mathcal{R}$-parameterized normal (resp. strong, weak) expansions, *i.e.* $\mathcal{A}$-$\mathcal{R}$-parameterized expansions where $\mathcal{F}'$ is (additionally) normal (resp. strong, weak).

This definition allows to take into account the arguments and attacks that are actually known by an agent that participates to a debate. We can show that a set of arguments that can be enforced with an arbitrary (strong) expansion (like in Example 8) may not be enforceable with parameterized expansions.

Example 9. *We continue Example 8. Suppose that the available arguments and attacks are $\mathcal{A} = \{f, g\}$ and $\mathcal{R} = \{(f, c), (g, b)\}$. Figure 5 depicts the agent's possible actions : say nothing (*i.e. *keep the initial AF, Fig. 5a), say "f attacks c" (Fig 5b), say "g attacks b" (Fig 5c), or both (Fig. 5d). We observe that in all the possible cases,*

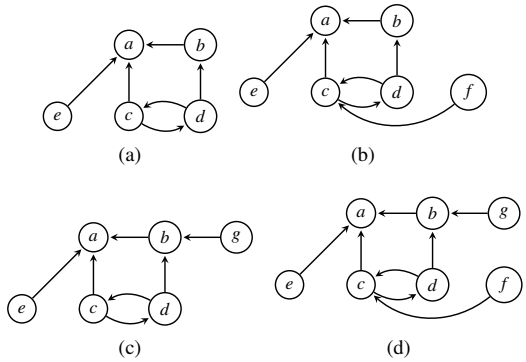


Figure 5 – The Agent's Possible Actions

$S = \{a, d\}$ is not enforced, since a is never defended against e.

What we call here the “possible actions" of the agent can actually be seen as the set of completions of a CIAF, and the possibility of enforcing a set of arguments correspond to the credulous acceptance of this set w.r.t. the CIAF.

Definition 14. *Given $\mathcal{F}$ an AF, $\mathcal{A}, \mathcal{R}$ a set of arguments and a set of attacks, and $X \in \{E, N, S, W\}$ denoting the type of expansion, we define $\mathfrak{F} = \{\mathcal{F}\} \cup \{\mathcal{F}' \mid \mathcal{F} \preceq_X^{\mathcal{A},\mathcal{R}} \mathcal{F}'\}$. Then, $C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}$ is a CIAF s.t.* $\text{comp}(C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}) = \mathfrak{F}$.

The existence of $C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}$ is guaranteed by Proposition 1. The construction given in the corresponding proof provides a suitable $C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}$. However, other CIAFs can be defined, for instance all the arguments and attacks from the initial $\mathcal{F}$ can be defined as certain elements (*i.e.* the sets A and R).

Example 10. *The agent's possible actions (Figure 5) are the completions of the following $C = \langle A, A^?, R, R^?, \top \rangle$, depicted in Figure 6.*

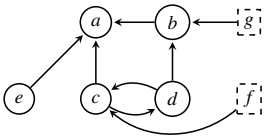


Figure 6 – CIAF C Corresponding to the Agent's Possible Actions

In this simple example, the constraint is a tautology, which means that all the completions of the underlying IAF $\mathcal{I}_C$ are completions of C. However, we can easily imagine scenarios where an agent has some particular constraint on the arguments that she can use, *e.g.* she cannot use both f and g because their is a high communication cost, which prevents using too many arguments. Then, $\phi = \neg(f \wedge g)$ leads to removing the completion depicted in Figure 5d from the CIAF completions.

Proposition 5. *Given an AF $\mathcal{F} = \langle A, R \rangle$, a set of arguments $S \subseteq A$, $X \in \{E, N, S, W\}$, $\mathcal{A}$ a set of arguments and $\mathcal{R}$ a set of attacks, and a semantics σ, S can be σ-enforced in $\mathcal{F}$ by the mean of a $\mathcal{A}$-$\mathcal{R}$-parameterized X-expansion iff S is credulously accepted in $C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}$ w.r.t. σ.*

Démonstration. The result is a consequence of the definitions : S can be enforced iff one of the “possible actions" yields an AF $\mathcal{F}'$ s.t. $S \subseteq S'$ for some $S' \in \sigma(\mathcal{F}')$. This is equivalent to saying that S belongs to some σ-extension of some completion of $C_{\mathcal{F},X}^{\mathcal{A},\mathcal{R}}$. □

Observe that this result holds for non-strict enforcement, as given in Definition 12. Strict enforcement requires, instead, the notion of extension verification [18] for CIAFs, *i.e.* we must check that the set of arguments S is actually an extension of one of the completions.

5 Related Work

Besides IAFs, our contribution is related to other previous works. Using propositional formulas as constraints in an argumentation framework has been originally proposed in [9], which defines Constrained Argumentation Frameworks (CAFs). In this setting, the propositional formula is a constraint on arguments that is used for selecting the best extensions. Intuitively, we use here the constraint in CIAFs in a similar way, but for selecting completions of a IAF instead of selecting the extensions of a (classical) AF.

In this paper, we have shown how to represent any set of extensions with a single CIAF. The question of representing sets of extensions has already arisen in classical AFs. This corresponds to the notion of realizability in the literature [4, 17], *i.e.* given a set of extensions E and a semantics σ, is there an AF $\mathcal{F}$ such that $\sigma(\mathcal{F}) = E$. Existing results show that this is not possible in general for the classical semantics. The non-realizability of some sets of extensions is the reason why some operations (like belief revision or merging) cannot be easily adapted to AFs, as mentioned in the introduction. With Proposition 2, we continue this line of research, by proving the realizability of any set of extensions by the mean of CIAFs.

Regarding extension enforcement, it has been proven that (non-strict) enforcement is also NP-complete [23] for another type of authorized change : argument-fixed enforcement [11], where the set of arguments cannot be modified, but all the attacks (or non-attacks) can be questions. Although this is out of the scope of this paper, we believe that this kind of enforcement can also be captured by the CIAF setting, which will allow to define a parameterized version of argument-fixed enforcement. The parameters $\mathcal{A}$ and $\mathcal{R}$ are also reminiscent of the “control part" of Control AFs [14], that allows to enforce a set of arguments in presence of uncertainty.

6 Conclusion

We have introduced Constrained Incomplete Argumentation Frameworks (CIAFs), that generalize IAFs by the addition of a constraint over the set of completions. This new framework increases the expressivity of IAFs without a gap of complexity, and paves the way for the definition of revision or merging operators for AFs that return a CIAF, *i.e.* a more compact result that a (potentially exponentially large) set of AFs or extensions. However, the CIAF that we have exhibited here to prove the representability of any set of AFs or extensions may not be a suitable solution in scenarios like belief revision or belieg merging, where the notion of minimal change is important. We will study how to generate a CIAF that is optimal in such contexts. Other interesting research tracks are the study of complexity for other decision problems (*e.g.* necessary credulous accep-

tance and possible skeptical acceptance, defined for IAFs), and the implementation of efficient algorithms (*e.g.* based on Boolean encoding, in the line of [22]). We will also study how to encode other extension enforcement operators as CIAF-based reasoning.

Références

[1] Alchourrón, Carlos E., Peter Gärdenfors et David Makinson: *On the Logic of Theory Change : Partial Meet Contraction and Revision Functions*. J. Symb. Log., 50(2) :510–530, 1985.

[2] Baroni, Pietro, Martin Caminada et Massimiliano Giacomin: *Abstract Argumentation Frameworks and Their Semantics*. Dans Baroni, Pietro, Dov Gabbay, Massimiliano Giacomin et Leendert van der Torre (rédacteurs) : *Handbook of Formal Argumentation*, pages 159–236. College Publications, 2018.

[3] Baumann, Ringo et Gerhard Brewka: *Expanding Argumentation Frameworks : Enforcing and Monotonicity Results*. Dans *Computational Models of Argument : Proceedings of COMMA 2010*, tome 216, pages 75–86, 2010.

[4] Baumann, Ringo, Wolfgang Dvorák, Thomas Linsbichler, Hannes Strass et Stefan Woltran: *Compact Argumentation Frameworks*. Dans *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, tome 263, pages 69–74, 2014.

[5] Baumeister, Dorothea, Daniel Neugebauer et Jörg Rothe: *Credulous and Skeptical Acceptance in Incomplete Argumentation Frameworks*. Dans *Proc. of COMMA'18*, pages 181–192, 2018.

[6] Baumeister, Dorothea, Daniel Neugebauer, Jörg Rothe et Hilmar Schadrack: *Complexity of Verification in Incomplete Argumentation Frameworks.* Dans *Proc. of AAAI'18*, 2018.

[7] Bonzon, Elise, Jérôme Delobelle, Sébastien Konieczny et Nicolas Maudet: *A Parametrized Ranking-Based Semantics for Persuasion*. Dans *Scalable Uncertainty Management - 11th International Conference, SUM 2017, Granada, Spain, October 4-6, 2017, Proceedings*, pages 237–251, 2017.

[8] Coste-Marquis, Sylvie, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex et Pierre Marquis: *On the merging of Dung's argumentation systems*. Artif. Intell., 171(10-15) :730–753, 2007.

[9] Coste-Marquis, Sylvie, Caroline Devred et Pierre Marquis: *Constrained Argumentation Frameworks*. Dans *Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006*, pages 112–122, 2006.

[10] Coste-Marquis, Sylvie, Sébastien Konieczny, Jean-Guy Mailly et Pierre Marquis: *On the Revision of Argumentation Systems : Minimal Change of Arguments Statuses*. Dans *Proc. of KR'14*, 2014.

[11] Coste-Marquis, Sylvie, Sébastien Konieczny, Jean-Guy Mailly et Pierre Marquis: *Extension Enforcement in Abstract Argumentation as an Optimization Problem*. Dans *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2876–2882, 2015.

[12] Coste-Marquis, Sylvie, Sébastien Konieczny, Jean Guy Mailly et Pierre Marquis: *A Translation-based Approach for Revision of Argumentation Frameworks*. Dans *Proc. of JELIA'14*, pages 77–85, 2014.

[13] Delobelle, Jérôme, Adrian Haret, Sébastien Konieczny, Jean Guy Mailly, Julien Rossit et Stefan Woltran: *Merging of Abstract Argumentation Frameworks*. Dans *Proc. of KR'16*, pages 33–42, 2016.

[14] Dimopoulos, Yannis, Jean Guy Mailly et Pavlos Moraitis: *Control Argumentation Frameworks*. Dans *Proc. of AAAI'18*, pages 4678–4685, 2018.

[15] Dimopoulos, Yannis, Jean-Guy Mailly et Pavlos Moraitis: *Argumentation-based Negotiation with Incomplete Opponent Profiles*. Dans *Proc. of AAMAS'19*, pages 1252–1260, 2019.

[16] Dung, Phan Minh: *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*. Art. Intel., 77 :321–357, 1995.

[17] Dunne, Paul E., Wolfgang Dvorák, Thomas Linsbichler et Stefan Woltran: *Characteristics of multiple viewpoints in abstract argumentation*. Artif. Intell., 228 :153–178, 2015.

[18] Fazzinga, Bettina, Sergio Flesca et Filippo Furfaro: *Revisiting the Notion of Extension over Incomplete Abstract Argumentation Frameworks*. Dans *Proc. of IJCAI'20*, pages 1712–1718, 2020.

[19] Katsuno, Hirofumi et Alberto O. Mendelzon: *On the Difference between Updating a Knowledge Base and Revising It*. Dans *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Cambridge, MA, USA, April 22-25, 1991*, pages 387–394, 1991.

[20] Katsuno, Hirofumi et Alberto O. Mendelzon: *Propositional Knowledge Base Revision and Minimal Change*. Artif. Intell., 52(3) :263–294, 1992.

[21] Konieczny, Sébastien et Ramón Pino Pérez: *Merging Information Under Constraints : A Logical Framework*. J. Log. Comput., 12(5) :773–808, 2002.

[22] Niskanen, Andreas, Daniel Neugebauer, Matti Järvisalo et Jörg Rothe: *Deciding Acceptance in Incomplete Argumentation Frameworks*. Dans *Proc. of AAAI'20*, pages 2942–2949, 2020.

[23] Wallner, Johannes Peter, Andreas Niskanen et Matti Järvisalo: *Complexity Results and Algorithms for Extension Enforcement in Abstract Argumentation*. J. Artif. Intell. Res., 60 :1–40, 2017.

Actes JIAF 2021

Contraction et effacement des croyances pour des fragments propositionnels

Nadia Creignou[1] **Raida Ktari**[2] **Odile Papini**[1]

[1] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[2] ISIMS,OLID LR 19ES21, Université de Sfax, Tunisie.
{nadia.creignou, odile.papini}@univ-amu.fr, raida.ktari@isims.usf.tn

Résumé

Le changement de croyances dans le cadre de fragments de la logique propositionnelle est l'objet d'un intérêt croissant. Le raffinement d'opérateurs de changement de croyances a été proposé pour adapter des opérateurs connus afin qu'ils opèrent dans des fragments et que leur résultat reste dans le fragment considéré. Alors que la notion de raffinement permet de définir des opérateurs concrets de révision et de mise à jour adaptés aux fragments propositionnels cette notion nécessite d'être adaptée pour la contraction et l'effacement. Nous proposons une notion plus spécifique de raffinement pour ces opérations, appelée *raffinement raisonnable*. Celle-ci nous permet de définir des opérateurs de contraction et d'effacement raffinés qui satisfont les postulats de base. Nous étudions les propriétés logiques de deux opérateurs concrets pour la contraction et pour l'effacement. Notre approche ne se limite pas au fragment de Horn mais s'applique à de nombreux fragments propositionnels.

Abstract

Recently, belief change within the framework of fragments of propositional logic has gained attention. In the context of belief change it has been proposed to refine existing operators so that they operate within propositional fragments, and that the result of belief change remains in the fragment under consideration. Whereas the notion of refinement allowed one to define concrete rational operators adapted to propositional fragments in the context of revision and update, it has to be specified for contraction and erasure. We propose a specific notion of refinement for contraction and erasure operators, called *reasonable refinement*. This allows us to provide refined contraction and erasure operators that satisfy the basic postulates. We study the logical properties of reasonable refinement of two model-based contraction (resp. erasure) operators. Our approach is not limited to the Horn fragment but applicable to many propositional fragments.

1 Introduction

Le changement de croyances est une thématique importante dans le domaine de la représentation des connaissances et du raisonnement en intelligence artificielle. Il s'agit d'étudier comment un agent rationnel modifie ses croyances en présence d'une nouvelle information. Le changement peut prendre différentes formes, introduction d'une information fraîchement acquise, retrait d'une croyance devenue obsolète, qui conduisent à différents types de changement de croyances. La révision de croyances consiste à introduire une nouvelle information dans un environnement statique, tandis que la mise à jour s'effectue dans un environnement évolutif où la nouvelle information reflète un changement qui s'est produit dans l'environnement de l'agent. La contraction de croyances porte sur le retrait de croyances dans un environnement statique alors que l'effacement a lieu dans un environnement dynamique où les croyances effacées ne sont plus valides après que l'environnement a changé.

Dans le cadre des approches symboliques, où les croyances d'un agent sont représentées par des théories, le paradigme AGM [1, 15] s'est imposé. Il fournit des postulats de rationalité que tout opérateur raisonnable de révision ou de contraction devrait satisfaire.

Lorsqu'une théorie est représentée par une formule propositionnelle, Katsuno and Mendelzon [18] ont reformulé les postulats AGM de la révision, certains postulats de la contraction, et proposé des postulats pour la mise à jour et l'effacement. Ils ont contribué à une meilleure compréhension de la différence entre révision et mise à jour, en caractérisant ces opérations dans un cadre logique commun. Lorsque les croyances d'un agent sont représentées par une formule logique, la révision fait évoluer globalement l'ensemble des modèles de cette formule vers les modèles les plus proches de la nouvelle information. En revanche, la

mise à jour fait évoluer localement chaque modèle de cette formule vers les modèles les plus proches de la nouvelle information. Plus récemment, Caridroit et al. [4] ont donné une complète reformulation des postulats AGM pour la contraction et ont proposé un théorème de représentation qui caractérise la contraction en termes de préordres totaux sur les interprétations.

Le changement des croyances dans le cadre de fragments propositionnels a donné lieu à des travaux dont la plupart portent sur la révision [3, 12, 30, 22, 5], la fusion [6], la mise à jour [8] et la contraction [11, 26, 2, 28, 10, 29]. Cependant, à notre connaissance, les travaux sur la contraction ont porté seulement sur le fragment de Horn et l'effacement n'a suscité aucune étude.

Nous proposons de nouveaux opérateurs de contraction et de nouveaux opérateurs d'effacement qui puissent opérer sur divers fragments propositionnels (le fragment de Horn inclus). En cela, nous poursuivons notre étude systématique des opérateurs de changement de croyances adaptés aux fragments propositionnels.

Les motivations de notre étude reposent sur les deux observations suivantes :

— Dans de nombreuses applications, le language est restreint *a priori*. Par exemple, des formalisations de connaissances génériques basées sur des règles sont plus faciles à manipuler pour des utilisateurs.

— On dispose de méthodes efficaces de raisonnement pour de nombreux fragments propositionnels, ainsi le résultat de la contraction et le résultat de l'effacement dans ces fragments peuvent être évalués efficacement.

Il semble donc naturel d'étudier dans quelle mesure les opérateurs de contraction et d'effacement connus peuvent être raffinés pour rester dans le langage d'un fragment considéré.

Soit $\mathcal{L}'$ un fragment de la logique propositionnelle et deux formules $\psi, \mu \in \mathcal{L}'$, l'obstacle principal est qu'il n'est pas garanti que le résultat de la contraction, noté $\psi - \mu$, et le résultat de l'effacement, noté $\psi \lhd \mu$, restent dans $\mathcal{L}'$.

Prenons l'exemple, inspiré de celui utilisé dans [18], où les croyances décrivent trois objets A, B, C, dans une pièce. Une table se trouve dans la pièce et les objets peuvent être ou pas sur la table. Soit les propositions a pour *"l'objet A est sur la table"*, b pour *"l'objet B est sur la table"* et c pour *"l'objet C est sur la table"* et supposons que les croyances de l'agent sont représentées par la formule $\psi = a$ qui exprime que l'agent croit que l'objet A est sur la table mais ignore si les objets B et C y sont également ou pas. Supposons que l'agent perde confiance en la croyance selon laquelle si B et C sont sur la table alors A s'y trouve aussi. Il sagit de contracter la formule ψ par la formule $\mu = a \vee \neg b \vee \neg c$. Les formules ψ et μ sont des formules de Horn, cependant le résultat de la contraction par des opérateurs de contraction reposant sur l'opérateur de Dalal [9] ou de Satoh [23] est une formule équivalente à $\phi = (a \vee b) \wedge (a \vee c)$ qui n'est pas une formule de Horn et n'est pas équivalente à une formule de Horn (l'ensemble de ses modèles n'est pas clos par intersection [17]).

Supposons maintenant que les croyances de l'agent sont toujours représentées par la formule $\psi = a$ et qu'un robot est envoyé dans la pièce avec pour instruction de conduire à une situation où les objets ne peuvent pas être tous les trois sur la table. Ce changement est représenté par la formule $\mu = a \wedge b \wedge c$ qui doit être effacée. Les formules ψ et μ sont des formules de Horn, cependant le résultat de l'effacement par des opérateurs d'effacement reposant sur l'opérateur de Forbus [14] ou de Winslett [25] est une formule équivalente à $\phi = (a \vee b) \wedge (a \vee c)$ qui n'est pas une formule de Horn et n'est pas équivalente à une formule de Horn.

La notion de raffinement définie initialement pour la révision [5], a été étendue dans [8] à tout opérateur de changement de croyances défini de $\mathcal{L} \times \mathcal{L}$ sur $\mathcal{L}$ où $\mathcal{L}$ représente la logique propositionnelle. Un raffinement adapte un opérateur de changement de croyances défini dans le cadre propositionnel afin qu'il soit applicable à un fragment propositionnel. Les propriétés de base d'un raffinement sont d'abord de garantir que le résultat de l'opération de changement de croyances reste dans le fragment considéré, ensuite d'approcher le comportement de l'opérateur initial de changement de croyances, en particulier, de conserver le comportement de l'opérateur initial si le résultat est déjà dans le fragment. Notre objectif est de définir des opérateurs raffinés rationnels de changement de croyances qui satisfont les postulats attendus. Contrairement à la révision et à la mise à jour [5, 8] les raffinements d'opérateurs rationnels de contraction (resp. d'effacement) ne satisfont pas nécessairement les postulats de base de la contraction (resp. de l'effacement). Pour que le raffinement d'opérateurs conduise à des opérateurs rationnels pour une classe plus large d'opérateurs, incluant la contraction et l'effacement, nous introduisons une notion spécifique de raffinement appelée *raffinement raisonnable*. Cette notion nous permet de définir des opérateurs rationnels concrets de contraction définis à partir des opérateurs de révision de Dalal et de Satoh et des opérateurs rationnels concrets d'effacement définis à partir des opérateurs de mise à jour de Forbus et de Winslett, adaptés aux fragments de Horn et Krom. Nous étudions les propriétés logiques de ces opérateurs raffinés en termes de satisfaction de postulats.

Une contribution importante de notre étude est la proposition de nouveaux opérateurs rationnels de contraction et de nouveaux opérateurs rationnels d'effacement qui opérent sur les fragments propositionnels.

2 Préliminaires

2.1 Logique propositionnelle

Nous considérons $\mathcal{L}$ le langage de la logique propositionnelle défini sur un ensemble infini dénombrable de variables (atomes) muni des connecteurs logiques usuels $\rightarrow$, $\oplus$, $\vee$, $\wedge$, $\neg$ et des constantes $\top$, $\bot$. Une clause est une disjonction de littéraux. Une clause est dite de *Horn* si elle comporte au plus un littéral positif, de *Krom* si elle comporte au plus deux littéraux. Nous identifions les sous-ensembles de $\mathcal{L}$ comme suit : $\mathcal{L}_{Horn}$ est l'ensemble des formules de $\mathcal{L}$ formées par la conjonction de clauses de Horn, $\mathcal{L}_{Krom}$ est l'ensemble des formules de $\mathcal{L}$ formées par la conjonction de clauses de Krom.

Soit $\mathcal{U}$ un ensemble fini d'atomes. Une interprétation sur $\mathcal{U}$ est représentée par un ensemble $\omega \subseteq \mathcal{U}$ d'atomes évalués à *vrai* ou par son vecteur caractéristique correspondant, de longueur $|\mathcal{U}|$. Par exemple, si nous considérons $\mathcal{U} = \{x_1, \ldots, x_6\}$, l'interprétation $x_1 = x_3 = x_6 = 1$ et $x_2 = x_4 = x_5 = 0$ sera représentée soit par $\{x_1, x_3, x_6\}$ soit par $(1, 0, 1, 0, 0, 1)$. On note $2^{\mathcal{U}}$ l'ensemble de toutes les interprétations sur $\mathcal{U}$. Si une interprétation ω satisfait une formule ϕ, nous l'appelons modèle de ϕ et nous notons $\mathrm{Mod}(\phi)$ l'ensemble des modèles de ϕ. De plus, $\psi \models \phi$ (ϕ est une conséquence logique de ψ) si $\mathrm{Mod}(\psi) \subseteq \mathrm{Mod}(\phi)$ et $\psi \equiv \phi$ (ψ et ϕ sont équivalentes) si $\mathrm{Mod}(\psi) = \mathrm{Mod}(\phi)$. Pour tout fragment $\mathcal{L}' \subseteq \mathcal{L}$, $T_{\mathcal{L}'}(\psi)$ désigne les conséquences logiques de ψ qui sont dans $\mathcal{L}'$, $T_{\mathcal{L}'}(\psi) = \{\phi \in \mathcal{L}' \mid \psi \models \phi\}$.

Soit ψ et μ deux formules propositionnelles, m et m' deux interprétations, nous notons $m \Delta m'$ la différence symétrique entre m et m', $|\Delta|_m^{min}(\mu)$ le nombre minimum de variables propositionnelles sur lesquelles les modèles de μ et m diffèrent et $\Delta_m^{min}(\mu)$ le sous-ensemble minimal de variables propositionnelles sur lesquelles les modèles de μ et m diffèrent.

Un opérateur de changement de croyances est une fonction de $\mathcal{L} \times \mathcal{L}$ dans $\mathcal{L}$.

2.2 Fragments caractérisables de la logique propositionnelle

Nous considérons $\mathcal{B}$ l'ensemble des fonctions booléennes $\beta : \{0,1\}^k \rightarrow \{0,1\}$ qui possèdent les propriétés suivantes : *(i) symétrie* : pour toute permutation σ, $\beta(x_1, \ldots, x_k) = \beta(x_{\sigma(1)}, \ldots, x_{\sigma(k)})$, *(ii)* 0- et 1-*reproduction* : pour tout $x \in \{0,1\}$, $\beta(x, \ldots, x) = x$.

Des exemples de ces fonctions sont la fonction binaire **et** notée $\wedge$, et la fonction ternaire **majorité**, $\mathrm{maj}_3(x, y, z) = 1$ si au moins deux des variables x, y, et z sont mises à 1.

Nous étendons les fonctions booléennes aux interprétations en appliquant ces fonctions aux coordonnées des vecteurs qui les représentent. Ainsi, si $m_1, \ldots m_k \in \{0,1\}^n$, alors $\beta(m_1, \ldots m_k)$ est défini par $(\beta(m_1[1], \ldots m_k[1]), \ldots, \beta(m_1[n], \ldots m_k[n]))$, où $m[i]$ est le i-ème coordonnée de l'interprétation m.

Définition 1 *Soit $\mathcal{M} \subseteq 2^{\mathcal{U}}$ un ensemble d'interprétations et $\beta \in \mathcal{B}$, nous définissons $Cl_\beta(\mathcal{M})$, la* clôture *de $\mathcal{M}$ pour β, comme le plus petit ensemble d'interprétations contenant $\mathcal{M}$ et clos sous l'action de β, c.-à-d., si $m_1, \ldots m_k \in Cl_\beta(\mathcal{M})$, alors $\beta(m_1, \ldots m_k) \in Cl_\beta(\mathcal{M})$.*

Définition 2 *Soit $\beta \in \mathcal{B}$. Un ensemble de formules propositionnelles $\mathcal{L}' \subseteq \mathcal{L}$ est un* β-fragment *(ou un* fragment caractérisable*) si :*

1. *Pour toute formule $\psi \in \mathcal{L}'$,* $\mathrm{Mod}(\psi) = Cl_\beta(\mathrm{Mod}(\psi))$
2. *Pour tout $\mathcal{M} \subseteq 2^{\mathcal{U}}$ avec $\mathcal{M} = Cl_\beta(\mathcal{M})$ il existe une formule $\psi \in \mathcal{L}'$ avec* $\mathrm{Mod}(\psi) = \mathcal{M}$
3. *Si $\phi, \psi \in \mathcal{L}'$, alors $\phi \wedge \psi \in \mathcal{L}'$.*

Les fragments les plus connus de la logique propositionnelle sont $\mathcal{L}_{Horn}$ qui est un $\wedge$-fragment, et $\mathcal{L}_{Krom}$ qui est un maj_3-fragment [17, 24].

2.3 Contraction

La contraction est une opération qui permet de retirer une croyance de l'ensemble des croyances d'un agent, tout en modifiant le moins possible ses croyances initiales. Les opérateurs de contraction que nous étudions sont des fonctions $- : \mathcal{L} \times \mathcal{L} \rightarrow \mathcal{L}$ qui à partir d'une formule ψ qui représente les croyances initiales d'un agent et d'une formule μ qui code l'information à retirer, renvoie une nouvelle formule $\psi - \mu$.

Des postulats que tout opérateur de contraction devrait satisfaire ont d'abord été proposés dans le cadre *A*GM [1], puis reformulés par Katsuno et Mandelzon [18]. Plus récemment Caridroit, Konieczny and Marquis [4] ont revisité ces postulats comme suit. Soit $\psi, \psi_1, \psi_2, \mu, \mu_1, \mu_2 \in \mathcal{L}$.

(C1) $\psi \models \psi - \mu$.
(C2) Si $\psi \not\models \mu$, alors $\psi - \mu \models \psi$.
(C3) Si $\psi - \mu \models \mu$, alors $\models \mu$.
(C4) Si $\psi_1 \equiv \psi_2$ et $\mu_1 \equiv \mu_2$, alors $\psi_1 - \mu_1 \equiv \psi_2 - \mu_2$.
(C5) Si $\psi \models \mu$, alors $(\psi - \mu) \wedge \mu \models \psi$.
(C6) $\psi - (\mu_1 \wedge \mu_2) \models (\psi - \mu_1) \vee (\psi - \mu_2)$.
(C7) Si $\psi - (\mu_1 \wedge \mu_2) \not\models \mu_1$, alors $\psi - \mu_1 \models \psi - (\mu_1 \wedge \mu_2)$.

Les postulats (C1)-(C4) ont été initialement proposés par Katsuno and Mendelzon, et les postulats (C5)-(C7) ont été introduits par Caridroit, Konieczny and Marquis. De plus, ces derniers ont montré qu'un opérateur de contraction correspond à un ensemble de préordres totaux sur les interprétations et ont proposé un théorème de représentation dans le même esprit que celui de Katsuno et Mendelzon pour la révision. Ce théorème utilise la notion *d'assignation fidèle* qui est une fonction qui associe à toute formule propositionnelle ψ un préordre total sur les interprétations $\leq_\psi$ tel

que : 1) Si $\omega \models \psi$ et $\omega' \models \varphi$, alors $\omega =_\psi \omega'$. 2) Si $\omega \models \psi$ et $\omega' \not\models \psi$, alors $\omega <_\psi \omega'$. 3) Si $\psi_1 \equiv \psi_2$ alors $\leq_{\psi_1} = \leq_{\psi_2}$.

Proposition 1 [4] *Un opérateur de contraction − satisfait* (C1)-(C7) *si et seulement si il existe une assignation fidèle qui à chaque formule ψ fait correspondre un préordre total $\leq_\psi$ tel que* $\text{Mod}(\psi - \mu) = \text{Mod}(\psi) \cup min(\text{Mod}(\neg\mu), \leq_\psi)$.

Les opérateurs de contraction et de révision sont étroitement liés. Un opérateur de révision peut être défini à partir d'un opérateur de contraction via l'identité de Levi [21] ($\psi \circ \mu = (\psi - \neg\mu) \wedge \mu$). De même, un opérateur de contraction peut être défini à partir d'un opérateur de révision via l'identité de Harper [16] ($\psi - \mu = \psi \vee (\psi \circ \neg\mu)$). Aussi nous pouvons définir un opérateur de contraction, noté $-_D$, défini par $\text{Mod}(\psi -_D \mu) = \text{Mod}(\psi) \cup \text{Mod}(\psi \circ_D \neg\mu)$ où $\circ_D$ est l'opérateur de révision de Dalal [9], $\text{Mod}(\psi \circ_D \mu) = \{m \in \text{Mod}(\mu) : \exists m' \in \text{Mod}(\psi) \text{ t.q. } |m\Delta m'| = |\Delta|_m^{min}(\mu)\}$. Nous pouvons également définir un opérateur de contraction, noté $-_S$, défini par $\text{Mod}(\psi -_S \mu) = \text{Mod}(\psi) \cup \text{Mod}(\psi \circ_S \neg\mu)$ où $\circ_S$ est l'opérateur de révision de Satoh [23], $\text{Mod}(\psi \circ_S \mu) = \{m \in \text{Mod}(\mu) : \exists m' \in \text{Mod}(\psi) \text{ t.q. } m\Delta m' \in \Delta_m^{min}(\mu)\}$. L'opérateur de contraction $-_D$ satisfait (C1)−(C7) [4] alors que l'opérateur de contraction $-_S$ satisfait (C1)−(C6), mais viole (C7) [19].

Exemple 1 *Nous reprenons l'exemple de l'introduction, $\psi = a$ et $Mod(\psi) = \{\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}\}$, $\mu = a \vee \neg b \vee \neg c$ et $Mod(\neg\mu) = \{\{b, c\}\}$. $Mod(\psi -_D \mu) = Mod(\psi) \cup Mod(\psi \circ_D \neg\mu) = \{\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \{b, c\}\}$. Remarquons que les formules ψ et μ sont dans le fragment de Horn, en revanche le résultat de la contraction n'est pas dans Horn puisque l'ensemble des modèles n'est pas clos par intersection.*

2.4 Effacement

L'effacement, introduit par Katsuno and Mendelzon [18], est à la contraction ce que la mise à jour est à la révision. Intuitivement, effacer une croyance signifie que l'environnement de l'agent a changé de telle sorte que cette croyance peut ne plus être valide. D'un point de vue logique lorsque les croyances d'un agent sont représentées par une formule ψ, effacer une croyance μ de ψ signifie ajouter localement les modèles de $\neg\mu$ aux modèles de ψ. Les opérateurs d'effacement que nous étudions sont des fonctions $\triangleleft : \mathcal{L} \times \mathcal{L} \rightarrow \mathcal{L}$ qui à partir d'une formule ψ qui représente les croyances initiales d'un agent et d'une formule μ à effacer, renvoie une formule $\psi \triangleleft \mu$.

Des postulats que tout opérateur d'effacement devrait satisfaire ont été proposés par Katsuno et Mendelzon [18] dans le même esprit que ceux proposés par la contraction et la mise à jour.

Soit $\psi, \psi_1, \psi_2, \mu, \mu_1, \mu_2 \in \mathcal{L}$.

(E1) $\psi \models \psi \triangleleft \mu$.
(E2) Si $\psi \not\models \mu$, alors $\psi \triangleleft \mu \equiv \psi$.
(E3) Si ψ est satisfaisable et $\not\models \mu$, alors $\psi \triangleleft \mu \not\models \mu$.
(E4) Si $\psi_1 \equiv \psi_2$ et $\mu_1 \equiv \mu_2$, alors $\psi_1 \triangleleft \mu_1 \equiv \psi_2 \triangleleft \mu_2$.
(E5) $(\psi \triangleleft \mu) \wedge \mu \models \psi$.
(E8) $(\psi_1 \vee \psi_2) \triangleleft \mu \equiv (\psi_1 \triangleleft \mu) \vee (\psi_2 \triangleleft \mu)$.

De façon similaire aux identités de Levi et Harper liant contraction et révision Katsuno and Mendelzon [18] ont proposé des identités reliant effacement et mise à jour :

(Id_1) $\psi \triangleleft \mu \equiv \psi \vee (\psi \diamond \neg\mu)$
(Id_2) $\psi \diamond \mu \equiv (\psi \triangleleft \neg\mu) \wedge \mu$

De plus, ils ont proposé le résultat suivant concernant la satisfaction des postulats.

Proposition 2 [18]

1. *Si un opérateur de mise à jour $\diamond$ satisfait (*U1)-(U4) et (U8), alors l'opérateur d'effacement $\triangleleft$ defini par l'identité (Id_1) satisfait (E1)-(E5) et (E8).
2. Si un opérateur d'effacement $\triangleleft$ satisfait (E1)-(E4) et (E8), alors l'opérateur de mis-à-jour $\diamond$ defini par l'identité (Id_2) satisfait (U1)-(U4) et (U8).
3. Si un opérateur de mise à jour $\diamond$ satisfait (U1)-(U4) et (U8), alors il est possible de définir un opérateur d'effacement grâce à (Id_1). L'opérateur de mise à jour obtenu à partir de cet opérateur d'effacement via (Id_2) est égal à l'opérateur de mise à jour initial $\diamond$.
4. Si un opérateur d'effacement $\triangleleft$ satisfait (E1)-(E5) et (E8), alors il est possible de définir un opérateur de mise à jour grâce à (Id_2). L'opérateur d'effacement obtenu à partir de cet opérateur de mise à jour via (Id_1) est égal à l'opérateur d'effacement initial $\triangleleft$.

L'identité (Id_1) nous permet de définir des opérateurs d'effacement à partir des opérateurs de mise à jour connus. Nous pouvons définir un opérateur d'effacement, noté $\triangleleft_F$, défini par $\text{Mod}(\psi \triangleleft_F \mu) = \text{Mod}(\psi) \cup \text{Mod}(\psi \diamond_F \neg\mu)$ où $\diamond_F$ est l'opérateur de mise à jour de Forbus [14], c.-à-d. $\text{Mod}(\psi \diamond_F \mu) = \bigcup_{m \in \text{Mod}(\psi)} \{m' \in \text{Mod}(\mu) : |m\Delta m'| = |\Delta|_m^{min}(\mu)\}$. Nous pouvons également définir un opérateur d'effacement, noté $\triangleleft_W$, défini par $\text{Mod}(\psi \triangleleft_W \mu) = \text{Mod}(\psi) \cup \text{Mod}(\psi \diamond_W \neg\mu)$ où $\diamond_W$ est l'opérateur de mise à jour de Winslett [25], c.-à-d. $\text{Mod}(\psi \diamond_W \mu) = \bigcup_{m \in \text{Mod}(\psi)} \{m' \in \text{Mod}(\mu) : m\Delta m' \in \Delta_m^{min}(\mu)\}$.

Selon la Proposition 2, les opérateurs d'effacement $\triangleleft_F$ et $\triangleleft_W$ satisfont (E1) − (E5) et (E8).

Exemple 2 *Nous reprenons l'exemple de l'introduction, $\psi = a$ et $Mod(\psi) = \{\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}\}$, $\mu = a \wedge b \wedge c$ et $Mod(\neg\mu) = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}\{b, c\}, \emptyset\}$. $Mod(\psi \triangleleft_F \mu) = Mod(\psi) \cup Mod(\psi \diamond_F \neg\mu) = \{\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \{b, c\}\}$. Remarquons que les formules ψ et μ sont dans le fragment de Horn, en revanche le résultat de l'effacement n'est pas dans Horn puisque l'ensemble des modèles n'est pas clos par intersection.*

3 Raffinements raisonnables d'opérateurs

3.1 Raffinements d'opérateurs

La notion de raffinement a d'abord été définie dans le contexte de la révision [5] avant d'être naturellement étendue à tout opérateur de changement de croyances [8]. L'idée est d'utiliser des opérateurs de changement de croyances bien connus afin d'obtenir des opérateurs adaptés aux fragments caractérisables. Etant donné un fragment propositionnel $\mathcal{L}'$ et un opérateur de changement de croyances $\triangle$, un raffinement de $\triangle$ est un nouvel opérateur $\blacktriangle$, construit à partir de $\triangle$ et pas très différent de $\triangle$, qui opère dans $\mathcal{L}'$ et est tel que le résultat du changement reste dans $\mathcal{L}'$. L'objectif est que la différence de comportement entre $\triangle$ et $\blacktriangle$ obéisse au principe du changement minimal dans le sens où si l'opérateur initial $\triangle$ fournit un résultat qui est dans le fragment alors l'opérateur raffiné ne fait rien de plus, et dans tous les cas n'accroît pas les conséquences logiques du résultat initial.

Définition 3 *Soit $\mathcal{L}'$ un fragment propositionnel et $\triangle: \mathcal{L} \times \mathcal{L} \to \mathcal{L}$ un opérateur de changement de croyances. Un opérateur $\blacktriangle : \mathcal{L}' \times \mathcal{L}' \to \mathcal{L}'$ est un $\triangle$-raffinement pour $\mathcal{L}'$ s'il satisfait les propriétés suivantes. Pour tous $\psi, \psi', \mu, \mu' \in \mathcal{L}'$:*

(i) cohérence : $\psi \blacktriangle \mu$ est satisfaisable ssi $\psi \triangle \mu$ est satisfaisable.

(ii) équivalence : Si $\psi \triangle \mu \equiv \psi' \triangle \mu'$, alors $\psi \blacktriangle \mu \equiv \psi' \blacktriangle \mu'$.

(iii) approximation : $T_{\mathcal{L}'}(\psi \triangle \mu) \subseteq T_{\mathcal{L}'}(\psi \blacktriangle \mu)$.

(iv) invariance : Si $\psi \triangle \mu \in \mathcal{L}'$, alors $T_{\mathcal{L}'}(\psi \blacktriangle \mu) = T_{\mathcal{L}'}(\psi \triangle \mu)$.

Partant d'opérateurs de changement de croyances bien connus nous pouvons obtenir de nouveaux opérateurs, des opérateurs raffinés, adaptés aux fragments caractérisables en utilisant des β-applications. Celles-ci opérent sur l'ensemble de modèles fourni comme résultat par l'opérateur de changement de croyances initial et le transforment en un ensemble de modèles caractéristique d'une formule du fragment considéré, tout en satisfaisant les exigences d'un raffinement comme défini ci-dessus.

Définition 4 *Etant donné $\beta \in \mathcal{B}$, nous définissons une β-*application*, f_β comme étant une application $f_\beta: 2^{2^{\mathcal{U}}} \longrightarrow 2^{2^{\mathcal{U}}}$, telle que pour tout ensemble de modèles $\mathcal{M} \subseteq 2^{\mathcal{U}}$:*

1. *$Cl_\beta(f_\beta(\mathcal{M})) = f_\beta(\mathcal{M})$, c.-à.-d. , $f_\beta(\mathcal{M})$ est clos sous β.*
2. *$f_\beta(\mathcal{M}) \subseteq Cl_\beta(\mathcal{M})$.*
3. *Si $\mathcal{M} = Cl_\beta(\mathcal{M})$, alors $f_\beta(\mathcal{M}) = \mathcal{M}$.*
4. *Si $\mathcal{M} \neq \emptyset$, alors $f_\beta(\mathcal{M}) \neq \emptyset$.*

Nous considérons ici deux β-applications particulières, à savoir la clôture Cl_β définie précédemment et Min_β définie comme suit

Définition 5 *Soit $\beta \in \mathcal{B}$ et $\leq$ un ordre total fixé sur l'ensemble d'interprétations $2^{\mathcal{U}}$. Nous définissons l'application Min_β comme $\text{Min}_\beta(\mathcal{M}) = \mathcal{M}$ si $Cl_\beta(\mathcal{M}) = \mathcal{M}$, et $\text{Min}_\beta(\mathcal{M}) = Min_\leq(\mathcal{M})$ sinon.*

Le concept de β-application nous permet de définir une famille d'opérateurs raffinés pour les fragments de la logique propositionnelle comme suit.

Définition 6 *Soit $\triangle: \mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}$ un opérateur de changement de croyances et $\mathcal{L}' \subseteq \mathcal{L}$ un β-fragment de la logique propositionnelle avec $\beta \in \mathcal{B}$. Etant donné une β-application f_β, nous définissons $\triangle^{f_\beta}: \mathcal{L}' \times \mathcal{L}' \longrightarrow \mathcal{L}'$ comme étant l'opérateur de changement de croyances pour $\mathcal{L}'$ défini par $\text{Mod}(\psi \triangle^{f_\beta} \mu) := f_\beta(\text{Mod}(\psi \triangle \mu))$. La classe $[\triangle, \mathcal{L}']$ est définie comme étant l'ensemble de tous les opérateurs $\triangle^{f_\beta}$ où f_β est une β-application.*

La proposition suivante est centrale puisqu'elle montre que la famille ci-dessus capture exactement tous les raffinements possibles d'un opérateur de changement de croyances que nous avions à l'esprit.

Proposition 3 [8] *Soit $\triangle: \mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}$ un opérateur de changement de croyances et $\mathcal{L}' \subseteq \mathcal{L}$ un fragment caractérisable de la logique propositionnelle. Alors $[\triangle, \mathcal{L}']$ est l'ensemble de tous les $\triangle$-raffinements pour $\mathcal{L}'$.*

Cette caractérisation des opérateurs raffinés permet de définir concrètement de nouveaux opérateurs dont nous pouvons étudier les propriétés logiques. Dans le contexte de la révision et de la mise à jour le raffinement d'un opérateur préserve les postulats basiques [5, 8]. En revanche dans le cas de la contraction et de l'effacement, alors que la notion de raffinement continue à exprimer une forme d'approximation de l'opérateur initial elle ne permet pas de préserver les postulats basiques. Pour cette raison les opérateurs raffinés de contraction et d'effacement ne vont pas nécessairement se conduire rationnellement. Afin de dépasser cette difficulté nous devons nous restreindre aux raffinements *raisonnables*, qui sont des raffinements qui ont deux propriétés supplémentaires.

Définition 7 *Soit $\mathcal{L}'$ un fragment de la logique propositionnelle et $\triangle: \mathcal{L} \times \mathcal{L} \to \mathcal{L}$ un opérateur de changement de croyances. Un opérateur $\blacktriangle : \mathcal{L}' \times \mathcal{L}' \to \mathcal{L}'$ est un $\triangle$-*raffinement raisonnable *si c'est un raffinement qui satisfait les deux conditions supplémentaires suivantes. Pour tous ψ, ψ', μ et $\mu' \in \mathcal{L}'$,*

(v) Si $T_{\mathcal{L}}(\psi \triangle \mu) \subseteq T_{\mathcal{L}}(\psi)$, alors $T_{\mathcal{L}'}(\psi \blacktriangle \mu) \subseteq T_{\mathcal{L}'}(\psi)$.

(vi) Si $T_{\mathcal{L}}(\mu) \nsubseteq T_{\mathcal{L}}(\psi \triangle \mu)$, alors $T_{\mathcal{L}'}(\mu) \nsubseteq T_{\mathcal{L}'}(\psi \blacktriangle \mu)$.

La propriété *(v)* exprime le fait que si aucune nouvelle information n'est ajoutée au résultat du changement de croyances par l'opérateur initial $\psi \triangle \mu$ alors aucune nouvelle information ne sera ajoutée non plus au résultat

du changement de croyances par l'opérateur raffiné. Enfin, la propriété *(vi)* stipule que si μ n'est pas une conséquence de $\psi \triangle \mu$, alors elle ne l'est pas non plus de $\psi \blacktriangle \mu$.

Il est clair que le raffinement par le minimum Min_β, n'est pas un raffinement raisonnable puisqu'il ne satisfait pas la propriété *(v)* . En revanche, la clôture est un tel raffinement.

Proposition 4 *Pour tout opérateur de changement de croyances* $\triangle$: $\mathcal{L} \times \mathcal{L} \to \mathcal{L}$ *et tout* β*-fragment* $\mathcal{L}' \subseteq \mathcal{L}$ *de la logique propositionnelle,* $\triangle^{Cl_\beta}$ *est un* $\triangle$*-raffinement raisonnable pour* $\mathcal{L}'$.

Preuve. L'opérateur$\triangle^{Cl_\beta}$ est $\triangle$-raffinement pour $\mathcal{L}'$, il reste à montrer qu'il est raisonnable.

(*v*) : Supposons que $T_\mathcal{L}(\psi \triangle \mu) \subseteq T_\mathcal{L}(\psi)$, soit $\text{Mod}(\psi) \subseteq \text{Mod}(\psi \triangle \mu)$. Par monotonie de la clôture, $Cl_\beta(\text{Mod}(\psi)) \subseteq Cl_\beta(\text{Mod}(\psi \triangle \mu))$. Puisque $\psi \in \mathcal{L}'$, nous avons $\text{Mod}(\psi) \subseteq \text{Mod}(\psi \triangle^{Cl_\beta} \mu)$, et donc $T_{\mathcal{L}'}(\psi \triangle^{Cl_\beta} \mu) \subseteq T_{\mathcal{L}'}(\psi)$.

(*vi*) : Supposons $T_\mathcal{L}(\mu) \nsubseteq T_\mathcal{L}(\psi \triangle \mu)$. Alors, $\text{Mod}(\psi \triangle \mu) \nsubseteq \text{Mod}(\mu)$, et *a fortiori* $Cl_\beta(\text{Mod}(\psi \triangle \mu)) \nsubseteq \text{Mod}(\mu)$, c.-à-d. $\text{Mod}(\psi \triangle^{Cl_\beta} \mu) \nsubseteq \text{Mod}(\mu)$. Puisque μ est dans $\mathcal{L}'$ nous avons donc $T_{\mathcal{L}'}(\mu) \nsubseteq T_{\mathcal{L}'}(\psi \triangle^{Cl_\beta} \mu)$.

3.2 Caractérisation des raffinements raisonnables

La caractérisation des raffinements raisonnables d'un opérateur passe par la notion de β-application raisonnable. Pour obtenir des raffinements raisonnables opérer sur le seul résultat fourni par l'opérateur initial est insuffisant. Il faut également prendre en compte l'ensemble des modèles de la formule initiale et celui de la formule par laquelle on effectue le changement pour pouvoir satisfaire les deux conditions supplémentaires de la Définition 7.

Définition 8 *Etant donné* $\beta \in \mathcal{B}$*, une* β-application raisonnable*,* f_β *est une application* f_β: $2^{2^\mathcal{U}} \times 2^{2^\mathcal{U}} \times 2^{2^\mathcal{U}} \longrightarrow 2^{2^\mathcal{U}}$*, telle que pour tous ensembles de modèles* $\mathcal{M}$*,* $\mathcal{M}_1$*,* $\mathcal{M}_2$ *:*

1. $Cl_\beta(f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)) = f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)$*,*
2. $f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) \subseteq Cl_\beta(\mathcal{M})$*,*
3. *Si* $\mathcal{M} = Cl_\beta(\mathcal{M})$*, alors* $f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = \mathcal{M}$*,*
4. *Si* $\mathcal{M} \neq \emptyset$*, alors* $f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) \neq \emptyset$*,*
5. *Si* $\mathcal{M}_1 \subseteq \mathcal{M}$*, alors* $\mathcal{M}_1 \subseteq f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)$*,*
6. *Si* $\mathcal{M} \nsubseteq \mathcal{M}_2$*, alors* $f_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) \nsubseteq \mathcal{M}_2$.

Observons que l'application Cl_β vue précedemment peut-être définie par $Cl_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = Cl_\beta(\mathcal{M})$. Il est alors facile de vérifier que c'est une β-application raisonnable.

Grâce à de telles β-applications raisonnables, on peut définir, comme précédemment, une famille d'opérateurs raffinés adaptés aux fragments de la logique propositionnelle.

Définition 9 *Soit* $\triangle$: $\mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}$ *un opérateur de changement de croyances et* $\mathcal{L}' \subseteq \mathcal{L}$ *un* β*-fragment de la logique propositionnelle avec* $\beta \in \mathcal{B}$*. Etant donné une* β*-application raisonnable,* f_β*, nous notons par* $\triangle^{f_\beta}$: $\mathcal{L}' \times \mathcal{L}' \longrightarrow \mathcal{L}'$ *l'opérateur pour* $\mathcal{L}'$ *défini par*

$$\text{Mod}(\psi \triangle^{f_\beta} \mu) := f_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu)).$$

On note $\langle \triangle, \mathcal{L}' \rangle$ *la classe de tous les opérateurs* $\triangle^{f_\beta}$ *où* f_β *est une* β*-application raisonnable.*

La proposition suivante montre que la famille ci-dessus capture exactement tous les raffinements raisonnables possibles d'un opérateur.

Proposition 5 *Soit* $\triangle$: $\mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}$ *un opérateur de changement de croyances et* $\mathcal{L}' \subseteq \mathcal{L}$ *un fragment caractérisable de la logique propositionnelle. Alors,* $\langle \triangle, \mathcal{L}' \rangle$ *est l'ensemble de tous les* $\triangle$*-raffinements raisonnables de* $\mathcal{L}'$.

Preuve. Montrons tout d'abord que tout opérateur de la classe $\triangle^{f_\beta} \in \langle \triangle, \mathcal{L}' \rangle$ est un $\triangle$-raffinement raisonnable pour $\mathcal{L}'$. Observons que bien qu'une β-application raisonnable possède trois paramètres, seul le premier est utilisé pour les quatre premières propriétés de la Définition 8 qui coïncident avec les propriétés d'une β-application. De ce fait, selon la Proposition 3 l'opérateur $\triangle^{f_\beta}$ est $\triangle$-raffinement pour $\mathcal{L}'$. Il reste à prouver qu'elle satisfait les deux propriétés supplémentaires de la Définition 7.

(v) Supposons que $T_\mathcal{L}(\psi \triangle \mu) \subseteq T_\mathcal{L}(\psi)$. Alors $\text{Mod}(\psi) \subseteq \text{Mod}(\psi \triangle \mu)$. Selon la Propriété 5 de la Définition 8, $\text{Mod}(\psi) \subseteq f_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu))$, et donc $\text{Mod}(\psi) \subseteq Cl_\beta(f_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu)))$ puisque $\psi \in \mathcal{L}'$. Il s'en suit que $T_{\mathcal{L}'}(f_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu))) \subseteq T_{\mathcal{L}'}(\psi)$, c.-à-d., $T_{\mathcal{L}'}(\psi \triangle^{f_\beta} \mu) \subseteq T_{\mathcal{L}'}(\psi)$.

(vi) Supposons que $T_\mathcal{L}(\mu) \nsubseteq T_\mathcal{L}(\psi \triangle \mu)$. Alors $\text{Mod}(\psi \triangle \mu) \nsubseteq \text{Mod}(\mu)$. Selon la Propriété 6 de la Définition 8, $f_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu)) \nsubseteq \text{Mod}(\mu)$, c.-à-d., $\text{Mod}(\psi \triangle^{f_\beta} \mu) \nsubseteq \text{Mod}(\mu)$. Donc $\text{Mod}(\psi \triangle^{f_\beta} \mu) \nsubseteq Cl_\beta(\text{Mod}(\mu))$ et puisque $\mu \in \mathcal{L}'$, $T_{\mathcal{L}'}(\mu) \nsubseteq T_{\mathcal{L}'}(\psi \triangle^{f_\beta} \mu)$.

Réciproquement, soit $\blacktriangle$ un $\triangle$-raffinement raisonnable pour $\mathcal{L}'$, montrons que $\blacktriangle \in \langle \triangle, \mathcal{L}' \rangle$. Soit f l'application définie comme suit : Pour tout triplet d'ensemble d'interprétations $(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)$: si $\mathcal{M} = \emptyset$, alors $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = \emptyset$. Considérons maintenant le cas où $\mathcal{M} \neq \emptyset$. S'il existe un couple de formules $(\psi_\mathcal{M}, \mu_\mathcal{M})$ dans $\mathcal{L}'$, tel que $\text{Mod}(\psi_\mathcal{M} \triangle \mu_\mathcal{M}) = \mathcal{M}$, $\text{Mod}(\psi_\mathcal{M}) = \mathcal{M}_1$ et $\text{Mod}(\mu_\mathcal{M}) = \mathcal{M}_2$, alors nous définissons $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = \text{Mod}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M})$, sinon on pose $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = Cl_\beta(\mathcal{M})$.

Remarquons que cette application est bien définie. En effet, puisque $\blacktriangle$ est un $\triangle$-raffinement (raisonnable) pour $\mathcal{L}'$, il ne dépend pas du choix du couple $(\psi_\mathcal{M}, \mu_\mathcal{M})$. De plus, une telle application vérifie les quatre premières propriétés de la Définition 8. Il ne nous reste à vérifier que les deux dernières.

(5) Supposons que $\mathcal{M}_1 \subseteq \mathcal{M}$ (le cas où $\mathcal{M} = \emptyset$ est trivial). Si $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = Cl_\beta(\mathcal{M})$, alors $\mathcal{M}_1 \subseteq f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)$. Considérons enfin le cas où $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = \text{Mod}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M})$, avec $\mathcal{M} = \text{Mod}(\psi \triangle \mu)$, $\mathcal{M}_1 = \text{Mod}(\psi)$ et $\mathcal{M}_2 = \text{Mod}(\mu)$. Puisque $\mathcal{M}_1 \subseteq \mathcal{M}$, $T_\mathcal{L}(\mathcal{M}) \subseteq T_\mathcal{L}(\mathcal{M}_1)$, c.-à-d. $T_\mathcal{L}(\psi_\mathcal{M} \triangle \mu_\mathcal{M}) \subseteq T_\mathcal{L}(\psi_\mathcal{M})$. Or $\blacktriangle$ satisfait la propriété (v) de la Définition 7, et donc $T_{\mathcal{L}'}(\psi_\mathcal{M} \blacktriangle \mu_M) \subseteq T_{\mathcal{L}'}(\psi_\mathcal{M})$. Nous obtenons $\text{Mod}(\psi_\mathcal{M}) \subseteq \text{Mod}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M})$ since $\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M} \in \mathcal{L}'$. Cela montre que $\mathcal{M}_1 \subseteq f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2)$.

(6) Supposons que $\mathcal{M} \nsubseteq \mathcal{M}_2$. Si $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = Cl_\beta(\mathcal{M})$, alors $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) \nsubseteq \mathcal{M}_2$. Si $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) = \text{Mod}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M})$, avec $\mathcal{M} = \text{Mod}(\psi_\mathcal{M} \triangle \mu_\mathcal{M})$, $\mathcal{M}_1 = \text{Mod}(\psi_\mathcal{M})$ et $\mathcal{M}_2 = \text{Mod}(\mu_\mathcal{M})$, alors $\text{Mod}(\psi_\mathcal{M} \triangle \mu_\mathcal{M}) \nsubseteq \text{Mod}(\mu_\mathcal{M})$, c.-à-d. $T_\mathcal{L}(\mu_\mathcal{M}) \nsubseteq T_\mathcal{L}(\psi_\mathcal{M} \triangle \mu_\mathcal{M})$. Donc d'après la Propriété (vi) de la Définition 7, $T_{\mathcal{L}'}(\mu_\mathcal{M}) \nsubseteq T_{\mathcal{L}'}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M})$. Or $\mu_\mathcal{M}$ est dans $\mathcal{L}'$ et donc $\text{Mod}(\psi_\mathcal{M} \blacktriangle \mu_\mathcal{M}) \nsubseteq \text{Mod}(\mu_\mathcal{M})$. Cela prouve que $f(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) \nsubseteq \mathcal{M}_2$.

Outre Cl_β on peut définir un autre exemple de β-application raisonnable p_β, qui est le "pendant raisonnable" de l'opérateur Min_β.

Définition 10 *Soit $\beta \in \mathcal{B}$ et $\leq$ un ordre total sur l'ensemble d'interprétations $2^\mathcal{U}$, $p_\beta(\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2) =$*

$$\begin{cases} \mathcal{M} & \text{si } \mathcal{M} = Cl_\beta(\mathcal{M}) \\ Cl_\beta(\mathcal{M}_1 \cup Min_\leq(\mathcal{M} \cap \overline{\mathcal{M}_2})) & \text{sinon et si } \mathcal{M}_1 \subseteq \mathcal{M} \\ & \text{et } \mathcal{M} \cap \overline{\mathcal{M}_2} \neq \emptyset \\ Cl_\beta(\mathcal{M}) & \text{dans les autres cas} \end{cases}$$

Il est facile de vérifier que l'application p_β vérifie les six propriétés de la Définition 8 et donc est bien une β-application raisonnable. En conséquence, selon la Proposition 5, étant donné $\mathcal{L}'$ un β-fragment et $\triangle$ un opérateur de changement de croyances, l'opérateur $\triangle^{p_\beta}$ défini par

$$\text{Mod}(\psi \triangle^{p_\beta} \mu) = p_\beta(\text{Mod}(\psi \triangle \mu), \text{Mod}(\psi), \text{Mod}(\mu)),$$

est un $\triangle$-raffinement raisonnable.

Exemple 3 *Soit ψ, μ dans $\mathcal{L}_{Horn}$ t.q.* $\text{Mod}(\psi) = \{\emptyset, \{a,b\}, \{c,d\}\}$, $\text{Mod}(\mu) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a,b\}, \{a,c\}, \{a,d\}, \{c,d\}, \{a,c,d\}\}$, *et donc* $\text{Mod}(\neg\mu) = \{\{b,c\}, \{b,d\}, \{a,b,c\}, \{a,b,d\}, \{b,c,d\}, \{a,b,c,d\}\}$.
Soit $\diamond_F$ l'opérateur de mise à jour de Forbus, $\text{Mod}(\psi \diamond_F \neg\mu) = \bigcup_{m \in \text{Mod}(\psi)} \{m' \in \text{Mod}(\neg\mu) : |m \Delta m'| = |\Delta|_m^{min}(\mu)\}$. *En utilisant la Table 1, on obtient aisément* $\text{Mod}(\psi \diamond_F \neg\mu) = \{\{a,b,c\}, \{a,b,d\}, \{b,c,d\}, \{b,c\}, \{b,d\}\}$.
L'opérateur d'effacement $\lhd_F$ est défini par :

$$\begin{aligned} \text{Mod}(\psi \lhd_F \mu) &= \text{Mod}(\psi) \cup \text{Mod}(\psi \diamond_F \neg\mu) \\ &= \{\{a,b\}, \{c,d\}, \{a,b,c\}, \{a,b,d\}, \{b,c,d\}, \\ &\quad \emptyset, \{b,c\}, \{b,d\}\}. \end{aligned}$$

$Mod(\psi)$	Mod($\neg\mu$)					
	{b,c}	{b,d}	{a,b,c}	{a,b,d}	{b,c,d}	{a,b,c,d}
{a,b}	2	2	**1**	**1**	3	2
{c,d}	2	2	3	3	**1**	2
$\emptyset$	**2**	**2**	3	3	3	4

TABLE 1 – Cardinalité de la différence symétrique entre les modèles de ψ et ceux de $\neg\mu$. Les minima sont en gras.

Les formules ψ et μ sont dans le fragment de Horn, cependant le résultat de l'effacement n'est pas dans Horn. L'opérateur raffiné $\lhd^{Cl_\wedge}$ est alors :

$$\begin{aligned} \text{Mod}(\psi \lhd_F^{Cl_\wedge} \mu) &= \{\{a,b\}, \{c,d\}, \emptyset, \{a,b,c\}, \{a,b,d\}, \\ &\quad \{b,c,d\}, \{b,c\}, \{b,d\}, \{b\}, \{c\}, \{d\}\}. \end{aligned}$$

Considérons l'ordre suivant sur les interprétations : $\{b,c\} < \{b,d\} < \{a,b,c\} < \{a,b,d\} < \{b,c,d\}$, l'opérateur raffiné $\lhd^{p_\wedge}$ est alors :

$$\begin{aligned} \text{Mod}(\psi \lhd_F^{p_\wedge} \mu) &= Cl_\wedge\Big(\text{Mod}(\psi) \cup Min_\leq(\{a,b,c\}, \{a,b,d\}, \\ &\quad \{b,c,d\}, \{b,c\}, \{b,d\})\Big) \\ &= Cl_\wedge\Big(\{\{a,b\}, \{c,d\}, \emptyset, \{b,c\}\}\Big). \\ &= \{\{a,b\}, \{c,d\}, \emptyset, \{b,c\}, \{b\}, \{c\}\}. \end{aligned}$$

On peut vérifier aisément que cet exemple illustre le raffinement de l'opérateur d'effacement $\lhd_F$ dans Krom. Il est également applicable aux opérateurs $\lhd_W$, $-_D$ et $-_S$.

4 Propriétés logiques

Dans cette partie nous étudions les propriétés logiques des opérateurs de contraction et d'effacement raffinés en terme de satisfaction des postulats KM.

Dans un premier temps nous montrons que les quatre postulats basiques sont préservés par tout raffinement raisonnable d'un opérateur de contraction (resp. d'effacement).

Proposition 6 *Soit $\mathcal{L}' \subseteq \mathcal{L}$ un fragment caractérisable.*
- *Soit $-$ un opérateur de contraction. Si $-$ satisfait le postulats KM* (C1) *(resp.* (C2)*,* (C3) *and* (C4)*), alors tout raffinement raisonnable de cet opérateur $\blacksquare \in \langle -, L' \rangle$ le satisfait également.*
- *Soit $\lhd$ un opérateur d'effacement. Si $\lhd$ satisfait le postulats KM* (E1) *(resp.* (E2)*,* (E3) *and* (E4)*), alors tout raffinement raisonnable de cet opérateur $\blacktriangleleft \in \langle \lhd, \mathcal{L}' \rangle$ le satisfait également.*

Preuve. Nous donnons la preuve dans le cas de l'effacement. Puisque $\mathcal{L}'$ est un un fragment caractérisable, $\mathcal{L}'$ est un β-fragment pour un certain $\beta \in \mathcal{B}$. D'après la Proposition 5 on peut supposer que $\blacktriangleleft \in [\lhd, \mathcal{L}']$ est un opérateur de la forme $\lhd^{f_\beta}$, où f_β est une β-application raisonnable.
(E1) : Etant donné que $\lhd$ satisfait (E1), $\text{Mod}(\psi) \subseteq \text{Mod}(\psi \lhd \mu)$. Selon la Propriété 5 de la Définition 8, nous

avons $\text{Mod}(\psi) \subseteq f_\beta(\text{Mod}(\psi \lhd \mu), \text{Mod}(\psi), \text{Mod}(\mu))$, c.-à-d., $\psi \models \psi \lhd^{f_\beta} \mu$. Donc $\psi \models \psi \blacktriangleleft \mu$.
(E2) : Supposons que $\psi \not\models \mu$, Etant donné que $\lhd$ satisfait (E2), $\text{Mod}(\psi \lhd \mu) \equiv \text{Mod}(\psi)$. On en déduit que $Cl_\beta(\text{Mod}(\psi \lhd \mu)) = Cl_\beta(\text{Mod}(\psi))$ et donc $Cl_\beta(\text{Mod}(\psi \lhd \mu)) = \text{Mod}(\psi)$ puisque $\psi \in \mathcal{L}'$ et que $\mathcal{L}'$ est un β-fragment. Selon la Propriété 2 de la Définition 4 $f_\beta(\text{Mod}(\psi \lhd \mu)) = Cl_\beta(\text{Mod}(\psi - \mu))$, d'où $f_\beta(\text{Mod}(\psi \lhd \mu)) = \text{Mod}(\psi)$. Par définition de $\blacktriangleleft$ cela signifie que $\psi \blacktriangleleft \mu \equiv \psi$.
(E3) : Supposons $\psi \blacktriangleleft \mu \models \mu$, c.-à-d., $\text{Mod}(\psi \lhd^{f_\beta} \mu) \subseteq \text{Mod}(\mu)$. Selon la Propriété 6 de la Définition 8, nous obtenons $\text{Mod}(\psi \lhd \mu) \subseteq \text{Mod}(\mu)$. Etant donné que $\lhd$ satisfait (E3), on a bien $\models \mu$.
(E4) : Soit ψ_1, ψ_2, μ_1 et μ_2 dans $\mathcal{L}'$ telles que $\psi_1 \equiv \psi_2$ et $\mu_1 \equiv \mu_2$. Etant donné que $\lhd$ satisfait (E4), on a $\psi_1 \lhd -\mu_1 \equiv \psi_2 \lhd \mu_2$. D'après la Propriété de l'équivalence dans la Définition 3, nous obtenons $\psi_1 \blacksquare \mu_1 \equiv \psi_2 \blacksquare \mu_2$.

Une question naturelle est de savoir s'il est possible d'obtenir par raffinement des opérateurs de contraction et d'effacement qui satisfont tous les postulats qui leur sont respectivement dédiés.

En fait la réponse à cette question a déjà été donnée pour la contraction dans le fragment de Horn. En effet, en 2004 Flouris et al. [13] ont étudié le changement de croyances dans un cadre plus général que celui classique de la logique propositionnelle. Lors de leur étude ils ont établi une condition nécessaire et suffisante pour qu'il existe un opérateur de contraction satisfaisant les cinq premiers postulats dans un fragment de la logique propositionnelle, en termes de décomposabilité de ce fragment. Or il a été ensuite démontré dans [20] que le fragment de Horn n'est pas décomposable. En conséquence il n'existe pas d'opérateur de contraction qui satisfait les cinq premiers postulats dans Horn, et donc *a fortiori* un tel opérateur raffiné n'existe pas. En particulier on obtient ce qui suit.

Proposition 7 *Soit $-$ un opérateur de contraction qui satisfait* C1 $-$ C4. *Alors tout raffinement raisonnable de cet opérateur $\blacksquare \in \langle -, \mathcal{L}_{Horn} \rangle$ viole le postulat* (C5) *dans $\mathcal{L}_{Horn}$.*

Autant que nous le sachions nous ne disposons pas d'un résultat si général pour le fragment de Krom. En effet la décomposabilité de ce fragment est une question encore ouverte. Néanmoins nous obtenons des résultats négatifs dans ce fragment pour les opérateurs de contraction de Satoh et de Dalal raffinés par les deux applications que nous considérons ici.

Proposition 8 *Soit $- \in \{-_D, -_S\}$. Alors les opérateurs raffinés $-^{Cl_{\text{maj}_3}}$ et $-^{p_{\text{maj}_3}}$ violent le postulat* (C5) *dans $\mathcal{L}_{Krom}$.*

Pour ces opérateurs raffinés nous obtenons un résultat négatif pour le postulat (C6) dans les fragments de Horn et de Krom.

Proposition 9 *Soit $- \in \{-_D, -_S\}$. Alors les opérateurs raffinés $-^{Cl_\wedge}$ et $-^{p_\wedge}$ violent le postulat* (C6) *dans $\mathcal{L}_{Horn}$, et l'opérateur raffiné $-^{Cl_{\text{maj}_3}}$ viole* (C6) *dans $\mathcal{L}_{Krom}$.*

En ce qui concerne le postulat (C7) nous obtenons des résultats plus mitigés, avec un résultat positif et un négatif.

Proposition 10 *Soit $-$ un opérateur de contraction et $\mathcal{L}'$ un β-fragment. Si $-$ satisfait le postulat* (C7)*, alors l'opérateur raffiné $-^{Cl_\beta}$ le satisfait également dans $\mathcal{L}'$.*

Proposition 11 *Soit $- \in \{-_D, -_S\}$ et $\mathcal{L}' \in \{\mathcal{L}_{Horn}, \mathcal{L}_{Krom}\}$. Alors l' opérateur raffiné $-^{p_\beta}$ viole le postulate* (C7) *dans $\mathcal{L}'$.*

La preuve des trois propositions ci-dessus est donnée dans [7] et nous nous tournons maintenant vers les propriétés logiques des opérateurs d'effacement.

Nous ne savons pas si la décomposabilité d'un fragment est une condition nécessaire pour l'existence d'opérateurs d'effacement sur ce fragment qui satisfont tous les postulats. Nous ne pouvons donc espérer un résultat aussi général que celui obtenu dans le cas de la contraction dans le fragment de Horn. Néanmoins nous obtenons des résultats négatifs pour des raffinements raisonnables des opérateurs d'effacement de Forbus et Winslett, à la fois dans le fragment de Horn et celui de Krom.

Proposition 12 *Soit $\lhd \in \{\lhd_F, \lhd_W\}$ et $\mathcal{L}' \in \{\mathcal{L}_{Horn}, \mathcal{L}_{Krom}\}$. Alors les opérateurs raffinés $\lhd^{Cl_\beta}$ et $\lhd^{p_\beta}$ violent le postulat* (E5) *dans $\mathcal{L}'$.*

Preuve. Considérons des formules de Horn (resp. de Krom) ψ et μ telles que $\text{Mod}(\psi) = \{\{a\}, \{b\}, \emptyset\}$ et $\text{Mod}(\mu) = \{\{a\}, \{b\}, \{c\}, \emptyset\}$. Observons que de telles formules existent puisque les ensembles de modèles correspondants sont à la fois $\wedge$-clos et maj_3-clos. Observons que $\psi \models \mu$. Prouvons maintenant que $(\psi -^{Cl_\beta} \mu) \wedge \mu \not\models \psi$. Nous avons $\text{Mod}(\psi \lhd \mu) = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{b, c\}\}$. Ainsi nous obtenons :

$\text{Mod}(\psi \lhd^{Cl_\beta} \mu) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$ si $\beta = \wedge$, et

$\text{Mod}(\psi \lhd^{Cl_\beta} \mu) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$ si $\beta = maj_3$.

Dans les deux cas nous avons $\text{Mod}(\psi \lhd^{Cl_\beta} \mu) \cap \text{Mod}(\mu) = \{\emptyset, \{a\}, \{b\}, \{c\}\}$ et donc $\text{Mod}(\psi \lhd^{Cl_\beta} \mu) \cap \text{Mod}(\mu) \not\subseteq \text{Mod}(\psi)$. On a donc bien montré que $\lhd^{Cl_\wedge}$ et $\lhd^{Cl_{\text{maj}_3}}$ violent le postulat (E5) dans $\mathcal{L}_{Horn}$ et $\mathcal{L}_{Krom}$.

Pour le raffinement $\lhd^{p_\beta}$, considérons ψ et μ deux formules de Horn (resp., de Krom) telles que $\text{Mod}(\psi) = \{\{a, b\}, \{c, d\}, \emptyset\}$ et $\text{Mod}(\mu) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{a, d\}, \{c, d\}, \{a, c, d\}\}$. De telles formules existent puisque les ensembles de modèles correspondants sont à la fois $\wedge$-clos and maj_3-clos.

Nous avons $\text{Mod}(\neg\mu) = \{\{b, c\}, \{b, d\}, \{a, b, c\}, \{a, b, d\}, \{b, c, d\}, \{a, b, c, d\}\}$.

Supposons que : $\{b,c\} < \{b,d\} < \{a,b,c\} < \{a,b,d\} < \{b,c,d\}$. On peut facilement vérifier que

$$\begin{aligned}\text{Mod}(\psi \lhd \mu) &= \text{Mod}(\psi) \cup min(\text{Mod}(\neg\mu, \leq_\psi)) \\ &= \{\{a,b\},\{c,d\},\emptyset,\{a,b,c\},\{a,b,d\},\\ &\quad \{b,c,d\},\{b,c\},\{b,d\}\}.\end{aligned}$$

Cet ensemble n'est clos ni sous $\wedge$ ni sous maj_3. D'après l'ordre choisi sur les interprétations, nous avons $\text{Mod}(\psi \lhd^{p_\beta} \mu) = Cl_\beta(\text{Mod}(\psi) \cup \{b,c\}) = \{\emptyset,\{a,b\},\{c,d\},\{b,c\},\{b\},\{c\}\}$. Ainsi, $\text{Mod}(\psi \lhd^{p_\beta} \mu) \cap \text{Mod}(\mu) = \{\emptyset,\{a,b\},\{c,d\},\{c\}\}$ et donc $\text{Mod}(\psi \lhd^{p_\beta} \mu) \cap \text{Mod}(\mu) \not\subseteq \text{Mod}(\psi)$. On en déduit que $\lhd^{p_\wedge}$et$\lhd^{p_{\text{maj}_3}}$ violent (E5) respectivement dans $\mathcal{L}_{Horn}$ et $\mathcal{L}_{Krom}$.

Le postulat (E8) est spécifique à l'effacement par rapport à la contraction. Il énonce le fait qu'un opérateur d'effacement devrait considérer équitablement chacun des modèles des croyances initiales. Malheureusement ce postulat ne joue pas ce rôle dans les fragments caractérisables car ceux-ci ne sont pas clos sous disjonction. En effet pour refléter le fait que chaque modèle est considéré équitablement il suffit dans le cadre de la logique propositionnelle de prendre l'union des ensembles de modèles obtenus par l'effacement vis-à-vis chacun des modèles des croyances initiales. Dans les fragments le problème est que même si chacun de ces ensembles est dans le fragment leur union ne l'est pas nécessairement. Néanmoins on peut observer que le principe de considération équitable de chacun des modèles des croyances intiales est maintenu. En effet par construction de nos opérateurs raffinés la première étape est de construire le résultat obtenu par un opérateur bien connu. Si celui-ci satisfait (E8), alors tous les modèles de la formule des croyances initiales contribueront équitablement au résultat, même si un traitement est ensuite requis sur cet ensemble.

5 Conclusion

Dans le cadre de notre étude systématique des opérateurs de changement de croyances adaptés aux fragments propositionnels, nous nous sommes intéressés à la contraction et à l'effacement. La notion de raffinement que nous avions définie précédemment, permet de définir des opérateurs concrets de révision et de mise à jour adaptés aux fragments propositionnels. Cependant dans le cas de la contraction et de l'effacement, une notion de raffinement plus spécifique, appelée *raffinement raisonnable*, est nécessaire afin d'obtenir des opérateurs de contraction et d'effacement rationnels adaptés aux fragments propositionnels. Cette notion nécessite de prendre en compte non seulement le résultat de l'opérateur initial de contraction ou d'effacement mais également deux autres paramètres, l'ensemble initial de croyances et l'information à retirer ou à effacer. Nous avons défini des opérateurs rationnels concrets de contraction à partir des opérateurs de révision de Dalal et de Satoh ainsi que des opérateurs rationnels concrets d'effacement à partir des opérateurs de mise à jour de Forbus et de Winslett. Nous avons montré que le raffinement raisonnable d'opérateurs de contraction, resp. d'effacement, préserve les postulats basiques de contraction, resp. d'effacement. En revanche la satisfaction des postulats de restauration (C5), resp. (E5), ainsi que celle des postulats capturant le changement minimal pour la contraction (C6) et (C7) sont plus problématiques.

Notre approche s'applique aux fragments propositionnels caractérisables. Dans le cas particulier du fragment de Horn, les opérateurs proposés peuvent être comparés à des opérateurs définis précédemment. Le raffinement par clôture coïncide avec la *"contraction basée sur les modèles dans Horn"* (MHC) [27] lorsque l'opérateur de contraction initial est défini par $\psi - \mu = \text{Mod}(\psi) \cup Min(\text{Mod}(\neg\mu), \leq_\psi)$ où $\leq_\psi$ est une assignation fidèle. C'est le cas des opérateurs de contraction de Dalal et de Satoh. Le p_β-raffinement peut coïncider sur certaines instances avec la *"contraction faible par intersection partielle dans Horn"* (MCHCWR) [10], mais ne coïncide pas toujours. En effet lorsque le résultat de la contraction avec l'opérateur initial n'est pas clos, on a $\text{Mod}(\psi -^{p_\beta} \mu) = Cl_\beta(\text{Mod}(\psi) \cup \{m\})$ où $m \in \text{Mod}(\neg\mu)$. Dans le cas de MCHCWR, le choix de m dans $\text{Mod}(\neg\mu)$ est arbitraire, en revanche dans le cas du p_β-raffinement, le choix de m doit se faire dans $\text{Mod}(\psi - \mu) \cap \text{Mod}(\neg\mu)$.

Cette étude soulève de nombreuses questions. Concernant la contraction, une question intéressante est la possibilité de définir pour certains fragments caractérisables des opérateurs raffinés qui satisfont tous les postulats. Flouris et al. ont donné une condition nécessaire pour l'existence d'opérateurs de contraction satisfaisant tous les postulats AGM en termes de décomposabilité du fragment [13]. Une question digne d'intérêt est de savoir s'il est possible de caractériser les fragments décomposables.

En ce qui concerne l'effacement, une question centrale à investiguer est la recherche d'un théorème de représentation. Cela nécessite en premier lieu la formulation de postulats capturant le changement minimal de l'effacement. Par ailleurs, dans le cadre des fragments il serait utile de savoir si, comme c'est le cas pour la contraction, la décomposabilité du fragment est une condition nécessaire pour l'existence d'opérateurs satisfaisant tous les postulats.

Finalement, il serait particulièrement intéressant d'étudier la complexité algorithmique de divers problèmes de décision, comme par exemple la vérification de modèle, pour les opérateurs de contraction et d'effacement raffinés.

Références

[1] Alchourrón, C.E., P. Gärdenfors et D. Makinson: *On*

the logic of theory change : Partial meet contraction and revision functions. Journal of Symbolic Logic, 50 :510–530, 1985.

[2] Booth, R., T.A. Meyer, I.J. Varzinczak et R. Wassermann: *On the Link between Partial Meet, Kernel, and Infra Contraction and its Application to Horn Logic*. Journal of Artificial Intelligence Research (JAIR), 42 :31–53, 2011.

[3] Cadoli, M. et F. Scarcello: *Semantical and computational aspects of Horn approximations*. Artificial Intelligence, 119(1-2) :1–17, 2000.

[4] Caridroit, T., S. Konieczny et P. Marquis: *Contraction in propositional logic*. Dans *Proceedings of ECSQARU'15*, pages 186–196, 2015.

[5] Creignou, N., O. Papini, R. Pichler et S. Woltran: *Belief revision within fragments of propositional logic*. Journal of Computer and System Sciences, 80(2) :427–449, 2014.

[6] Creignou, N., O. Papini, S. Rümmele et S. Woltran: *Belief Merging within Fragments of Propositional Logic*. ACM Transactions on Computational Logic, 17(3) :20, 2016. A preliminary version appeared in Proc. of ECAI'2014.

[7] Creignou, Nadia, Raïda Ktari et Odile Papini: *Belief Contraction Within Fragments of Propositional Logic*. Dans *Proceedings of ECAI'2016*, pages 390–398, 2016.

[8] Creignou, Nadia, Raïda Ktari et Odile Papini: *Belief Update within Propositional Fragments*. Journal of Artificial Intelligence Research (JAIR), 61 :807–834, 2018.

[9] Dalal, M.: *Investigations into Theory of Knowledge Base Revision.* Dans *Proceedings of AAAI'88*, pages 449–479, 1988.

[10] Delgrande, J. P. et R. Wassermann: *Horn Clause Contraction Functions*. Journal of Artificial Intelligence Research (JAIR), 48 :475–511, 2013.

[11] Delgrande, J.P.: *Horn Clause Belief Change : Contraction Functions*. Dans *Proceedings of KR'08*, pages 156–165, 2008.

[12] Delgrande, J.P. et P. Peppas: *Belief revision in Horn theories*. Artificial Intelligence, 218 :1–22, 2015.

[13] Flouris, G., D. Plexousakis et G. Antoniou: *Generalizing the AGM postulates : preliminary results and applications*. Dans *Proceedings of NMR'04*, pages 171–179, 2004.

[14] Forbus, K. D.: *Introducing Actions into Qualitative Simulation*. Dans *Proceedings of IJCAI'89*, pages 1273–1278, 1989.

[15] Gärdenfors, P.: *Knowledge in flux*. Dans *Cambridge University Press, Cambridge UK*, 1988.

[16] Harper, W.L.: *Rational Conceptual Change*. Dans *PSA Proceedings of the Biennial Meeting of the philosophy of Science Association*, tome 2 : Symposia and invited Papers, pages 462–494, 1977.

[17] Horn, A.: *On sentences which are true of direct unions of algebras*. Journal of Symbolic Logic, 16 :14–21, 1951.

[18] Katsuno, H. et A.O. Mendelzon: *On the difference between updating a knowledge base and revising it*. Dans Gärdenfors, P. (rédacteur) : *Belief revision*, pages 183–203. Cambridge University Press, 1992.

[19] Ktari, R.: *Changement de croyances dans des fragments de la logique propositionnelle*. Thèse de doctorat, Aix-Marseille Université, mai 2016.

[20] Langlois, M., R.H. Sloan, B. Szörényi et G. Turán: *Horn Complements : Towards Horn-to-Horn Belief Revision*. Dans *Proceedings of AAAI'08*, pages 466–471, 2008.

[21] Levi, I.: *Subjunctives, dispositions of chances.* Dans *Synthese*, pages 423–455, 1977.

[22] Putte, F. Van De: *Prime implicates and relevant belief revision*. Journal of Logic and Computation, 23(1) :109–119, 2013.

[23] Satoh, K.: *Nonmonotonic reasoning by minimal belief revision*. Dans *Proceedings of FGCS'88*, pages 455–462, Tokyo, 1988.

[24] Schaefer, T. J.: *The Complexity of Satisfiability Problems*. Dans *Proceedings of STOC'78*, pages 216–226, 1978.

[25] Winslett, M.: *Reasoning about Action Using a Possible Models Approach*. Dans *Proceedings of AAAI*, pages 89–93, 1988.

[26] Wu, M., D. Zhang et M. Zhang: *Language Splitting and Relevance-Based Belief Change in Horn Logic*. Dans *Proceedings of AAAI'11*, pages 268–273, 2011.

[27] Zhuang, Z. Q. et M. Pagnucco: *Model Based Horn Contraction*. Dans *Proceedings of KR'12*, pages 169–178, 2012.

[28] Zhuang, Z.Q. et M. Pagnucco: *Transitively Relational Partial Meet Horn Contraction*. Dans *Proceedings of IJCAI'11*, pages 1132–1138, 2011.

[29] Zhuang, Z.Q. et M. Pagnucco: *Entrenchment-Based Horn Contraction*. Journal of Artificial Intelligence Research (JAIR), 51 :227–254, 2014.

[30] Zhuang, Z.Q., M. Pagnucco et Y. Zhang: *Definability of Horn Revision from Horn Contraction*. Dans *Proceedings of IJCAI'13*, 2013.

Actes JIAF 2021

Remarks about roles and entitlements

Philippe Balbiani *

Toulouse Institute of Computer Science Research
CNRS — Toulouse University, Toulouse, France
philippe.balbiani@irit.fr

Abstract

We introduce an hybrid access control model where abstract pairs consisting of objects and access rights are considered. In this model, an access control matrix is a binary relation of permission between subjects and abstract pairs. Treating sets of subjects as instances of a concept called role and sets of abstract pairs as instances of a concept called entitlement, we introduce in each matrix the binary relations of existential permission and universal permission between roles and entitlements, we analyse their properties, we present the duality between subjects and abstract pairs on one hand and roles and entitlements on the other hand, we extend this duality to the setting with obligations and we demonstrate how this duality is useful by proving the completeness of a propositional modal logic for permissions and obligations.

Résumé

Nous introduisons un modèle hybride de contrôle d'accès dans lequel des paires abstraites constituées d'objets et de droits d'accès sont considérés. Dans ce modèle, une matrice de contrôle d'accès est une relation binaire de permission entre sujets et paires abstraites. Traitant les ensembles de sujets comme les instances d'un concept appelé role et les ensembles de paires abstraites comme les instances d'un concept appelé autorisation, nous introduisons dans chaque matrice les relations binaires de permission existentielle et de permission universelle entre roles et autorisations, nous analysons leurs propriétés, nous présentons la dualité entre sujets et paires abstraites d'une part et roles et autorisations d'autre part, nous étendons cette dualité au modèle avec obligations et nous démontrons comment cette dualité est utile en prouvant la complétude d'une logique modale propositionnelle pour permissions et obligations.

*Address: Toulouse Institute of Computer Science Research, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

1 Introduction

Accesses of subjects to objects in a computer system are permitted in accordance with a security policy embodied in an access control database. Many computer systems use the access control matrix model to represent security policies [9]. An access control matrix

	o_1	o_2	...
s_1	$\{\mathtt{r}\}$	$\emptyset$	...
s_2	$\emptyset$	$\{\mathtt{r},\mathtt{w}\}$	...
s_3	$\{\mathtt{r},\mathtt{w}\}$	$\{\mathtt{r},\mathtt{w}\}$	...
s_4	$\emptyset$	$\{\mathtt{r},\mathtt{w}\}$	...
s_5	$\{\mathtt{r},\mathtt{w}\}$	$\{\mathtt{r},\mathtt{w}\}$	...
...	...	...	...

TABLE 1 – Classical form of an access control matrix.

is a relational structure consisting of a nonempty set of subjects (users, processes, etc), a nonempty set of objects (files, tables, etc), access rights (`r`, `w`, etc) and a ternary relation between subjects, objects and access rights represented as in Table 1 where subject s_1 has access right `r` on object o_1, subject s_2 has access rights `r` and `w` on object o_2, etc. For computer systems with a lot of subjects and objects, the access control matrices will become very large and most of their entries will be empty. Moreover, many subjects will have the same access rights to the same objects. In order to reduce the cost of security administration for such computer systems, other access control models have been introduced, for instance role-based access control [12] and organization-based access control [1].

Within the context of role-based access control (**RBAC**), it has been proposed that administrators treat sets of subjects as instances of a concept cal-

led role. Formally, an **RBAC**-structure consists of a

	o_1	o_2	...
r_a	$\{\mathtt{r}\}$	$\emptyset$	...
r_b	$\emptyset$	$\{\mathtt{r},\mathtt{w}\}$	...
r_c	$\{\mathtt{r},\mathtt{w}\}$	$\{\mathtt{r},\mathtt{w}\}$	...
...	...	...	...

TABLE 2 – A role-based access control matrix.

nonempty set of roles representing sets of subjects, a nonempty set of objects, access rights and a ternary relation between roles, objects and access rights represented as in Table 2 where role r_a has access right $\mathtt{r}$ on object o_1, role r_b has access rights $\mathtt{r}$ and $\mathtt{w}$ on object o_2, etc.

Within the context of organization-based access control (**OrBAC**), sets of subjects are still treated as instances of a concept called role. Moreover, sets of objects are treated as instances of a concept called view and sets of access rights are treated as instances of a concept called activity. Formally, an **OrBAC**-structure consists of a nonempty set of roles representing sets of subjects, a nonempty set of views representing sets of objects, a nonempty set of activities representing sets of access rights, a nonempty set of organizations, a nonempty set of contexts and a 5-ary relation between roles, views, activities, organizations and contexts specifying, given an organization and a context, how permissions — i.e. triples consisting of a role, a view and an activity — are granted.

In this paper, we introduce an hybrid access control model where abstract pairs consisting of objects and access rights are considered. Within its context, sets of subjects are still treated as instances of a concept called role. Moreover, sets of abstract pairs are treated as instances of a concept called entitlement. In our setting, an access control matrix is simply a binary relation of permission between subjects and abstract pairs. In each access control matrix, we introduce the binary relations of existential permission and universal permission between roles and entitlements, we analyse their properties, we show how to build access control matrices from role-entitlement frames satisfying these properties, we present the basic dualities that exist between matrices and frames, we extend these dualities to the setting where obligations are added to permissions and we demonstrate how these dualities are useful by proving the completeness of a propositional modal logic talking about permissions and obligations.

It is not so much the hybrid access control model that we want to develop in this paper as the 2 following ideas : the *duality between matrices and frames* developed in Sections 2–5 and the *propositional modal logic interpreted over matrices and frames* presented in Sections 6–9.

From now on in this paper, for all $n\in\mathbb{N}$, let $(n)=\{i \in \mathbb{N} : 1\leq i\leq n\}$ and $[n)=\{i \in \mathbb{N} : n<i\}$. The proofs of some of our results can be found in [2].

2 Access control matrices

The accesses of subjects to objects in a computer system can be presented under the form of a matrix where rows represent subjects and columns represent objects. In that case, the entries in a concrete access control matrix specify the access rights that each subject has on each object. The accesses of subjects to objects in a computer system can also be presented under the alternative form of a matrix where rows represent subjects and columns represent abstract pairs consisting of objects and access rights. In that case, the entries in a concrete access control matrix are bits as in Table 3. The above discussion suggests to consider *abstract access control matrices,* i.e. structures of the form (S,Π,P) where S is a nonempty set of *subjects* (with typical members denoted s, t, etc), Π is a nonempty set of *abstract pairs* (with typical members denoted π, ρ, etc) and P is a binary relation of *permission* between subjects and abstract pairs [1].

P	$(o_1,\mathtt{r})$	$(o_1,\mathtt{w})$	$(o_2,\mathtt{r})$	$(o_2,\mathtt{w})$	...
s_1	1	0	0	0	...
s_2	0	0	1	1	...
s_3	1	1	1	1	...
s_4	0	0	1	1	...
s_5	1	1	1	1	...
...	...	...	...	...	...

TABLE 3 – Alternative form of an access control matrix.

Let (S,Π,P) be an abstract access control matrix. For all $s\in S$, let $P(s)=\{\pi\in\Pi : sP\pi\}$. For all $\pi\in\Pi$, let $P^{-1}(\pi)=\{s\in S : sP\pi\}$. Let $R=\mathcal{P}(S)$ and $E=\mathcal{P}(\Pi)$. Elements of R shall be called *roles* and elements of E shall be called *entitlements.* For all $a\in R$ and for all $\alpha\in E$, we shall say that entitlement α is *existentially permitted to* role a (in symbols $aC_\exists\alpha$) if P intersects $a\times\alpha$, i.e. there exists $s\in a$ and there exists $\pi\in\alpha$ such that $sP\pi$ [2]. For all $a\in R$ and for all $\alpha\in E$, we shall say that entitlement α is

1. In some logical models of deontic systems [10], every subject has the permission to perform some action on some object. Since it may be the case in a computer system that a subject has access to no object at all, we do not require in an abstract access control matrix (S,Π,P) that for all $s\in S$, there exists $\pi\in\Pi$ such that $sP\pi$.

2. "Someone in a has the permission to do something in α".

universally permitted to role a (in symbols $aC_\forall\alpha$) if P contains $a\times\alpha$, i.e. for all $s\in a$ and for all $\pi\in\alpha$, $sP\pi$ [3].

Example: If (S,Π,P) is the restriction to $\{s_1,s_2,s_3\}$ and to $\{o_1,o_2\}$ of the abstract access control matrix rep resented in Table 3 then R consists of 8 roles (the subsets of $\{s_1,s_2,s_3\}$) and 16 entitlements (the subsets of $\{(o_1,\mathtt{r}),(o_1,\mathtt{w}),(o_2,\mathtt{r}),(o_2,\mathtt{w})\}$). In particular, entitlement $\{(o_1,\mathtt{r}),(o_1,\mathtt{w})\}$ is existentially permitted to role $\{s_1,s_2\}$ and entitlement $\{(o_2,\mathtt{r}),(o_2,\mathtt{w})\}$ is universally permitted to role $\{s_2,s_3\}$.

It is also of interest to consider the relations $C_{\exists\forall}^{\rightarrow}$, $C_{\forall\exists}^{\rightarrow}$, $C_{\exists\forall}^{\leftarrow}$ and $C_{\forall\exists}^{\leftarrow}$ between roles and entitlements such that for all $a\in R$ and for all $\alpha\in E$,

- $aC_{\exists\forall}^{\rightarrow}\alpha$ if and only if there exists $s\in a$ such that for all $\pi\in\alpha$, $sP\pi$ [4],
- $aC_{\forall\exists}^{\rightarrow}\alpha$ if and only if for all $s\in a$, there exists $\pi\in\alpha$ such that $sP\pi$ [5],
- $aC_{\forall\exists}^{\leftarrow}\alpha$ if and only if there exists $\pi\in\alpha$ such that for all $s\in a$, $sP\pi$ [6],
- $aC_{\exists\forall}^{\leftarrow}\alpha$ if and only if for all $\pi\in\alpha$, there exists $s\in a$ such that $sP\pi$ [7].

We leave the development of the model-theoretic viewpoint about these relations to future investigations.

The properties of existential permissiveness and universal permissiveness are illustrated by the 2 following obvious results.

Lemma 1 *For all $a,b\in R$ and for all $\alpha,\beta\in E$,*

- *if $aC_\exists\alpha$ then $a\neq\emptyset$ and $\alpha\neq\emptyset$,*
- *if $a=\emptyset$ or $\alpha=\emptyset$ then $aC_\forall\alpha$,*
- *$(a\cup b)C_\exists\alpha$ if and only if $aC_\exists\alpha$ or $bC_\exists\alpha$,*
- *$(a\cup b)C_\forall\alpha$ if and only if $aC_\forall\alpha$ and $bC_\forall\alpha$,*
- *$aC_\exists(\alpha\cup\beta)$ if and only if $aC_\exists\alpha$ or $aC_\exists\beta$,*
- *$aC_\forall(\alpha\cup\beta)$ if and only if $aC_\forall\alpha$ and $aC_\forall\beta$.*

Lemma 2 *For all $a\in R$ and for all $\alpha\in E$,*

- *$aC_\exists\alpha$ if and only if there exists $b\in R$ and there exists $\beta\in E$ such that $b\neq\emptyset$, $b\subseteq a$, $\beta\neq\emptyset$, $\beta\subseteq\alpha$ and $bC_\forall\beta$,*
- *$aC_\forall\alpha$ if and only if for all $b\in R$ and for all $\beta\in E$, if $b\neq\emptyset$, $b\subseteq a$, $\beta\neq\emptyset$ and $\beta\subseteq\alpha$ then $bC_\exists\beta$.*

The reader is invited to compare the above properties of existential permissiveness and universal permissiveness to the properties of the contact relations usually considered between regular closed subsets in topological spaces [3, 5, 6, 7] or the properties of the inference relations usually considered between propositions in relational syllogistics [8, 11].

3. "Everyone in a has the permission to do everything in α".
4. "Someone in a has the permission to do everything in α".
5. "Everyone in a has the permission to do something in α".
6. "Something in α is permitted to everyone in a".
7. "Everything in α is permitted to someone in a".

3 Role-entitlement frames

A *role-entitlement frame* is a structure of the form $(R,E,C_\exists,C_\forall)$ where R is a Boolean algebra of *roles* (with typical members denoted a, b, etc), E is a Boolean algebra of *entitlements* (with typical members denoted α, β, etc) and $C_\exists$ and $C_\forall$ are binary relations between roles and entitlements satisfying the properties illustrated in Lemmas 1 and 2.

Let $(R,E,C_\exists,C_\forall)$ be a role-entitlement frame. Obviously,

Lemma 3 *For all atomic roles a and for all atomic entitlements α, $aC_\exists\alpha$ if and only if $aC_\forall\alpha$* [8].

4 Duality matrices/frames

For all abstract access control matrices $\bar{M}=(S,\Pi,P)$, let $\mathtt{f}(\bar{M})=(R,E,C_\exists,C_\forall)$ where R is the set of all subsets of S, E is the set of all subsets of Π and $C_\exists$ and $C_\forall$ are the binary relations between subsets of S and subsets of Π such that for all $a\in R$ and for all $\alpha\in E$,

- $aC_\exists\alpha$ if and only if there exists $s\in a$ and there exists $\pi\in\alpha$ such that $sP\pi$,
- $aC_\forall\alpha$ if and only if for all $s\in a$ and for all $\pi\in\alpha$, $sP\pi$.

Obviously,

Proposition 1 *For all abstract access control matrices $\bar{M}$, $\mathtt{f}(\bar{M})$ is a role-entitlement frame.*

For all role-entitlement frames $\bar{F}=(R,E,C_\exists,C_\forall)$, let $\mathtt{m}(\bar{F})=(S,\Pi,P)$ where S is the set of all ultrafilters of R, Π is the set of all ultrafilters of E and P is the binary relation between ultrafilters of R and ultrafilters of E such that for all $s\in S$ and for all $\pi\in\Pi$ [9],

- $sP\pi$ if and only if for all $a\in s$ and for all $\alpha\in\pi$, $aC_\exists\alpha$.

Obviously, for all $s\in S$ and for all $\pi\in\Pi$, $sP\pi$ if and only if there exists $a\in s$ and there exists $\alpha\in\pi$ such that $aC_\forall\alpha$. Moreover,

Proposition 2 *For all role-entitlement frames $\bar{F}$, $\mathtt{m}(\bar{F})$ is an abstract access control matrix.*

The duality between abstract access control matrices and role-entitlement frames is illustrated by the

8. Remind that for all Boolean algebras A and for all $a\in A$, a is *atomic* if $a\neq 0_A$ and for all $b\in A$, if $b\neq 0_A$ and $b\leq_A a$ then $b=a$.
9. Remind that for all Boolean algebras A and for all $U\subseteq A$, U is a *proper filter* if $0_A\notin U$, for all $a,b\in A$, if $a\in U$ and $a\leq_A b$ then $b\in U$ and for all $a,b\in A$, if $a\in U$ and $b\in U$ then $a\cdot_A b\in U$. Moreover, U is an *ultrafilter* if U is a maximal proper filter.

2 following results [10].

Proposition 3 *Let* $\bar{M}=(S,\Pi,P)$ *be an abstract access control matrix,* $\mathtt{f}(\bar{M})=(R',E',C'_\exists,C'_\forall)$ *and* $\mathtt{m}(\mathtt{f}(\bar{M}))=(S'',\Pi'',P'')$. *The function* $h : \bar{M} \longrightarrow \mathtt{m}(\mathtt{f}(\bar{M}))$ *such that for all* $s\in S$, $h(s)=\{a\in R' : s\in a\}$ *and for all* $\pi\in\Pi$, $h(\pi)=\{\alpha\in E' : \pi\in\alpha\}$ *is an isomorphism.*

Proposition 4 *Let* $\bar{F}=(R,E,C_\exists,C_\forall)$ *be a role-entitlement frame,* $\mathtt{m}(\bar{F})=(S',\Pi',P')$ *and* $\mathtt{f}(\mathtt{m}(\bar{F}))=(R'',E'',C''_\exists,C''_\forall)$. *The function* $h : \bar{F} \longrightarrow \mathtt{f}(\mathtt{m}(\bar{F}))$ *such that for all* $a\in R$, $h(a)=\{s\in S' : a\in s\}$ *and for all* $\alpha\in E$, $h(\alpha)=\{\pi\in\Pi' : \alpha\in\pi\}$ *is an embedding.*

5 Introducing obligations

In most logical models of deontic systems [10], if a subject has the obligation to perform an action on some object then it also has the permission to perform that action on that object. An *extended access control matrix* is a structure of the form (S,Π,P,O) where (S,Π,P) is an abstract access control matrix and O is a binary relation of *obligation* between subjects in S and abstract pairs in Π such that for all $s\in S$ and for all $\pi\in\Pi$, if $sO\pi$ then $sP\pi$.

Example: If (S,Π,P) is the abstract access control matrix represented in Table 3 and O is the binary relation between elements of S and elements of Π represented in Table 4 then (S,Π,P,O) is an extended access control matrix.

O	$(o_1,\mathtt{r})$	$(o_1,\mathtt{w})$	$(o_2,\mathtt{r})$	$(o_2,\mathtt{w})$	...
s_1	1	0	0	0	...
s_2	0	0	1	1	...
s_3	0	0	1	1	...
s_4	0	0	1	1	...
s_5	0	0	1	1	...
...	...	...	...	...	...

TABLE 4 – Representation of a binary relation between the subjects and the abstract pairs of the abstract access control matrix represented in Table 3.

Let (S,Π,P,O) be an extended access control matrix. For all $s\in S$, let $O(s)=\{\pi\in\Pi : sO\pi\}$. For all $\pi\in\Pi$, let $O^{-1}(\pi)=\{s\in S : sO\pi\}$. Let $R=\mathcal{P}(S)$ and $E=\mathcal{P}(\Pi)$. For all $a\in R$ and for all $\alpha\in E$, we shall say that α is *existentially obligatory to* a (in symbols $aD_\exists\alpha$) if O intersects $a\times\alpha$, i.e. there exists $s\in a$ and there exists $\pi\in\alpha$ such that $sO\pi$ [11]. For all $a\in R$ and for all $\alpha\in E$, we shall say that α is *universally obligatory to* a (in symbols $aD_\forall\alpha$) if O contains $a\times\alpha$, i.e. for all $s\in a$ and for all $\pi\in\alpha$, $sO\pi$ [12].

Example: If (S,Π,P,O) is the restriction to $\{s_1,s_2,s_3\}$ and $\{o_1,o_2\}$ of the extended abstract access control matrix represented in Tables 3 and 4 then R consists of 8 roles (the subsets of $\{s_1,s_2,s_3\}$) and 16 entitlements (the subsets of $\{(o_1,\mathtt{r}),(o_1,\mathtt{w}),(o_2,\mathtt{r}),(o_2,\mathtt{w})\}$). In particular, entitlement $\{(o_1,\mathtt{r}),(o_1,\mathtt{w})\}$ is existentially obligatory to role $\{s_1,s_2\}$ and entitlement $\{(o_2,\mathtt{r}),(o_2,\mathtt{w})\}$ is universally obligatory to role $\{s_2,s_3\}$.

It is also of interest to consider the relations $D_{\exists\forall}^{\rightarrow}$, $D_{\forall\exists}^{\rightarrow}$, $D_{\exists\forall}^{\leftarrow}$ and $D_{\forall\exists}^{\leftarrow}$ between roles and entitlements such that for all $a\in R$ and for all $\alpha\in E$,

- $aD_{\exists\forall}^{\rightarrow}\alpha$ if and only if there exists $s\in a$ such that for all $\pi\in\alpha$, $sO\pi$ [13],
- $aD_{\forall\exists}^{\rightarrow}\alpha$ if and only if for all $s\in a$, there exists $\pi\in\alpha$ such that $sO\pi$ [14],
- $aD_{\forall\exists}^{\leftarrow}\alpha$ if and only if there exists $\pi\in\alpha$ such that for all $s\in a$, $sO\pi$ [15],
- $aD_{\exists\forall}^{\leftarrow}\alpha$ if and only if for all $\pi\in\alpha$, there exists $s\in a$ such that $sO\pi$ [16].

We leave the development of the model-theoretic viewpoint about these relations to future investigations.

The properties of existential obligation and universal obligation are illustrated by the 3 following obvious results.

Lemma 4 *For all* $a,b\in R$ *and for all* $\alpha,\beta\in E$,
- *if* $aD_\exists\alpha$ *then* $a\neq\emptyset$ *and* $\alpha\neq\emptyset$,
- *if* $a=\emptyset$ *or* $\alpha=\emptyset$ *then* $aD_\forall\alpha$,
- $(a\cup b)D_\exists\alpha$ *if and only if* $aD_\exists\alpha$ *or* $bD_\exists\alpha$,
- $(a\cup b)D_\forall\alpha$ *if and only if* $aD_\forall\alpha$ *and* $bD_\forall\alpha$,
- $aD_\exists(\alpha\cup\beta)$ *if and only if* $aD_\exists\alpha$ *or* $aD_\exists\beta$,
- $aD_\forall(\alpha\cup\beta)$ *if and only if* $aD_\forall\alpha$ *and* $aD_\forall\beta$.

Lemma 5 *For all* $a\in R$ *and for all* $\alpha\in E$,
- $aD_\exists\alpha$ *if and only if there exists* $b\in R$ *and there exists* $\beta\in E$ *such that* $b\neq\emptyset$, $b\subseteq a$, $\beta\neq\emptyset$, $\beta\subseteq\alpha$ *and* $bD_\forall\beta$,
- $aD_\forall\alpha$ *if and only if for all* $b\in R$ *and for all* $\beta\in E$, *if* $b\neq\emptyset$, $b\subseteq a$, $\beta\neq\emptyset$ *and* $\beta\subseteq\alpha$ *then* $bD_\exists\beta$.

10. Remind that for all Boolean algebras A,B and for all functions $f : A \longrightarrow B$, f is a *Boolean homomorphism* if $f(0_A)=0_B$, for all $a\in A$, $f(a^{*_A})=f(a)^{*_B}$ and for all $a,b\in A$, $f(a+_A b)=f(a)+_B f(b)$. A *Boolean isomorphism* is a bijective Boolean homomorphism. A *Boolean embedding* is an injective Boolean homomorphism.

11. "Someone in a has the obligation to do something in α".
12. "Everyone in a has the obligation to do everything in α".
13. "Someone in a has the obligation to do everything in α".
14. "Everyone in a has the obligation to do something in α".
15. "Something in α is mandatory to everyone in a".
16. "Everything in α is mandatory to someone in a".

Lemma 6 *For all $a \in R$ and for all $\alpha \in E$,*
- *if $aD_\exists\alpha$ then $aC_\exists\alpha$,*
- *if $aD_\forall\alpha$ then $aC_\forall\alpha$.*

An *extended role-entitlement frame* is a structure of the form $(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ where $(R, E, C_\exists, C_\forall)$ is a role-entitlement frame and $D_\exists$ and $D_\forall$ are binary relations between roles in R and entitlements in E satisfying the properties illustrated in Lemmas 4, 5 and 6.

Let $(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ be an extended role-entitlement frame. Obviously,

Lemma 7 *For all atomic roles a and for all atomic entitlements α, $aD_\exists\alpha$ if and only if $aD_\forall\alpha$.*

For all extended access control matrices $\bar{M}=(S, \Pi, P, O)$, let $\mathtt{f}(\bar{M})=(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ where $(R, E, C_\exists, C_\forall)=\mathtt{f}((S, \Pi, P))$ and $D_\exists$ and $D_\forall$ are the binary relations between subsets of S and subsets of Π such that for all $a \in R$ and for all $\alpha \in E$,
- $aD_\exists\alpha$ if and only if there exists $s \in a$ and there exists $\pi \in \alpha$ such that $sO\pi$,
- $aD_\forall\alpha$ if and only if for all $s \in a$ and for all $\pi \in \alpha$, $sO\pi$.

Obviously,

Proposition 5 *For all extended access control matrices $\bar{M}$, $\mathtt{f}(\bar{M})$ is an extended role-entitlement frame.*

For all extended role-entitlement frames $\bar{F}=(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$, let $\mathtt{m}(\bar{F})=(S, \Pi, P, O)$ where $(S, \Pi, P)=\mathtt{m}((R, E, C_\exists, C_\forall))=$ and O is the binary relation between ultrafilters of R and ultrafilters of E such that for all $s \in S$ and for all $\pi \in \Pi$,
- $sO\pi$ if and only if for all $a \in s$ and for all $\alpha \in \pi$, $aD_\exists\alpha$.

Obviously, for all $s \in S$ and for all $\pi \in \Pi$, $sO\pi$ if and only if there exists $a \in s$ and there exists $\alpha \in \pi$ such that $aD_\forall\alpha$. Moreover,

Proposition 6 *For all extended role-entitlement frames $\bar{F}$, $\mathtt{m}(\bar{F})$ is an extended access control matrix.*

The duality between extended access control matrices and extended role-entitlement frames is illustrated by the 2 following results.

Proposition 7 *Let $\bar{M}=(S, \Pi, P, O)$ be an extended access control matrix, $\mathtt{f}(\bar{M})=(R', E', C'_\exists, C'_\forall, D'_\exists, D'_\forall)$ and $\mathtt{m}(\mathtt{f}(\bar{M}))=(S'', \Pi'', P'', O'')$. The function $h : \bar{M} \longrightarrow \mathtt{m}(\mathtt{f}(\bar{M}))$ such that for all $s \in S$, $h(s)=\{a \in R' : s \in a\}$ and for all $\pi \in \Pi$, $h(\pi)=\{\alpha \in E' : \pi \in \alpha\}$ is an isomorphism.*

Proposition 8 *Let $\bar{F}=(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ be an extended role-entitlement frame, $\mathtt{m}(\bar{F})=(S', \Pi', P', O')$ and $\mathtt{f}(\mathtt{m}(\bar{F}))=(R'', E'', C''_\exists, C''_\forall, D''_\exists, D''_\forall)$. The function $h : \bar{F} \longrightarrow \mathtt{f}(\mathtt{m}(\bar{F}))$ such that for all $a \in R$, $h(a)=\{s \in S' : a \in s\}$ and for all $\alpha \in E$, $h(\alpha)=\{\pi \in \Pi' : \alpha \in \pi\}$ is an embedding.*

6 REL : syntax and semantics

6.1 Syntax

Let **RVAR** be a countable set of *role variables* (with typical members denoted X, Y, etc) and **EVAR** be a countable set of *entitlement variables* (with typical members denoted x, y, etc). Let $(X_1, X_2, \ldots)$ be an enumeration without repetition of **RVAR** and $(x_1, x_2, \ldots)$ be an enumeration without repetition of **EVAR**.

The set **RTER** of all *role terms* (with typical members denoted a, b, etc) is inductively defined as follows :
- $a, b ::= X \mid 0 \mid a^* \mid (a + b)$.

The Boolean constructs 1 and $\cdot$ are defined for role terms by the usual abbreviations : $1 ::= 0^*$ and $(a \cdot b) ::= (a^* + b^*)^*$. The set **ETER** of all *entitlement terms* (with typical members denoted α, β, etc) is inductively defined as follows :
- $\alpha, \beta ::= x \mid 0 \mid \alpha^* \mid (\alpha + \beta)$.

The Boolean constructs 1 and $\cdot$ are defined for entitlement terms by the usual abbreviations : $1 ::= 0^*$ and $(\alpha \cdot \beta) ::= (\alpha^* + \beta^*)^*$.

The set **FOR** of all *formulas* (with typical members denoted φ, ψ, etc) is inductively defined as follows :
- $\varphi, \psi ::= \mathbf{C}_\exists(a, \alpha) \mid \mathbf{C}_\forall(a, \alpha) \mid \mathbf{D}_\exists(a, \alpha) \mid \mathbf{D}_\forall(a, \alpha) \mid a \equiv b \mid \alpha \equiv \beta \mid \bot \mid \neg\varphi \mid (\varphi \vee \psi)$.

We adopt the standard rules for omission of the parenthesis. The Boolean connectives $\top$, $\wedge$, $\rightarrow$ and $\leftrightarrow$ are defined by the usual abbreviations. Let $\mathbf{FOR}_\exists$ be the set of all $\mathbf{C}_\forall$-free $\mathbf{D}_\forall$-free formulas.

Example: Here are examples of formulas :
- $\mathbf{D}_\exists(X, y) \rightarrow \mathbf{C}_\forall(X, y)$,
- $\mathbf{D}_\exists(X_1, y_1) \wedge \mathbf{D}_\forall(X_2, y_2) \rightarrow \mathbf{C}_\forall(X_1 \cdot X_2, y_1 + y_2)$,

the former formula being read "if someone in X has the obligation to do something in y then everyone in X has the permission to do everything in y" and the latter formula being read "if someone in X_1 has the obligation to do something in y_1 and everyone in X_2 has the obligation to do everything in y_2 then everyone in $X_1 \cdot X_2$ has the permission to do everything in $y_1 + y_2$".

On one hand, $\mathbf{C}_\exists$, $\mathbf{C}_\forall$, $\mathbf{D}_\exists$ and $\mathbf{D}_\forall$ can be seen as first-order predicates in a first-order language accepting as arguments pairs consisting of a role term and an entitlement term. On the other hand, $\mathbf{C}_\exists$, $\mathbf{C}_\forall$, $\mathbf{D}_\exists$ and $\mathbf{D}_\forall$ can be seen as diamonds in a propositional

modal language accepting as arguments pairs of Boolean expressions.

6.2 Algebraic semantics

A *valuation* on the extended role-entitlement frame $(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ is a pair (V, v) consisting of a homomorphism $V : (\mathbf{RTER}, 0, *, +) \longrightarrow R$ and a homomorphism $v : (\mathbf{ETER}, 0, *, +) \longrightarrow E$. *Role-entitlement models* are tuples of the form $(R, E, C_\exists, C_\forall, D_\exists, D_\forall, V, v)$ where $(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$ is an extended role-entitlement frame and (V, v) is a valuation on $(R, E, C_\exists, C_\forall, D_\exists, D_\forall)$.

The Boolean connectives $\bot$, $\neg$ and $\vee$ being interpreted as usual, the unary relation of *satisfiability* of formulas with respect to a role-entitlement model $\mathcal{M}$=$(R, E, C_\exists, C_\forall, D_\exists, D_\forall, V, v)$ (in symbols $\models_\mathcal{M}$) is inductively defined as follows :

— $\models_\mathcal{M} \mathbf{C}_\exists(a, \alpha)$ if and only if $V(a) C_\exists v(\alpha)$,
— $\models_\mathcal{M} \mathbf{C}_\forall(a, \alpha)$ if and only if $V(a) C_\forall v(\alpha)$,
— $\models_\mathcal{M} \mathbf{D}_\exists(a, \alpha)$ if and only if $V(a) D_\exists v(\alpha)$,
— $\models_\mathcal{M} \mathbf{D}_\forall(a, \alpha)$ if and only if $V(a) D_\forall v(\alpha)$,
— $\models_\mathcal{M} a \equiv b$ if and only if $V(a)$=$V(b)$,
— $\models_\mathcal{M} \alpha \equiv \beta$ if and only if $v(\alpha)$=$v(\beta)$.

We shall say that formula φ is *valid* in the extended role-entitlement frame $\bar{F}$ (in symbols $\models_{\bar{F}} \varphi$) if for all role-entitlement models $\mathcal{M}$ based on $\bar{F}$, $\models_\mathcal{M} \varphi$. We shall say that formula φ is *valid* in a class $\mathcal{C}$ of extended role-entitlement frames (in symbols $\models_\mathcal{C} \varphi$) if for all extended role-entitlement frames $\bar{F}$ in $\mathcal{C}$, $\models_{\bar{F}} \varphi$.

6.3 Relational semantics

A *valuation* on the extended access control matrix (S, Π, P, O) is a pair (V, v) consisting of a homomorphism $V : (\mathbf{RTER}, 0, *, +) \longrightarrow \mathcal{P}(S)$ and a homomorphism $v : (\mathbf{ETER}, 0, *, +) \longrightarrow \mathcal{P}(\Pi)$. *Access control models* are tuples of the form (S, Π, P, O, V, v) where (S, Π, P, O) is an extended access control matrix and (V, v) is a valuation on (S, Π, P, O).

The Boolean connectives $\bot$, $\neg$ and $\vee$ being interpreted as usual, the unary relation of *satisfiability* of formulas with respect to an access control model $\mathcal{M}$=(S, Π, P, O, V, v) (in symbols $\models_\mathcal{M}$) is inductively defined as follows :

— $\models_\mathcal{M} \mathbf{C}_\exists(a, \alpha)$ if and only if there exists $s \in V(a)$ and there exists $\pi \in v(\alpha)$ such that $sP\pi$,
— $\models_\mathcal{M} \mathbf{C}_\forall(a, \alpha)$ if and only if for all $s \in V(a)$ and for all $\pi \in v(\alpha)$, $sP\pi$,
— $\models_\mathcal{M} \mathbf{D}_\exists(a, \alpha)$ if and only if there exists $s \in V(a)$ and there exists $\pi \in v(\alpha)$ such that $sO\pi$,
— $\models_\mathcal{M} \mathbf{D}_\forall(a, \alpha)$ if and only if for all $s \in V(a)$ and for all $\pi \in v(\alpha)$, $sO\pi$,
— $\models_\mathcal{M} a \equiv b$ if and only if $V(a)$=$V(b)$,
— $\models_\mathcal{M} \alpha \equiv \beta$ if and only if $v(\alpha)$=$v(\beta)$.

We shall say that *formula φ is valid in the extended access control matrix $\bar{M}$* (in symbols $\models_{\bar{M}} \varphi$) if for all access control models $\mathcal{M}$ based on $\bar{M}$, $\models_\mathcal{M} \varphi$. We shall say that *formula φ is valid in a class $\mathcal{C}$ of extended access control matrices* (in symbols $\models_\mathcal{C} \varphi$) if for all extended access control matrices $\bar{M}$ in $\mathcal{C}$, $\models_{\bar{M}} \varphi$.

Example: Let (S, Π, P, O) be the extended abstract access control matrix represented in Tables 3 and 4. If (V, v) is a valuation on it such that $V(X)$=$\{s_1, s_2\}$ and $v(y)$=$\{(o_1, \mathtt{r}), (o_1, \mathtt{w})\}$ then $\models_\mathcal{M} \mathbf{D}_\exists(X, y)$ and $\not\models_\mathcal{M} \mathbf{C}_\forall(X, y)$. If (V, v) is a valuation on it such that $V(X_1)$=$\{s_1, s_2\}$, $V(X_2)$=$\{s_2, s_3\}$, $v(y_1)$=$\{(o_1, \mathtt{r}), (o_1, \mathtt{w})\}$ and $v(y_2)$=$\{(o_2, \mathtt{r}), (o_2, \mathtt{w})\}$ then $\models_\mathcal{M} \mathbf{D}_\exists(X_1, y_1) \wedge \mathbf{D}_\forall(X_2, y_2)$ and $\not\models_\mathcal{M} \mathbf{C}_\forall(X_1 \cdot X_2, y_1 + y_2)$.

7 REL : definability

The 4 following results illustrate the fact that $\mathbf{C}_\exists$, $\mathbf{C}_\forall$, $\mathbf{D}_\exists$ and $\mathbf{D}_\forall$ are not interdefinable when one considers the semantics introduced in Section 6.3. When we prove Proposition 11, they will also illustrate the fact that $\mathbf{C}_\exists$, $\mathbf{C}_\forall$, $\mathbf{D}_\exists$ and $\mathbf{D}_\forall$ are not interdefinable when one considers the semantics introduced in Section 6.2.

Lemma 8 *There exists no $\mathbf{C}_\exists$-free formula φ such that for all extended access control matrices $\bar{M}$, $\models_{\bar{M}} \mathbf{C}_\exists(X, x) \leftrightarrow \varphi$.*

Lemma 9 *There exists no $\mathbf{C}_\forall$-free formula φ such that for all extended access control matrices $\bar{M}$, $\models_{\bar{M}} \mathbf{C}_\forall(X, x) \leftrightarrow \varphi$.*

Lemma 10 *There exists no $\mathbf{D}_\exists$-free formula φ such that for all extended access control matrices $\bar{M}$, $\models_{\bar{M}} \mathbf{D}_\exists(X, x) \leftrightarrow \varphi$.*

Lemma 11 *There exists no $\mathbf{D}_\forall$-free formula φ such that for all extended access control matrices $\bar{M}$, $\models_{\bar{M}} \mathbf{D}_\forall(X, x) \leftrightarrow \varphi$.*

The expressive capacity of the language introduced in Section 6.1 with respect to the semantics introduced in Section 6.3 is illustrated by the 2 following results.

Proposition 9 *For all extended access control matrices $\bar{M}$=(S, Π, P, O),*

— $\models_{\bar{M}} \mathbf{C}_\exists(1, 1)$ *if and only if* $P \neq \emptyset$ [17],
— $\models_{\bar{M}} \neg\mathbf{C}_\forall(1, 1)$ *if and only if* $P \neq S \times \Pi$ [18],

17. "Someone has the permission to do something".
18. "Someone has not the permission to do everything".

— $\models_{\bar{M}} X \not\equiv 0 \to \mathbf{C}_\exists(X,1)$ *if and only if for all* $s\in S$, $P(s)\neq\emptyset$ [19],
— $\models_{\bar{M}} \mathbf{C}_\forall(X,1) \to X \equiv 0$ *if and only if for all* $s\in S$, $P(s)\neq\Pi$ [20],
— $\models_{\bar{M}} x \not\equiv 0 \to \mathbf{C}_\exists(1,x)$ *if and only if for all* $\pi\in\Pi$, $P^{-1}(\pi)\neq\emptyset$ [21],
— $\models_{\bar{M}} \mathbf{C}_\forall(1,x) \to x \equiv 0$ *if and only if for all* $\pi\in\Pi$, $P^{-1}(\pi)\neq S$ [22],
— $\models_{\bar{M}} \mathbf{D}_\exists(1,1)$ *if and only if* $O\neq\emptyset$ [23],
— $\models_{\bar{M}} \neg\mathbf{D}_\forall(1,1)$ *if and only if* $O\neq S\times\Pi$ [24],
— $\models_{\bar{M}} X \not\equiv 0 \to \mathbf{D}_\exists(X,1)$ *if and only if for all* $s\in S$, $O(s)\neq\emptyset$ [25],
— $\models_{\bar{M}} \mathbf{D}_\forall(X,1) \to X \equiv 0$ *if and only if for all* $s\in S$, $O(s)\neq\Pi$ [26],
— $\models_{\bar{M}} x \not\equiv 0 \to \mathbf{D}_\exists(1,x)$ *if and only if for all* $\pi\in\Pi$, $O^{-1}(\pi)\neq\emptyset$ [27],
— $\models_{\bar{M}} \mathbf{D}_\forall(1,x) \to x \equiv 0$ *if and only if for all* $\pi\in\Pi$, $O^{-1}(\pi)\neq S$ [28].

Proposition 10 *For all extended access control matrices* $\bar{M}=(S,\Pi,P,O)$,
— $\models_{\bar{M}} X_1\neq 0 \wedge X_2\neq 0 \to \mathbf{C}_\exists(X_1,x) \vee \mathbf{C}_\exists(X_2,x^*)$ *if and only if for all* $s_1,s_2\in S$, *there exists* $\pi\in\Pi$ *such that* $s_1P\pi$ *and* $s_2P\pi$,
— $\models_{\bar{M}} x_1\neq 0 \wedge x_2\neq 0 \to \mathbf{C}_\exists(X,x_1) \vee \mathbf{C}_\exists(X^*,x_2)$ *if and only if for all* $\pi_1,\pi_2\in\Pi$, *there exists* $s\in S$ *such that* $sP\pi_1$ *and* $sP\pi_2$,
— $\models_{\bar{M}} X_1\neq 0 \wedge X_2\neq 0 \to \mathbf{D}_\exists(X_1,x) \vee \mathbf{D}_\exists(X_2,x^*)$ *if and only if for all* $s_1,s_2\in S$, *there exists* $\pi\in\Pi$ *such that* $s_1O\pi$ *and* $s_2O\pi$,
— $\models_{\bar{M}} x_1\neq 0 \wedge x_2\neq 0 \to \mathbf{D}_\exists(X,x_1) \vee \mathbf{D}_\exists(X^*,x_2)$ *if and only if for all* $\pi_1,\pi_2\in\Pi$, *there exists* $s\in S$ *such that* $sO\pi_1$ *and* $sO\pi_2$.

8 REL : axiomatization

A *logic* is a set $\mathbf{L}$ of formulas such that
(**TAU**) $\mathbf{L}$ contains all tautologies,
— $\mathbf{L}$ contains all formulas of the form
(**A1**) $\mathbf{C}_\exists(a,\alpha) \to a\neq 0 \wedge \alpha\neq 0$,
(**A2**) $a\equiv 0 \vee \alpha\equiv 0 \to \mathbf{C}_\forall(a,\alpha)$,
(**A3**) $\mathbf{C}_\exists(a+b,\alpha) \leftrightarrow \mathbf{C}_\exists(a,\alpha) \vee \mathbf{C}_\exists(b,\alpha)$,
(**A4**) $\mathbf{C}_\forall(a+b,\alpha) \leftrightarrow \mathbf{C}_\forall(a,\alpha) \wedge \mathbf{C}_\forall(b,\alpha)$,
(**A5**) $\mathbf{C}_\exists(a,\alpha+\beta) \leftrightarrow \mathbf{C}_\exists(a,\alpha) \vee \mathbf{C}_\exists(a,\beta)$,
(**A6**) $\mathbf{C}_\forall(a,\alpha+\beta) \leftrightarrow \mathbf{C}_\forall(a,\alpha) \wedge \mathbf{C}_\forall(a,\beta)$,
(**A7**) $a\neq 0 \wedge \alpha\neq 0 \wedge \mathbf{C}_\forall(a,\alpha) \to \mathbf{C}_\exists(a,\alpha)$,
(**A8**) $\mathbf{D}_\exists(a,\alpha) \to a\neq 0 \wedge \alpha\neq 0$,
(**A9**) $a\equiv 0 \vee \alpha\equiv 0 \to \mathbf{D}_\forall(a,\alpha)$,
(**A10**) $\mathbf{D}_\exists(a+b,\alpha) \leftrightarrow \mathbf{D}_\exists(a,\alpha) \vee \mathbf{D}_\exists(b,\alpha)$,
(**A11**) $\mathbf{D}_\forall(a+b,\alpha) \leftrightarrow \mathbf{D}_\forall(a,\alpha) \wedge \mathbf{D}_\forall(b,\alpha)$,
(**A12**) $\mathbf{D}_\exists(a,\alpha+\beta) \leftrightarrow \mathbf{D}_\exists(a,\alpha) \vee \mathbf{D}_\exists(a,\beta)$,
(**A13**) $\mathbf{D}_\forall(a,\alpha+\beta) \leftrightarrow \mathbf{D}_\forall(a,\alpha) \wedge \mathbf{D}_\forall(a,\beta)$,
(**A14**) $a\neq 0 \wedge \alpha\neq 0 \wedge \mathbf{D}_\forall(a,\alpha) \to \mathbf{D}_\exists(a,\alpha)$,
(**A15**) $\mathbf{D}_\exists(a,\alpha) \to \mathbf{C}_\exists(a,\alpha)$,
(**A16**) $\mathbf{D}_\forall(a,\alpha) \to \mathbf{C}_\forall(a,\alpha)$,
(**MP**) $\mathbf{L}$ is closed under modus ponens,
(**US**) $\mathbf{L}$ is closed under uniform substitution,
— $\mathbf{L}$ is closed under all rules of the form
(**R1**) from $X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \mathbf{C}_\forall(X,x) \to \varphi$, infer $\mathbf{C}_\exists(a,\alpha) \to \varphi$,
(**R2**∞) from $X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \neg\mathbf{C}_\exists(X,x) \to \varphi$, infer $\neg\mathbf{C}_\forall(a,\alpha) \to \varphi$,
(**R3**∞) from $X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \mathbf{D}_\forall(X,x) \to \varphi$, infer $\mathbf{D}_\exists(a,\alpha) \to \varphi$,
(**R4**∞) from $X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \neg\mathbf{D}_\exists(X,x) \to \varphi$, infer $\neg\mathbf{D}_\forall(a,\alpha) \to \varphi$,
where neither X, nor x occur in a, α or φ.

We write **REL** for the least logic. For all logics $\mathbf{L}$ and for all $\varphi\in\mathbf{FOR}$, we write $\mathbf{L}\oplus\varphi$ for the least logic containing $\mathbf{L}$ and φ. For all logics $\mathbf{L}$ and for all $\Sigma\subseteq\mathbf{FOR}$, we write $\mathbf{L}\oplus\Sigma$ for the least logic containing $\mathbf{L}$ and Σ. For all logics $\mathbf{L}$, an $\mathbf{L}$-*theory* is a set Σ of formulas such that
— Σ contains $\mathbf{L}$,
— Σ is closed under modus ponens,
— Σ is closed under all infinitary rules of the form
(**R1**∞) from $\{X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \mathbf{C}_\forall(X,x) \to \varphi : X\in\mathbf{RVAR}, x\in\mathbf{EVAR}\}$, infer $\mathbf{C}_\exists(a,\alpha) \to \varphi$,
(**R2**∞) from $\{X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \neg\mathbf{C}_\exists(X,x) \to \varphi : X\in\mathbf{RVAR}, x\in\mathbf{EVAR}\}$, infer $\neg\mathbf{C}_\forall(a,\alpha) \to \varphi$,
(**R3**∞) from $\{X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \mathbf{D}_\forall(X,x) \to \varphi : X\in\mathbf{RVAR}, x\in\mathbf{EVAR}\}$, infer $\mathbf{D}_\exists(a,\alpha) \to \varphi$,
(**R4**∞) from $\{X\neq 0 \wedge X\leq a \wedge x\neq 0 \wedge x\leq\alpha \wedge \neg\mathbf{D}_\exists(X,x) \to \varphi : X\in\mathbf{RVAR}, x\in\mathbf{EVAR}\}$, infer $\neg\mathbf{D}_\forall(a,\alpha) \to \varphi$.

As is well-known, $\mathbf{L}$ is the least $\mathbf{L}$-theory and **FOR** is the greatest $\mathbf{L}$-theory. Obviously, the importance of the infinitary rules (**R1**∞), (**R2**∞), (**R3**∞) and (**R4**∞) is to allow the proof of the following result.

Lemma 12 *Let* $\mathbf{L}$ *be a logic. For all consistent* $\mathbf{L}$-*theories* Σ,
— *for all* $a\in\mathbf{RTER}$ *and for all* $\alpha\in\mathbf{ETER}$, *if* $\mathbf{C}_\exists(a,\alpha)\in\Sigma$ *then there exists* $X\in\mathbf{RVAR}$ *and there exists* $x\in\mathbf{EVAR}$ *such that* $X\neq 0\in\Sigma$, $X\leq a\in\Sigma$, $x\neq 0\in\Sigma$, $x\leq\alpha\in\Sigma$ *and* $\mathbf{C}_\forall(X,x)\in\Sigma$,
— *for all* $a\in\mathbf{RTER}$ *and for all* $\alpha\in\mathbf{ETER}$, *if* $\neg\mathbf{C}_\forall(a,\alpha)\in\Sigma$ *then there exists* $X\in\mathbf{RVAR}$ *and*

19. "Everyone has the permission to do something".
20. "Everyone has not the permission to do everything".
21. "Everything is permitted to someone".
22. "Everything is not permitted to everyone".
23. "Someone has the obligation to do something".
24. "Someone has not the obligation to do everything".
25. "Everyone has the obligation to do something".
26. "Everyone has not the obligation to do everything".
27. "Everything is mandatory to someone".
28. "Everything is not mandatory to everyone".

there exists $x \in \mathbf{EVAR}$ *such that* $X \not\equiv 0 \in \Sigma$, $X \leq a \in \Sigma$, $x \not\equiv 0 \in \Sigma$, $x \leq \alpha \in \Sigma$ *and* $\neg \mathbf{C}_\exists(X, x) \in \Sigma$,
- *for all* $a \in \mathbf{RTER}$ *and for all* $\alpha \in \mathbf{ETER}$, *if* $\mathbf{D}_\exists(a, \alpha) \in \Sigma$ *then there exists* $X \in \mathbf{RVAR}$ *and there exists* $x \in \mathbf{EVAR}$ *such that* $X \not\equiv 0 \in \Sigma$, $X \leq a \in \Sigma$, $x \not\equiv 0 \in \Sigma$, $x \leq \alpha \in \Sigma$ *and* $\mathbf{D}_\forall(X, x) \in \Sigma$,
- *for all* $a \in \mathbf{RTER}$ *and for all* $\alpha \in \mathbf{ETER}$, *if* $\neg \mathbf{D}_\forall(a, \alpha) \in \Sigma$ *then there exists* $X \in \mathbf{RVAR}$ *and there exists* $x \in \mathbf{EVAR}$ *such that* $X \not\equiv 0 \in \Sigma$, $X \leq a \in \Sigma$, $x \not\equiv 0 \in \Sigma$, $x \leq \alpha \in \Sigma$ *and* $\neg \mathbf{D}_\exists(X, x) \in \Sigma$.

For all logics $\mathbf{L}$, the $\mathbf{L}$-theory Σ is *consistent* if $\bot \notin \Sigma$. As is well-known, an $\mathbf{L}$-theory Σ is consistent if and only if $\Sigma \neq \mathbf{FOR}$. For all logics $\mathbf{L}$, for all $\mathbf{L}$-theories Σ and for all formula φ, let $\Sigma + \varphi = \{\psi \in \mathbf{FOR} : \varphi \rightarrow \psi \in \Sigma\}$. As is well-known,

Lemma 13 *For all logics* $\mathbf{L}$, *for all* $\mathbf{L}$*-theories* Σ *and for all formula* φ, $\Sigma + \varphi$ *is an* $\mathbf{L}$*-theory. Moreover,* $\Sigma + \varphi$ *is consistent if and only if* $\neg\varphi \notin \Sigma$.

The main instrument in proofs of completeness is the so-called Lindenbaum Lemma. Its proof is standard and can be found in many papers or textbooks [29].

Lemma 14 (Lindenbaum Lemma) *Let* $\mathbf{L}$ *be a logic. For all consistent* $\mathbf{L}$*-theories* Σ, *there exists a maximal consistent* $\mathbf{L}$*-theory* Δ *such that* $\Sigma \subseteq \Delta$.

The next result states the completeness of **REL** with respect to validity in the class of all extended role-entitlement frames and with respect to validity in the class of all extended access control matrices.

Proposition 11 *For all* $\varphi \in \mathbf{FOR}$, *the following conditions are equivalent :*

1. $\varphi \in \mathbf{REL}$,
2. φ *is valid in the class of all extended role-entitlement frames,*
3. φ *is valid in the class of all extended access control matrices.*

The next result states the completeness of some extensions of **REL** with respect to validity in restricted classes of extended access control matrices.

Proposition 12 *For all* $\varphi \in \mathbf{FOR}$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{C}_\exists(1,1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that* $P \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \neg\mathbf{C}_\forall(1,1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that* $P \neq S \times \Pi$,
- $\varphi \in \mathbf{REL} \oplus X \not\equiv 0 \rightarrow \mathbf{C}_\exists(X, 1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $s \in S$, $P(s) \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{C}_\forall(X, 1) \rightarrow X \equiv 0$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $s \in S$, $P(s) \neq \Pi$,
- $\varphi \in \mathbf{REL} \oplus x \not\equiv 0 \rightarrow \mathbf{C}_\exists(1, x)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $\pi \in \Pi$, $P^{-1}(\pi) \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{C}_\forall(1, x) \rightarrow x \equiv 0$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $\pi \in \Pi$, $P^{-1}(\pi) \neq S$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{D}_\exists(1,1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that* $O \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \neg\mathbf{D}_\forall(1,1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that* $O \neq S \times \Pi$,
- $\varphi \in \mathbf{REL} \oplus X \not\equiv 0 \rightarrow \mathbf{D}_\exists(X, 1)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $s \in S$, $O(s) \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{D}_\forall(X, 1) \rightarrow X \equiv 0$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $s \in S$, $O(s) \neq \Pi$,
- $\varphi \in \mathbf{REL} \oplus x \not\equiv 0 \rightarrow \mathbf{D}_\exists(1, x)$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $\pi \in \Pi$, $O^{-1}(\pi) \neq \emptyset$,
- $\varphi \in \mathbf{REL} \oplus \mathbf{D}_\forall(1, x) \rightarrow x \equiv 0$ *if and only if* φ *is valid in the class of all extended access control matrices* (S, Π, P, O) *such that for all* $\pi \in \Pi$, $O^{-1}(\pi) \neq S$.

9 REL : decidability

Let $\mathbf{DER}_\exists$ be the decision problem defined as follows :

input : $\varphi, \psi \in \mathbf{FOR}_\exists$,

output : determine whether $\psi \in \mathbf{REL} \oplus \varphi$.

The following result will be crucial for the proof of the decidability of $\mathbf{DER}_\exists$.

Proposition 13 *For all* $\varphi, \psi \in \mathbf{FOR}_\exists$, *the following conditions are equivalent :*

1. $\psi \in \mathbf{REL} \oplus \varphi$,
2. ψ *is valid in the class of all finite extended access control matrices validating* φ,

29. For example, see [3, Lemma 3.2].

3. *ψ is valid in the class of all extended access control matrices validating φ.*

Now, we are in a position to prove the main result of this section.

Proposition 14 $\mathbf{DER}_\exists$ *is decidable.*

Let **DER** be the decision problem defined as follows :

input : $\varphi, \psi \in \mathbf{FOR}$,

output : determine whether $\psi \in \mathbf{REL} \oplus \varphi$.

It is still unknown whether **DER** is decidable.

10 Conclusion

In this paper, the propositional modal logic **REL** for permissions and obligations has been introduced together with some of its extensions. Interpreted over extended role-entitlement frames or extended access control matrices, it constitutes a decidable setting for reasoning about the access rights of subjects in computer systems.

An axiomatization of **REL** has been proposed and its completeness has been proved. The decision problem of derivability has been presented in the case of $\mathbf{C}_\forall$-free $\mathbf{D}_\forall$-free formulas and its decidability has been proved. The decidability of the decision problem of derivability in the general case where the connectives $\mathbf{C}_\forall$ and $\mathbf{D}_\forall$ are allowed is still unknown.

It is not so much the hybrid access control model that we want to develop in this paper as the 2 following ideas : the *duality between matrices and frames* developed in Sections 2–5 and the *propositional modal logic interpreted over matrices and frames* presented in Sections 6–9. We believe our reasoning can be adapted to other categories of relational structures.

Acknowledgements

Special acknowledgement is heartily granted to our colleagues of the Toulouse Institute of Computer Science Research (Toulouse, France) for many stimulating discussions about the subject of this paper. We also make a point of strongly thanking the referees for their feedback : their useful suggestions have been essential for improving the correctness and the readability of a preliminary version of this paper.

Références

[1] Abou El Kalam, A., R. El Baida, P. Balbiani, S. Benferhat, F. Cuppens, Y. Deswarte, A. Miège, C. Saurel, and G. Troues-sin, 'Organization based access control', In : *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks,* IEEE (2003) 120–131.

[2] Balbiani, P., 'Remarks about roles and entitlements', hal-03244488.

[3] Balbiani, P., T. Tinchev, and D. Vakarelov, 'Modal logics for region-based theories of space', *Fundamenta Informaticæ* **81** (2007) 29–82.

[4] Davey, B., and H. Priestley, *Introduction to Lattices and Order. Second Edition,* Cambridge University Press (2002).

[5] Dimov, G., and D. Vakarelov, 'Contact algebras and region-based theory of space : a proximity approach — I', *Fundamenta Informaticæ* **74** (2006) 209–249.

[6] Dimov, G., and D. Vakarelov, 'Contact algebras and region-based theory of space : proximity approach — II', *Fundamenta Informaticæ* **74** (2006) 251–282.

[7] Düntsch, I., and D. Vakarelov, 'Region-based theory of discrete spaces : a proximity approach', *Annals of Mathematics and Artificial Intelligence* **49** (2007) 5–14.

[8] Ivanov, N., and D. Vakarelov, 'A system of relational syllogistic including full Boolean reasoning', *Journal of Logic, Language and Information* **21** (2012) 433–459.

[9] Lampson, B., 'Protection', *Operating Systems Review* **8** (1974) 18–24.

[10] Parent, X., and L. van der Torre, *Introduction to Deontic Logic and Normative Systems,* College Publications (2018).

[11] Pratt-Hartmann, I., and L. Moss, 'Logics for the relational syllogistic', *The Review of Symbolic Logic* **2** (2009) 647–683.

[12] Sandhu, R., E. Coyne, H. Feinstein, and C. Youman, 'Role-based access control models', *IEEE Computer* **29** (1996) 38–47.

[13] Wechler, W., *Universal Algebra for Computer Scientists,* Springer (1992).

Actes JIAF 2021

Quand l'adaptation des cas par révision des croyances et l'extrapolation analogique se rencontrent *

Jean Lieber[1] **Emmanuel Nauer**[1] **Henri Prade**[2]

[1] Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
[2] IRIT, CNRS, Université de Toulouse, France
lieber@loria.fr nauer@loria.fr prade@irit.fr

Résumé

Le raisonnement à partir de cas vise à résoudre des problèmes en s'appuyant sur une base de cas, où un cas est un couple problème-solution $(\mathtt{x}, \mathtt{y})$. D'un côté, l'élaboration d'une solution au problème à résoudre (ou problème cible) peut être vue comme une forme de révision des croyances de la solution du cas remémoré (dont la partie problème est similaire au problème cible), contrainte par les connaissances du domaine. D'un autre côté, un mécanisme d'extrapolation fondé sur les proportions analogiques a été proposé récemment. Une proportion analogique est une relation entre 4 termes a, b, c, d qui se lit « a est à b ce que c est à d ». L'extrapolation analogique exploite des triplets de cas dont les parties problèmes forment une proportion analogique avec le problème cible. Puis, l'inférence consiste à résoudre une équation analogique entre les 3 solutions des cas du triplet et l'inconnue de cette équation. Cet article étudie une relation entre ces deux approches de l'adaptation. En plus de constituer une passerelle inattendue entre ces deux domaines, il propose un nouvel éclairage sur le mécanisme d'adaptation en raisonnement à partir de cas. L'article est illustré à l'aide d'un exemple suivi.

Mots-clefs : inférence analogique, proportion analogique, révision des croyances, raisonnement à partir de cas, extrapolation

Abstract

Case-based reasoning, where cases are described in terms of problem-solution pairs $\mathtt{case} = (\mathtt{x}, \mathtt{y})$, amounts to propose a solution to a new problem on the basis of past experience made of stored cases. On the one hand, the building of the solution to a new problem may be viewed as a form of belief revision of the solution of a retrieved case (whose problem part is similar to the new problem) constrained by domain knowledge. On the other hand, an extrapolation mechanism based on analogical proportions has been proposed. It exploits triplets of cases $(\mathtt{case}^a, \mathtt{case}^b, \mathtt{case}^c)$ whose descriptions of problem parts $\mathtt{x}^a$, $\mathtt{x}^b$, $\mathtt{x}^c$ form an analogical proportion with the new problem $\mathtt{x}^{\mathtt{tgt}}$, in such a way that "$\mathtt{x}^a$ is to $\mathtt{x}^b$ as $\mathtt{x}^c$ is to $\mathtt{x}^{\mathtt{tgt}}$". Then, the analogical inference amounts to compute a solution $\mathtt{y}^{\mathtt{tgt}}$ of $\mathtt{x}^{\mathtt{tgt}}$ by solving (when possible) an equation expressing that "$\mathtt{y}^a$ is to $\mathtt{y}^b$ as $\mathtt{y}^c$ is to $\mathtt{y}^{\mathtt{tgt}}$" (where $\mathtt{y}^a$, $\mathtt{y}^b$ and $\mathtt{y}^c$ are respectively the solution parts of $\mathtt{case}^a$, $\mathtt{case}^b$ and $\mathtt{case}^c$). The paper investigates how the belief revision view and analogical extrapolation relate. Besides that it constitutes an unexpected bridge between areas which ignore each other, it casts some light on the adaptation mechanism in case-based reasoning. The paper is illustrated by a running example.

Keywords : analogical inference, analogical proportion, belief revision, case-based reasoning, extrapolation

1 Introduction

La révision de croyances [2] et le raisonnement à partir de cas [1] (RàPC) sont deux domaines de l'intelligence artificielle qui sont généralement considérés comme assez éloignés l'un de l'autre et sans rapport entre eux, puisque le premier se développe principalement dans le cadre de la logique, tandis que le second traite de données et est basé sur la similarité. De plus, la révision des croyances vise à rétablir la cohérence après avoir reçu un nouvel élément d'information qui entre en conflit avec l'état courant des croyances. Le RàPC a un programme tout à fait différent puisqu'il s'agit plutôt de faire face à des informations manquantes en tirant parti de la similarité pour compléter un nouveau problème avec une solution plausible. Cependant, notez que dans les deux approches les conclusions dérivées par la révision des croyances ou par le RàPC ne sont que plausibles.

En dépit de cet état de fait apparent, il existe une vision du RàPC basée sur la révision des croyances [16, 4].

*Les auteurs tiennent à remercier les relecteurs de cet article pour leurs remarques encourageantes et pour les suggestions qu'ils ont faites et qui seront éclairantes pour leurs travaux futurs.

En effet, compléter le nouveau problème par une solution plausible peut être considéré comme une adaptation de la solution d'un cas récupéré afin d'être cohérent avec les spécificités du nouveau cas, ce qui est une forme de révision. Ce qui est révisé n'est clairement pas la base de cas, mais c'est une copie d'un cas traitant d'un problème similaire. En d'autres termes, ce qui est réellement révisé, c'est l'allégation selon laquelle le cas récupéré peut être *appliqué (directement)* pour résoudre le problème cible.

Un mécanisme d'extrapolation basé sur des proportions analogiques (qui sont des énoncés de la forme « a est à b comme c est à d ») a été proposé pour exploiter les cas en RàPC [18]. Sur la base de 3 cas dont les parties de descriptions de problèmes sont en proportion analogique avec le nouveau problème, on induit une solution pour celui-ci à partir des solutions des 3 cas. Cela signifie qu'un processus d'adaptation a lieu à l'intérieur de l'inférence, ce qui est rendu possible par le fait que les proportions analogiques sont à la fois une question de similarité et de dissimilarité. En effet, lorsque les parties problèmes des cas a, b, c, d sont décrites par des vecteurs de traits caractéristiques, la proportion analogique est valide entre les vecteurs si a diffère de b de la même façon que c diffère de d (et vice-versa) [19, 23]. À cet égard, un couple de cas peut être considéré comme codant une sorte de règle d'adaptation au sens du RàPC [6].

Les points ci-dessus suggèrent qu'il existe une passerelle entre l'approche du RàPC basée sur la révision des croyances et l'extrapolation analogique basée sur les proportions. Cette investigation est le sujet de cet article. Il est organisé comme suit. Dans la section 2, après un bref rappel sur la logique propositionnelle et l'introduction des notations, deux ensembles de rappels sont fournis respectivement sur l'extrapolation analogique des cas et sur la vision du RàPC en termes de révision de croyances. Cette mise en place du cadre nécessaire à l'étude de la problématique traité dans cet article rend la section 2 assez longue, d'autant plus que l'exemple courant y est également introduit. Afin d'établir un pont entre l'extrapolation et l'adaptation basée sur la révision, la section 3 reformule d'abord l'adaptation par extrapolation comme une adaptation d'un seul cas. Ensuite, un opérateur de révision basé sur la compétence des couples de cas est défini, ce qui permet d'établir que l'adaptation par extrapolation peut être obtenue de manière équivalente par une révision constituée de couples de cas compétents exprimant la connaissance de l'adaptation. La section 4 présente des travaux connexes et des remarques finales, ainsi qu'une présentation de travaux futurs.

2 Présentation du problème et de l'exemple suivi

Cette section présente les notions liées au problème que cet article entend résoudre, i.e., comment l'adaptation par extrapolation analogique et l'adaptation par révision peuvent se rencontrer. Dans cette optique, les notions et notations utilisées sont présentées, et l'exemple suivi dans le domaine culinaire est présenté.

2.1 Un rapide rappel sur la logique propositionnelle

Le formalisme de représentation des cas et des connaissances du domaine est, dans cet article, la logique propositionnelle. Soit $\mathcal{V}$ un ensemble fini de symboles, appelés variables. Une formule est soit une variable soit une expression d'une des formes suivantes : $\top$, $\bot$, $\neg\varphi$, $\varphi_1 \wedge \varphi_2$, $\varphi_1 \vee \varphi_2$, $\varphi_1 \rightarrow \varphi_2$ et $\varphi_1 \leftrightarrow \varphi_2$ où φ, φ_1 et φ_2 sont des formules.

Une interprétation $\mathcal{I}$ est une fonction de $\mathcal{V}$ dans $\{0, 1\}$ où 0 et 1 dénotent les valeurs booléennes « faux » et « vrai ». L'ensemble des interprétations est dénoté par Ω. Une interprétation $\mathcal{I} \in \Omega$ est étendue sur l'ensemble de toutes les formules comme suit : $\mathcal{I}(\top) = 1$, $\mathcal{I}(\bot) = 0$, $\mathcal{I}(\neg\varphi) = \textbf{non}\,\mathcal{I}(\varphi)$, $\mathcal{I}(\varphi_1 \wedge \varphi_2) = \mathcal{I}(\varphi_1)$ **et** $\mathcal{I}(\varphi_2)$, $\mathcal{I}(\varphi_1 \vee \varphi_2) = \mathcal{I}(\varphi_1)$ **ou** $\mathcal{I}(\varphi_2)$, $\mathcal{I}(\varphi_1 \rightarrow \varphi_2) = \mathcal{I}(\neg\varphi_1 \vee \varphi_2)$ et $\mathcal{I}(\varphi_1 \leftrightarrow \varphi_2) = \mathcal{I}((\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_2 \rightarrow \varphi_1))$ où **non**, **et** et **ou** sont les opérations classiques sur les booléens. Un modèle d'une formule φ est une interprétation $\mathcal{I}$ telle que $\mathcal{I}(\varphi) = 1$ et l'ensemble des modèles de φ est dénoté par $\mathcal{M}(\varphi)$. Une formule φ_1 entraîne une formule φ_2, noté $\varphi_1 \models \varphi_2$, si $\mathcal{M}(\varphi_1) \subseteq \mathcal{M}(\varphi_2)$. Les formules φ_1 et φ_2 sont équivalentes, noté $\varphi_1 \equiv \varphi_2$, si $\mathcal{M}(\varphi_1) = \mathcal{M}(\varphi_2)$. Une formule φ est *cohérente* si $\mathcal{M}(\varphi) \neq \emptyset$.

Soit $\Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_p\}$ un ensemble fini de formules. $\bigvee \Phi$ dénote $\varphi_1 \vee \varphi_2 \vee \ldots \vee \varphi_p$. $\bigwedge \Phi$ dénote $\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_p$.

Dans cet article, dans un but de simplification, les notations booléennes et les notations propositionnelles sont parfois utilisées les unes pour les autres (p. ex., 0 et 1 à la place de $\bot$ et $\top$ et utilisation des connecteurs propositionnels entre valeurs booléennes).

Un« flip [1] » est une opération d'édition sur une interprétation consistant à changer exactement l'interprétation d'une variable (on peut effectuer $|\mathcal{V}|$ *flips* sur une interprétation).

2.2 Notions et notations liées au RàPC

Le RàPC a pour objectif de résoudre des problèmes grâce à une base de cas, où un cas est la représentation d'un épisode de résolution de problème. Soit $\mathcal{P}$ et $\mathcal{S}$ les espaces de problèmes et de solutions : un *problème* $\mathtt{x}$ (resp., une *solution* $\mathtt{y}$) est, par définition, un élément de $\mathcal{P}$ (resp.,

1. Le terme anglais a été conservé à défaut de trouver une meilleure traduction que « chiquenaude ».

de $\mathcal{S}$). Une relation sur $\mathcal{P} \times \mathcal{S}$ est supposée exister qui se lit « a pour solution », mais qui n'est, pour la majorité des applications, pas complètement connue du système de RàPC. Un *cas* est un couple $\mathtt{cas} = (\mathtt{x}, \mathtt{y}) \in \mathcal{P} \times \mathcal{S}$ tel que $\mathtt{x}$ a pour solution $\mathtt{y}$. Un *cas source* $\mathtt{cas}^s = (\mathtt{x}^s, \mathtt{y}^s)$ est un élément de la base de cas, laquelle est notée $\mathtt{BC}$. Le problème à résoudre est appelé *problème cible*, dénoté par $\mathtt{x}^{\mathtt{cible}}$.

Le modèle de processus classique du RàPC s'appuie sur les étapes de remémoration et d'adaptation [26] (également appelées *retrieve* et *reuse* dans [1], i.e., *retrouver* et *réutiliser*). D'autres étapes suivent l'adaptation mais ne sont pas considérées dans cet article. La remémoration sélectionne un ou plusieurs cas source(s) dans l'optique de la résolution de $\mathtt{x}^{\mathtt{cible}}$. L'adaptation vise à utiliser le(s) cas remémoré(s) pour proposer une solution plausible $\mathtt{y}^{\mathtt{cible}}$ de $\mathtt{x}^{\mathtt{cible}}$. Une *adaptation simple* est une adaptation d'un seul cas remémoré, sinon, on parlera d'*adaptation multiple*.

Les *connaissances du domaine* forment une base de connaissances $\mathtt{CD}$ qui peut être comprise comme un ensemble de contraintes d'intégrité : si un problème $\mathtt{x}$ (resp., une solution $\mathtt{y}$ ou un cas $(\mathtt{x}, \mathtt{y})$) n'est pas cohérent avec $\mathtt{CD}$ alors on peut déduire qu'il n'est pas licite.

Dans cet article, il est supposé qu'une séparation claire entre problèmes et solutions est faite. Cela n'est pas toujours vrai : pour certaines applications du RàPC, chaque cas constitue un tout et le problème cible est considéré comme un cas incomplet. Dans ce cas, la séparation entre partie problème et partie solution d'un cas remémoré peut se faire au moment de l'adaptation. Étant donné que cet article concerne principalement l'adaptation, cette hypothèse de séparation entre problèmes et solutions n'est pas une grande restriction. Par conséquent, dans le cadre propositionnel, $\mathcal{V}$ est partitionné en $\{\mathcal{V}_\mathcal{P}, \mathcal{V}_\mathcal{S}\}$ et les variables apparaissant dans un problème (resp., dans une solution) sont des éléments de $\mathcal{V}_\mathcal{P}$ (resp., de $\mathcal{V}_\mathcal{S}$). Cette idée est illustrée dans la section suivante.

De plus, il est supposé que chaque cas source $(\mathtt{x}^s, \mathtt{y}^s)$ et le problème cible sont complètement décrits : étant donné les connaissances du domaine, la valeur de vérité de toute variable $a \in \mathcal{V}_\mathcal{P}$ est connue pour $\mathtt{x}^s$ et pour $\mathtt{x}^{\mathtt{cible}}$ (i.e., $\mathtt{CD} \wedge \mathtt{x}^s \models a$ ou $\mathtt{CD} \wedge \mathtt{x}^s \models \neg a$, et de même pour $\mathtt{x}^{\mathtt{cible}}$) et une contrainte similaire est vérifiée par $\mathtt{y}^s$. Un cas $(\mathtt{x}^s, \mathtt{y}^s)$ est représenté en logique propositionnelle par une seule formule $\mathtt{x}^s \wedge \mathtt{y}^s$.

Selon le modèle des connaissances classique du RàPC (voir, p. ex., [25]), la base de connaissances d'un système de RàPC consiste en quatre *conteneurs de connaissances*. Deux d'entre eux ont déjà été mentionnés : la base de cas $\mathtt{BC}$ et les connaissances du domaine $\mathtt{CD}$. Les deux autres sont les connaissances de remémoration et les connaissances d'adaptation, utilisées respectivement durant les étapes de remémoration et d'adaptation.

Une dernière remarque peut être faite ici à propos de la notion de similarité en RàPC : on dit en général que la remémoration vise à trouver des cas sources *similaires* au problème cible. Cette notion de similarité ne doit pas nécessairement être comprise comme une égalité approximative : cela peut aller plus loin. En fait, $(\mathtt{x}^s, \mathtt{y}^s)$ peut être considéré comme similaire à $\mathtt{x}^{\mathtt{cible}}$ même si les descriptions de $\mathtt{x}^s$ et $\mathtt{x}^{\mathtt{cible}}$ sont très différentes, pourvu que l'adaptation de $(\mathtt{x}^s, \mathtt{y}^s)$ retourne une solution plausible de $\mathtt{x}^{\mathtt{cible}}$([2]). Cela est lié à ce que Barry Smyth et Mark Keane appellent la remémoration guidée par l'adaptabilité [27].

2.3 Spécification de l'exemple suivi

L'exemple utilisé tout au long de cet article se situe dans le domaine de la cuisine. Une recette de cuisine est simplement représentée par les types des ingrédients qu'elle contient. Les variables sont des noms d'aliments représentant des classes de recettes. Par exemple, $\mathtt{fruit}$ représente la classe des recettes ayant au moins un fruit comme ingrédient. Les connaissances du domaine sont représentées par

$$\begin{aligned}\mathtt{CD} = {} & \mathtt{ananas} \rightarrow \mathtt{fruit} \ \wedge \ \mathtt{crème} \rightarrow \mathtt{sauce} \\ & \wedge \mathtt{pesto} \rightarrow \mathtt{sauce} \ \wedge \ \mathtt{Saint\text{-}Pierre} \rightarrow \mathtt{poisson} \\ & \wedge \mathtt{saumon} \rightarrow \mathtt{poisson} \ \wedge \ \neg(\mathtt{fruit} \wedge \mathtt{pesto})\end{aligned}$$

Le dernier terme de cette conjonction signifie qu'une recette ne devrait pas contenir à la fois du pesto (qui, lui-même, contient de l'ail) et un fruit (où la notion de fruit est considérée dans un sens culinaire, pas botanique).

Les cas sources considérés dans cet exemple sont :

$$\begin{aligned}\mathtt{cas}^a = {} & \mathtt{Saint\text{-}Pierre} \wedge \mathtt{crème} \wedge \mathtt{vanille} \wedge \mathtt{curry} \\ & \wedge \textit{Rien d'autre} \qquad (1)\\ \mathtt{cas}^b = {} & \mathtt{saumon} \wedge \mathtt{crème} \wedge \mathtt{vanille} \wedge \mathtt{pesto} \wedge \\ & \wedge \textit{Rien d'autre} \qquad (2)\\ \mathtt{cas}^s = {} & \mathtt{Saint\text{-}Pierre} \wedge \mathtt{crème} \wedge \mathtt{ananas} \wedge \\ & \wedge \textit{Rien d'autre} \qquad (3)\end{aligned}$$

où, pour une formule φ, $\varphi \wedge \textit{Rien d'autre}$ dénote la formule $\varphi \wedge \Gamma$, avec Γ la conjonction des littéraux négatifs $\neg a$ tels que $a \in \mathcal{V}$ et $\mathtt{CD} \wedge \varphi \not\models a$ (on peut voir cela comme une application de l'hypothèse du monde clos dans la représentation des cas sources). Par exemple, $\mathtt{cas}^s$ contient le littéral $\neg\mathtt{saumon}$.

Le problème cible représente la requête d'une recette avec du poisson, du pesto, mais pas de vanille :

$$\mathtt{x}^{\mathtt{cible}} = \mathtt{poisson} \wedge \mathtt{pesto} \wedge \neg\mathtt{vanille} \qquad (4)$$

2. L'exemple suivant illustre cette idée. Supposons qu'un problème soit une représentation d'un réel $\mathtt{x}$ et que sa solution soit une représentation de $\sin(\mathtt{x})$. Soit à présent $\mathtt{x}^{\mathtt{cible}} = -2$ et $(\mathtt{x}^s, \mathtt{y}^s) \in \mathtt{BC}$ tel que $\mathtt{x}^s = 2$. En utilisant la connaissance du domaine indiquant que sin est une fonction impaire ($\sin(-x) = -\sin(x)$ pour tout $x \in \mathbb{R}$), il apparaît que l'adaptation du cas source en $\mathtt{y}^{\mathtt{cible}} = -\mathtt{y}^s$ donne une bonne solution de $\mathtt{x}^{\mathtt{cible}}$. Par conséquent, dans un sens relatif au RàPC et pour cette application, 2 et -2 sont similaires.

L'ensemble des variables pour décrire les problèmes est donc $\mathcal{V}_{\mathcal{P}} = \{\texttt{poisson}, \texttt{pesto}, \texttt{vanille}\}$. Les autres variables sont utilisées pour décrire les solutions.

2.4 Proportions analogiques et RàPC

Les proportions analogiques sont des énoncés de la forme « a est à b ce que c est à d », notée $a:b::c:d$. Leur origine remonte (au moins) à Aristote [3], et a été inspirée par un parallèle avec la proportion numérique (géométrique) : $\frac{a}{b} = \frac{c}{d}$; voir [22].

En accord avec ce parallèle, elles sont supposées obéir aux postulats suivants (e.g., [15]), pour a, b, c et d, des éléments de l'univers considéré $\mathcal{U}$:

1. $a:b::a:b$ (*réflexivité*) ;
2. si $a:b::c:d$ alors $c:d::a:b$ (*symétrie*) ;
3. si $a:b::c:d$ alors $a:c::b:d$ (*permutation centrale*).

L'unique modèle booléen minimal [24] obéissant à ces 3 postulats est un connecteur quaternaire de la logique propositionnelle, lorsque $\mathcal{U}$ est l'ensemble des booléens $\mathbb{B} = \{0, 1\}$ [19] :

$$a:b::c:d = ((a \wedge \neg b) \leftrightarrow (c \wedge \neg d)) \wedge ((\neg a \wedge b) \leftrightarrow (\neg c \wedge d))$$

Cela rend explicite que « a diffère de b comme c diffère de d (et vice-versa) ». Il est facile de vérifier que cette formule n'est valable que pour les 6 valuations suivantes $0:0::0:0$, $1:1::1:1$, $0:1::0:1$, $1:0::1:0$, $0:0::1:1$, et $1:1::0:0$.

On peut voir que 1 et 0 jouent un rôle symétrique, ce qui rend la définition *indépendante du code* : ceci s'exprime formellement avec l'opérateur de négation comme suit si $a:b::c:d$ alors $\neg a:\neg b::\neg c:\neg d$([3]). Pour traiter des éléments, par exemple des cas, représentés par des *vecteurs* de valeurs booléennes, la définition de la proportion analogique est étendue composante par composante de $\mathcal{U}$ à $\mathcal{U}^n$:

$$a:b::c:d \quad \text{ssi} \quad \text{pour tout } i \in \{1, \ldots, n\}, a_i:b_i::c_i:d_i$$

C'est la base d'un principe d'inférence, proposé pour la première fois dans [20] pour des valeurs nominales, qui peut s'énoncer comme suit :

$$\frac{\forall i \in \{1, ..., m\},\ \ a_i:b_i::c_i:d_i}{\forall j \in \{m+1, ..., n\},\ \ a_j:b_j::c_j:d_j}$$

Comme on peut le voir, la connaissance de certaines composantes des vecteurs sources est transférée à leurs autres composantes, en supposant implicitement que les valeurs des premières composantes déterminent les valeurs des autres.

3. Cette propriété est importante en pratique car elle indique qu'entre deux propriétés contraires, par exemple `marié` et `célibataire`, peu importe celle qui est codée par 1 ou par 0.

Cela requiert de trouver ? tel que $a:b::c:?$ soit valide. La solution peut ne pas exister (e.g., pour $0:1::1:?$). Il y a une solution si et seulement si $a = b$ ou $a = c$, dans le cas booléen. L'*unique* solution est alors donnée par $? = c$ si $a = b$ et $? = b$ si $a = c$. Ainsi, nous avons la propriété suivante dans le cas booléen

$$a:a::b:? \text{ si et seulement si } ? = b \tag{5}$$

qui est quelquefois prise comme un postulat supplémentaire, et qui n'est pas une conséquence des 3 postulats. Ceci est la base de l'*extrapolation* analogique entre cas proposée dans [18].

Étant donné un couple de vecteurs décrivant des problèmes $(\mathbf{x}^a, \mathbf{x}^b)$, avec n composantes, leur comparaison donne une partition des n caractéristiques en deux sous-ensembles : le sous-ensemble des caractéristiques $\mathcal{E}_{(\mathbf{x}^a,\mathbf{x}^b)}$ pour lesquelles les valeurs de $\mathbf{x}^a$ et $\mathbf{x}^b$ sont égales et le sous-ensemble $\mathcal{D}_{(\mathbf{x}^a,\mathbf{x}^b)}$ pour lesquelles elles sont différentes. Considérons deux couples de vecteurs décrivant des problèmes $(\mathbf{x}^a, \mathbf{x}^b)$ et $(\mathbf{x}^c, \mathbf{x}^d)$ tels que $\mathcal{E}_{(\mathbf{x}^a,\mathbf{x}^b)} = \mathcal{E}_{(\mathbf{x}^c,\mathbf{x}^d)}$ et tels que $\forall i \in \mathcal{D}_{(\mathbf{x}^a,\mathbf{x}^b)}, \mathbf{x}^{a_i} = \mathbf{x}^{c_i}$ et $\mathbf{x}^{b_i} = \mathbf{x}^{d_i}$. Il est clair alors $\mathbf{x}^a:\mathbf{x}^b::\mathbf{x}^c:\mathbf{x}^d$ est valide, puisque

(i) pour chaque $j \in \mathcal{E}_{(\mathbf{x}^a,\mathbf{x}^b)} = \mathcal{E}_{(\mathbf{x}^c,\mathbf{x}^d)}$, $(\mathbf{x}^{a_j}, \mathbf{x}^{b_j}, \mathbf{x}^{c_j}, \mathbf{x}^{d_j})$ est de la forme (u, u, v, v) ou (u, u, u, u) pour des valeurs u et v où $u \neq v$ et $u, v \in \{0, 1\}$,

(ii) pour chaque $k \in \mathcal{D}_{(\mathbf{x}^a,\mathbf{x}^b)} = \mathcal{D}_{(\mathbf{x}^c,\mathbf{x}^d)}$, $(\mathbf{x}^{a_k}, \mathbf{x}^{b_k}, \mathbf{x}^{c_k}, \mathbf{x}^{d_k})$ est de la forme (u, v, u, v).

L'idée de considérer des couples de cas peut être reliée à la lecture d'un couple $((\mathbf{x}^a, \mathbf{y}^a), (\mathbf{x}^b, \mathbf{y}^b))$ comme une règle virtuelle exprimant soit que le changement de $\mathbf{x}^a$ à $\mathbf{x}^b$ induit le changement de $\mathbf{y}^a$ à $\mathbf{y}^b$, quel que soit le contexte du problème (encodé par les caractéristiques où $\mathbf{x}^a$ et $\mathbf{x}^b$ sont égaux), soit que le changement de $\mathbf{x}^a$ à $\mathbf{x}^b$ ne modifie pas la solution (dans le cas $\mathbf{y}^a = \mathbf{y}^b$).

Cependant, de telles règles virtuelles peuvent avoir des exceptions dans la base de cas. En effet, il peut exister $(\mathbf{x}^a, \mathbf{x}^b)$, $(\mathbf{x}^c, \mathbf{x}^d)$ et $(\mathbf{x}^{a'}, \mathbf{x}^{b'})$ tels que

— $\mathbf{x}^a:\mathbf{x}^b::\mathbf{x}^c:\mathbf{x}^d$ et $\mathbf{x}^{a'}:\mathbf{x}^{b'}::\mathbf{x}^c:\mathbf{x}^d$ sont vraies. Cela signifie que $\mathcal{D}_{(\mathbf{x}^a,\mathbf{x}^b)} = \mathcal{D}_{(\mathbf{x}^c,\mathbf{x}^d)} = \mathcal{D}_{(\mathbf{x}^{a'},\mathbf{x}^{b'})}$ et que les changements de $\mathbf{x}^a$ à$\mathbf{x}^b$, de $\mathbf{x}^c$ à $\mathbf{x}^d$, de $\mathbf{x}^{a'}$ à $\mathbf{x}^{b'}$ sont les mêmes.

— $\mathbf{y}^a:\mathbf{y}^b::\mathbf{y}^c:\mathbf{y}^d$ est vraie.

— $\mathbf{y}^{a'}:\mathbf{y}^{b'}::\mathbf{y}^{c'}:\mathbf{y}^{d'}$ *ne tient pas* pour une caractéristique i.

Cela peut arriver par exemple quand $\mathbf{y}^{a_i} \neq \mathbf{y}^{b_i}$ tandis que $\mathbf{y}^{a'_i} = \mathbf{y}^{b'_i}$. Dans une telle situation, les deux règles virtuelles associées à $((\mathbf{x}^a, \mathbf{y}^{a_i}), ((\mathbf{x}^b, \mathbf{y}^{b_i}))$ et à $((\mathbf{x}^{a'}, \mathbf{y}^{a'_i}), (\mathbf{x}^{b'}, \mathbf{y}^{b'_i}))$ sont en désaccord. On a alors $\mathbf{y}^{a_i} = \mathbf{y}^{c_i}$ et $\mathbf{y}^{b_i} = \mathbf{y}^{d_i}$, mais la solution de l'équation $\mathbf{y}^{a'_i}:\mathbf{y}^{b'_i}::\mathbf{y}^{c_i}:?\mathbf{y}_i$ est $?\mathbf{y}_i = \mathbf{y}^{c_i} \neq \mathbf{y}^{d_i}$.

Ainsi le taux d'exceptions de la règle virtuelle associée à un couple $((\mathbf{x}^a, \mathbf{y}^{a_i}), (\mathbf{x}^b, \mathbf{y}^{b_i}))$ est une indication de l'intérêt du couple pour l'inférence analogique. C'est ce qu'on appelle la compétence du couple de cas [17]. Notez que

chaque règle se rapporte à une caractéristique particulière utilisée dans la description des solutions. En effet, il n'y a pas toujours une règle unique qui calcule l'adaptation de $\mathtt{y}^a$ en $\mathtt{y}^b$ à partir du même contexte de problème. Appliquons l'inférence analogique à l'exemple suivi. Il est facile de vérifier que son expression propositionnelle donne naissance au tableau de la figure 1. En considérant les 3 cas $(\mathtt{cas}^a, \mathtt{cas}^b, \mathtt{cas}^s)$, il apparaît que $\mathtt{x}^a : \mathtt{x}^b :: \mathtt{x}^s : \mathtt{x}^{\mathtt{cible}}$ est valide pour les caractéristiques `poisson`, `pesto` et `vanille`. De plus, l'équation analogique $\mathtt{x}^a : \mathtt{x}^b :: \mathtt{x}^s : \mathtt{x}^{\mathtt{cible}}$ peut être résolue pour les caractéristiques `sauce`, `fruit`, `ananas`, `Saint-Pierre`, `saumon`, et `crème`. On obtient ainsi

$$\begin{array}{rcl} \mathtt{cas}^{\mathtt{cible}}_{\text{extrap}} & = & \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}_{\text{extrap}} \\ \text{avec } \mathtt{y}^{\mathtt{cible}}_{\text{extrap}} & \equiv & \mathtt{sauce} \wedge \mathtt{fruit} \wedge \mathtt{ananas} \\ & \wedge & \neg\mathtt{Saint\text{-}Pierre} \wedge \mathtt{saumon} \\ & \wedge & \mathtt{crème} \end{array} \quad (6)$$

Quelques remarques méritent d'être mentionnées :

1. L'équation analogique ne peut pas être résolue pour la caractéristique `curry`, donc rien n'est inféré concernant sa présence ou son absence à partir des cas $(\mathtt{cas}^a, \mathtt{cas}^b, \mathtt{cas}^s)$. Dans un exemple plus réaliste (où la base de cas serait plus riche), il pourrait exister un triplet de cas nous permettant de conclure sur `curry`.
2. Cependant, même si une résolution exacte pour `curry` n'existe pas, il serait possible de minimiser une mesure de dissimilarité analogique $AD(a,b,c,d)$ pour calculer une solution « approchée ». $AD(a,b,c,d)$ est égal au nombre minimal de *flips* pour passer de (a,b,c,d) à un quadruplet correspondant à une proportion analogique : $AD(a,b,c,d)$ est maximal (et égal à 2) pour $(1,0,0,1)$ (ou $(0,1,1,0)$) [19]. Ainsi ici, la solution approchée serait 0 (c'est-à-dire pas de `curry`) puisque $AD(1,0,0,0) = 1$.
3. En général, de nombreux triplets de cas peuvent être appliqués à un même $\mathtt{x}^{\mathtt{cible}}$. Rappelons que l'on ne doit utiliser dans de telles situations que les triplets construits sur les couples les plus compétents (dans le cas où ils mènent à des conclusions différentes, un vote doit avoir lieu).
4. Comme on peut le voir dans le calcul de $\mathtt{cas}^{\mathtt{cible}}$, une modification de $\mathtt{cas}^s$ a lieu : à savoir, dans le contexte `poisson` et $\neg$`vanille` avec l'ajout de `pesto`, $\mathtt{cas}^s$ est adapté en changeant `Saint-Pierre` en `saumon`.
5. Dans cet exemple, certaines caractéristiques sont liées par des implications. Il convient de noter que si pour deux caractéristiques mutuellement exclusives i et j, on a $\mathtt{x}^{a_i} : \mathtt{x}^{b_i} :: \mathtt{x}^{c_i} : \mathtt{x}^{d_i}$ et $\mathtt{x}^{a_j} : \mathtt{x}^{b_j} :: \mathtt{x}^{c_j} : \mathtt{x}^{d_j}$, cela implique que c'est également vrai pour une caractéristique k pour laquelle i et j sont des sous-classes, comme on peut le voir sur l'exemple.
6. On peut observer que les rôles de $\mathtt{cas}^b$ et $\mathtt{cas}^s$ pourraient être échangés, puisque les proportions analogiques sont stables par permutation centrale.

2.5 La révision des croyances et le RàPC

La révision des croyances. Considérons un agent ayant un ensemble de croyances ψ et qui est confronté à un autre ensemble de croyances μ, lequel est supposé être prioritaire par rapport à ψ. Dans cet article, ψ et μ sont supposés être représentés en logique propositionnelle. La question soulevée par la révision des croyances est « Comment les croyances de l'agent évoluent-elles par incorporation de μ ? » Quand le nouvel ensemble de croyances n'est pas en contradiction avec l'ancien — i.e., $\psi \wedge \mu$ est cohérent — alors la révision donne simplement la conjonction $\psi \wedge \mu$. Sinon, selon le principe de changement minimal de la théorie AGM (nommée d'après les initiales des auteurs de [2]), la révision des croyances consiste à faire un « changement minimal » de ψ en ψ' de telle sorte que $\psi' \wedge \mu$ soit cohérent, le résultat de la révision est alors $\psi' \wedge \mu$, noté $\psi \dot{+} \mu$ dans la suite. Cependant, la notion de minimalité du changement n'est pas définie de façon unique et dépend de la manière dont le changement est évalué. Par conséquent, de nombreux *opérateurs de révision* $\dot{+}$ existent. Les postulats AGM proposent un ensemble de postulats qu'un tel opérateur $\dot{+}$ devrait respecter. Ces postulats ont été formulés en logique propositionnelle par [12] de la façon suivante (pour toutes formules $\varphi, \psi, \varphi_1, \varphi_2, \psi_1, \psi_2$ et χ) :

($\dot{+}1$) Si μ est cohérent alors $\psi \dot{+} \mu$ est cohérent.

($\dot{+}2$) Si $\psi \wedge \mu$ est cohérent alors $\psi \dot{+} \mu \equiv \psi \wedge \mu$.

($\dot{+}3$) $\psi \dot{+} \mu \models \mu$.

($\dot{+}4$) Si $\psi_1 \equiv \psi_2$ et $\mu_1 \equiv \mu_2$ alors $\psi_1 \dot{+} \mu_1 \equiv \psi_2 \dot{+} \mu_2$.

($\dot{+}5$) Si $(\psi \dot{+} \mu) \wedge \chi$ est cohérent
alors $\psi \dot{+} (\mu \wedge \chi) \models (\psi \dot{+} \mu) \wedge \chi$.

($\dot{+}6$) $(\psi \dot{+} \mu) \wedge \chi \models \psi \dot{+} (\mu \wedge \chi)$.

($\dot{+}1$) affirme qu'un agent vise à avoir des croyances cohérentes (à moins d'accepter un ensemble de croyances μ incohérent). ($\dot{+}2$) est lié au principe du changement minimal : si ψ est cohérent avec μ alors le changement minimal $\psi \mapsto \psi'$ est $\psi' = \psi$ (i.e., pas de changement). ($\dot{+}3$) est lié au fait que μ est prioritaire par rapport à ψ, i.e., les seuls changements de croyances effectuées le sont sur ψ : μ est inchangé. ($\dot{+}4$) affirme que la révision doit respecter le principe d'indépendance à la syntaxe (substituer une formule par une formule équivalente doit conserver le résultat de l'inférence, à l'équivalence logique près). On peut montrer que la conjonction des postulats ($\dot{+}5$) et ($\dot{+}6$) est équivalente à l'assertion suivante : si $(\psi \dot{+} \mu) \wedge \chi$ est cohérent alors $\psi \dot{+} (\mu \wedge \chi) \equiv (\psi \dot{+} \mu) \wedge \chi$. En d'autres termes, s'il n'y a pas de nécessité de modifier les croyances après la révision de ψ par μ pour incorporer χ, alors aucune modification additionnelle n'est effectuée : les nouvelles croyances

	poisson	pesto	vanille	sauce	fruit	ananas	Saint-Pierre	saumon	crème	curry
$\mathtt{cas}^a$	1	0	1	1	0	0	1	0	1	1
$\mathtt{cas}^b$	1	1	1	1	0	0	0	1	1	0
$\mathtt{cas}^s$	1	0	0	1	1	1	1	0	1	0
$\mathtt{x}^{\mathtt{cible}}$	1	1	0	?	?	?	?	?	?	?
$\mathtt{cas}^{\mathtt{cible}}_{\mathrm{extrap}}$	1	1	0	1	1	1	0	1	1	?

FIGURE 1 – Inférence analogique dans l'exemple suivi.

χ sont simplement ajoutées aux croyances $\psi \dot{+} \mu$. C'est également lié au principe du changement minimal : quand aucun changement n'est nécessaire pour restaurer la cohérence, alors aucun changement n'est effectué.

Malgré ces postulats, l'ensemble des opérateurs de révision des croyances reste très grand et dépend de la façon dont le changement est évalué. En particulier, ce changement peut être évalué grâce à une mesure de similarité entre interprétations, ce qui permet de définir la famille d'opérateurs de révision des croyances présentée ci-dessous.

Opérateurs de révision fondée sur des mesures de similarité. Une mesure de similarité sur un ensemble E est définie dans cet article comme étant une fonction $\mathtt{sim} : E \times E \to [0,1]$ telle que $\mathtt{sim}(a,b) = 1$ ssi $a = b$ (pour $a, b \in E$). Un opérateur de révision des croyances $\dot{+}^{\mathtt{sim}}$ satisfaisant les postulats AGM peut être défini pour toute mesure de similarité $\mathtt{sim}$ sur Ω, l'ensemble des interprétations, de la façon suivante :

$$\text{avec } \mathtt{sim}^* = \max\left\{\mathtt{sim}(\mathcal{I},\mathcal{J}) \;\middle|\; \begin{array}{l}\mathcal{I} \in \mathcal{M}(\psi) \text{ et}\\ \mathcal{J} \in \mathcal{M}(\mu)\end{array}\right\}$$
$$\mathcal{M}(\psi \dot{+}^{\mathtt{sim}} \mu) = \left\{\mathcal{J} \in \mathcal{M}(\mu) \;\middle|\; \max_{\mathcal{I}\in\mathcal{M}(\psi)} \mathtt{sim}(\mathcal{I},\mathcal{J}) = \mathtt{sim}^*\right\} \quad (7)$$

En d'autres termes, les modèles de $\psi \dot{+}^{\mathtt{sim}} \mu$ sont les modèles de μ les plus similaires aux modèles de ψ, selon $\mathtt{sim}$. Notons que cela ne permet de définir $\dot{+}^{\mathtt{sim}}$ que modulo l'équivalence logique, ce qui ne soulève pas de problème : toute formule ϱ telle que $\varrho \equiv \psi \dot{+}^{\mathtt{sim}} \mu$ constitue une $\dot{+}^{\mathtt{sim}}$-révision de ψ par μ selon le principe d'indépendance à la syntaxe.

Étant donné une fonction $\mathtt{dist}$ sur Ω, une mesure de similarité $\mathtt{sim}$ peut être définie par $\mathtt{sim}(\mathcal{I},\mathcal{J}) = 1/(1 + \mathtt{dist}(\mathcal{I},\mathcal{J}))$ pour $\mathcal{I},\mathcal{J} \in \Omega$. En particulier, soit H la distance de Hamming entre interprétations : $H(\mathcal{I},\mathcal{J})$ est le nombre de variables a telles que $\mathcal{I}(a) \neq \mathcal{J}(a)$. Soit à présent $\mathtt{sim}_H$ la mesure de similarité associée à H. L'opérateur de révision des croyances $\dot{+}^{\mathtt{sim}_H}$ coincide avec l'opérateur de Dalal [5] et sera noté $\dot{+}_{\mathrm{Dalal}}$ dans la suite. La distance de Hamming pondère chaque variable de la même façon et considère les variables indépendamment les unes des autres : elle peut être vue comme une distance d'édition sur des interprétations fondée sur l'opérateur *flip*. Pour cette raison, quand il n'y a aucune connaissance explicitant comment les changements doivent être évalués, la distance de Hamming étant « neutre » est utilisée pour mesurer ces changements. C'est pourquoi $\dot{+}_{\mathrm{Dalal}}$ est utilisé comme un opérateur « non informé », avec une connaissance sur le changement vide.

L'adaptation par révision est une approche d'adaptation simple fondée sur un opérateur de révision $\dot{+}$. L'intuition est que la modification du cas remémoré $\mathtt{x}^s \wedge \mathtt{y}^s$ afin de proposer une solution au problème cible $\mathtt{x}^{\mathtt{cible}}$ est effectuée par $\dot{+}$. Le cas remémoré et le problème cible sont interprétés en accord avec les connaissances du domaine, la révision qui doit être effectuée est donc celle de $\mathtt{CD} \wedge \mathtt{x}^s \wedge \mathtt{y}^s$ par $\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}$. Le résultat de cette révision est une formule ϱ qui entraîne $\mathtt{x}^{\mathtt{cible}}$ (selon le postulat ($\dot{+}3$)) et donc, ϱ est équivalent à une formule $\mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}$ où toutes les variables de $\mathtt{y}^{\mathtt{cible}}$ appartiennent à $\mathcal{V}_S$. $\mathtt{y}^{\mathtt{cible}}$ est la solution proposée pour $\mathtt{x}^{\mathtt{cible}}$. Formellement, cela s'écrit comme suit :

$$(\mathtt{CD} \wedge \mathtt{x}^s \wedge \mathtt{y}^s) \;\dot{+}\; (\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}) \quad\equiv\quad \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}} \quad (8)$$

Afin d'appliquer l'adaptation par révision, un opérateur de révision doit être choisi. Ce choix est lié à la façon dont le changement est évalué, c'est-à-dire, en terme de RàPC, à la connaissance d'adaptation. En particulier, quand aucune connaissance d'adaptation n'est disponible, l'opérateur de révision de Dalal est utilisé.

On peut noter que la solution $\mathtt{y}^{\mathtt{cible}}$ donnée par adaptation par révision est nécessairement cohérente pour peu que $\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}$ le soit, mais elle n'est pas nécessairement complètement décrite : $\mathcal{M}\left(\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}\right)$ peut contenir plusieurs interprétations. Dans une telle situation, cela signifie que l'adaptation par révision affirme qu'il existe une solution plausible $\mathtt{y}$ de $\mathtt{x}^{\mathtt{cible}}$ vérifiant $\mathtt{y} \models \mathtt{y}^{\mathtt{cible}}$. Dans le cas extrême, $\mathtt{y}^{\mathtt{cible}} \equiv \top$, ce qui signifie que l'adaptation

par révision ne donne aucune information sur une solution potentielle de $\mathtt{x}^{\mathtt{cible}}$.

L'exemple suivi peut être résolu en utilisant l'adaptation par révision. On suppose que le cas remémoré est $\mathtt{cas}^s$ et que le problème cible est $\mathtt{x}^{\mathtt{cible}}$, définis par (3) et (4). On considérera dans cette section que l'adaptation utilise $\dot{+}_{\text{Dalal}}$. On peut alors montrer que le résultat $\mathtt{y}^{\mathtt{cible}}$ de cette révision est

$$\begin{array}{rcl} \mathtt{y}^{\mathtt{cible}}_{\text{Dalal}} & \equiv & \mathtt{Saint\text{-}Pierre} \wedge \mathtt{crème} \wedge \mathtt{sauce} \\ & \wedge & \neg\mathtt{fruit} \wedge \neg\mathtt{ananas} \wedge \neg\mathtt{saumon} \\ & \wedge & \neg\mathtt{curry} \end{array} \quad (9)$$

Ainsi, cette adaptation consiste à enlever de $\mathtt{x}^s \wedge \mathtt{y}^s$ les fruits, puisque leur présence serait incohérente avec $\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}$ (à cause du pesto).

3 Un pont entre l'extrapolation et l'adaptation par révision

Les deux approches de l'adaptation des cas présentées ci-dessus — celle fondée sur l'extrapolation analogique et celle fondée sur la révision des croyances — apparaissent de façon assez différente : la première est une adaptation multiple de cas alors que la seconde est une application simple de cas (elles s'appuient respectivement sur la remémoration de cas sources par triplets et par singletons). Néanmoins, le but de cette section est de montrer comment ces deux approches peuvent se rencontrer. Premièrement, l'approche fondée sur l'extrapolation est reformulée comme une adaptation simple. Deuxièmement, un opérateur de révision fondé sur des couples de cas est défini et il est montré comment, sous certaines hypothèses, les deux approches de l'adaptation coïncident. Cela permet de définir une approche de l'adaptation fondée à la fois sur l'extrapolation et la révision qui prend en compte, d'un côté, la base de cas et les compétences des couples de cas et, d'un autre côté, les connaissances du domaine.

3.1 Reformuler l'adaptation par extrapolation comme une adaptation simple

Dans la présentation faite ci-dessus de l'extrapolation analogique, on considérait qu'un triplet de cas $(\mathtt{cas}^a, \mathtt{cas}^b, \mathtt{cas}^c)$ était remémoré puis réutilisé pour résoudre $\mathtt{x}^{\mathtt{cible}}$. À présent, cela va être reformulé en considérant que seul le cas $\mathtt{cas}^c$ est remémoré, et que les autres, $\mathtt{cas}^a$ et $\mathtt{cas}^b$, sont sélectionnés durant le processus d'adaptation lui-même. Cette « brisure de symétrie » a deux avantages. D'abord, elle peut être utilisée pour une implantation efficace de l'extrapolation analogique (ce point a été détaillé dans les algorithmes d'extrapolation décrits dans [18]). Ensuite, cela facilite l'association des deux approches de l'adaptation. Pour cette raison, le cas remémoiré $\mathtt{cas}^c$ est renommé $\mathtt{cas}^s = (\mathtt{x}^s, \mathtt{y}^s)$, pour mieux correspondre avec les notations de l'adaptation simple.

Par conséquent, la reformulation de l'extrapolation analogique en une adaptation simple se fait comme suit :

Entrée : un cas à adapter $(\mathtt{x}^s, \mathtt{y}^s)$, le problème cible $\mathtt{x}^{\mathtt{cible}}$, la base de cas $\mathtt{BC}$, la relation de préférence entre couples de cas ;

Sortie : un ensemble $\mathtt{Y}$ de solutions proposées pour $\mathtt{x}^{\mathtt{cible}}$

1. Soit $\mathtt{CouplesCandidats}$ l'ensemble des $(\mathtt{cas}^a, \mathtt{cas}^b) \in \mathtt{BC} \times \mathtt{BC}$ tels que $\mathtt{x}^a : \mathtt{x}^b :: \mathtt{x}^s : \mathtt{x}^{\mathtt{cible}}$ et tels que l'équation analogique $\mathtt{y}^a : \mathtt{y}^b :: \mathtt{y}^s : ?\mathtt{y}$ admet une solution.
2. Soit $\mathtt{MeilleursCouplesCandidats}$ l'ensemble des couples de cas les plus compétents de $\mathtt{CouplesCandidats}$.
3. Soit $\mathtt{Y} = \left\{ \mathtt{y} \;\middle|\; \begin{array}{l} \mathtt{y} \text{ est la solution de } \mathtt{y}^a : \mathtt{y}^b :: \mathtt{y}^s : ?\mathtt{y} \\ \text{pour } (\mathtt{cas}^a, \mathtt{cas}^b) \in \\ \quad \mathtt{MeilleursCouplesCandidats} \end{array} \right\}$.
4. $\mathtt{Y}$ est retourné comme ensemble de solutions candidates pour $\mathtt{x}^{\mathtt{cible}}$.

3.2 Un opérateur de révision fondé sur la compétence de couples de cas

En s'appuyant sur la compétence des couples de cas et sous certaines hypothèses, on peut définir une mesure de similarité $\mathtt{sim}_{\mathtt{comp}}$ sur Ω, ce qui permet d'introduire l'opérateur de révision $\dot{+}^{\mathtt{sim}_{\mathtt{comp}}}$. Il est notable que pour les préordres de compétences présentés dans [17], ces hypothèses sont vérifiées.

La première hypothèse est que le préordre de compétence entre couples de cas peut être caractérisé grâce à un *niveau de compétence*, i.e., une fonction $\mathtt{nivComp}$ qui associe à un couple de cas sources une valeur de [0, 1] telle que $(\mathtt{cas}^1, \mathtt{cas}^2)$ est considéré comme étant strictement plus compétent que $(\mathtt{cas}^3, \mathtt{cas}^4)$ ssi $\mathtt{nivComp}(\mathtt{cas}^1, \mathtt{cas}^2) > \mathtt{nivComp}(\mathtt{cas}^3, \mathtt{cas}^4)$.

La deuxième hypothèse traduit simplement le fait que deux couples de cas reliés par des proportions analogiques (dans l'espace des problèmes et l'espace des solutions) ont la même compétence et donc le même niveau de compétence :

$$\begin{array}{rl} \textbf{si} & \mathtt{x}^a : \mathtt{x}^b :: \mathtt{x}^c : \mathtt{x}^d \textbf{ et } \mathtt{y}^a : \mathtt{y}^b :: \mathtt{y}^c : \mathtt{y}^d \\ \textbf{alors} & \mathtt{nivComp}(\mathtt{cas}^a, \mathtt{cas}^b) = \mathtt{nivComp}(\mathtt{cas}^c, \mathtt{cas}^d) \end{array} \quad (10)$$

La troisième hypothèse est que le niveau maximum de compétence est 1 et est atteint seulement par les couples $(\mathtt{cas}^a, \mathtt{cas}^a)$ pour $\mathtt{cas}^a \in \mathtt{BC}$. Cette troisième hypothèse peut être justifiée par le fait que si $\mathtt{x}^a : \mathtt{x}^a :: \mathtt{x}^s : \mathtt{x}^{\mathtt{cible}}$ alors $\mathtt{x}^s = \mathtt{x}^{\mathtt{cible}}$ (d'après (5)) et donc que l'équation analogique $\mathtt{y}^a : \mathtt{y}^a :: \mathtt{y}^s : ?\mathtt{y}$ a exactement une solution $?\mathtt{y} = \mathtt{y}^s$ qui résout $\mathtt{x}^{\mathtt{cible}} = \mathtt{x}^s$.

La quatrième hypothèse est que le niveau de compétence est minimal pour $(\mathtt{cas}^a, \mathtt{cas}^b)$ si et seulement si ces deux cas ne sont en analogie avec aucun couple de cas $(\mathtt{cas}^c, \mathtt{cas}^d)$ de la base de cas (en particulier, au moins un cas parmi $\mathtt{cas}^a$ et $\mathtt{cas}^b$ n'est pas un cas de la base de cas).

Une bijection $\mathtt{casDe}$ entre Ω et l'ensemble des cas complètement décrits peut être définie, pour $\mathcal{I} \in \Omega$, par $\mathtt{casDe}(\mathcal{I}) = (\mathtt{x}, \mathtt{y})$ est tel que $\mathcal{M}(\mathtt{x} \wedge \mathtt{y}) = \{\mathcal{I}\}$. On peut alors définir une mesure de similarité $\mathtt{sim}_{\mathtt{comp}}$ utilisant cette bijection et la fonction donnant le niveau de compétence :

$$\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}, \mathcal{J}) = \mathtt{nivComp}(\mathtt{casDe}(\mathcal{I}), \mathtt{casDe}(\mathcal{J})) \quad \text{(pour } \mathcal{I}, \mathcal{J} \in \Omega)$$

La rencontre entre les deux approches de l'adaptation peut être exprimée par le résultat suivant (étant donné un cas source $(\mathtt{x}^s, \mathtt{y}^s)$ et un problème cible $\mathtt{x}^{\mathtt{cible}}$) :

si	$\mathtt{y}^{\mathtt{cible}}$ est le résultat de l'adaptation par révision avec $\dotplus = \dotplus^{\mathtt{sim}_{\mathtt{comp}}}$, les connaissances du domaine sont vides ($\mathtt{CD} = \top$), $\mathtt{Y}$ est l'ensemble des solutions obtenues par extrapolation analogique et $\mathtt{Y} \neq \emptyset$	(11)
alors	$\mathtt{y}^{\mathtt{cible}} \equiv \bigvee \mathtt{Y}$	

En d'autres termes, en l'absence de connaissances du domaine, l'adaptation par révision utilisant l'opérateur de révision $\dotplus^{\mathtt{sim}_{\mathtt{comp}}}$ qui s'appuie sur la compétence de couples de cas donne le même résultat que l'extrapolation analogique, à moins que cette dernière ne donne aucune solution. De plus, on peut montrer que si cet ensemble est vide, alors $\mathtt{y}^{\mathtt{cible}} \equiv \top$, i.e., dans ce cas, l'adaptation par révision ne donne aucune information sur la solution.

La preuve de (11) est donnée en figure 2.

3.3 Une approche de l'adaptation fondée sur l'extrapolation et la révision

À présent, considérons l'application de l'adaptation par $\dotplus^{\mathtt{sim}_{\mathtt{comp}}}$ pour l'exemple suivi et, dans un premier temps, avec $\mathtt{CD}$ vide. À cause de la variable $\mathtt{curry}$, l'extrapolation analogique ne donne aucune solution [4] ($\mathtt{Y} = \emptyset$) et, dans cette situation, $\mathtt{y}^{\mathtt{cible}} \equiv \top$ (aucune information sur la solution n'est proposée). Si la variable $\mathtt{curry}$ est supprimée de $\mathcal{V}$ alors, selon (11), la solution proposée vérifie

$$\mathtt{y}^{\mathtt{cible}} = \mathtt{y}^{\mathtt{cible}}_{\dotplus^{\mathtt{sim}_{\mathtt{comp}}}\text{-adaptation sans } \mathtt{curry}} \equiv \mathtt{y}^{\mathtt{cible}}_{\mathtt{extrap}} \tag{14}$$

À présent, afin de définir une approche qui considère toutes les variables, même celles telles que $\mathtt{curry}$ pour laquelle l'extrapolation ne s'applique pas exactement, une mesure de similarité $\mathtt{sim}_{\mathtt{E\&D}}$ peut être définie qui combine $\mathtt{sim}_{\mathtt{comp}}$ et $\mathtt{sim}_H$ (pour un α donné, $0 < \alpha < 1$, et $\mathcal{I}, \mathcal{J} \in \Omega$) :

$$\mathtt{sim}_{\mathtt{E\&D}}(\mathcal{I}, \mathcal{J}) = (1 - \alpha)\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}, \mathcal{J}) + \alpha \mathtt{sim}_H(\mathcal{I}, \mathcal{J}) \tag{15}$$

d'où l'opérateur de révision $\dotplus^{\mathtt{sim}_{\mathtt{E\&D}}}$ (E&D pour « extrapolation et Dalal »). Donc, pour un α suffisamment petit, l'adaptation par $\dotplus^{\mathtt{sim}_{\mathtt{E\&D}}}$ consiste à faire un nombre minimal de *flips* de variables sur $\mathtt{cas}^s$ pour rendre l'extrapolation possible puis l'appliquer [5].

L'application de l'adaptation par $\dotplus^{\mathtt{sim}_{\mathtt{E\&D}}}$ sur l'exemple suivi donne :

$$\begin{aligned} \mathtt{y}^{\mathtt{cible}}_{\mathtt{E\&D}} \equiv\ & \mathtt{sauce} \wedge \mathtt{fruit} \wedge \mathtt{ananas} \\ \wedge\ & \neg\mathtt{Saint\text{-}Pierre} \wedge \mathtt{saumon} \\ \wedge\ & \mathtt{crème} \wedge \neg\mathtt{curry} \end{aligned} \tag{16}$$

On peut expliquer cela comme suit. L'équation analogique $1\mathtt{:}0\mathtt{::}0\mathtt{:?}\mathtt{y}_{\mathtt{curry}}$ pour trouver la valeur de l'attribut $\mathtt{curry}$ dans $\mathtt{y}^{\mathtt{cible}}$ par extrapolation de $(\mathtt{cas}^a, \mathtt{cas}^b, \mathtt{cas}^s)$ n'a pas de solution, donc un *flip* de cet attribut pour $\mathtt{cas}^s$ donne $\mathtt{cas}^{s\prime}$ et le triplet $(\mathtt{cas}^a, \mathtt{cas}^b, \mathtt{cas}^{s\prime})$ peut être utilisé par extrapolation pour résoudre $\mathtt{x}^{\mathtt{cible}}$ en $\mathtt{y}^{\mathtt{cible}} = \mathtt{y}^{\mathtt{cible}}_{\mathtt{E\&D}}$, puisque $1\mathtt{:}0\mathtt{::}1\mathtt{:?}\mathtt{y}_{\mathtt{curry}}$ a une solution unique $\mathtt{?y}_{\mathtt{curry}} = 0$ (d'où le $\neg\mathtt{curry}$ dans la solution proposée de $\mathtt{x}^{\mathtt{cible}}$).

Cette adaptation n'a pas tenu compte des connaissances du domaine $\mathtt{CD}$ définies dans la section 2.3 et, de fait, $\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}_{\mathtt{E\&D}}$ est incohérent avec ces connaissances du domaine (à cause du conflit pesto-fruit). Par conséquent, l'approche de l'adaptation proposée consiste à faire une adaptation par $\dotplus^{\mathtt{sim}_{\mathtt{E\&D}}}$ prenant en compte $\mathtt{CD}$, selon (8). Dans le cadre de l'exemple suivi, cela donne :

$$\begin{aligned} \mathtt{y}^{\mathtt{cible}}_{\mathtt{E\&D\ avec\ CD}} \equiv\ & \mathtt{sauce} \wedge \neg\mathtt{fruit} \wedge \neg\mathtt{ananas} \\ \wedge\ & \neg\mathtt{Saint\text{-}Pierre} \wedge \mathtt{saumon} \\ \wedge\ & \mathtt{crème} \wedge \neg\mathtt{curry} \end{aligned} \tag{17}$$

qui consiste à enlever les fruits de $\mathtt{y}^{\mathtt{cible}}_{\mathtt{E\&D}}$.

3.4 Synthèse

La figure 3 décrit les cas sources et le problème cible de l'exemple suivi, ainsi que les cas proposés $\mathtt{cas}^s = \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}$ après les différents processus d'adaptation.

Cet exemple illustre comment les atouts des deux approches d'adaptation peuvent être combinés :

— La force de l'extrapolation analogique est d'exploiter les variations dans la base de cas (variations pouvant être considérées comme des règles d'adaptation spécifiques) : la variation « Saint-Pierre vers saumon » de $\mathtt{cas}^a$ à $\mathtt{cas}^b$ est appliquée à $\mathtt{cas}^s$ dans un contexte similaire.

4. Plus précisément, l'équation analogique $\mathtt{y}^a\mathtt{:}\mathtt{y}^b\mathtt{::}\mathtt{y}^s\mathtt{:y}$ n'a pas de solution : cette équation est résoluble attribut par attribut sur tous les attributs à l'exception de $\mathtt{curry}$, donc une solution peut être proposé pour tous les autres attributs.

5. On peut montrer qu'il est suffisant d'avoir $\alpha < (2\,|\mathcal{V}|)^{-1}$.

Cette preuve repose essentiellement sur l'application directe des définitions. Cependant, elle nécessite l'introduction de notations pour « naviguer » entre la logique propositionnelle et les ensembles d'interprétations.

On suppose les prémisses de l'implication (11).

Pour quatre interprétations $\mathcal{I}^a$, $\mathcal{I}^b$, $\mathcal{I}^c$ et $\mathcal{I}^d$, on note $\mathcal{I}^a : \mathcal{I}^b :: \mathcal{I}^c : \mathcal{I}^d$ si $\mathcal{I}^a(v) : \mathcal{I}^b(v) :: \mathcal{I}^c(v) : \mathcal{I}^d(v)$ pour toute variable v.

Pour $(\mathtt{x}^\sigma, \mathtt{y}^\sigma) \in \mathtt{BC}$, on note $\mathcal{I}^\sigma$ l'interprétation telle que $\mathcal{M}(\mathtt{CD} \wedge \mathtt{x}^\sigma \wedge \mathtt{y}^\sigma) = \{\mathcal{I}^\sigma\}$ (ce qui est possible puisque les cas sources sont supposés complètement décrits). En particulier, $\mathcal{I}^s$ est l'unique modèle de $\mathtt{CD} \wedge \mathtt{x}^s \wedge \mathtt{y}^s$, pour $(\mathtt{x}^s, \mathtt{y}^s)$ le cas à adapter. On note $M^{\mathtt{BC}} = \{\mathcal{I}^\sigma \mid (\mathtt{x}^\sigma, \mathtt{y}^\sigma) \in \mathtt{BC}\}$.

Soit $M^{\mathtt{cible}} = \mathcal{M}\left(\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}\right)$, $M^{\mathtt{cible}}_{\dot{+}} = \mathcal{M}\left(\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}\right)$ et $M^{\mathtt{cible}}_{\text{extrap}} = \bigcup\left\{\mathcal{M}\left(\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}\right) \mid \mathtt{y} \in \mathtt{Y}\right\}$. L'objectif de la preuve est de montrer que $M^{\mathtt{cible}}_{\dot{+}} = M^{\mathtt{cible}}_{\text{extrap}}$.

Soit $\mathtt{sim}^*_{\mathtt{comp}} = \max\left\{\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^s, \mathcal{J}) \mid \mathcal{J} \in \mathcal{M}\left(\mathtt{CD} \wedge \mathtt{x}^{\mathtt{cible}}\right)\right\}$.

Pour $\mathcal{J} \in M^{\mathtt{cible}}$, on a :

$$\mathcal{J} \in M^{\mathtt{cible}}_{\dot{+}} \quad \text{ssi} \quad \mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^s, \mathcal{J}) = \mathtt{sim}^*_{\mathtt{comp}} \tag{12}$$

$$\mathcal{J} \in M^{\mathtt{cible}}_{\text{extrap}} \quad \text{ssi} \quad \left| \begin{array}{l} \text{il existe } (\mathcal{I}^a, \mathcal{I}^b) \in M^{\mathtt{BC}} \times M^{\mathtt{BC}} \text{ tel que} \\ \mathcal{I}^a : \mathcal{I}^b :: \mathcal{I}^s : \mathcal{J} \text{ et } \mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^a, \mathcal{I}^b) = \mathtt{sim}^*_{\mathtt{comp}} \end{array} \right. \tag{13}$$

Par conséquent, si $\mathcal{J} \in M^{\mathtt{cible}}_{\text{extrap}}$, d'après (13) et (10), on a $\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^s, \mathcal{J}) = \mathtt{sim}^*_{\mathtt{comp}}$. Donc, d'après (12), $\mathcal{J} \in M^{\mathtt{cible}}_{\dot{+}}$. On peut donc conclure que $M^{\mathtt{cible}}_{\text{extrap}} \subseteq M^{\mathtt{cible}}_{\dot{+}}$.

Réciproquement, si $\mathcal{J} \in M^{\mathtt{cible}}_{\dot{+}}$, on a $\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^s, \mathcal{J}) = \mathtt{sim}^*_{\mathtt{comp}}$. Or $\mathtt{Y} \neq \emptyset$, donc il existe $(\mathcal{I}^a, \mathcal{I}^b) \in M^{\mathtt{BC}} \times M^{\mathtt{BC}}$ tel que $\mathcal{I}^a : \mathcal{I}^b :: \mathcal{I}^s : \mathcal{J}$. Pour un tel couple, d'après (10), $\mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^a, \mathcal{I}^b) = \mathtt{sim}_{\mathtt{comp}}(\mathcal{I}^s, \mathcal{J}) = \mathtt{sim}^*_{\mathtt{comp}}$. Donc, selon (13), on a $\mathcal{J} \in M^{\mathtt{cible}}_{\text{extrap}}$. Par conséquent $M^{\mathtt{cible}}_{\dot{+}} \subseteq M^{\mathtt{cible}}_{\text{extrap}}$.

On a donc $M^{\mathtt{cible}}_{\dot{+}} = M^{\mathtt{cible}}_{\text{extrap}}$, ce qui permet de conclure. □

FIGURE 2 – Preuve de (11).

— La force de l'adaptation par $\dot{+}_{\text{Dalal}}$ est de prendre en compte les connaissances du domaine afin d'ajuster le cas remémoré pour proposer une solution au problème cible : le pesto étant incompatible avec les fruits, l'ananas est retiré de la recette [6].

— L'adaptation par $\dot{+}^{\mathtt{sim}_{\mathtt{comp}}}$ combine les deux approches d'adaptation : elle applique une extrapolation analogique sur chaque caractéristique pour laquelle c'est à la fois possible et cohérent avec les connaissances du domaine, et ajuste les autres caractéristiques avec l'adaptation par $\dot{+}_{\text{Dalal}}$.

4 Travaux proches et remarques finales

Cet article a examiné deux approches très différentes pour l'adaptation des cas — l'extrapolation analogique et l'adaptation fondée sur la révision — et a étudié la question de savoir comment elles peuvent se rencontrer. Il a été montré que, dans certaines circonstances (dans le cadre de la logique propositionnelle, sans connaissance du domaine, etc.), elles coïncident (cf. (11)) et que l'approche peut être étendue lorsque des connaissances du domaine sont ajoutées et/ou lorsque les proportions analogiques peuvent être trouvées seulement pour certaines composantes de la solution. L'idée est que les couples de cas et leurs compétences — utilisées dans extrapolation analogique — peuvent être utilisés pour « remodeler l'espace d'adaptation » en rendant plus similaire le cas source et le problème cible, et que cette similarité est utilisée par l'opérateur de révision.

Travaux connexes. Il existe une riche littérature sur la révision des croyances suite aux travaux fondateurs d'Alchourrón, Gärdenfors et Makinson [2], et son expression dans une logique propositionnelle [12]. Cependant, l'idée d'appliquer la révision des croyances au RàPC, telle que réaffirmée dans les préliminaires de cet article, ne peut être trouvée que dans quelques travaux (voir en particulier [4]).

L'idée d'un modèle d'inférence analogique fondé sur des proportions analogiques remonte à [20]. Son application au RàPC est suggérée dans [23], et étudiée plus en profondeur dans [18].

Il n'y a eu jusqu'à présent aucun travail reliant l'extrapolation analogique et la révision des croyances. En gardant à l'esprit que la révision des croyances et la logique non monotone sont, dans un certain sens, les deux faces d'une même pièce [10], nous pouvons cependant mentionner une discussion [21] opposant raisonnement non monotone et raisonnement analogique, mais aussi donnant des

6. Il est à noter que l'adaptation $\dot{+}_{\text{Dalal}}$ peut faire plus que supprimer les faits positifs car elle peut remplacer une classe par une classe sœur dans la taxonomie (voir, par exemple, [16]).

n°§		poisson	pesto	vanille	sauce	fruit	ananas	Saint-Pierre	saumon	crème	curry
2.3	$\mathtt{cas}^a$	1	0	1	1	0	0	1	0	1	1
	$\mathtt{cas}^b$	1	1	1	1	0	0	0	1	1	0
	$\mathtt{cas}^s$	1	0	0	1	1	1	1	0	1	0
	$\mathtt{x}^{\mathtt{cible}}$	1	1	0	?	?	?	?	?	?	?
2.4	$\mathtt{cas}^{\mathtt{cible}}_{\text{extrap}}$	1	1	0	1	1	1	0	1	1	?
2.5	$\mathtt{cas}^{\mathtt{cible}}_{\text{Dalal}}$	1	1	0	1	0	0	1	0	1	0
3.3	$\mathtt{cas}^{\mathtt{cible}}_{\text{E\&D}}$	1	1	0	1	1	1	0	1	1	0
3.3	$\mathtt{cas}^{\mathtt{cible}}_{\text{E\&D avec CD}}$	1	1	0	1	0	0	0	1	1	0

FIGURE 3 – L'exemple suivi : spécification du problème et résultats $\mathtt{cas}^{\mathtt{cible}} = \mathtt{x}^{\mathtt{cible}} \wedge \mathtt{y}^{\mathtt{cible}}$ des processus d'adaptation présentés dans l'article. La colonne n°§ indique le numéro de la section pertinente.

passerelles entre les deux.

Perspectives. L'étude présentée dans cet article montre que deux modèles d'adaptation peuvent se rencontrer, mais une question restante concerne l'utilité pratique de cette rencontre, au-delà de l'exemple courant. Dans ce but, une première perspective consiste à mener une expérimentation pour comparer les différentes approches présentées.

Les opérateurs de révision $\dot{+}^{\mathtt{sim}_{\mathtt{comp}}}$ et $\dot{+}^{\mathtt{sim}_{\text{E\&D}}}$ peuvent changer avec l'ajout d'un nouveau cas dans la base de cas, par exemple, le cas $(\mathtt{x}^{\mathtt{cible}}, \mathtt{y}^{\mathtt{cible}})$ lorsque ce nouveau cas est validé. En effet, le niveau de compétence des couples de cas est calculé sur la base de BC. L'évolution des opérateurs de révision au fil du temps est un problème lié à une révision itérée (voir, par exemple, [7]) ; il serait alors intéressant d'étudier $\dot{+}^{\mathtt{sim}_{\mathtt{comp}}}$ et $\dot{+}^{\mathtt{sim}_{\text{E\&D}}}$ à la lumière des postulats de la révision itérée.

Les opérateurs de révision des croyances considérés dans cet article sont fondés sur des mesures de similarité ou, de façon équivalente, sur des (pseudo)-distances. Par exemple, l'opérateur de révision de Dalal peut se définir sur la base de la distance de Hamming. L'étude d'autres distances entre interprétations pour la définition d'autres opérateurs de révision pourrait être profitable pour ce travail, en particulier les pseudo-distances *topic-decomposable* qui s'appuient sur une agrégation de pseudo-distances sur des sous-ensembles des variables propositionnelles (voir, notamment [14, 13]). Cela pourrait être utile en particulier pour incorporer des dépendances connues entre variables (à la différence de la distance de Hamming qui considère les variables comme étant indépendantes).

Cette rencontre entre adaptation par proportions analogiques et adaptation fondée sur la révision des croyances peut également être étudiée plus globalement pour la résolution de problèmes par RàPC (i.e., remémoration et adaptation) : cela constitue la troisième direction de travail.

Ce travail contribue à l'idée selon laquelle l'adaptation par révision recouvre une grande variété d'approches de l'adaptation : en l'occurrence, on peut choisir un opérateur de révision des croyances pour que l'adaptation par extrapolation analogique rencontre l'adaptation par révision. D'autres travaux en cours montrent que d'autres approches de l'adaptation sont également couvertes par l'adaptation par révision. Cela ne rend pas ces travaux caduques : l'approche de l'adaptation par révision ne dit pas grand chose sur le choix de l'opérateur de révision.

Enfin, on sait que la révision des croyances peut être encodée en logique possibiliste [9], puisque la révision des croyances repose au niveau sémantique sur des relations d'enracinement épistémiques [11], qui ne sont que des relations de nécessité qualitative au sens de la théorie des possibilités [8]. Comment traiter les différents types de révision/adaptation envisagés dans cet article dans le cadre possibiliste est un autre sujet de recherche future.

Références

[1] Aamodt, A. et E. Plaza: *Case-based Reasoning : Foundational Issues, Methodological Variations, and System Approaches*. AI Communications, 7(1) :39–59, 1994.

[2] Alchourrón, C. E., P. Gärdenfors et D. Makinson: *On the logic of theory change : partial meet functions for contraction and revision*. J. Symbolic Logic, 50 :510–530, 1985.

[3] Aristotle: *Nicomachean Ethics*. Univ. of Chicago Press, 2011. Trans. by R. C. Bartlett and S. D. Collins.

[4] Cojan, J. et J. Lieber: *Applying belief revision to case-based reasoning*. Dans Prade, H. et G. Richard (rédacteurs) : *Computational Approaches to Analogical Reasoning : Current Trends*, tome 548 de *Studies in*

Computational Intelligence, pages 133–161. Springer, 2014.

[5] Dalal, M.: *Investigations into a theory of Knowledge Base Revision : Preliminary Report*. Dans *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI)*, pages 475–479, 1988.

[6] d'Aquin, M., F. Badra, S. Lafrogne, J. Lieber, A. Napoli et L. Szathmary: *Case base mining for adaptation knowledge acquisition*. Dans Veloso, M. M. (rédacteur) : *IJCAI 2007, Proc. of the 20th Int. Joint Conf. on Artificial Intelligence, Hyderabad, Jan. 6-12*, pages 750–755, 2007.

[7] Darwiche, A. et J. Pearl: *On the logic of iterated belief revision*. Artificial intelligence, 89(1-2) :1–29, 1997.

[8] Dubois, D. et H. Prade: *Epistemic entrenchment and possibilistic logic*. Artif. Intell., 50(2) :223–239, 1991.

[9] Dubois, D. et H. Prade: *Possibilistic logic - An overview*. Dans Siekmann, J. H. (rédacteur) : *Computational Logic*, tome 9 de *Handbook of the History of Logic*, pages 283–342. Elsevier, 2014.

[10] Gärdenfors, P.: *Belief revision and nonmonotonic logic : Two sides of the same coin ?* Dans *Proc. 9th Europ. Conf. on Artificial Intelligence (ECAI'90), Stockholm*, pages 768–773, 1990.

[11] Gärdenfors, P. et D. Makinson: *Revisions of knowledge systems using epistemic entrenchment*. Dans Vardi, M. Y. (rédacteur) : *Proc. 2nd Conf. on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove*, pages 83–95. Morgan Kaufmann, 1988.

[12] Katsuno, H. et A. Mendelzon: *Propositional knowledge base revision and minimal change*. Artificial Intelligence, 52(3) :263–294, 1991.

[13] Konieczny, S., J. M. Lagniez et P. Marquis: *Boosting distance-based revision using SAT encodings*. Dans *International Workshop on Logic, Rationality and Interaction*, pages 480–496. Springer, 2017.

[14] Lafage, C. et J. Lang: *Propositional distances and compact preference representation*. Eur. J. Oper. Res., 160(3) :741–761, 2005. https://doi.org/10.1016/j.ejor.2003.06.037.

[15] Lepage, Y.: *Analogy and formal languages*. Electr. Notes Theor. Comput. Sci., 53, 2001.

[16] Lieber, J.: *Application of the revision theory to adaptation in case-based reasoning : The conservative adaptation*. Dans Weber, R. et M. M. Richter (rédacteurs) : *Case-Based Reasoning Research and Development, 7th Int. Conf. on Case-Based Reasoning, ICCBR 2007, Belfast, Aug. 13-16, Proceedings*, tome 4626 de *LNCS*, pages 239–253. Springer, 2007.

[17] Lieber, J., E. Nauer et H. Prade: *Improving analogical extrapolation using case pair competence*. Dans *Case-Based Reasoning Research and Development, 27th International Conference (ICCBR-2019)*, Otzenhausen, France, septembre 2019. https://hal.inria.fr/hal-02370747.

[18] Lieber, J., E. Nauer, H. Prade et G. Richard: *Making the best of cases by approximation, interpolation and extrapolation*. Dans *ICCBR 2018 - 26th International Conference on Case-Based Reasoning*, Stockholm, Sweden, juillet 2018. Springer. https://hal.inria.fr/hal-01905058.

[19] Miclet, L. et H. Prade: *Handling analogical proportions in classical logic and fuzzy logics settings*. Dans *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, pages 638–650. Springer, LNCS 5590, 2009.

[20] Pirrelli, V. et F. Yvon: *Analogy in the lexicon : a probe into analogy-based machine learning of language*. Dans *Proc. 6th Int. Symp. on Human Communication*, Santiago de Cuba, 6 p.,1999.

[21] Prade, H. et G. Richard: *Cataloguing/analogizing : A nonmonotonic view*. Int. J. Intell. Syst., 26(12) :1176–1195, 2011.

[22] Prade, H. et G. Richard: *From Analogical Proportion to Logical Proportions*. Logica Universalis, 7(4) :441–505, 2013.

[23] Prade, H. et G. Richard: *Analogical proportions and analogical reasoning - An introduction*. Dans Aha, D. W. et J. Lieber (rédacteurs) : *Proc. 25th Int. Conf. on Case-Based Reasoning Research and Development (ICCBR'17), Trondheim, Norway, June 26-28*, tome 10339 de *LNAI*, pages 16–32. Springer, 2017.

[24] Prade, H. et G. Richard: *Analogical proportions : From equality to inequality*. Int. J. of Approximate Reasoning, 101 :234 – 254, 2018.

[25] Richter, M. M. et R. O. Weber: *Case-based reasoning, a textbook*. Springer, 2013.

[26] Riesbeck, C. K. et R. C. Schank: *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1989. Available on line.

[27] Smyth, B. et M. T. Keane: *Using adaptation knowledge to retrieve and adapt design cases*. Knowledge-Based Systems, 9(2) :127–135, 1996.

Actes JIAF 2021

Reaching Individually Stable Coalition Structures

Felix Brandt[1] **Martin Bullinger**[1] **Anaëlle Wilczynski**[2]

[1] Institut für Informatik, Technische Universität München
[2] MICS, CentraleSupélec, Université Paris-Saclay
brandtf@in.tum.de bullinge@in.tum.de anaelle.wilczynski@centralesupelec.fr

Résumé

L'étude formelle des processus de formation de coalitions dans les systèmes multi-agents passe commmunément par le concept économique des jeux hédoniques. La recherche autour des jeux hédoniques s'est principalement focalisée sur l'existence—et la complexité algorithmique du problème d'existence—de structures de coalitions satisfaisant divers critères de stabilité. Le processus réel de formation de coalitions basé sur des comportements individuels a été relativement peu exploré. Dans cet article, on se propose d'étudier la convergence d'une dynamique de jeu simple aboutissant à des partitions stables, dans différentes classes de jeux hédoniques. La dynamique considérée est basée sur la stabilité dite individuelle : un agent va rejoindre une autre coalition seulement si cela améliore sa satisfaction sans détériorer la satisfaction d'aucun membre de la nouvelle coalition. On identifie les conditions permettant la convergence de cette dynamique dans les classes de jeux hédoniques étudiées, en fournissant des preuves de convergence mais aussi des contre-exemples non-triviaux à la convergence de la dynamique, voire même à l'existence de partitions stables. On analyse également la complexité computationnelle de problèmes liés à la dynamique de formation de coalitions. En particulier, dans cet article, on répond à des problèmes ouverts suggérés par Bogomolnaia and Jackson [7], Brandl et al. [8] et Boehmer and Elkind [6].

Abstract

The formal study of coalition formation in multiagent systems is typically realized using so-called hedonic games, which originate from economic theory. The main focus of this branch of research has been on the existence and the computational complexity of deciding the existence of coalition structures that satisfy various stability criteria. The actual process of forming coalitions based on individual behavior has received little attention. In this paper, we study the convergence of simple dynamics leading to stable partitions in a variety of classes of hedonic games, including anonymous, dichotomous, fractional, and hedonic diversity games. The dynamics we consider is based on individual stability: an agent will join another coalition if she is better off and no member of the welcoming coalition is worse off. We identify conditions for convergence, provide elaborate counterexamples of existence of individually stable partitions, and study the computational complexity of problems related to the coalition formation dynamics. In particular, we settle open problems suggested by Bogomolnaia and Jackson [7], Brandl et al. [8], and Boehmer and Elkind [6].

1 Introduction

Coalitions and coalition formation are central concerns in the study of multiagent systems as well as cooperative game theory. Typical real-world examples include individuals joining clubs or societies such as orchestras, choirs, or sport teams, countries organizing themselves in international bodies like the European Union (EU) or the North Atlantic Treaty Organization (NATO), students living together in shared flats, or employees forming unions. The formal study of coalition formation is often realized using so-called hedonic games, which originate from economic theory and focus on coalition structures (henceforth partitions) that satisfy various stability criteria based on the agents' preferences over coalitions. A partition is defined to be stable if single agents or groups of agents cannot gain by deviating from the current partition by means of leaving their current coalition and joining another coalition or forming a new one. Which kinds of deviations are permitted depends on the underlying notion of stability. Two important and well-studied questions in this context concern the existence of stable partitions in restricted classes of hedonic games and the computational complexity of finding a stable partition. However, stability is only concerned with the end-state of the coalition formation process and ignores how these desirable partitions can actually be reached. Essentially, an underlying assumption in most of the existing work is that there is a central authority that receives the preferences of all agents, computes a stable partition as an end-state, and has the means to enforce this partition on the

agents. By contrast, our work focuses on simple dynamics, where starting with some partition (e.g., the partition of singletons), agents deliberately decide to join and leave coalitions based on their individual preferences. We study the convergence of such a process and the stable partitions that can arise from it. For example, in some cases the only partition satisfying a certain stability criterion is the grand coalition consisting of all agents, while the dynamics based on the agents' individual decisions can never reach this partition and is doomed to cycle.

The dynamics we consider is based on *individual stability*, a natural notion of stability going back to Drèze and Greenberg [13]: an agent joins another coalition if she is better off and no member of the welcoming coalition is worse off. Individual stability is suitable to model the situations mentioned above. By Article 49 of the Treaty on EU, admitting new members to the EU requires the unanimous approval of the current members. Similarly, by Article 10 of the founding treaty, unanimous agreement of all parties is necessary to become a member of the NATO. Also, for joining a choir or orchestra it is often necessary to audition successfully, and joining a shared flat requires the consent of all current residents. This distinguishes individual stability from Nash stability, which ignores the consent of members of the welcoming coalition.

The analysis of coalition formation processes provides more insight in the natural behavior of agents and the conditions that are required to guarantee that desirable social outcomes can be reached without a central authority. Similar dynamic processes have been studied in the special domain of matching, which only allows coalitions of size 2 [e.g., 1, 9, 17]. More recently, the dynamics of coalition formation have also come under scrutiny in the context of hedonic games [5, 11, 15]. While coalition formation dynamics are an object of study worthy for itself, they can also be used as a means to design algorithms that compute stable outcomes, and have been implicitly used for this purpose before. For example, for finding an individually stable partition in hedonic diversity games, the algorithm by Boehmer and Elkind [6] predefines a promising partition and then runs the dynamics from there. Similarly, for finding an individually stable partition on games with ordered characteristics, a generalization of anonymous hedonic games, the algorithm by Bogomolnaia and Jackson [7] runs the dynamics using a specific sequence of deviations starting from the singleton partition.

In many cases, the convergence of the dynamics follows from the existence of potential functions, whose local optima form individually stable states. Generalizing a result by Bogomolnaia and Jackson [7], Suksompong [18] has shown via a potential function argument that an individually stable—and even a Nash stable—partition always exists in subset-neutral hedonic games, a generalization of symmetric additively-separable hedonic games. Using the same potential function, it can straightforwardly be shown that the dynamics converge. [1] Another example are hedonic games with the common ranking property, a class of hedonic games where preferences are induced by a common global order [14]. The dynamics associated with core-stable deviations is known to converge to a core-stable partition that is also Pareto-optimal, thanks to a potential function argument [12]. The same potential function implies convergence of the dynamics based on individual stability.

In this paper, we study the coalition formation dynamics based on individual stability for a variety of classes of hedonic games, including anonymous hedonic games (AHGs), hedonic diversity games (HDGs), fractional hedonic games (FHGs), and dichotomous hedonic games (DHGs). Whether we obtain positive or negative results often depends on the initial partition and on restrictions imposed on the agents' preferences. Computational questions related to the dynamics are investigated in two ways: the existence of a *path to stability*, that is the existence of a sequence of deviations that leads to a stable state, and the *guarantee of convergence* where every sequence of deviations should lead to a stable state. The former gives an optimistic view on the behavior of the dynamics and may be used to motivate the choice of reachable stable partitions (we can exclude "artificial" stable partitions that may never naturally form). If such a sequence can be computed efficiently, it enables a central authority to coordinate the deviations towards a stable partition. However, since this approach does not give any guarantee on the outcome of the dynamics, we also study the latter, more pessimistic, problem. Our main results are as follows.

— In AHGs, the dynamics converges for (naturally) single-peaked strict preferences. We provide a 15-agent example showing the non-existence of individually stable partitions in general AHGs. The previous known counterexample [7] requires 63 agents and the existence of smaller examples was an acknowledged open problem [see 4, 6].

— In HDGs, we show that the dynamics converges for (naturally) singled-peaked strict preferences when starting from the singleton partition. In contrast to empirical evidence reported by Boehmer and Elkind [6], these preference restrictions are not sufficient to guarantee convergence from an arbitrary initial partition.

— In FHGs, the dynamics converges for simple symmetric preferences when starting from the singleton partition or when preferences form an acyclic digraph. We show that individually stable partitions need not exist in general symmetric FHGs, which was left as an open problem by Brandl et al. [8].

1. By inclusion, convergence also holds for symmetric additively-separable hedonic games. Symmetry is essential for this result to hold since an individually stable partition may not exist in additively-separable hedonic games, even under additional restrictions [7].

— For each of these classes, including DHGs, we show that deciding whether there is a sequence of deviations leading to an individually stable partition is NP-hard while deciding whether all sequences of deviations lead to an individually stable partition is co-NP-hard. Some of the results hold under preference restrictions and even when starting from the singleton partition.

Due to space restrictions, we omit some of the proofs or provide only proof sketches.

2 Preliminaries

Let $N = [n] = \{1, \dots, n\}$ be a set of n agents. The goal of a coalition formation problem is to partition the agents into different disjoint coalitions according to their preferences. A solution is then a partition $\pi : N \to 2^N$ such that $i \in \pi(i)$ for every agent $i \in N$ and either $\pi(i) = \pi(j)$ or $\pi(i) \cap \pi(j) = \emptyset$ holds for every agents i and j, where $\pi(i)$ denotes the coalition to which agent i belongs. Two prominent partitions are the *singleton partition* π given by $\pi(i) = \{i\}$ for every agent $i \in N$, and the *grand coalition* π given by $\pi = \{N\}$.

Since we focus on dynamics of deviations, we assume that there exists an initial partition π_0, which could be a natural initial state (such as the singleton partition) or the outcome of a previous coalition formation process.

2.1 Classes of Hedonic Games

In a hedonic game, the agents only express preferences over the coalitions to which they belong, i.e., there are no externalities. Let $\mathcal{N}_i$ denote all possible coalitions containing agent i, i.e., $\mathcal{N}_i = \{C \subseteq N : i \in C\}$. A hedonic game is defined by a tuple $(N, (\succsim_i)_{i \in N})$ where $\succsim_i$ is a weak order over $\mathcal{N}_i$ which represents the preferences of agent i. Since $|\mathcal{N}_i| = 2^{n-1}$, the preferences are rarely given explicitly, but rather in some concise representation. These representations give rise to several classes of hedonic games:

— *Anonymous hedonic games (AHGs)* [7]: The agents only care about the size of the coalition they belong to, i.e., for each agent i, there exists a weak order $\succsim_i$ over integers in $[n]$ such that $\pi(i) \succsim_i \pi'(i)$ iff $|\pi(i)| \succsim_i |\pi'(i)|$.

— *Hedonic diversity games (HDGs)* [10]: The agents are divided into two different types, red and blue agents, represented by the subsets R and B, respectively, such that $N = R \cup B$ and $R \cap B = \emptyset$. Each agent only cares about the proportion of red agents present in her own coalition, i.e., for each agent $i \in N$, there exists a weak order $\succsim_i$ over $\{\frac{p}{q} : p \in [|R|] \cup \{0\}, q \in [n]\}$ such that $\pi(i) \succsim_i \pi'(i)$ iff $\frac{|R \cap \pi(i)|}{|\pi(i)|} \succsim_i \frac{|R \cap \pi'(i)|}{|\pi'(i)|}$.

— *Fractional Hedonic Games (FHGs)* [2]: The agents evaluate a coalition according to how much they like each of its members on average, i.e., for each agent i, there exists a utility function $v_i : N \to \mathbb{R}$ where $v_i(i) = 0$ such that $\pi(i) \succsim_i \pi'(i)$ iff $\frac{\sum_{j \in \pi(i)} v_i(j)}{|\pi(i)|} \geq \frac{\sum_{j \in \pi'(i)} v_i(j)}{|\pi'(i)|}$. An FHG can be represented by a weighted complete directed graph $G = (N, E)$ where the weight of arc (i, j) is equal to $v_i(j)$. An FHG is *symmetric* if $v_i(j) = v_j(i)$ for every pair of agents i and j, i.e., it can be represented by a weighted complete undirected graph with weights $v(i, j)$ on each edge $\{i, j\}$. An FHG is simple if $v_i : N \to \{0, 1\}$ for every agent i, i.e., it can be represented by an unweighted directed graph where $(i, j) \in E$ iff $v_i(j) = 1$. We say that a simple FHG is *asymmetric* if, for every pair of agents i and j, $v_i(j) = 1$ implies $v_j(i) = 0$, i.e., it can be represented by an asymmetric directed graph.

— *Dichotomous hedonic games (DHGs)*: The agents only approve or disapprove coalitions, i.e., for each agent i there exists a utility function $v_i : \mathcal{N}_i \to \{0, 1\}$ such that $\pi(i) \succsim_i \pi'(i)$ iff $v_i(\pi(i)) \geq v_i(\pi'(i))$. When the preferences are represented by a propositional formula, such games are called *Boolean hedonic games* [3].

An anonymous game (resp., hedonic diversity game) is *generally single-peaked* if there exists a linear order $>$ over integers in $[n]$ (resp., over ratios in $\{\frac{p}{q} : p \in [|R|] \cup \{0\}, q \in [n]\}$) such that for each agent $i \in N$ and each triple of integers $x, y, z \in [n]$ (resp., $x, y, z \in \{\frac{p}{q} : p \in |R| \cup \{0\}, q \in [n]\}$) with $x > y > z$ or $z > y > x$, $x \succsim_i y$ implies $y \succsim_i z$. The obvious linear order $>$ that comes to mind is, of course, the natural order over integers (resp., over rational numbers). We refer to such games as *naturally single-peaked*. Clearly, a naturally single-peaked preference profile is generally single-peaked but the converse is not true.

2.2 Dynamics of Individually Stable Deviations

Starting from the initial partition, agents can leave and join coalitions in order to improve their well-being. We focus on unilateral deviations, which occur when a single agent decides to move from one coalition to another. A *unilateral deviation* performed by agent i transforms a partition π into a partition π' where $\pi(i) \neq \pi'(i)$ and, for all agents $j \neq i$,

$$\pi'(j) = \begin{cases} \pi(j) \setminus \{i\} & \text{if } j \in \pi(i) \\ \pi(j) \cup \{i\} & \text{if } j \in \pi'(i) \\ \pi(j) & \text{otherwise} \end{cases}.$$

Since agents are assumed to be rational, agents only engage in a unilateral deviation if it makes them better off, i.e., $\pi'(i) \succ_i \pi(i)$. Any partition in which no such deviation is possible is called *Nash stable (NS)*.

This type of deviation can be refined by additionally requiring that no agent in the welcoming coalition is worse off when agent i joins. A partition in which no such deviation is possible is called *individually stable (IS)*. Formally, a unilateral deviation performed by agent i who moves from coalition $\pi(i)$ to $\pi'(i)$ is an IS-deviation if $\pi'(i) \succ_i \pi(i)$

and $\pi'(i) \succsim_j \pi(j)$ for all agents $j \in \pi'(i)$. Clearly, an NS partition is also IS. [2] In this article, we focus on dynamics based on IS-deviations. By definition, all terminal states of the dynamics have to be IS partitions.

We are mainly concerned with whether sequences of IS-deviations can reach or always reach an IS partition. If there exists a sequence of IS-deviations leading to an IS partition, i.e., a path to stability, then although the agents perform myopic deviations, they can optimistically reach (or can be guided towards) a stable partition. The corresponding decision problem is described as follows.

∃-IS-SEQUENCE-[HG]	
Input:	Instance of a particular class of hedonic games [HG], initial partition π_0
Question:	Does there exist a sequence of IS-deviations starting from π_0 leading to an IS partition?

In order to provide some guarantee, we also examine whether *all* sequences of IS-deviations terminate. Whenever this is the case, we say that the dynamics *converges*. The corresponding decision problem is described below.

∀-IS-SEQUENCE-[HG]	
Input:	Instance of a particular class of hedonic games [HG], initial partition π_0
Question:	Does every sequence of IS-deviations starting from π_0 reach an IS partition?

We mainly investigate this problem via the study of its complement: given a hedonic game and an initial partition, does there exist a sequence of IS-deviations that cycles?

A common idea behind hardness reductions concerning these two problems is to exploit the existence of instances without an IS partition or instances which allow for cycling starting from a certain partition. These can be used to create prohibitive subconfigurations in reduced instances.

3 Anonymous Hedonic Games (AHGs)

Bogomolnaia and Jackson [7] showed that IS partitions always exist in AHGs under naturally single-peaked preferences, and proved that this does not hold under general preferences, by means of a 63-agent counterexample. Here, we provide a counterexample that only requires 15 agents and additionally satisfies general single-peakedness.

Proposition 3.1. *There may not exist an IS partition in AHGs even when $n = 15$ and the agents have strict and generally single-peaked preferences.*

Sketch of proof. Let us consider an AHG with 15 agents with the following (incompletely specified) preferences ([...] is an arbitrary order over the remaining sizes).

$$\begin{aligned}
1 &: 2 \succ 3 \succ 13 \succ 12 \succ 1 \succ [\dots] \\
2 &: 13 \succ 3 \succ 2 \succ 1 \succ 12 \succ [\dots] \\
3,4 &: 3 \succ 2 \succ 1 \succ [\dots] \\
5,\dots,15 &: 13 \succ 12 \succ 15 \succ 14 \succ 11 \succ 10 \succ \dots \succ 1
\end{aligned}$$

They can be completed to be generally single-peaked w.r.t. axis $1 > 2 > 3 > 13 > 12 > 15 > 14 > 11 > 10 > \dots > 4$.

One can prove that in an IS partition,

(i) agents 3 and 4 are in a coalition of size at most 3;
(ii) agents 5 to 15 are in the same coalition;
(iii) agents 3 and 4 are in the same coalition;
(iv) agents 1 and 2 cannot be both alone.

Therefore, agents 3 and 4 must be together, as well as agents 5 to 15, but not in the same coalition. It remains to identify the coalitions of agents 1 and 2. By *(i)*, they cannot be both with agents 3 and 4. If one agent among them is alone and the other one with agents 5 to 15, then the alone agent can deviate to join them, a contradiction. The remaining possible partitions are present in the cycle of IS-deviations below (the deviating agent is written on top of the arrows).

$$\begin{array}{ccccc}
\{\{1\},\{2,3,4\},\{5,\dots,15\}\} & \xrightarrow{1} & \{\{2,3,4\},\{1,5,\dots,15\}\} & \xrightarrow{2} & \{\{3,4\},\{1,2,5,\dots,15\}\} \\
{\scriptstyle 2}\uparrow & & & & \downarrow{\scriptstyle 1} \\
\{\{1,2\},\{3,4\},\{5,\dots,15\}\} & \xleftarrow{1} & \{\{2\},\{1,3,4\},\{5,\dots,15\}\} & \xleftarrow{2} & \{\{1,3,4\},\{2,5,\dots,15\}\}
\end{array}$$

Hence, there is no IS partition in this instance. □

We conjecture that the counterexample given in Proposition 3.1 is minimal and that an IS partition always exists when $n < 15$. However, even when $n < 15$ and IS partitions do exist, there may still be cycles in the dynamics.

Proposition 3.2. *The dynamics of IS-deviations may cycle in AHGs even when starting from the singleton partition or grand coalition, for strict generally single-peaked preferences, and for $n < 15$.*

Proof. Let us consider an AHG with 7 agents with the following (incompletely specified) preferences ([...] is an arbitrary order over the remaining sizes).

$$\begin{aligned}
1 &: 2 \succ 3 \succ 5 \succ 4 \succ 1 \succ [\dots] \\
2 &: 5 \succ 3 \succ 2 \succ 1 \succ 4 \succ [\dots] \\
3,4 &: 3 \succ 2 \succ 1 \succ [\dots] \\
5,6,7 &: 5 \succ 4 \succ 3 \succ 2 \succ 1 \succ [\dots]
\end{aligned}$$

They can be completed to be generally single-peaked w.r.t. axis $1 > 2 > 3 > 5 > 4 > 6 > 7$. We represent below a cycle in IS-deviations that can be reached from the singleton partition or the grand coalition.

$$\begin{array}{ccccc}
\{1,2\},\{3,4\},\{5,6,7\} & \xrightarrow{2} & \{1\},\{2,3,4\},\{5,6,7\} & \xrightarrow{1} & \{2,3,4\},\{1,5,6,7\} \\
{\scriptstyle 1}\uparrow & & & & \downarrow{\scriptstyle 2} \\
\{2\},\{1,3,4\},\{5,6,7\} & \xleftarrow{2} & \{1,3,4\},\{2,5,6,7\} & \xleftarrow{1} & \{3,4\},\{1,2,5,6,7\}
\end{array}$$

□

Note that $\{\{1\},\{3,5,6\},\{2,4,7\}\}$ is an IS partition in the example of the previous proposition.

2. It is possible to weaken the notion of individual stability even further by also requiring that no member of the *former* coalition of agent i is worse off. The resulting stability notion is called contractual individual stability and guarantees convergence of our dynamics.

We know that it is NP-complete to recognize instances for which an IS partition exists in AHGs, even for strict preferences [4]. We prove that both checking the existence of a sequence of IS-deviations ending in an IS partition and checking convergence are hard.

Theorem 3.3. ∃-IS-Sequence-AHG *is* NP*-hard and* ∀-IS-Sequence-AHG *is* co-NP*-hard, even for strict preferences.*

However, this hardness result does not hold under strict naturally single-peaked preferences, since we show in the next proposition that *every* sequence of IS-deviations is finite under such a restriction.

Proposition 3.4. *The dynamics of IS-deviations always converges to an IS partition in AHGs for strict naturally single-peaked preferences.*

Proof. Assume for contradiction that there exists a cycle of IS-deviations. The key idea is to construct an infinite sequence of agents $(a_k)_{k\geq 1}$ that perform deviations from coalitions $(C_k)_{k\geq 1}$, which are strictly increasing in size. Let a_1 be an agent that deviates within this cycle towards a larger coalition by an IS-deviation. This transforms, say, partition π_1 into partition π_1^1. Set $C_1 = \pi_1(a_1)$ and $\hat{C}_1 = \pi_1^1(a_1)$. We will now observe how the coalition $\hat{C}_1$ evolves during the cycle. After possibly some agents outside $\hat{C}_1$ joined it or some left it, some agent b originally in $\hat{C}_1$ must deviate from the coalition evolved from $\hat{C}_1$. Otherwise, we cannot reach partition π_1 again in the cycle. If $b \neq a_1$, we assume that the deviation transforms partition π_2 into partition π_2^1 and we set $a_2 = b$, $C_2 = \pi_2(b)$, and $\hat{C}_2 = \pi_2^1(b)$. Note that $|\hat{C}_2| > |C_2| \geq |\hat{C}_1|$, by single-peakedness and the fact that $|\hat{C}_2| \succ_b |C_2| \succ_b |C_2| - 1 \succ_b \cdots \succ_b |\hat{C}_1| \succ_b |\hat{C}_1| - 1$ (where all preferences but the first follow from the assumption of strictness when some other agent joined the coalition of b). In particular, $|C_2| > |C_1|$.

If $b = a_1$, assume that the deviation transforms partition π_1^2 into π_1^3, where possibly $\pi_1^2 = \pi_1^1$. We update $\hat{C}_1 = \pi_1^3(a_1)$. Note that still $|\hat{C}_1| > |C_1|$ by single-peakedness, because the original deviation of a_1 performed in partition π_1 was towards a larger coalition and $|\pi_1^2(a_1)| \succeq_{a_1} |\pi_1^1(a_1)|$ (equality if the partitions are the same). We consider again the next deviation from $\hat{C}_1$ until it is from an agent $b \neq a_1$, in which case we proceed as in the first case. This must eventually happen, because every time the deviation is again performed by agent a_1 she gets closer to her peak. We proceed in the same manner. In step k, we are given a coalition $\hat{C}_k$ with $|\hat{C}_k| > |C_k|$ which was just joined by an agent. When the next agent originally in $\hat{C}_k$ deviates from the coalition evolved from $\hat{C}_k$, it is either an agent different from a_k and we call it a_{k+1}, and find coalitions C_{k+1} and $\hat{C}_{k+1}$ with $|\hat{C}_{k+1}| > |C_{k+1}| \geq |\hat{C}_k|$; or this agent is a_k, she moves towards an updated coalition $\hat{C}_k$ which maintains $|\hat{C}_k| > |C_k|$. We have thus constructed an infinite sequence of coalitions $(C_k)_{k\geq 1}$ occurring in the cycle with $|C_{k+1}| > |C_k|$ for all $k \geq 1$, a contradiction. □

An open question is whether convergence still holds under naturally single-peaked preferences with indifference.

Additionally, convergence is guaranteed under other constrained anonymous games, called *neutral anonymous games*, which are subset-neutral, as defined by Suksompong [18], thanks to the use of the potential function by Suksompong.

4 Hedonic Diversity Games (HDGs)

Hedonic diversity games take into account more information about the identity of the agents, changing the focus from coalition sizes to proportions of given types of agents. We obtain more positive results regarding the existence of IS partitions. Indeed, there always exists an IS partition in a hedonic diversity game, even with preferences that are not single-peaked [6]. However, we prove that there may exist cycles in IS-deviations, even under some strong restrictions. This stands in contrast to empirical evidence for convergence based on extensive computer simulations by Boehmer and Elkind [6].

Proposition 4.1. *The dynamics of IS-deviations may cycle in HDGs even*

1. *when preferences are strict and naturally single-peaked,*
2. *when preferences are strict and the initial partition is the singleton partition or the grand coalition, or*
3. *when preferences are naturally single-peaked and the initial partition is the singleton partition.*

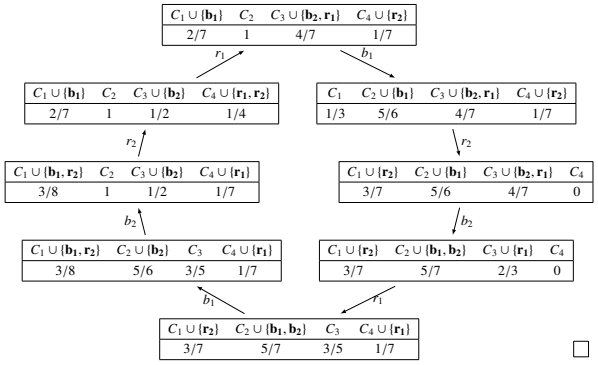
Sketch of proof. We only provide the counterexample for an HDG with strict and naturally single-peaked preferences (restriction 1). Let us consider an HDG with 26 agents: 12 red agents and 14 blue agents. There are four deviating agents: red agents r_1 and r_2 and blue agents b_1 and b_2, and four fixed coalitions C_1, C_2, C_3 and C_4 such that:

— C_1 contains 2 red agents and 4 blue agents;
— C_2 contains 5 red agents;
— C_3 contains 3 red agents and 2 blue agents;
— C_4 contains 6 blue agents.

The relevant part of the preferences is given below.

$$b_1: \ \tfrac{3}{8} \succ \tfrac{5}{7} \succ \tfrac{5}{6} \succ \tfrac{2}{7} \qquad C_1: \ \tfrac{3}{8} \succ \tfrac{3}{7} \succ \tfrac{1}{3}$$
$$b_2: \ \tfrac{5}{7} \succ \tfrac{4}{7} \succ \tfrac{1}{2} \succ \tfrac{5}{6} \qquad C_2: \ \tfrac{5}{7} \succ \tfrac{5}{6} \succ 1$$
$$r_1: \ \tfrac{4}{7} \succ \tfrac{1}{4} \succ \tfrac{1}{7} \succ \tfrac{2}{3} \qquad C_3: \ \tfrac{4}{7} \succ \tfrac{1}{2} \succ \tfrac{3}{5}$$
$$r_2: \ \tfrac{1}{4} \succ \tfrac{3}{8} \succ \tfrac{3}{7} \succ \tfrac{1}{7} \qquad C_4: \ \tfrac{1}{4} \succ \tfrac{1}{7} \succ 0$$

Consider the following sequence of IS-deviations that describe a cycle in the dynamics. The four deviating agents of the cycle b_1, b_2, r_1 and r_2 are marked in bold and the specific deviating agent between two states is indicated next to the arrows.



This example does not show the impossibility to reach an IS partition since the IS partition $\{C_1 \cup \{b_1, r_2\}, C_2, C_3 \cup \{r_1, b_2\}, C_4\}$ can be reached via IS-deviations from some partitions in the cycle. Thus, starting in these partitions, a path to stability may still exist. Nevertheless, it may be possible that every sequence of IS-deviations cycles, even for strict or naturally single-peaked preferences (with indifference), as the next proposition shows. An interesting open question is whether strict and single-peaked preferences allow for the existence of a path to stability.

Proposition 4.2. *The dynamics of IS-deviations may never reach an IS partition in HDGs, whatever the chosen path of deviations, even for (1) strict preferences or (2) naturally single-peaked preferences with indifference.*

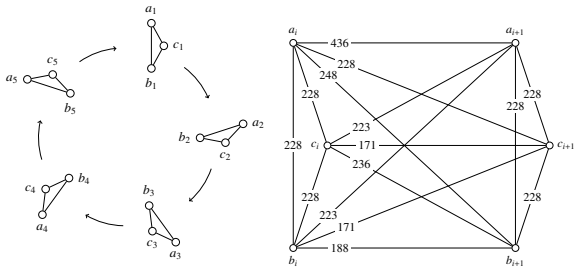
However, convergence is guaranteed by combining all previous restrictions, as stated in the next proposition.

Proposition 4.3. *The dynamics of IS-deviations starting from the singleton partition always converges to an IS partition in HDGs for strict naturally single-peaked preferences.*

Sketch of proof. One can prove that at any step of the dynamics, a coalition is necessarily of the form $\{r_1, b_1, \ldots, b_k\}$ or $\{b_1, r_1, \ldots, r_{k'}\}$ or $\{b_1\}$ or $\{r_1\}$ where $r_i \in R$ and $b_j \in B$ for every $i \in [k']$, $j \in [k]$ and $k \leq |B|$ and $k' \leq |R|$. Therefore, the ratio of a coalition can only be equal to $\frac{1}{k+1}$, $\frac{k'}{k'+1}$, 0 or 1. Let us define as $\rho(C)$ the modified ratio of a valid coalition C formed by the dynamics where $\rho(C) = \begin{cases} \frac{|R \cap C|}{|C|} & \text{if } C = \{b_1, r_1, \ldots, r_{k'}\} \text{ for } k' \geq 1 \\ 1 - \frac{|R \cap C|}{|C|} & \text{if } C = \{r_1, b_1, \ldots, b_k\} \text{ for } k \geq 2 \\ 0 & \text{otherwise, i.e., } C = \{r_1\} \text{ or } C = \{b_1\} \end{cases}$.

For each partition π in a sequence of IS-deviations, consider the vector composed of the ratios $\rho(C)$ for all coalitions C in π. One can prove that for each sequence of IS-deviations, either this vector strictly increases lexicographically at each deviation or there is an equivalent sequence of deviations, leading to the same partition, where it does. □

Under strict preferences, checking the existence of a path to stability and convergence are hard.



(a) Five triangles are ordered in a cycle such that there is a tendency of agents in N_i to deviate to coalitions in N_{i+1}.

(b) The transition weights between the triangles allow for infinite loops of deviations.

Figure 1 – Description of the graph associated with the constructed symmetric FHG without an IS partition.

Theorem 4.4. ∃-IS-Sequence-HDG *is* NP*-hard and* ∀-IS-Sequence-HDG *is* co-NP*-hard, even for strict preferences.*

5 Fractional Hedonic Games (FHGs)

Next, we study fractional hedonic games, which are closely related to hedonic diversity games, but instead of agent types, utilities rely on a cardinal valuation function of the other agents. The first part of the section deals with symmetric games, the second part with simple games.

An open problem for symmetric FHGs was whether they always admit an IS partition [8]. Here, we provide a counterexample using 15 agents.

Theorem 5.1. *There exists a symmetric FHG without an IS partition.*

Sketch of proof. Define the sets of agents $N_i = \{a_i, b_i, c_i\}$ for $i \in \{1, \ldots, 5\}$ and consider the FHG on the agent set $N = \bigcup_{i=1}^{5} N_i$ where symmetric weights are given as in Figure 1b. All weights not specified in this figure are set to -2251. The FHG consists of five triangles that form a cycle. Its structure is illustrated in Figure 1a.

There is an infinite cycle of deviations starting with the partition $\{N_5 \cup N_1, N_2, N_3, N_4\}$. First, a_1 deviates by joining N_2. Then, b_1 joins this new coalition, then c_1. Afterwards, we are in an isomorphic state as in the initial partition. One can show that this instance contains no IS partition. □

Using this counterexample, the methods of Brandl et al. [8], which come from hardness constructions of Sung and Dimitrov [19], can be adapted to show the NP-hardness to decide the existence of IS partitions in symmetric FHGs.

Corollary 5.2. *Deciding whether there exists an individually stable partition in symmetric FHGs is* NP*-hard.*

If we consider symmetric, non-negative utilities, the grand coalition forms an NS, and therefore IS, partition.

However, deciding about the convergence of the IS dynamics starting with the singleton partition is NP-hard. The reduction is similar to the one in Corollary 5.2 but avoids negative weights (needed in the Theorem 5.1's counterexample) because, due to the symmetry of the weights, in a dynamics starting with the singleton partition, all coalitions that can be obtained in the process must have strictly positive mutual utility for all pairs of agents in the coalition.

Theorem 5.3. ∃-IS-Sequence-FHG *is* NP*-hard and* ∀-IS-Sequence-FHG *is* co-NP*-hard, even in symmetric FHGs with non-negative weights. The former is even true if the initial partition is the singleton partition.*

From now on, we consider simple FHGs. We start with the additional assumption of symmetry.

Proposition 5.4. *The dynamics of IS-deviations starting from the singleton partition always converges to an IS partition in simple symmetric FHGs in at most $O(n^2)$ steps. The dynamics may take $\Omega(n\sqrt{n})$ steps.*

Sketch of proof. We only prove the upper bound. Note that all coalitions formed through the deviation dynamics are cliques. Hence, every deviation step will increase the total number of edges in all coalitions. More precisely, the dynamics will increase the potential $\Lambda(\pi) = \sum_{C\in\pi} |C|(|C|-1)/2$ in every step by at least 1. Since the total number of edges is bounded by $n(n-1)/2$, this proves the upper bound. □

There is a simple way to converge in a linear number of steps starting with the singleton partition by forming largest cliques and removing them from consideration.

If we allow for asymmetries, the dynamics is not guaranteed to converge anymore. For instance, the IS dynamics on an FHG induced by a directed triangle will not converge for any initial partition except for the grand coalition. We can, however, characterize convergence on asymmetric FHGs. Tractability highly depends on the initial partition. First, we assume that we start from the singleton partition.

Proposition 5.5. *The dynamics of IS-deviations starting from the singleton partition converges in asymmetric FHGs if and only if the underlying graph is acyclic. Moreover, under acyclicity, it converges in $O(n^4)$ steps.*

Sketch of proof. Let $G = (N, A)$ be an asymmetric graph. If the graph contains a cycle, it is easy to find a cycle in IS-deviations. Assume that the graph is acyclic. We observe that in every step of the dynamics, the subgraphs induced by coalitions are always transitive tournaments, i.e., linear orders (on their vertices). Indeed, by induction, in a deviation, the coalition that is left still induces a transitive tournament and the new coalition induced a transitive tournament before the deviation. Hence, every agent except one has at least one outgoing edge and will only accept the new agent if she likes her. Since the deviating agent must have non-negative utility after the deviation, she needs to approve the single agent without outgoing edge. Hence, the newly formed coalition induces a transitive tournament.

We define two potentials based on the agents that receive 0 utility in a partition, and on the coalition sizes. The first potential is monotonically decreasing and bounded. The second potential is strictly increasing whenever the first potential is not strictly decreasing, and bounded. Hence, we establish convergence of the dynamics.

First, fix a topological order of the agents, i.e., a bijection $\sigma : N \to [n]$ such that for all $(v, w) \in A$, $\sigma(v) < \sigma(w)$. For a given partition π of the agents, we define the vector $v^\sigma(\pi)$ of length $|\pi|$ that sorts the numbers $\max_{i\in C} \sigma(i)$ for $C \in \pi$ in decreasing order. This is exactly the unique agent in every coalition receiving 0 utility. In addition, we define the vector $w(\pi)$ of length $|\pi|$ that sorts the coalition sizes in increasing order. Note that this vector does not depend on the underlying topological order. We write $v >_{lex} w$ if a vector v is lexicographically greater than a vector w.

The key insight is that, for π' formed from π by an IS-deviation, $v^\sigma(\pi') <_{lex} v^\sigma(\pi)$, or $v^\sigma(\pi') =_{lex} v^\sigma(\pi)$ and $w(\pi') >_{lex} w(\pi)$. For a proof, assume that π' is formed from π by an IS-deviation of agent i. Note that $\max_{j\in\pi'(i)} \sigma(j) = \max_{j\in\pi'(i)\setminus\{i\}} \sigma(j)$. We distinguish two cases. Either $i = \arg\max_{j\in\pi(i)} \sigma(j)$ and it follows $v^\sigma(\pi') <_{lex} v^\sigma(\pi)$. Otherwise, $v_i(\pi(i)) \geq \frac{1}{|\pi(i)|}$, and because i is improving her utility, $\frac{1}{|\pi'(i)|} = v_i(\pi') > \frac{1}{|\pi(i)|}$. It follows that $|\pi(i)| > |\pi'(i)|$. Hence, $|\pi'(i)|-1 < \min\{|\pi'(i)|, |\pi(i)|\}$, and thus $w(\pi') >_{lex} w(\pi)$. One can show that in total there are at most n^4 deviations. □

The previous statement shows convergence of the dynamics for asymmetric, acyclic FHGs. Moreover, there is always a sequence converging after n steps, starting with the singleton partition. One can use a topological order of the agents and allow agents to deviate in decreasing topological order towards a best possible coalition.

There are two interesting further directions. One can weaken either the restriction on the initial partition or on asymmetry. If we allow for general initial partitions, we immediately obtain hardness results that apply in particular to the broader class of simple FHGs.

Theorem 5.6. ∃-IS-Sequence-FHG *is* NP*-hard and* ∀-IS-Sequence-FHG *is* co-NP*-hard, even in asymmetric FHGs.*

Sketch of proof. We only provide the proof for the NP-hardness part with a reduction from Exact-3-Cover. An instance of Exact-3-Cover consists of a tuple (R, S), where R is a ground set together with a set S of 3-element subsets of R. A 'yes'-instance is an instance so that there exists a subset $S' \subseteq S$ that partitions R.

Let (R, S) be an instance of Exact-3-Cover. We may assume that every $r \in R$ occurs in at least one set of S. Let $m_r := |\{s \in S : r \in s\}| - 1 \geq 0$, and

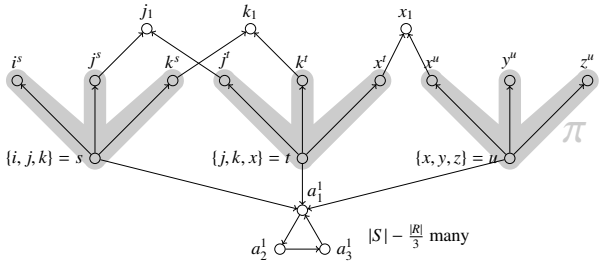


Figure 2 – Asymmetric directed graph corresponding to the instance $(\{i, j, k, x, y, z\}, \{\{i, j, k\}, \{j, k, x\}, \{x, y, z\})$ of the hardness construction. The non-singleton coalitions of the initial partition π are depicted in gray.

$l = |S| - |R|/3$. Define the simple, asymmetric FHG based on the directed graph $G = (V, A)$, where $V = \bigcup_{r\in R}\{r_1, \ldots, r_{m_r}\} \cup S \cup \bigcup_{s\in S}\{r^s : r \in s\} \cup \bigcup_{v=1}^{l}\{a_1^v, a_2^v, a_3^v\}$ and $A = \bigcup_{s\in S}(\{(s, r^s), (r^s, r_1), \ldots, (r^s, r_{m_r}) : r \in S\} \cup \{(s, a_1^1), \ldots, (s, a_1^l)\}) \cup \bigcup_{v=1}^{l}\{(a_1^v, a_2^v), (a_2^v, a_3^v), (a_3^v, a_1^v)\}$. Finally, define $\pi = \bigcup_{a\in V\setminus(S\cup\{r^s\,:\,s\in S, r\in s\})}\{\{a\}\} \cup \{\{s, i_1^s, j_1^s, k_1^s\} : \{i, j, k\} = s \in S\}$. The reduction is illustrated in Figure 2.

We claim that (R, S) is a 'yes'-instance if and only if the IS dynamics starting with π can converge.

Assume first that (R, S) is a 'yes'-instance and let $S' \subseteq S$ be a partition of R by sets in S. Let the agents in $\bigcup_{s\in S\setminus S'}\{r^s : r \in s\}$ join one by one the agents in $\bigcup_{r\in R}\{r_1, \ldots, r_{m_r}\}$ to form coalitions of size 2. Since S' covers every element of R exactly once, this step can be performed. Next, the agents $\{s \in S \setminus S'\}$ join the agents $\{a_1^1, \ldots, a_1^l\}$ in an arbitrary bijective way. Finally, agents a_2^v join agents a_3^v. One can check that the resulting partition is IS.

Conversely, assume that there exists a converging sequence of deviations starting with the partition π_0 and terminating in partition π^*. Then, one agent of every set $\{a_1^v, a_2^v, a_3^v\}$ must form a coalition with an agent outside of this set. The only possibility for this is if a_1^v is joined by an agent corresponding to a set in S. Every such agent can only perform a deviation if all the other agents in her initial coalition have deviated before. Define $S' = \{s \in S : \pi_0(s) = \pi^*(s)\}$. It can only happen that $3|S| - |R|$ agents corresponding to the sets in $S \setminus S'$ deviate if S' forms a partition of R. Hence, (R, S) is a 'yes'-instance. □

On the other hand, if we transition to simple FHGs while maintaining the initial partition, the problem of deciding whether a path to stability exists becomes hard.

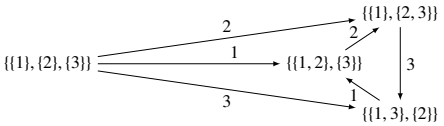
Theorem 5.7. ∃-IS-Sequence-FHG *is* NP*-hard even in simple FHGs when starting from the singleton partition.*

6 Dichotomous Hedonic Games (DHGs)

By taking into account the identity of other agents in the preferences, it can be more complicated to get positive results regarding individual stability (see, e.g., Theorem 5.1). However, by restricting the evaluation of coalitions to dichotomous preferences, the existence of an IS partition is guaranteed [16], as well as convergence of the dynamics of IS-deviations, when starting from the grand coalition [6]. Yet the convergence of the dynamics is not guaranteed for an arbitrary initial partition and it is possible that no sequence of IS-deviations may ever reach an IS partition.

Proposition 6.1. *There are DHGs in which the dynamics of IS-deviations may never reach an IS partition, whatever the chosen path of deviations, even when starting from the singleton partition.*

Proof. Let us consider a DHG with three agents where the only coalition approved by agent i is $\{i, i + 1\}$ for $i \in \{1, 2\}$ and $\{1, 3\}$ for agent 3. There is a unique IS partition which consists of the grand coalition $\{1, 2, 3\}$. We represent below all possible IS-deviations between all other possible partitions. An IS-deviation between two partitions is indicated by an arrow mentioning the name of the deviating agent.



One can check that the described deviations are IS-deviations. A cycle is necessarily reached when starting from a partition different from the unique IS partition, which can be reached only if it is the initial partition. □

Moreover, it is hard to decide on the existence of a sequence of IS-deviations ending in an IS partition, even when starting from the singleton partition, as well as checking convergence.

Theorem 6.2. ∃-IS-Sequence-DHG *is* NP*-hard even when starting from the singleton partition, and* ∀-IS-Sequence-DHG *is* co-NP*-hard.*

Sketch of proof. We only provide a sketch of proof for the co-NP-hardness part, where we prove the NP-hardness of the complement problem, which asks whether there exists a cycle in IS-deviations. Let us perform a reduction from the Satisfiability problem which asks the satisfiability of a CNF propositional formula φ given by a set of clauses $C_1, \ldots, C_m$ over variables $x_1, \ldots, x_p$. We construct an instance of a DHG with initial partition as follows.

For each clause C_j, for $1 \leq j \leq m$, we create two clause-agents k_j and k'_j. Let us denote by p_i^+ (resp., p_i^-) the number of positive (resp., negative) literals of variable x_i in formula φ. For each k^{th} occurrence of literal x_i (resp., $\overline{x_i}$) of variable x_i, we create a literal-agent y_i^k (resp., $\overline{y}_i^k$). The initial partition is given by $\pi^0 := \{\{x_i^1, \ldots, x_i^{p_i^+}, \overline{x}_i^1, \ldots, \overline{x}_i^{p_i^-}\}_{1\leq i\leq p}, \{k_j, k'_j\}_{1\leq j\leq m}\}$. The dichotomous preferences of the agents are summarized below.

— Each literal-agent y_i^k (resp., $\overline{y}_i^k$), for $1 \leq i \leq p$ and $1 \leq k \leq p_i^+$ (resp., $1 \leq k \leq p_i^-$), gives utility 1 to the coalitions where agent k'_j belongs, where k'_j refers to the clause C_j to which the k^{th} occurrence of literal x_i (resp., $\overline{x}_i$) belongs, and to all coalitions only composed of literal-agents associated with variable x_i where some literal-agents associated with $\overline{x_i}$ (resp., x_i) are missing. All the other coalitions are valued 0.
— Each clause-agent k_j, for $1 \leq j \leq m$, only gives utility 1 to the coalitions which contain clause-agent k'_{j+1} and one literal-agent associated with a literal belonging to clause C_{j+1} (where $m+1$ refers to 1).
— Each clause-agent k'_j, for $1 \leq j \leq m$, only gives utility 1 to the coalitions which contain agent k_j.

We claim that there exists a cycle of IS-deviations iff formula φ is satisfiable.

Suppose that formula φ is satisfiable by a truth assignment of the variables denoted by ϕ. For each clause C_j, for $1 \leq j \leq m$, we choose a literal-agent ℓ_j associated with a literal true in ϕ which belongs to clause C_j. By satisfiability of formula φ, there always exists such a literal-agent. Then, each literal-agent ℓ_j deviates, from her coalition of literal-agents associated with the same variable, to coalition $\{k_j, k'_j\}$. After all these IS-deviations, we reach a partition π which contains the coalitions $\{k_j, k'_j, \ell_j\}$ for every $1 \leq j \leq m$. Then, for each $1 \leq j \leq m$, by increasing order of the indices, clause-agent k_j deviates to coalition $\{k_{j+1}, k'_{j+1}, \ell_{j+1}\}$ (where $m+1$ refers to 1). After all these IS-deviations, we reach a partition which contains the coalitions $\{k_j, k'_{j+1}, \ell_{j+1}\}$ for every $1 \leq j \leq m$ (where $m+1$ refers to 1). At this point, for each $1 \leq j \leq m$, by increasing order of the indices, clause-agent k'_j deviates to coalition $\{k_j, k'_{j+1}, \ell_{j+1}\}$ (where $m+1$ refers to 1), in order to recover utility 1 by belonging to the same coalition as clause-agent k_j. After all these IS-deviations, we reach a partition which contains the coalitions $\{k_j, k'_j, \ell_{j+1}\}$ for every $1 \leq j \leq m$ (where $m+1$ refers to 1). Then, for each $1 \leq j \leq m$, by increasing order of the indices, literal-agent ℓ_j deviates to coalition $\{k_j, k'_j, \ell_{j+1}\}$, in order to recover utility 1 by belonging to the same coalition as clause-agent k'_j. After all these deviations, we reach again partition π and thus there is a cycle in the sequence of IS-deviations. For proving the reverse part of the equivalence proof, one can show that no other type of cycle can occur. □

The counterexample provided in the proof of Proposition 6.1 exhibits a global cycle in the preferences of the agents: $\{1,2\} \rhd \{1,3\} \rhd \{2,3\} \rhd \{1,2\}$. However, by considering dichotomous preferences with *common ranking property*, that is, each agent has a threshold for acceptance in a given global order, we obtain convergence thanks to the same potential function argument used by Caskurlu and Kizilkaya [12], for proving the existence of a core-stable partition in hedonic games with common ranking property.

When assuming that if a coalition is approved by one agent, then it must be approved by all the members of the coalition (so-called *symmetric dichotomous preferences*), we obtain a special case of preferences with common ranking property where all the approved coalitions are at the top of the global order. Therefore, convergence is also guaranteed under symmetric dichotomous preferences.

7 Conclusion

We have investigated dynamics of deviations based on individual stability in hedonic games. The two main questions we considered were whether there exists *some* sequence of deviations terminating in an IS partition, and whether *all* sequences of deviations terminate in an IS partition, i.e., the dynamics converges. Our results are mostly negative with examples of cycles in dynamics or even non-existence of IS partitions under rather strong preference restrictions. In particular, we have answered a number of open problems proposed in the literature. On the other hand, we have identified natural conditions for convergence that are mostly based on preferences relying on a common scale for the agents, like the common ranking property, single-peakedness or symmetry. An overview of our results can be found in Table 1.

Class	Convergence	Hardness
AHGs	✓ strict & nat. SP (single-peaked) (Prop. 3.4) ✓ neutral (derived from Suksompong [18]) ○ strict & gen. SP; singletons / grand coalition (Prop. 3.2)	∃ strict (Th. 3.3) ∀ strict (Th. 3.3)
HDGs	✓ strict & nat. SP; singletons (Prop. 4.3) ○ strict & nat. SP (Prop. 4.1) ○ strict; singletons / grand coalition (Prop. 4.1) ○ nat. SP; singletons (Prop. 4.1)	∃ strict (Th. 4.4) ∀ strict (Th. 4.4)
FHGs	✓ simple & sym.; singletons (Prop. 5.4) ✓ acyclic digraph (Th. 5.5) ○ sym. (Th. 5.1)	∃ sym. (Th. 5.3) ∃ simple; singletons (Th. 5.7) ∃ asym. (Th. 5.6) ∀ sym. (Th. 5.3) ∀ asym. (Th. 5.6)
DHGs	✓ grand coalition (Boehmer and Elkind [6]) ✓ common ranking property or symmetric (derived from Caskurlu and Kizilkaya [12]) ○ singletons (Prop. 6.1)	∃ singletons (Th. 6.2) ∀ general (Th. 6.2)

Table 1 – Convergence and hardness results for the dynamics of IS-deviations in various classes of hedonic games. Symbol ✓ marks convergence under the given preference restrictions and initial partition (if applicable) while ○ marks non-convergence, i.e., cycling dynamics. Symbol ∃ (resp., ∀) denotes that problem ∃-IS-Sequence-HG (resp., ∀-IS-Sequence-HG) is NP-hard (resp., co-NP-hard).

For all hedonic games under study, it turned out that the existence of cycles for IS-deviations is sufficient to prove the hardness of recognizing instances for which there exists a finite sequence of deviations or whether all sequences of deviations are finite, i.e., the dynamics converges. While

our results cover a broad range of hedonic games considered in the literature, there are still promising directions for further research. First, even though our hardness results hold under strong restrictions, the complexity of these questions remains open for other interesting preference restrictions, some of which do not guarantee convergence. Following our work, the most intriguing cases are AHGs under single-peaked weak preferences, simple symmetric FHGs with arbitrary initial partitions, and HDGs under single-peaked preferences. Secondly, one could investigate more specific rules of IS-deviations that quickly terminate in IS partitions, even in classes of hedonic games that allow for cyclic IS-deviations. For instance, for simple symmetric FHGs, there is the possibility of convergence such that each agent deviates at most once, but the selection of the deviating agents in this approach requires to solve a maximum clique problem (cf. the discussion after Proposition 5.4). Finally, the dynamics we consider only guarantee individual stability. One could also aim at reaching outcomes that satisfy Pareto optimality or other desirable properties on top of individual stability.

References

[1] Abeledo, H. et U. G. Rothblum: *Paths to marriage stability*. Discrete Applied Mathematics, 63(1) :1–12, 1995.

[2] Aziz, H., F. Brandl, F. Brandt, P. Harrenstein, M. Olsen et D. Peters: *Fractional Hedonic Games*. ACM Transactions on Economics and Computation, 7(2) :1–29, 2019.

[3] Aziz, H., P. Harrenstein, J. Lang et M. Wooldridge: *Boolean Hedonic Games*. Dans *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 166–175, 2016.

[4] Ballester, C.: *NP-completeness in hedonic games*. Games and Economic Behavior, 49(1) :1–30, 2004.

[5] Bilò, V., A. Fanelli, M. Flammini, G. Monaco et L. Moscardelli: *Nash Stable Outcomes in Fractional Hedonic Games : Existence, Efficiency and Computation*. Journal of Artificial Intelligence Research, 62 :315–371, 2018.

[6] Boehmer, N. et E. Elkind: *Individual-Based Stability in Hedonic Diversity Games*. Dans *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 1822–1829, 2020.

[7] Bogomolnaia, A. et M. O. Jackson: *The Stability of Hedonic Coalition Structures*. Games and Economic Behavior, 38(2) :201–230, 2002.

[8] Brandl, F., F. Brandt et M. Strobel: *Fractional Hedonic Games : Individual and Group Stability*. Dans *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1219–1227. IFAAMAS, 2015.

[9] Brandt, F. et A. Wilczynski: *On the Convergence of Swap Dynamics to Pareto-Optimal Matchings*. Dans *Proceedings of the 15th International Conference on Web and Internet Economics (WINE)*, pages 100–113, 2019.

[10] Bredereck, R., E. Elkind et A. Igarashi: *Hedonic Diversity Games*. Dans *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 565–573. IFAAMAS, 2019.

[11] Carosi, R., G. Monaco et L. Moscardelli: *Local Core Stability in Simple Symmetric Fractional Hedonic Games*. Dans *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 574–582. IFAAMAS, 2019.

[12] Caskurlu, B. et F. E. Kizilkaya: *On Hedonic Games with Common Ranking Property*. Dans *Proceedings of the 11th International Conference on Algorithms and Complexity (CIAC)*, pages 137–148. Springer International Publishing, 2019.

[13] Drèze, J. H. et J. Greenberg: *Hedonic Coalitions : Optimality and Stability*. Econometrica, 48(4) :987–1003, 1980.

[14] Farrell, J. et S. Scotchmer: *Partnerships*. Quarterly Journal of Economics, 103 :279–297, 1988.

[15] Hoefer, M., D. Vaz et L. Wagner: *Dynamics in matching and coalition formation games with structural constraints*. Artificial Intelligence, 262 :222–247, 2018.

[16] Peters, D.: *Complexity of Hedonic Games with Dichotomous Preferences*. Dans *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 579–585, 2016.

[17] Roth, A. E. et J. H. Vande Vate: *Random Paths to Stability in Two-Sided Matching*. Econometrica, 58(6) :1475–1480, 1990.

[18] Suksompong, W.: *Individual and Group Stability in Neutral Restrictions of Hedonic Games*. Mathematical Social Sciences, 78 :1–5, 2015.

[19] Sung, S. C. et D. Dimitrov: *Computational Complexity in Additive Hedonic Games*. European Journal of Operational Research, 203(3) :635–639, 2010.

Actes JIAF 2021

Apprentissage de structure de réseau bayésien sous contraintes en présence de connaissances d'experts

Christophe Gonzales[1] **Axel Journe**[2]
Ahmed Mabrouk[2]

[1] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[2] CSAI, ENGIE Lab CRIGEN, France
christophe.gonzales@lis-lab.fr axel.journe@engie.com
ahmed.mabrouk@engie.com

Résumé

Durant l'apprentissage de la structure d'un réseau bayésien (RB), l'introduction de connaissances d'experts peut améliorer significativement les résultats. Néanmoins, en pratique, les experts n'ont pas toujours une certitude absolue sur ces connaissances, lesquelles pouvant même être contradictoires lorsque plusieurs experts interviennent en même temps. L'introduction de connaissances incertaines dans des algorithmes d'apprentissage est une tâche délicate. Dans la littérature, il existe quelques algorithmes fondés sur des scores capables d'exploiter ces connaissances. Mais, pour autant que nous le sachions, il n'existe pas d'équivalent pour les algorithmes sous contraintes. Dans cet article, nous proposons un nouveau test d'indépendance capable de gérer les connaissances incertaines au moment de l'apprentissage. Nous présentons également un nouvel algorithme d'apprentissage sous contraintes utilisant ce test, ainsi qu'une étude expérimentale afin d'évaluer l'efficacité et la robustesse de notre approche par rapport aux méthodes de l'état de l'art.

Abstract

Exploiting experts' knowledge can significantly increase the quality of the Bayesian network (BN) structures produced by learning algorithms. However, in practice, experts may not be 100% confident about the opinions they provide. Worst, the latter can also be conflicting. Including such specific knowledge in learning algorithms is therefore complex. In the literature, there exist a few score-based algorithms that can exploit both data and the knowledge about the existence/absence of arcs in the BN. But, as far as we know, no constraint-based learning algorithm is capable of exploiting such knowledge. In this paper, we fill this gap by introducing the mathematical foundations for new independence tests including this kind of information. We provide a new constraint-based algorithm relying on these tests as well as experiments that highlight the robustness of our method and its benefits compared to other constraint-based learning algorithms.

1 Introduction

Les réseaux bayésiens (RBs) figurent parmi les modèles graphiques probabilistes les plus populaires pour représenter des distributions de probabilité. Ils offrent un cadre efficace pour la représentation des connaissances et le raisonnement probabiliste. La construction de ces modèles peut se faire manuellement lorsque le nombre de variables n'est pas trop important ou par apprentissage à partir de données. Malgré le caractère NP-difficile de l'apprentissage de structure [5], il existe dans la littérature de nombreux algorithmes d'apprentissage exacts ou approchés. Ces méthodes peuvent être divisées en quatre grandes catégories : i) les approches à base de score ; ii) les approches d'apprentissage sous contraintes ; iii) les approches utilisant un ordre sur les variables ; iv) les approches hybrides qui ont pour but d'exploiter les avantages des méthodes précédentes. Dans cet article, nous nous focalisons sur les approches d'apprentissage sous contraintes. Ces dernières visent à construire la structure du RB par le biais de l'ensemble des indépendances conditionnelles contenues dans la distribution associée au modèle représenté. Ces méthodes possèdent des propriétés intéressantes en termes de précision des structures apprises, de découverte causale ainsi qu'une bonne gestion des problèmes liés au surapprentissage.

Bien que les RBs soient rarement construits entièrement à partir des connaissances d'experts, il est souvent intéressant d'exploiter ces dernières au sein des algorithmes d'apprentissage automatique afin d'améliorer les résultats

obtenus. En un sens, les approches à base de score incluent naturellement des connaissances expertes via l'utilisation d'*a prioris* sur les distributions des paramètres des tables de probabilités des RBs appris. Cependant, ces méthodes ne sont pas particulièrement adaptées pour représenter correctement l'ensemble des connaissances que peuvent exprimer les experts. Pour pallier ce problème, d'autres méthodes plus sophistiquées ont été proposées par [2, 4]. Ces approches utilisent des connaissances sous forme de contraintes structurelles au moment de la construction du RB. D'autres connaissances de natures incertaines et conflictuelles, portant sur la présence et/ou l'absence des arcs, ont également été exploitées dans [10, 1] grâce à de nouvelles fonctions de score.

Dans cet article, nous nous intéressons à l'exploitation de connaissances incertaines et potentiellement conflictuelles au sein d'une approche d'apprentissage sous contraintes. Pour autant que nous le sachions, seules les connaissances certaines concernant la présence et/ou l'absence d'arcs ont été considérées dans ce type d'approche [3]. En revanche, les connaissances exprimées par les experts sont souvent entachées d'incertitude ainsi que de conflits dans la plupart des domaines d'application. Les approches sous contraintes, telles que décrites dans la littérature, utilisent des tests d'indépendances statistiques pour identifier l'ensemble des indépendances conditionnelles en se basant uniquement sur les données d'observation. Notre objectif ici consiste à incorporer les connaissances au moment de l'utilisation de ces tests d'indépendances. Dans la section 2, nous introduisons les notions mathématiques nécessaires pour ce type de tests. Nous proposons également un nouveau test d'indépendance de type G^2 permettant d'introduire d'une manière efficace l'ensemble des connaissances incertaines sur les indépendances et/ou les dépendances conditionnelles entre les variables. La certitude des experts ainsi que notre confiance en chacun d'eux sont également des paramètres qui sont pris en compte. Dans la section 3, nous proposons une variante de l'algorithme PC-stable [6] utilisant le nouveau test d'indépendance. Dans la section 4, nous présentons une étude expérimentale afin de souligner la pertinence de l'approche proposée en terme d'efficacité et de robustesse. Enfin la section 5 conclut cet article.

2 Un nouveau test d'indépendance

Dans cette partie, nous introduisons un nouveau test d'indépendance capable de prendre en compte des connaissances incertaines d'experts. Ce nouveau test est fondé sur le test du G^2, qui s'appuie lui-même sur un rapport de vraisemblances. Pour développer les calculs de vraisemblances, nous nous appuyons sur l'exploitation de RBs dont la définition est la suivante :

Définition 1 (Réseau bayésien) *Un RB est constitué d'un couple* $(G, \mathbf{\Theta})$ *où* $G = (\mathbf{V}, \mathbf{E})$ *est un graphe orienté sans circuit (DAG) avec* $\mathbf{V}$ *un ensemble de variables aléatoires* [1]*,* $\mathbf{E}$ *un ensemble d'arc, et* $\mathbf{\Theta} = \{P(X|\boldsymbol{Pa}(X))\}_{X\in\mathbf{V}}$ *l'ensemble des distributions de probabilités conditionnelles (CPD) des nœuds / variables aléatoires* X *dans* G *sachant leurs parents* $\boldsymbol{Pa}(X)$ *dans* G*. Le RB encode la probabilité jointe sur* $\mathbf{V}$ *définie par* $P(\mathbf{V}) = \prod_{X\in\mathbf{V}} P(X|\boldsymbol{Pa}(X))$*.*

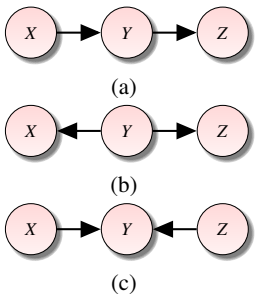


FIGURE 1 – Trois sous-graphes de base : (1a) en série, (1b) divergent et (1c) en v-structure

Il est possible de résumer toutes les structures possibles avec les trois sous-structures de la figure 1. Les structures en série (1a) et divergentes (1b) indiquent une dépendance marginale entre X et Z mais une indépendance conditionnelle entre X et Z sachant Y. La v-structure (1c) indique au contraire une indépendance marginale entre X et Z mais une dépendance conditionnelle entre X et Z sachant Y. Cette description est généralisable grâce à un critère graphique, appelé d-séparation, permettant de lire l'ensemble des indépendances conditionnelles de la distribution associée au graphe dans la structure de ce dernier.

Définition 2 (D-séparation) *Deux ensembles de variables* $\boldsymbol{X}$ *et* $\boldsymbol{Y}$ *sont d-séparés par un ensemble de variables* $\boldsymbol{Z}$ *si et seulement si :*

- *les chemins entre* $\boldsymbol{X}$ *et* $\boldsymbol{Y}$ *qui contiennent une v-structure dont aucune variable parmi celles au centre de la v-structure et ses descendantes ne sont dans* $\boldsymbol{Z}$
- *et les autres chemins entre* $\boldsymbol{X}$ *et* $\boldsymbol{Y}$ *ne contiennent pas de v-structure mais au moins une variable sur chaque chemin est dans* $\boldsymbol{Z}$

Dans ce qui suit, nous allons présenter une approche à base de réseau bayésien permettant de décrire l'ensemble des éléments requis pour le calcul du nouveau test d'indépendance G^2. Considérons deux variables aléatoires $\mathbb{X}$ et $\mathbb{Y}$ avec leurs domaines respectifs $\Omega_{\mathbb{X}} = \{x_1, \ldots, x_r\}$ et $\Omega_{\mathbb{Y}} = \{y_1, \ldots, y_s\}$ et dont la distribution jointe est $P(\mathbb{X}, \mathbb{Y}) = \{\theta_{x_i y_j} : x_i \in \Omega_{\mathbb{X}}, y_j \in \Omega_{\mathbb{Y}}\}$. Pour tout $x_i \in \Omega_{\mathbb{X}}$, notons $\theta_{x_i.} = \sum_{y_j\in\Omega_{\mathbb{Y}}} \theta_{x_i y_j}$ et, pour tout $y_j \in \Omega_{\mathbb{Y}}$, notons

1. Par abus de notation, nous utiliserons $X \in \mathbf{V}$ sans distinction pour parler d'un nœud dans le graphe et sa variable aléatoire correspondante.

$\theta_{\cdot y_j} = \sum_{x_i \in \Omega_{\mathbb{X}}} \theta_{x_i y_j}$. Notons également $\mathbf{S} = \{(XY)^{(n)}\}_{n=1}^N$ un ensemble de variables mutuellement indépendantes, distribuées selon $P(\mathbb{X}, \mathbb{Y})$. Ainsi, $P((XY)^{(n)} = x_i y_j) = \theta_{x_i y_j}$ pour tout $x_i y_j \in \Omega_{\mathbb{XY}} = \Omega_{\mathbb{X}} \times \Omega_{\mathbb{Y}}$. Un ensemble $\mathbf{s} = \{(xy)^{(n)}\}_{n=1}^N$ d'observations de $(XY)^{(n)}$ est donc un échantillon d'observations i.i.d. de $(\mathbb{X}, \mathbb{Y})$. Enfin, notons $\mathbb{I}$ une variable aléatoire indiquant si $\mathbb{X}$ et $\mathbb{Y}$ sont indépendants ($\mathbb{I} = 1$) ou dépendants ($\mathbb{I} = 0$). Nous représentons la relation entre $(XY)^{(n)}$ et $\mathbb{I}$ en utilisant le RB de la figure 2. Ce RB montre que l'ensemble des $(XY)^{(n)}$ sont mutuellement indépendants mais deviennent dépendants lorsque $\mathbb{I}$ est observé. En effet, si $\mathbb{I} = 1$ et si on observe un sous-ensemble $\mathbf{S}' \subset \mathbf{S}$ en adéquation avec $\mathbb{I} = 0$, alors il est probable que l'observation de $\mathbf{S} \backslash \mathbf{S}'$ compensera cela pour, globalement, représenter l'indépendance entre $\mathbb{X}$ et $\mathbb{Y}$.

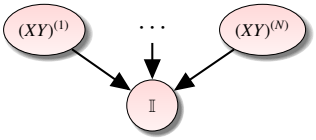


FIGURE 2 – RB modélisant la relation entre les variables $(XY)^{(n)}$ et la variable $\mathbb{I}$.

Proposition 1 *Soit* $\mathbf{s} = \{(xy)^{(n)}\}_{n=1}^N$ *un échantillon d'observations de* $\{(XY)^{(n)}\}_{n=1}^N$. *Notons* $p_{\mathbb{I}} = P(\mathbb{I} = 1)$. *Supposons que la distribution jointe conditionnelle empirique de* $\{(XY)^{(n)}\}_{n=1}^N$ *sachant* $\mathbb{I} = 1$ *est approximativement égale à* $\prod_{n=1}^N P((XY)^{(n)} | \mathbb{I} = 1)$. *Alors :*

$$P(\mathbb{I} = i | \mathbf{s}) \approx \begin{cases} 1 - P(\mathbb{I} = 1 | \mathbf{s}) & \text{si } i = 0, \\ p_{\mathbb{I}} \times \prod_{xy \in \Omega_{\mathbb{XY}}} \left[\frac{\theta_{x\cdot} \times \theta_{\cdot y}}{\theta_{xy}} \right]^{N_{xy}} & \text{si } i = 1, \end{cases} \quad (1)$$

où N_{xy} *est le nombre d'éléments dans* $\mathbf{s}$ *égaux à* xy.

Notons que l'hypothèse dans la proposition 1 correspond précisément à l'hypothèse faite habituellement dans les tests d'indépendance du χ^2 et du G^2. La figure 2, les valeurs de $p_{\mathbb{I}}$ et de $\{\theta_{xy}\}$ ainsi que la proposition 1 caractérisent complètement le RB de la distribution jointe de $\mathbb{I}$ et $\{(XY)^{(n)}\}_{n=1}^N$. Les paramètres $\{\theta_{xy}\}$ peuvent être estimés simplement par Maximum de Vraisemblance. De plus, on peut constater que l'équation (1) ne restreint pas la valeur de $p_{\mathbb{I}}$. Cette dernière devrait donc refléter nos connaissances initiales sur l'indépendance entre $\mathbb{X}$ et $\mathbb{Y}$. Pour des raisons expliquées dans la partie Expérimentations, nous la fixons à 0.95.

Maintenant, supposons que K experts apportent leurs connaissances à propos de l'état de $\mathbb{I}$ et donc l'indépendance entre $\mathbb{X}$ et $\mathbb{Y}$, avec des assertions e_k telles que "Je pense que $\mathbb{X}$ et $\mathbb{Y}$ sont indépendants" ou "Je crois que $\mathbb{X}$ influence directement $\mathbb{Y}$". Soit O_k une variable aléatoire qui représente l'opinion de l'expert k (ici, $O_k = 0$ et $O_k = 1$ signifient respectivement une dépendance et une indépendance). Ainsi, e_k est une observation de la variable O_k et $P(e_k|O_k)$ est un vecteur de taille $|O_k|$ avec une valeur égale à 1 et toutes les autres égales à 0. En effet, dans l'expression $P(e_k|O_k)$, seul O_k est une variable aléatoire, e_k n'étant qu'une observation, une information et, à ce titre, e_k n'a pas de dimension. Ainsi $P(e_k|O_k = 0) = 1$ s'interprète comme la probabilité que le $k$ème expert nous ait informé qu'il penche pour une dépendance entre $\mathbb{X}$ et $\mathbb{Y}$ sachant qu'il pense qu'il y a effectivement une dépendance. Évidemment, les experts ne sont pas forcément certains de leurs expertises et ils accordent donc une certaine confiance dans leur jugement. Nous supposons ici qu'ils sont capables d'estimer cette confiance dans leur identifications de l'état de $\mathbb{I}$. Notons γ_k cette confiance. Par exemple, l'expert k peut estimer être sûr à 70% que son jugement est correct. Dans ce cas, $\gamma_k = 0.7$ et donc $P(O_k = i | \mathbb{I} = i) = \gamma_k$ pour $i \in \{0, 1\}$. Les opinions des experts sont fondées sur leurs expériences avec $\mathbb{X}$ et $\mathbb{Y}$. Par conséquent, elles ne sont pas indépendantes. Cependant, du moment que la valeur de $\mathbb{I}$ est connue, les connaissances d'un ou plusieurs experts n'apportent plus aucune information supplémentaire sur les opinions des autres experts. D'après cela, les relations entre $\mathbb{I}$ et les variables O_k peuvent être encodées par un RB contenant uniquement des arcs $\mathbb{I} \rightarrow O_k$. On pourrait rajouter ces arcs directement dans le RB de la figure 2 afin de tenir compte du savoir d'experts dans l'exploitation statistique des échantillons. Toutefois, les distributions $P(O_k|\mathbb{I})$, autrement dit les γ_k, ne sont pas nécessairement aisées à estimer par les experts. En effet, les difficultés d'estimation des probabilités de la part des humains sont connues [11]. Pour corriger ce biais, nous proposons d'appliquer une transformation aux γ_k. Pour éviter toute ambiguïté, nous introduisons une nouvelle variable E_k utilisant cette transformation : $P(E_k = i | \mathbb{I} = i) = \varphi_k(\gamma_k)$ et c'est cette variable que nous exploiterons dans nos tests statistiques. Notons que $\varphi_k(\cdot)$ peut également inclure notre propre perception quant à la justesse du savoir de l'expert k. Dans cet article, nous suggérons d'utiliser la fonction logistique suivante de $[0, 1] \mapsto [0, 1]$:

$$\varphi_k(\gamma_k) = \frac{1}{\beta_k - 1} \left[\frac{1 + \beta_k}{1 + \beta_k^{1-2\gamma_k}} - 1 \right], \quad (2)$$

avec β_k un paramètre strictement positif. Cette transformation est plutôt générique, comme le montre la figure 4 dans laquelle, pour simplifier, $\rho_k = 2 \log \beta_k$.

A partir de ce qui précède, pour prendre en compte les connaissances des experts et leurs incertitudes, nous pouvons étendre le RB de la figure 2 comme indiqué sur la figure 3. Les évidences e_k peuvent être soumises en tant que $P(e_k|E_k) = P(e_k|O_k)$.

En utilisant le RB de la figure 3, il est possible de calculer le rapport de vraisemblances utilisé dans les tests d'indépendance de type G^2. Rappelons qu'en statistiques, le

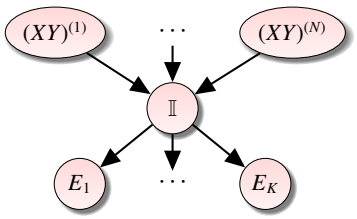


FIGURE 3 – RB modélisant la relation entre les variables $(XY)^{(n)}$ et la variable $\mathbb{I}$, en incluant les connaissances d'experts.

test d'indépendance du G^2 est un test très populaire fondé sur le ratio de vraisemblances $\mathcal{LR}(\mathbf{s}) = P(\mathbf{s})/P(\mathbf{s}|\mathbb{I} = 1)$, où $\mathbf{s}$ représente l'échantillon observé $\{(XY)^{(n)}\}_{n=1}^{N}$. Si $\mathbf{S}$ est la variable aléatoire correspondant aux échantillons possibles (avant que l'on observe effectivement celui servant pour notre apprentissage), alors, sous l'hypothèse que la vraisemblance sous la contre-hypothèse $P(\mathbf{s}|\mathbb{I} = 0)$ est proche de celle sous l'hypothèse nulle $P(\mathbf{s}|\mathbb{I} = 1)$, la variable aléatoire $2\ln(\mathcal{LR}(S))$ suit une loi du chi2 à k degrés de liberté (χ^2_k), où $k = (|\Omega_{\mathbb{X}}| - 1) \times (|\Omega_{\mathbb{Y}}| - 1)$. On peut donc en déduire avec un niveau de risque fixé si l'on peut estimer que $\mathbb{X}$ et $\mathbb{Y}$ sont indépendants ou non.

Proposition 2 *soit* $\mathbf{s} = \{(xy)^{(n)}\}_{n=1}^{N}$ *un échantillon d'observations de* $\{(XY)^{(n)}\}_{n=1}^{N}$ *et* $\mathbf{e} = \{e_k\}_{k=1}^{K}$ *des connaissances d'experts. Alors, les vraisemblances* $\mathcal{L}(\mathbf{s}|\mathbb{I} = 1, \mathbf{e}) = P(\mathbf{s}|\mathbb{I} = 1, \mathbf{e}) = \mathcal{L}_{P_1}(\mathbf{s}|\mathbf{e})$ *et* $\mathcal{L}(\mathbf{s}|\mathbf{e}) = P(\mathbf{s}|\mathbf{e})$ *s'obtiennent de la manière suivante :*

$$\mathcal{L}(\mathbf{s}|\mathbf{e}) = \frac{\epsilon_0}{P(\mathbf{e})} \times \mathcal{L}_{P_0}(\mathbf{s}|\mathbf{e}) + \left[1 - \frac{\epsilon_0}{P(\mathbf{e})}\right] \times \mathcal{L}_{P_1}(\mathbf{s}|\mathbf{e}), \quad (3)$$

$$\text{avec } \mathcal{L}_{P_0}(\mathbf{s}|\mathbf{e}) = \prod_{xy \in \Omega_{\mathbb{XY}}} \theta_{xy}^{N_{xy}},\ \mathcal{L}_{P_1}(\mathbf{s}|\mathbf{e}) = \prod_{xy \in \Omega_{\mathbb{XY}}} \left(\theta_{x.}\theta_{.y}\right)^{N_{xy}},$$

$$P(\mathbf{e}) = \epsilon_0 + P(\mathbb{I} = 1) \times [\epsilon_1 - \epsilon_0],$$

$$\epsilon_0 = P(\mathbf{e}|\mathbb{I} = 0) = \sum_{E_1,\dots,E_K} \prod_{k=1}^{K} P(E_k|\mathbb{I} = 0) P(e_k|E_k),$$

$$\epsilon_1 = P(\mathbf{e}|\mathbb{I} = 1) = \sum_{E_1,\dots,E_K} \prod_{k=1}^{K} P(E_k|\mathbb{I} = 1) P(e_k|E_k),$$

N_{xy} *= le nombre d'éléments de* $\mathbf{s}$ *égaux à* xy.

On peut interpréter l'équation (3) de la manière suivante : il y a une incertitude concernant la distribution de probabilité P de la population, qui peut être égale soit à $P_0(\mathbb{X}, \mathbb{Y}) = \{\theta_{xy} : x \in \Omega_{\mathbb{X}}, y \in \Omega_{\mathbb{Y}}\}$, soit à $P_1(\mathbb{X}, \mathbb{Y}) = \{\theta_{x.}\theta_{.y} : x \in \Omega_{\mathbb{X}}, y \in \Omega_{\mathbb{Y}}\}$. $\mathcal{L}$ peut ainsi être interprété comme un mélange des vraisemblances de deux population, pondérées par la certitude que les experts ont dans leurs connaissances ainsi que par notre perception de leur justesse.

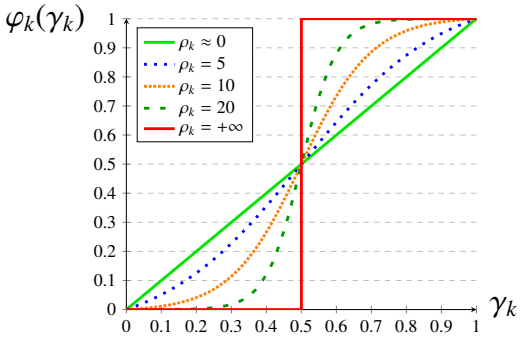


FIGURE 4 – Transformation de probabilité $\varphi_k(\cdot)$.

Lorsque qu'aucune connaissance e_k n'est disponible, nous retrouvons $\epsilon_0 = \epsilon_1 = P(\mathbf{e}) = 1$, ce qui correspond aux vraisemblances traditionnellement utilisées dans les tests du G^2. De la même manière, lorsque nous avons les mêmes degrés de confiances pour chaque expert et quand les experts n'ont pas de connaissances et donnent juste une estimation 50-50% sur l'indépendance et la dépendance, alors $\epsilon_0 = \epsilon_1$. Notons que dès que $\epsilon_0 = \epsilon_1$, $P(\mathbf{s}, \mathbf{e}) = \mathcal{L}_{P_0}(\mathbf{s}|\mathbf{e})$, ce qui correspond à la vraisemblance utilisée au numérateur du test G^2 classique.

Notons $\mathcal{LR}(\mathbf{s}|\mathbf{e})$ le rapport de vraisemblance $\mathcal{L}(\mathbf{s}|\mathbf{e})/\mathcal{L}(\mathbf{s}|\mathbb{I} = 1, \mathbf{e})$. Alors d'après la proposition 2, on a :

$$\mathcal{LR}(\mathbf{s}|\mathbf{e}) = \frac{\epsilon_0}{P(\mathbf{e})} \times \mathcal{LR}_{P_0}(\mathbf{s}|\mathbf{e}) + \left[1 - \frac{\epsilon_0}{P(\mathbf{e})}\right] \times \mathcal{LR}_{P_1}(\mathbf{s}|\mathbf{e}),$$

avec $\mathcal{LR}_{P_0}(\mathbf{s}|\mathbf{e}) = \mathcal{L}_{P_0}(\mathbf{s}|\mathbf{e})/\mathcal{L}_{P_1}(\mathbf{s}|\mathbf{e})$ et $\mathcal{LR}_{P_1}(\mathbf{s}|\mathbf{e}) = 1$.

Ce rapport de vraisemblances est un mélange de deux vraisemblances, l'une sur la population liée à P_0 et l'autre sur la population liée à P_1. Avec le test du G^2 classique, si on suppose que la vraisemblance sous la contre-hypothèse est proche de celle sous l'hypothèse nulle, leur rapport devrait être proche de 1. Par conséquent, sous cette hypothèse, à un niveau de risque (de première espèce) α, le test du G^2 classique consiste à calculer le seuil au-dessus duquel seul un pourcentage α de l'ensemble de tous les échantillons possibles va produire un rapport plus grand que ce seuil. Dans notre cas, le pourcentage α est logiquement séparé en un pourcentage $\alpha \times \eta$ sur la population liée à P_0 et un pourcentage $\alpha \times (1 - \eta)$ sur la population liée à P_1, avec $\eta = \epsilon_0/P(\mathbf{e})$. Le rapport de vraisemblances $\mathcal{LR}_{P_1}(\mathbf{s}|\mathbf{e})$ est toujours égal à 1, donc le pourcentage $\alpha \times (1 - \eta)$ n'apporte aucune information pour le test. Ainsi, ce nouveau test repose essentiellement sur $\mathcal{LR}_{P_0}(\mathbf{s}|\mathbf{e})$.

Proposition 3 *En supposant que la vraisemblance sous la contre-hypothèse est proche de celle sous l'hypothèse nulle, on a :*

$$2\ln(\mathcal{LR}_{P_0}(\mathbf{S}|\mathbf{e})) \sim \chi^2_k, \text{ où } k = (|\Omega_{\mathbb{X}}| - 1) \times (|\Omega_{\mathbb{Y}}| - 1).$$

La proposition précédente permet donc de définir le test d'indépendance suivant :

Règle 1 *À un niveau de risque α, notre test d'indépendance sur un échantillon* $\mathbf{s}$ *et avec des connaissances d'experts* $\mathbf{e}$ *consiste à accepter l'indépendance entre $\mathbb{X}$ et $\mathbb{Y}$ si et seulement si $2\ln(\mathcal{LR}_{P_0}(\mathbf{s}|\mathbf{e})) < c_{k,\alpha\eta}$ où $\eta = \epsilon_0/P(\mathbf{e})$ et, pour tout δ, $c_{k,\delta}$ représente le $(1-\delta)$ quantile de la distribution χ^2 à k degrés de liberté.*

De manière équivalente, on peut énoncer notre test de la manière suivante : on accepte l'indépendance si la p-valeur de $2\ln(\mathcal{LR}_{P_0}(\mathbf{s}|\mathbf{e}))$ est plus grande que $\alpha \times \eta$.

En d'autres termes, notre test consiste à appliquer le test G^2 classique avec un niveau de signification $\alpha \times \eta$ au lieu de α. Ce test peut être aisément étendu aux tests d'indépendance conditionnelle sachant un ensemble $\mathbb{Z}$ de variables aléatoires. Pour cela, il suffit d'étendre le RB de la figure 3 comme indiqué sur la figure 5. Dans ce nouveau RB, $Z^{(1)}, \ldots, Z^{(N)}$ sont des variables aléatoires distribuées selon $P(\mathbb{Z}) = \{\theta_z : z \in \Omega_{\mathbb{Z}}\}$ et $\mathbb{I}$ est une variable aléatoire indiquant la dépendance ou l'indépendance conditionnelle de $\mathbb{X}$ et $\mathbb{Y}$ sachant $\mathbb{Z}$.

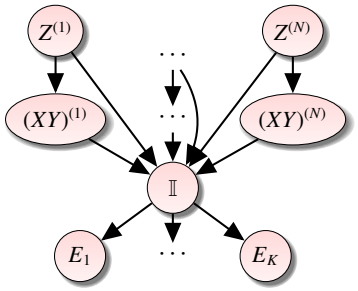


FIGURE 5 – Le RB permettant de tester l'indépendance conditionnelle entre $\mathbb{X}$ et $\mathbb{Y}$ sachant $\mathbb{Z}$, en tenant compte des savoirs incertains d'experts.

Proposition 4 *Soit* $\mathbf{s} = \{((xy)^{(n)}, z^{(n)})\}_{n=1}^{N}$ *un échantillon d'observations de* $\mathbf{S} = \{((XY)^{(n)}, Z^{(n)})\}_{n=1}^{N}$. *Pour tout $z \in \Omega_{\mathbb{Z}}$, supposons que $P(Z^{(n)}|\mathbb{I}) \approx P(Z^{(n)})$ et que :*

$$P(\mathbb{I}=1|\mathbf{s}) \approx p(\mathbb{I}=1) \times \prod_{xyz \in \Omega_{\mathbb{XYZ}}} \left[\frac{\theta_{x.|z}\theta_{.y|z}}{\theta_{xy|z}}\right]^{N_{xyz}}, \quad (4)$$

où $\theta_{x.|z} = \sum_{y\in\Omega_{\mathbb{Y}}} \theta_{xy|z}$, $\theta_{.y|z} = \sum_{x\in\Omega_{\mathbb{X}}} \theta_{xy|z}$ et $\theta_{xy|z} = P(\mathbb{X}=x, \mathbb{Y}=y|\mathbb{Z}=z)$. En notant $\mathcal{L}_{|\mathbf{z}}(\cdot)$ la vraisemblance conditionnelle sachant $\{z^{(n)}\}_{n=1}^{N}$, i.e., $\mathcal{L}_{|\mathbf{z}}(\mathbf{s}|\mathbf{e}) = P(\{(xy)^{(n)}\}_{n=1}^{N}|\mathbf{e}, \{z^{(n)}\}_{n=1}^{N})$, les propriétés suivantes sont vérifiées :

$$\begin{aligned}
\mathcal{LR}_{|\mathbf{z}}(\mathbf{s}|\mathbf{e}) &= \frac{\mathcal{L}_{|\mathbf{z}}(\mathbf{s}|\mathbf{e})}{\mathcal{L}_{|\mathbf{z}}(\mathbf{s}|\mathbb{I}=1,\mathbf{e})} \\
&= \frac{\epsilon_0}{P(\mathbf{e})} \times \mathcal{LR}_{P_0|\mathbf{z}}(\mathbf{s}|\mathbf{e}) + \left[1 - \frac{\epsilon_0}{P(\mathbf{e})}\right],
\end{aligned}$$

où $P(\mathbf{e}) = \epsilon_0 + P(\mathbb{I}=1) \times [\epsilon_1 - \epsilon_0]$,

$$\mathcal{LR}_{P_0|\mathbf{z}}(\mathbf{s}|\mathbf{e}) = \mathcal{L}_{P_0|\mathbf{z}}(\mathbf{s}|\mathbf{e}) / \mathcal{LR}_{P_1|\mathbf{z}}(\mathbf{s}|\mathbf{e}), \quad (5)$$

$$\mathcal{L}_{P_0|\mathbf{z}}(\mathbf{s}|\mathbf{e}) = \prod_{xyz\in\Omega_{\mathbb{XYZ}}} \theta_{xy|z}^{N_{xyz}},$$

$$\mathcal{L}_{P_1|\mathbf{z}}(\mathbf{s}|\mathbf{e}) = \prod_{xyz\in\Omega_{\mathbb{XYZ}}} \left(\theta_{x.|z}\theta_{.y|z}\right)^{N_{xyz}},$$

N_{xyz} est le nombre d'éléments dans $\mathbf{s}$ *égaux à xyz.*

En supposant que la vraisemblance sous la contre-hypothèse et celle sous l'hypothèse nulle sont proches, $2\ln(\mathcal{LR}_{P_0|\mathbf{z}}(\mathbf{S}|\mathbf{e})) \sim \chi^2_k$, *où* $k = (|\Omega_{\mathbb{X}}|-1) \times (|\Omega_{\mathbb{Y}}|-1) \times \Omega_{\mathbb{Z}}$.

Notre test d'indépendance conditionnelle est donc le même que notre test d'indépendance marginale, à l'exception que ce sont les probabilités conditionnelles, et non les probabilités marginales qui sont utilisées. Notons que tous les θ peuvent être estimés par Maximum de Vraisemblance, comme dans le test du G^2 classique.

3 Notre algorithme d'apprentissage avec connaissances d'experts

En théorie, nous pouvons directement utiliser notre nouveau test d'indépendance au sein des algorithmes d'apprentissage de structure de réseaux bayésiens sous contraintes (PC, IC, PC-stable, etc.). Toutefois, l'utilisation des connaissances incertaines soulève deux problèmes majeurs : i) la présence des bases de données de taille réduite et ii) la divergence entre les connaissances d'experts et les résultats obtenus avec les tests d'indépendance. Concernant le premier problème, les algorithmes d'apprentissage ignorent généralement les résultats retournés par le test d'indépendance lorsque la base de données est trop petite, car les résultats obtenues ne sont pas assez fiables. En revanche, l'utilisation des savoirs d'experts dans ce genre de situations demeure tout de même possible, et on peut donc se baser sur ces connaissances pour conclure sur l'indépendance étudiée. Pour cela, notons d la décision de considérer que « $\mathbb{X}$ et $\mathbb{Y}$ sont indépendants sachant $\mathbb{Z}$ » et soit u_1 et u_0 les utilités que l'on associe à une telle décision lorsque l'expert a respectivement raison et tort. En d'autres termes, u_1 et u_0 représentent notre perception du gain et de la perte associés à la prise de décision d lorsque celle-ci se fonde uniquement sur le savoir d'expert (nous supposons que $u_1 > 0$ et $u_0 < 0$). En général, l'élicitation d'utilités telles que u_1 et u_0 est difficile à réaliser. En revanche, leur ratio $T = -u_0/u_1$ est beaucoup plus aisé à estimer [8]. Enfin, notons par ϵ_1 et ϵ_0 les probabilités que les experts aient

respectivement raison et tort. Si notre critère pour prendre nos décisions est la maximisation de l'espérance d'utilité, nous devrions considérer $\mathbb{X}$ et $\mathbb{Y}$ conditionnellement indépendants sachant $\mathbb{Z}$ si et seulement si $\epsilon_1 u_1 + \epsilon_0 u_0 > 0$ ou, de manière équivalente, si $\epsilon_1/\epsilon_0 > T$. Nous utilisons précisément cette règle dans l'algorithme 1.

Concernant le deuxième problème, notons que retirer ou garder une arête à une étape d'un algorithme d'apprentissage sous contraintes peut avoir un impact significatif sur les étapes suivantes. De telles décisions doivent donc être prises avec prudence. Quand les conclusions avec et sans connaissances d'experts sont identiques, nous pouvons être confiant dans la décision qui en résulte. Au contraire, quand les conclusions diffèrent, nous devrions suivre la conclusion du test qui inclut les savoirs d'experts, puisque ce test prend en compte plus d'informations. Cependant, il ne serait pas prudent de complètement ignorer la conclusion obtenue sans les connaissances d'experts. Ainsi, dans l'algorithme 1, nous construisons un graphe G qui sera le squelette appris, mais nous proposons également de maintenir en parallèle un autre graphe G_{adj} utilisé pour déterminer les *sepset*, c'est-à-dire les ensembles de séparation qui rendent des variables indépendantes conditionnellement à cet ensemble. Le graphe G_{adj} prend en compte les désaccords entre les conclusions avec et sans connaissances d'experts.

Ces variations combinées dans l'algorithme PC-stable [6] résultent en l'algorithme 1, que l'on nomme PCSe. Outre la structure graphique apprise, celui-ci renvoie la liste des « sepset ». En effet c'est cette information qui permet, en exploitant la notion de d-séparation, d'orienter les arêtes du graphe G afin d'obtenir un réseau bayésien qui représente les indépendances conditionnelles apprises. Pour plus de détails sur cette phase d'orientation, le lecteur pourra se référer en particulier aux travaux de Meek dans [9].

Algorithm 1 Algorithme PCSe

Entrée : base de données $\mathbf{D}$, connaissances d'experts $\mathbf{K}$, niveau de risque α, connaissances*a priori* $P(\mathbb{I})$, seuil T
Sortie : un graphe non orienté G et un ensemble de $Sepsets$

$\mathbf{V} \leftarrow$ toutes les variables/colonnes de $\mathbf{D}$
$G = (\mathbf{V}, \mathbf{E}) \leftarrow$ graphe complet non orienté
$G_{adj} = (\mathbf{V}, \mathbf{E}_{adj}) \leftarrow$ graphe complet non orienté
$m \leftarrow -1$; $Sepset \leftarrow \emptyset$
répéter
  $m \leftarrow m + 1$
  pour tout X_i dans $\mathbf{V}$ **faire**
    $\mathbf{adj}_i \leftarrow \{X_j \in \mathbf{V} : (X_i, X_j) \in \mathbf{E}_{adj}\}$
  fin pour
  pour tout $(X_i, X_j) \in \mathbf{E}_{adj}$ t.q. $|\mathbf{adj}_i| \geq m + 1$ **faire**
    répéter
      Choisir un ensemble $\mathbf{Z} \subseteq \mathbf{adj}_i \backslash \{X_j\}$ t.q. $|\mathbf{Z}| = m$
      Extraire de $\mathbf{K}$ la connaissance sur $(X_i, X_j|\mathbf{Z})$
      Créer le RB de la figure 5 et calculer ses paramètres θ, ϵ_0, ϵ_1, $P(\mathbf{e})$ et $\eta = \epsilon_0/P(\mathbf{e})$
      si $\mathbf{D}$ est trop petit pour faire le test du G^2 **alors**
        si $\epsilon_1/\epsilon_0 > T$ **alors**
          Supprimer (X_i, X_j) dans $\mathbf{E}$ et $\mathbf{E}_{adj}$
        fin si
      sinon
        $p \leftarrow$ p-valeur de $2\ln(\mathcal{LR}_{P_0|\mathbf{z}}(\mathbf{s}|\mathbf{e}))$ de l'équation (5)
        si $p \geq \alpha$ **alors**
          Supprimer (X_i, X_j) dans $\mathbf{E}_{adj}$
        fin si
        si $p \geq \alpha \times \eta$ **alors**
          Supprimer (X_i, X_j) dans $\mathbf{E}$
          $Sepset(X_i, X_j) \leftarrow \mathbf{Z}$
          $Sepset(X_j, X_i) \leftarrow \mathbf{Z}$
        fin si
      fin si
    jusqu'à ce que $p \geq \alpha$ et $p \geq \alpha \times \eta$, ou tous les $\mathbf{Z}$ t.q. $|\mathbf{Z}| = m$ ont été considérés
  fin pour
jusqu'à ce que toutes les variables $X_i \in \mathbf{V}$ sont telles que $|\mathbf{adj}_i| \leq m$
retourne G, $Sepset$

4 Expérimentations

Dans cette section, nous démontrons l'efficacité de PCSe en le comparant avec PC-stable et l'algorithme proposé par [3], appelés respectivement par la suite PCS et PCSDC. Les algorithmes PCS et PCSDC sont deux variantes de l'algorithme classique PC. Grâce à l'algorithme PCS, l'ordre sur les variables fourni en entrée n'a aucun impact sur le résultat final de l'algorithme d'apprentissage. Concernant l'algorithme PCSDC, il permet d'introduire les connaissances d'experts sous forme déterministe. Afin d'optimiser l'analyse comparative entre les algorithmes, nous avons décidé d'incorporer la gestion de l'ordre sur les variables (de la même manière que PCS) au sein de l'algorithme PSDC. Pour cette comparaison, nous avons généré aléatoirement des bases de données à partir de RBs de référence sélectionnés depuis un dépôt [2] classique de RBs.

2. https ://www.bnlearn.com/bnrepository/

Les bases de données générées ont des tailles allant de 500 à 10000 observations. Pour les connaissances d'experts, nous avons d'abord sélectionné 50% des informations structurelles contenues dans le graphe original (présence et absence d'arêtes), puis nous avons fourni une partie de ces informations directement sous la forme de connaissances d'experts, alors que l'autre partie a été modifiée pour être fournie sous la forme de connaissances erronées, c'est-à-dire que les présences (resp. absences)

d'arêtes ont été introduites en tant qu'absences (resp. présences) d'arêtes, cela dans le but de tester la robustesse de notre algorithme en présence de connaissances erronées. Pour permettre la comparaison avec l'algorithme PCSDC, nous nous sommes restreints à un seul expert. De plus, pour toutes les expériences, nous avons fixé $P(\mathbb{I}) = 0.95$. Cela retranscrit le fait qu'en pratique, les RBs ont souvent des structures assez peu denses, c'est-à-dire qu'ils ont peu d'arcs, induisant ainsi beaucoup d'indépendances. Enfin, le seuil T a été fixé à 0.05.

PCSe, PCS et PCSDC sont comparés sur la base des squelettes qu'ils produisent, c'est-à-dire le graphe du RB appris sans les orientations. Pour cela, les squelettes obtenus sont comparés à celui du RB original utilisé pour générer la base de données. Nous avons utilisé le F-score pour la comparaison. Ce dernier est obtenu de la manière suivante : $(2 \times \text{Précision} \times \text{Rappel})/(\text{Precision} + \text{Recall})$, où le Rappel et la Précision sont deux métriques définies respectivement par les rapports $\text{TP}/(\text{TP} + \text{FN})$ et $\text{TP}/(\text{TP} + \text{FP})$, et TP, FP et FN correspondent respectivement aux nombres de vrais positifs, de faux positifs et de faux négatifs. Ainsi, la métrique du F-score fournit un bon aperçu de la fiabilité des algorithmes étudiés. Pour rendre la comparaison des résultats plus complète, nous avons fait varier les paramètres suivants :

— la taille de la base de données,
— le pourcentage d'assertions correctes venant de l'expert (que l'on appelle véracité ci-après),
— la confiance γ_k de l'expert pour donner une assertion correcte,
— le paramètre ρ_k pour la transformation de la confiance de l'expert (voir la figure 4 et l'équation (2)).

Enfin, pour chaque ensemble de paramètres ci-dessus, 100 itérations des expériences ont été menées et les tableaux 1 à 4 montrent les moyennes des F-scores obtenus sur ces 100 itérations.

Toutes les expériences ont été menées sur un ordinateur muni d'un processeur Intel Core i7 de 1.9GHz avec 16Go de mémoire RAM tournant sur Windows 10. Les algorithmes ont été implémentés en utilisant la bibliothèque pyAgrum [7].

Le tableau 1 montre le F-score moyen pour le RB Alarm, avec $\gamma_k = 0.8$ et $\rho_k = 14$, en fonction de la taille de la base de données et du pourcentage de véracité. Dans le cas de PCSDC, malgré $\gamma_k = 0.8$, toutes les connaissances de l'expert sont prises en compte, du fait que PCSDC considère systématiquement les assertions comme totalement vraies. Comme on peut l'observer, excepté lorsque la véracité est réellement égale à 100%, notre algorithme obtient toujours de meilleurs résultats que la méthode PCSDC, quelle que soit la taille de la base de données utilisée. Avec une taille de 1000 observations, le F-score de notre algorithme est même supérieur à celui de PCSDC de 24% à 69%. Et plus la base de données est grande, plus la différence entre les résultats des deux algorithmes augmente. Cela vient du fait que notre nouveau test d'indépendance est capable d'agréger l'information contenue dans les connaissances des experts avec celle contenue dans les données. À l'inverse, en contraignant l'algorithme à prendre en compte les connaissances d'expert sans leurs incertitudes, PCSDC devient forcément moins efficace lorsque l'expert fait trop d'erreurs. Les différences par rapport à PCS sont un peu moins prononcées que celles par rapport à PCSDC, mais nous pouvons toujours observer que les résultats de PCSe restent meilleurs que ceux de PCS lorsque le pourcentage de véracité est supérieur à 50%. Entre 25% et 50%, alors que le nombre d'erreurs devient important, l'algorithme PCSe parvient encore à produire des résultats supérieurs à ceux de PCS. Enfin, en dessous de 25%, le nombre d'erreurs devient trop important, ce qui a un impact négatif sur les résultats de PCSe, lesquels deviennent légèrement inférieurs à ceux de PCS. Ces observations montrent que l'algorithme PCSe permet d'améliorer les résultats grâce aux connaissances justes, tout en étant robuste aux erreurs. Les mêmes résultats peuvent être observés avec d'autres RBs de référence, comme le montre le tableau 2, dans lequel γ_k et ρ_k ont été fixés respectivement à 0.8 et 14.

La précision de notre algorithme peut également être démontrée en faisant varier les paramètres γ_k et ρ_k. Dans le tableau 3, nous avons fixé le pourcentage de véracité à 75% et ρ_k à 14. Nous avons fait varier γ_k puis calculé le F-score avec des bases de données de 1000 observations générées à partir du RB de référence Insurance. Comme nous pouvons le constater, les résultats de PCSe sont significativement meilleurs que ceux des autres algorithmes. Notons que, pour $\gamma_k = 0.5$, c'est-à-dire lorsque l'expert n'a aucune idée si son assertion est vraie ou fausse (ce qui correspond à une absence de connaissance), nous avons obligé PCSDC à ne pas prendre en compte les connaissances de l'expert. Pour PCSe, $\gamma_k = 0.5$ implique que $\epsilon_0 = \epsilon_1$, et donc que la connaissance de l'expert ne doit pas être prise en compte non plus. Au final, pour cette valeur de γ_k, aucune connaissance n'est fournie, ce qui explique le score identique pour les trois algorithmes. Notons aussi que, pour $\gamma_k = 0.8$, le F-score des algorithmes PCS et PCSe est plus bas que dans le tableau 2, pour le RB Insurance et un pourcentage de véracité de 75%, bien que les bases de données utilisées dans le tableau 3 soient deux fois plus petites que celles utilisées dans le tableau 2. Cela s'explique par le fait que les tests du G^2 ne sont effectués que lorsqu'il y a assez de données (voir lignes 15–18 de l'algorithme 1).

Enfin, dans le tableau 4, le pourcentage de véracité est fixé à 100%, avec des bases de données dont la taille a été fixée à 1000 observations générées à partir du RB Alarm. Plus la valeur de ρ_k est grande, plus l'opinion de l'expert est prise en compte. Ainsi, si ses assertions sont correctes, la précision atteinte par l'algorithme augmente. Pour $\gamma_k < 1$,

Taille	1000			5000			10000		
Véracité (%)	PCS	PCSDC	PCSe	PCS	PCSDC	PCSe	PCS	PCSDC	PCSe
0	**0.703**	0.000	0.695	**0.836**	0.000	0.822	**0.889**	0.000	0.873
25	0.703	0.174	**0.719**	0.836	0.174	**0.837**	**0.889**	0.174	0.881
50	0.703	0.347	**0.742**	0.836	0.347	**0.848**	0.889	0.347	**0.891**
75	0.703	0.521	**0.762**	0.836	0.521	**0.862**	0.889	0.521	**0.897**
100	0.703	**0.863**	0.779	0.836	**0.913**	0.859	0.889	**0.939**	0.909

TABLE 1 – F-scores de PCS, PCSDC et PCSe pour le RB Alarm en fonction du pourcentage de véracité.

	PCS	PCSDC	PCSe	PCS	PCSDC	PCSe	PCS	PCSDC	PCSe
Véracité (%)	Asia			Andes			Alarm		
0	**0.662**	0.000	0.591	**0.639**	0.000	0.599	**0.623**	0.000	0.615
25	**0.662**	0.166	0.624	0.639	0.164	**0.635**	0.623	0.166	**0.634**
50	0.662	0.399	**0.671**	0.639	0.326	**0.667**	0.623	0.333	**0.663**
75	0.662	0.674	**0.729**	0.639	0.487	**0.706**	0.623	0.503	**0.694**
100	0.662	**0.844**	0.778	0.639	**0.838**	0.752	0.623	**0.831**	0.736
Véracité (%)	Child			Hailfinder			Insurance		
0	**0.833**	0.000	0.768	**0.522**	0.000	0.495	**0.608**	0.000	0.602
25	**0.833**	0.199	0.785	**0.522**	0.117	0.499	**0.608**	0.177	0.634
50	**0.833**	0.350	0.802	0.522	0.233	**0.523**	0.608	0.329	**0.640**
75	0.833	0.503	**0.854**	0.522	0.358	**0.534**	0.608	0.506	**0.646**
100	0.833	**0.914**	0.877	0.522	**0.725**	0.540	0.608	**0.815**	0.664

TABLE 2 – F-scores de PCS, PCSDC et PCSe pour les RBs de référence avec des bases de données de 500 enregistrements en fonction du pourcentage de véracité.

γ_k	PCS	PCSDC	PCSe
0.5	**0.581**	**0.581**	**0.581**
0.6	0.581	0.513	**0.592**
0.7	0.581	0.513	**0.606**
0.8	0.581	0.513	**0.620**
0.9	0.581	0.513	**0.639**

TABLE 3 – F-scores de PCS, PCSDC and PCSe pour le RB Insurance en fonction de γ_k.

γ_k	$\rho_k = 5$	$\rho_k = 10$	$\rho_k = 15$	$\rho_k = 20$
0.6	0.711	0.718	0.726	0.738
0.7	0.720	0.739	0.757	0.774
0.8	0.736	0.760	0.783	0.802
0.9	0.755	0.782	0.805	0.822
1.0	0.873	0.873	0.873	0.873

TABLE 4 – F-scores de PCSe pour le RB Alarm en fonction de γ_k et ρ_k.

lorsque le paramètre de transformation de la confiance ρ_k augmente de 5 à 20, on observe une augmentation de 2% à 7% du F-score. Tous ces résultats montrent clairement la performance de notre approche, qui a été conçue pour fonctionner efficacement dans de nombreuses situations.

5 Conclusion

Dans cet article, nous proposons un nouvel algorithme d'apprentissage sous contraintes pour déterminer la structure de réseaux bayésiens en présence de connaissances incertaines d'experts. Cet algorithme fait intervenir les connaissances directement dans un nouveau test d'indépendance dont le fonctionnement est mathématiquement prouvé. Comme le montrent les expérimentations, notre approche permet une amélioration significative des performances d'apprentissage par rapport à d'autres algorithmes fonctionnant sous contraintes. Dans de futurs travaux, nous prévoyons d'étendre notre algorithme pour permettre l'apprentissage de réseaux bayésiens causaux.

Références

[1] Amirkhani, Hossein, Mohammad Rahmati, Peter J. F. Lucas et Arjen Hommersom: *Exploiting Experts' Knowledge for Structure Learning of Bayesian Networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11) :2154–2170, novembre 2017.

[2] Borboudakis, Giorgos et Ioannis Tsamardinos: *Incorporating Causal Prior Knowledge as Path-Constraints in Bayesian Networks and Maximal Ancestral Graphs*. Dans *Proceedings of the 29th International Conference on Machine Learning*, pages 427–434, Edinburgh, Scotland, juin 2012. `http://arxiv.org/abs/1206.6390`, visité le 2020-07-29.

[3] Campos, Luis M. de et Javier G. Castellano: *Bayesian network learning algorithm using structural restrictions*. International Journal of Approximate Reasoning, 45 :233–254, 2007.

[4] Chen, Eunice Yuh Jie, Yujia Shen, Arthur Choi et Adnan Darwiche: *Learning Bayesian networks with ancestral constraints*. Dans *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 2333–2341, 2016.

[5] Chickering, D.M.: *Learning Bayesian Networks is NP-Complete*. Dans Fisher D., Lenz HJ. (rédacteur) : *Learning from Data*, tome 112 de *Lecture Notes in Statistics*, pages 121–130. Springer, 1996.

[6] Colombo, Diego et Marloes H. Maathuis: *Order-Independent Constraint-Based Causal Structure Learning*. Journal of Machine Learning Research, 15 :3921–3962, 2014.

[7] Gonzales, Christophe, Lionel Torti et Pierre Henri Wuillemin: *aGrUM : A Graphical Universal Model Framework*. Dans *Proceedings of the 30th International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems*, pages 171–177, 2017.

[8] Keeney, Ralph L et Howard Raiffa: *Decisions with Multiple Objectives - Preferences and Value Tradeoffs*. Cambridge university Press, 1976.

[9] Meek, Christopher: *Causal inference and causal explanation with background knowledge*. Dans *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 403–410, 1995.

[10] Richardson, Matthew et Pedro Domingos: *Learning with Knowledge from Multiple Experts*. Proceedings of the 20th International Conference on Machine Learning, 2 :624–631, octobre 2003.

[11] Tversky, Amos et Daniel Kahneman: *Advances in Prospect Theory : Cumulative Representation of Uncertainty*. Journal of Risk and Uncertainty, 5(4) :297–323, 1992.

Actes JIAF 2021

Systèmes de preuve pour les logiques de "Bringing-it-About" *

Tiziano Dalmonte[1,2] **Charles Grellois**[2] **Nicola Olivetti**[2]

[1] TU Wien, Autriche
[2] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

tiziano@logic.at
charles.grellois@univ-amu.fr
nicola.olivetti@univ-amu.fr

Résumé

La logique dite de Bringing-it-About ("faire en sorte que") a été introduite par Elgesem en 1997 pour modéliser ce qu'un agent fait (brings about) et ce qu'un agent peut faire (can bring about). La logique contient deux familles de modalités indexées par les agents. Nous introduisons tout d'abord une nouvelle sémantique de voisinages, définie à l'aide de modèles dont les voisinages sont des couples : les deux composantes représentent intuitivement des conditions permettant et interdisant une action. Cette sémantique est particulièrement adaptée à la construction de contre-modèles, une tâche difficile si l'on utilise les sémantiques précédemment définies dans la littérature. Nous définissons ensuite un calcul d'hyperséquents pour cette logique, qui permet de donner une procédure de décision et permet la construction effective de contre-modèles.

Abstract

The logic of Bringing-it-About was introduced by Elgesem to formalise the notions of agency and capability. It contains two families of modalities indexed by agents, the first one expressing what an agent brings about (does), and the second expressing what she *can* bring about (can do). We first introduce a new neighbourhood semantics, defined in terms of bi-neighbourhood models for this logic, which is more suited for countermodel construction than the semantics defined in the literature. We then introduce a hypersequent calculus for this logic, which leads to a decision procedure allowing for a practical countermodel extraction. We finally extend both the semantics and the calculus to a coalitional version of Elgesem logic proposed by Troquard.

Cet article est une version abrégée, et traduite en français, d'un travail accepté à DEON 2020-2021 [3].

1 Introduction

La logique de "Bringing-It-About" (*faire en sorte que*) a été définie par Elgesem [6]. Elle propose une modélisation des actions d'un ensemble d'agents : un agent effectue une action (il *fait en sorte que* cette action ait lieu), ou il peut être en capacité de la faire (il *peut faire en sorte que* cette action ait lieu). Le système logique introduit par Elgesem contient ainsi deux

*Ce travail a été partiellement financé par le projet ANR TICAMORE ANR-16-CE91-0002-01. La recherche de Dalmonte à TU Wien est financée par une bourse Ernst Mach gérée par OeAD et financée par BMBWF.

modalités [1] $\mathbb{E}_i$ et $\mathbb{C}_i$, indexées par un agent i. La première exprime l'*effectivité*, c'est-à-dire le fait qu'un agent fait en sorte qu'un évènement se produise : $\mathbb{E}_{Lucie}$ *VirementBancaire* signifie ainsi que Lucie fait un virement bancaire. La seconde exprime la *capacité* à agir : $\mathbb{C}_{Lucie}$ *VirementBancaire* modélise le fait que Lucie *peut* effectuer un virement bancaire.

La logique d'Elgesem vise ainsi à capturer les notions d'effectivité (un agent fait en sorte qu'une action se produise) et de capacité (un agent peut faire en sorte qu'une action ait lieu), sans utiliser de notion temporelle ou de modélisation par des jeux. Cette logique définit ainsi un cadre logique succint pour étudier les notions d'effectivité et de capacité, cadre qui est devenu une référence, notamment par sa simplicité relativement à d'autres logiques, telles que par exemple les logiques STIT [2, 8]. La logique d'Elgesem permet naturellement d'exprimer des notions de *responsabilité* et de formaliser les notions de contrôle, de pouvoir, et de délégation. La phrase "Sarah empêche Lucie de faire un virement bancaire" se représente ainsi, très simplement, par la formule $\mathbb{E}_{sarah}\,\neg\mathbb{E}_{lucie}$ *VirementBancaire*. L'utilisation conjointe de modalités déontiques permet d'exprimer simplement, entre autres, la notion d'obligation (l'agent est obligé d'effectuer une action).

Elgesem a proposé une axiomatisation de sa logique et une sémantique (presque équivalente) basée sur des modèles de fonctions de sélection. Remarquons que la notion de capacité modélisée est assez faible, puisque ses seuls axiomes sont (i) que l'effectivité implique la capacité : $\mathbb{E}_i A \rightarrow \mathbb{C}_i A$ et (ii) $\neg\mathbb{C}_i\top$, ce qui exprime le fait qu'un agent i ne peut pas effectuer une action correspondant à une tautologie ; en miroir, Elgesem pose l'axiome $\neg\mathbb{E}_i\top$: un agent ne peut pas faire quelque chose qui se produira même sans lui, quel que soit son niveau d'investissement et de responsabilité relativement à cette action.

L'étude de la logique d'Elgesem a été poursuivie par Governatori et Rotolo [7]. Ils ont proposé une sémantique alternative utilisant des modèles de voisinages. Dans leur sémantique, les modèles contiennent deux fonctions de voisinage, une pour chacun des opérateurs $\mathbb{E}_i$ et $\mathbb{C}_i$; ces fonctions donnent (pour chaque agent i) les propositions, identifiées à leurs ensembles de vérité, que les agents font (effectivité) ou peuvent faire (capacité). Governatori et Rotolo prouvent également que la sémantique d'Elgesem valide l'axiome additionnel $\neg\mathbb{C}_i\bot$, qui signifie qu'un agent ne peut pas être en mesure de réaliser une contradiction.

Là où la sémantique de la logique d'Elgesem a été étudiée en détails, la théorie de la preuve de cette logique n'a été que très peu explorée : le seul système de preuve connu pour la logique d'Elgesem a été proposé par Lellmann [9]. Il n'y a, en particulier, aucun système de preuve à ce jour qui relie la syntaxe à la sémantique : aucun système de preuves ne permet de construire de contre-modèles pour les formules invalides.

Le but de cet article est précisément de combler cette lacune. Dans ce but, nous proposons une sémantique alternative de la logique d'Elgesem : nous considérons des modèles à voisinages doubles, une variante des modèles de voisinages donnés par Governatori et Rotolo [7]. Comme les modèles de Governatori et Rotolo, nos modèles contiennent, pour chaque agent i, deux fonctions de voisinage correspondant aux opérateurs $\mathbb{E}_i$ et $\mathbb{C}_i$. Mais ces fonctions diffèrent par le fait qu'elles assignent à chaque monde un ensemble de *couples* de voisinages (α,β). Intuitivement, un couple (α,β) s'interprète comme suit : étant donné une proposition A représentant le résultat d'une action d'un agent i, les deux composantes de la paire représentent (de façon indépendante) un ensemble de situations α permettant à i d'effectuer A, et de situations β empêchant i de réaliser A. Les paires (α,β) peuvent être vues comme des bornes inférieures et supérieures approximant les propositions qu'un agent effectue ou est en capacité d'effectuer. Il est à noter qu'un modèle à voisinages doubles peut être transformé en un modèle de voisinages standard, et que la transformation inverse est également possible. L'avantage de la sémantique à voisinages doubles est qu'elle permet de calculer des contre-modèles des formules invalides bien plus aisément, et directement, que la sémantique à voisinages usuelle car elle permet d'éviter le calcul exact des ensembles de vérité des formules.

Nous passons ensuite à la théorie de la preuve en proposant un calcul d'hyperséquents. On peut voir un hyperséquent comme une disjonction de

1. Ces notations ne sont pas celles utilisées originellement par Elgesem.

séquents usuels. La structure d'hyperséquent n'est pas strictement nécessaire à l'obtention d'un calcul complet (comme Lellmann le remarque dans [9]), mais elle permet la définition d'un calcul dans lequel les règles sont inversibles. Ceci n'est pas le cas dans le calcul de séquents de Lellmann [9]. Cette inversibilité des règles permet d'extraire d'*un seul* hyperséquent invalide, apparaissant comme feuille d'*un seul* arbre de dérivation, un contre-modèle de la formule considérée dans la sémantique à voisinages doubles. Cela implique que notre calcul d'hyperséquents ne donne pas seulement une procédure de décision pour la logique d'Elgesem : il offre également la première procédure pratique de construction de contre-modèles pour cette logique. Remarquons que la construction directe de contre-modèles à partir de séquents usuels est impossible : comme les règles ne sont pas inversibles, il ne suffit pas d'une dérivation qui échoue pour invalider la prouvabilité d'un séquent, mais il faut essayer et analyser toutes les dérivations possibles.

Le calcul d'hyperséquents a également de bonnes propriétés relativement à la théorie de la preuve. Une preuve syntaxique de l'élimination des coupures est possible, et implique la complétude vis-à-vis de l'axiomatisation.

2 Sémantique à voisinages doubles pour la logique d'Elgesem

Dans cette section, nous présentons tout d'abord la logique d'effectivité et de capacité d'Elgesem, que nous notons **ELG**. Nous définissons ensuite des modèles à voisinages doubles pour cette logique.

Soit $\mathcal{A} = \{a, b, c, ...\}$ un ensemble d'agents, la logique **ELG** est alors définie à partir d'un langage propositionnel $\mathcal{L}_{Elg}$ contenant, pour chaque $i \in \mathcal{A}$, deux modalités unaires $\mathbb{E}_i$ et $\mathbb{C}_i$, appelées respectivement modalité d'*effectivité* et de *capacité*. Les formules de $\mathcal{L}_{Elg}$ sont définies par la grammaire suivante :

$$A \quad := \quad p \mid \bot \mid \top \mid \neg A \mid A \wedge B \mid A \vee B \mid A \rightarrow B \mid \mathbb{E}_i A \mid \mathbb{C}_i A$$

où $\mathbb{E}_i A$ et $\mathbb{C}_i A$ signifient respectivement que "l'agent *i* fait en sorte que *A* survienne", et "l'agent *i* est en capacité de réaliser l'action *A*".

La logique **ELG** est définie en étendant la logique classique propositionnelle (formulée dans le langage $\mathcal{L}_{Elg}$) avec les axiomes modaux et les règles donnés en Figure 1. [2]

Remarquons que $\neg\mathbb{E}_i\bot$ et $\neg\mathbb{E}_i\top$ sont dérivables dans la logique **ELG**. En revanche, les axiomes C et T ne valent que pour la modalité $\mathbb{E}$. Ceci signifie, pour le premier axiome, que si un agent réalise deux actions, alors il peut réaliser les deux à la fois ; le second axiome exprime le fait que si un agent effectue *A*, alors *A* est vrai. Observons que la condition C ne s'applique pas à l'opérateur de capacité $\mathbb{C}$: en effet, si l'agent peut faire *A* et *B*, on ne peut en déduire qu'il peut faire à la fois *A* et *B*. Par exemple, avec 1€50, je suis en capacité d'acheter un café ou un croissant, mais pas les deux.

Des caractérisations sémantiques de la logique **ELG** existent. Elgesem [6] en donne une à l'aide de fonctions de sélection, et Governatori et Rotolo [7] utilisent pour leur part des modèles de voisinages, en utilisant deux fonctions de voisinages différentes pour les modalités $\mathbb{E}$ et $\mathbb{C}$. Nous proposons ici une sémantique alternative, utilisant les modèles à voisinages doubles [5]. Après avoir défini cette sémantique alternative, nous expliquerons ses avantages pour l'extraction de contre-modèles.

Définition 1 *Un* modèle à voisinages doubles *pour* **ELG** *est un quadruplet* $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}_i^{\mathbb{E}}, \mathcal{N}_i^{\mathbb{C}}, \mathcal{V} \rangle$*, où* $\mathcal{W}$ *est un ensemble non vide,* $\mathcal{V}$ *est une fonction de valuation, et, pour chaque agent i,* $\mathcal{N}_i^{\mathbb{E}}$ *et* $\mathcal{N}_i^{\mathbb{C}}$ *sont deux fonctions à voisinages doubles* $\mathcal{W} \longrightarrow \mathcal{P}(\mathcal{P}(\mathcal{W}) \times \mathcal{P}(\mathcal{W}))$ *satisfaisant les conditions suivantes :*

($\mathrm{C}_{\mathbb{E}}$)	*Si* $(\alpha, \beta), (\gamma, \delta) \in \mathcal{N}_i^{\mathbb{E}}(w)$, *alors* $(\alpha \cap \gamma, \beta \cup \delta) \in \mathcal{N}_i^{\mathbb{E}}(w)$.
($\mathrm{T}_{\mathbb{E}}$)	*Si* $(\alpha, \beta) \in \mathcal{N}_i^{\mathbb{E}}(w)$, *alors* $w \in \alpha$.
($\mathrm{Q}_{\mathbb{C}}$)	*Si* $(\alpha, \beta) \in \mathcal{N}_i^{\mathbb{C}}(w)$, *alors* $\beta \neq \emptyset$.
($\mathrm{P}_{\mathbb{C}}$)	*Si* $(\alpha, \beta) \in \mathcal{N}_i^{\mathbb{C}}(w)$, *alors* $\alpha \neq \emptyset$.
($\mathrm{Int}_{\mathbb{EC}}$)	$\mathcal{N}_i^{\mathbb{E}}(w) \subseteq \mathcal{N}_i^{\mathbb{C}}(w)$.

2. Une variante de la logique d'Elgesem ne contenant pas l'axiome $\mathrm{P}_{\mathbb{C}}$ est étudiée dans [7, 9]. Tous les résultats présentés dans cet article peuvent être étendus à cette variante en ôtant la règle correspondant à cet axiome dans le calcul, et la condition sémantique correspondante dans les modèles à voisinages doubles.

$\mathrm{RE}_{\mathbb{E}}$	$\dfrac{A \leftrightarrow B}{\mathbb{E}_i A \leftrightarrow \mathbb{E}_i B}$	$\mathrm{RE}_{\mathbb{C}}$	$\dfrac{A \leftrightarrow B}{\mathbb{C}_i A \leftrightarrow \mathbb{C}_i B}$
$\mathrm{C}_{\mathbb{E}}$	$\mathbb{E}_i A \wedge \mathbb{E}_i B \rightarrow \mathbb{E}_i(A \wedge B)$	$\mathrm{Q}_{\mathbb{C}}$	$\neg\mathbb{C}_i \top$
$\mathrm{T}_{\mathbb{E}}$	$\mathbb{E}_i A \rightarrow A$	$\mathrm{P}_{\mathbb{C}}$	$\neg\mathbb{C}_i \bot$
$\mathrm{Int}_{\mathbb{EC}}$	$\mathbb{E}_i A \rightarrow \mathbb{C}_i A$		

FIGURE 1 – Axiomes modaux et règles de la logique d'Elgesem **ELG**.

La relation de forcing $\Vdash$ *est définie de la manière usuelle pour les formules atomiques et les connecteurs booléens. Pour les formules dont le connecteur principal est* $\mathbb{E}$ *ou* $\mathbb{C}$*, le forcing est définit comme suit :*

$$\begin{gathered}\mathcal{M}, w \Vdash \mathbb{E}_i A \\ \textit{ssi} \\ \textit{il existe } (\alpha,\beta) \in \mathcal{N}_i^{\mathbb{E}}(w) \textit{ tel que} \\ \textit{pour tout } v \in \alpha,\ \mathcal{M}, v \Vdash A, \\ \textit{et pour tout } u \in \beta,\ \mathcal{M}, u \nVdash A.\end{gathered}$$

$$\begin{gathered}\mathcal{M}, w \Vdash \mathbb{C}_i A \\ \textit{ssi} \\ \textit{il existe } (\alpha,\beta) \in \mathcal{N}_i^{\mathbb{C}}(w) \textit{ tel que} \\ \textit{pour tout } v \in \alpha,\ \mathcal{M}, v \Vdash A, \\ \textit{et pour tout } u \in \beta,\ \mathcal{M}, u \nVdash A.\end{gathered}$$

Notons $[\![A]\!]$ l'ensemble $\{v \mid \mathcal{M}, v \Vdash A\}$, alors les clauses ci-dessus peuvent être réécrites comme suit : $\mathcal{M}, w \Vdash \mathbb{E}_i A$ si et seulement s'il existe $(\alpha,\beta) \in \mathcal{N}_i^{\mathbb{E}}(w)$ tel que $\alpha \subseteq [\![A]\!]$ et $\beta \subseteq [\![\neg A]\!]$. De même pour les formules dont le connecteur principal est $\mathbb{C}$. Nous suivons la convention usuelle, et n'indiquons pas le modèle $\mathcal{M}$ lorsque le contexte suffit à le retrouver. Nous écrivons alors simplement $w \Vdash A$.

La principale raison d'être de la sémantique à voisinages doubles est la simplicité qu'elle offre pour extraire des contre-modèles à partir de preuves qui échouent. Dans la sémantique de voisinages usuelle, l'ensemble des mondes où A est vraie doit appartenir exactement à $\mathcal{N}_i^{\mathbb{E}}(w)$ pour que w satisfasse $\mathbb{E}_i A$. En revanche, dans la sémantique à voisinages doubles, il suffit de trouver un couple (α,β) tel que $\alpha \subseteq [\![A]\!]$ et que $\beta \subseteq [\![\neg A]\!]$. Le point crucial est donc le suivant : pour trouver des contre-modèles dans la sémantique à voisinages doubles, il n'est pas nécessaire de calculer exactement les ensembles de mondes où les formules sont vraies, ce qui est délicat si l'on s'appuie seulement sur l'information *partielle* que donne une preuve qui a échoué.

De plus, la sémantique à voisinages doubles permet également d'expliquer l'effectivité et la capacité en termes de conditions permettant la réalisation d'actions, ou les empêchant. Ces conditions sont représentées respectivement par les éléments α et β des couples apparaissant dans la sémantique à voisinages doubles. Les conditions ($\mathrm{P}_{\mathbb{C}}$) et ($\mathrm{Q}_{\mathbb{C}}$), i.e. $\alpha \neq \emptyset$ et $\beta \neq \emptyset$, correspondent respectivement au fait qu'il n'est pas possible de réaliser une contradiction, ni une tautologie : il doit toujours y avoir un moyen de réaliser ou d'empêcher une action. Remarquons aussi que la validité de $\neg\mathbb{E}_i\top$ et de l'axiome $\mathrm{T}_{\mathbb{E}}$ impliquent que les formules de la forme $\mathbb{E}_i A$ ne sont jamais valides dans les modèles de la logique **ELG** : ceci caractérise sémantiquement le fait qu'une action doit toujours pouvoir être évitée.

Théorème 1 (Caractérisation) *A est dérivable dans* **ELG** *si et seulement si elle est valide dans tous les modèles à voisinages doubles de la logique* **ELG**.

Un modèle à voisinages doubles pour **ELG** peut être transformé en modèle de voisinages pour cette même logique en suivant la transformation décrite dans [4, 5].

Proposition 1 (Transformation de modèles) *Soit $\mathcal{M}_{bi} = \langle \mathcal{W}, \mathcal{N}_{bi}, \mathcal{V}\rangle$ un modèle à voisinages doubles de* **ELG***, et $\mathcal{M}_n = \langle \mathcal{W}, \mathcal{N}_n, \mathcal{V}\rangle$ le modèle de voisinages défini en considérant les mêmes $\mathcal{W}$ et $\mathcal{V}$ et en posant, pour tout $w \in \mathcal{W}$,*

$$\begin{aligned}\mathcal{N}_n(w) = \{\gamma \subseteq \mathcal{W} \mid &\text{il existe } (\alpha,\beta) \in \\ &\mathcal{N}_b(w) \text{ tel que } \alpha \subseteq \gamma \subseteq \mathcal{W} \setminus \beta\}.\end{aligned}$$

Alors, pour tout $A \in \mathcal{L}$ et tout $w \in \mathcal{W}$, $\mathcal{M}_n, w \Vdash A$ si et seulement si $\mathcal{M}_{bi}, w \Vdash A$.

Ainsi que le montre la transformation ci-dessus, les modèles à voisinages doubles sont en général décrits par des fonctions plus petites que celles obtenues pour les modèles de voisinages correspondants. Cela provient du fait que toute paire de voisinages (α, β) – dont on peut considérer les éléments comme des bornes inférieures et supérieures de voisinages – peuvent valider plus d'une formule modale.

3 Calcul d'hyperséquents

Nous nous intéressons maintenant à la théorie de la preuve pour la logique **ELG**. A notre connaissance, la seule étude de la logique d'Elgesem considérant la théorie de la preuve a été menée par Lellman [9], qui définit un calcul de séquents sans coupures. Nous définissons ici un calcul d'hyperséquents [1] en suivant la méthodologie proposée dans [4], qui permet la définition de calculs pour les logiques modales non-normales. Notre calcul contient, en plus de la structure d'hyperséquents, des *blocs* qui permettent de collecter les formules dont les connecteurs principaux sont $\mathbb{E}$ et $\mathbb{C}$.

L'utilisation combinée d'hyperséquents et de blocs est essentiellement motivée, comme nous allons le voir dans la suite, par le fait que cela permet de construire aisément des contre-modèles pour les formules non dérivables. Les blocs permettent également d'encoder simplement (et analytiquement) les relations entre les modalités $\mathbb{E}$ et $\mathbb{C}$. Commençons par quelques définitions.

Définition 2 ***(Bloc, séquent, hyperséquent)*** *Un* bloc *est une structure $\langle\Sigma\rangle_i^{\mathbb{E}}$ ou $\langle\Sigma\rangle_i^{\mathbb{C}}$, où i est un agent, et Σ un multiensemble de formules de $\mathcal{L}_{Elg}$. Un* séquent *est un couple $\Gamma \Rightarrow \Delta$, où Γ est un multiensemble de formules et de blocs, et Δ est un multiensemble de formules. Nous notons* $\mathsf{set}(\Gamma)$ *le* support *du multiensemble Γ, i.e. l'ensemble de ses éléments sans multiplicités. Un* hyperséquent *est un multiensemble de séquents $S_1, ..., S_n$, noté $S_1 \mid ... \mid S_n$: nous appelons $S_1, ..., S_n$ les* composants *de l'hyperséquent.*

Définition 3 ***(Interprétation des formules)*** *Les séquents sont interprétés comme des formules de la logique* **ELG** *comme suit :*

$$i(A_1, ..., A_n, \langle\Sigma_1\rangle_{a_1}^{\mathbb{E}}, ..., \langle\Sigma_m\rangle_{a_m}^{\mathbb{E}}, \langle\Pi_1\rangle_{b_1}^{\mathbb{C}}, ..., \langle\Pi_k\rangle_{b_k}^{\mathbb{C}} \Rightarrow B_1, ..., B_\ell)$$
$$=$$
$$\bigwedge_{i\leq n} A_i \wedge \bigwedge_{j\leq m} \mathbb{E}_{a_j} \bigwedge \Sigma_j \wedge \bigwedge_{s\leq k} \mathbb{C}_{a_s} \bigwedge \Pi_s \rightarrow \bigvee_{t\leq \ell} B_t.$$

Définition 4 ***(Interprétation sémantique)*** *Un séquent S est* valide *dans un modèle à voisinages doubles $\mathcal{M}$, ce que l'on note $\mathcal{M} \models S$, si pour tout $w \in \mathcal{M}$, $\mathcal{M}, w \Vdash i(S)$. Un hyperséquent H est* valide *dans $\mathcal{M}$, ce que l'on note $\mathcal{M} \models H$, si $\mathcal{M} \models S$ pour un séquent $S \in H$.*

Les règles du calcul d'hyperséquents $\mathbf{HS_{ELG}}$ sont données Fig. 2. Nous les exprimons de façon cumulative : les formules principales ou les blocs sont recopiés dans les prémisses. On peut ainsi extraire un contre-modèle à partir d'un seul séquent saturé. Les règles propositionnelles sont simplement une traduction en hyperséquents des règles ordinaires du calcul des séquents. Nous omettons les règles pour $\neg$, $\vee$, et $\rightarrow$, qui sont définies de façon standard. Suivant la coutume, les séquents initiaux init sont restreints aux variables propositionnelles. On peut cependant aisément constater que $G \mid A, \Gamma \Rightarrow \Delta, A$ est dérivable pour tout A. Comme pour les connecteurs propositionnels, les formules dont le connecteur principal est $\mathbb{E}$ ou $\mathbb{C}$ sont introduites à gauche ou à droite par des règles différentes. Les règles $\mathsf{R}_{\mathbb{E}}$ et $\mathsf{R}_{\mathbb{C}}$ ont de multiples prémisses, mais leur nombre est contraint par la cardinalité des blocs principaux $\langle\Sigma\rangle_i^{\mathbb{E}}$ et $\langle\Sigma\rangle_i^{\mathbb{C}}$. Il y a dans le calcul, pour tout axiome de **ELG**, une règle qui lui correspond. Les blocs ont un rôle crucial dans la définition des règles modales. Remarquons en particulier que les blocs pour $\mathbb{E}$ peuvent être fusionnés par la règle $\mathsf{C}_{\mathbb{E}}$, mais qu'il n'y a pas de règle similaire pour les blocs pour $\mathbb{C}$. Notons cependant qu'il est possible, après la création de blocs potentiellement complexes pour $\mathbb{E}$, de les convertir en blocs pour $\mathbb{C}$. Nous donnons un exemple de dérivation dans le calcul $\mathbf{HS_{ELG}}$ en Figure 3.

Proposition 2 (Adéquation) *Si H est dérivable dans le calcul $\mathbf{HS_{ELG}}$, alors il est valide dans tous les modèles à voisinages doubles de* **ELG**.

Nous étudions maintenant les propriétés structurelles de notre calcul, et montrons en particulier qu'il est complet pour l'axiomatisation. Il est à remarquer que la preuve de complétude est purement

$$\mathsf{init}\ \frac{}{G \mid \Gamma, p \Rightarrow p, \Delta} \qquad \mathsf{L}\bot\ \frac{}{G \mid \Gamma, \bot \Rightarrow \Delta} \qquad \mathsf{R}\top\ \frac{}{G \mid \Gamma \Rightarrow \top, \Delta}$$

$$\mathsf{L}\wedge\ \frac{G \mid \Gamma, A \wedge B, A, B \Rightarrow \Delta}{G \mid \Gamma, A \wedge B \Rightarrow \Delta} \qquad \mathsf{R}\wedge\ \frac{G \mid \Gamma \Rightarrow A, A \wedge B, \Delta \qquad G \mid \Gamma \Rightarrow B, A \wedge B, \Delta}{G \mid \Gamma \Rightarrow A \wedge B, \Delta}$$

$$\mathsf{L}_{\mathbb{E}}\ \frac{G \mid \Gamma, \mathbb{E}_i A, \langle A \rangle_i^{\mathbb{E}} \Rightarrow \Delta}{G \mid \Gamma, \mathbb{E}_i A \Rightarrow \Delta} \qquad \mathsf{L}_{\mathbb{C}}\ \frac{G \mid \Gamma, \mathbb{C}_i A, \langle A \rangle_i^{\mathbb{C}} \Rightarrow \Delta}{G \mid \Gamma, \mathbb{C}_i A \Rightarrow \Delta}$$

$$\mathsf{Int}_{\mathbb{EC}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}}, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \Delta}$$

$$\mathsf{R}_{\mathbb{E}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \mathbb{E}_i A, \Delta \mid \Sigma \Rightarrow A \qquad \{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \mathbb{E}_i A, \Delta \mid A \Rightarrow B\}_{B \in \Sigma}}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \mathbb{E}_i A, \Delta}$$

$$\mathsf{R}_{\mathbb{C}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A, \Delta \mid \Sigma \Rightarrow A \qquad \{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A, \Delta \mid A \Rightarrow B\}_{B \in \Sigma}}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A, \Delta}$$

$$\mathsf{C}_{\mathbb{E}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}}, \langle \Pi \rangle_i^{\mathbb{E}}, \langle \Sigma, \Pi \rangle_i^{\mathbb{E}} \Rightarrow \Delta}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}}, \langle \Pi \rangle_i^{\mathbb{E}} \Rightarrow \Delta} \qquad \mathsf{T}_{\mathbb{E}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}}, \Sigma \Rightarrow \Delta}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \Delta}$$

$$\mathsf{Q}_{\mathbb{C}}\ \frac{\{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta \mid \Rightarrow B\}_{B \in \Sigma}}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta} \qquad \mathsf{P}_{\mathbb{C}}\ \frac{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta \mid \Sigma \Rightarrow}{G \mid \Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta}$$

FIGURE 2 – Le calcul d'hyperséquents $\mathbf{HS_{ELG}}$.

$$\dfrac{\dfrac{\dfrac{\mathbb{E}_i A, \langle A \rangle_i^{\mathbb{E}}, \langle A \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A \mid A \Rightarrow A \qquad \mathbb{E}_i A, \langle A \rangle_i^{\mathbb{E}}, \langle A \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A \mid A \Rightarrow A}{\mathbb{E}_i A, \langle A \rangle_i^{\mathbb{E}}, \langle A \rangle_i^{\mathbb{C}} \Rightarrow \mathbb{C}_i A}\,\mathsf{R}_{\mathbb{C}}}{\mathbb{E}_i A, \langle A \rangle_i^{\mathbb{E}} \Rightarrow \mathbb{C}_i A}\,\mathsf{Int}_{\mathbb{EC}}}{\mathbb{E}_i A \Rightarrow \mathbb{C}_i A}\,\mathsf{L}_{\mathbb{E}}$$

FIGURE 3 – Un exemple de dérivation dans $\mathbf{HS_{ELG}}$.

syntaxique, et qu'elle est à ce titre indépendante du choix d'une sémantique spécifique. Cette preuve requiert, comme c'est usuellement le cas, l'admissibilité de la règle de coupure, que nous définissons ainsi :

$$\mathsf{cut}\ \frac{G \mid \Gamma \Rightarrow \Delta, A \qquad G \mid A, \Gamma \Rightarrow \Delta}{G \mid \Gamma \Rightarrow \Delta}$$

Ceci signifie que lorsque les prémisses de cut sont dérivables, sa conclusion l'est également. L'admissibilité de la coupure dépend à son tour de l'admissibilité des règles structurelles d'affaiblissement et de contraction. Pour ce calcul d'hyperséquents, on peut les formuler de façon interne ou externe, comme suit :

Proposition 3 (Règles structurelles) *Soit ϕ une formule A ou un bloc $\langle \Sigma \rangle_i^{\mathbb{E}}$ ou $\langle \Sigma \rangle_i^{\mathbb{C}}$. Alors les règles suivantes sont admissibles dans* $\mathbf{HS_{ELG}}$ *:*

$$\mathsf{Lwk}\ \frac{G \mid \Gamma \Rightarrow \Delta}{G \mid \phi, \Gamma \Rightarrow \Delta} \qquad \mathsf{Lctr}\ \frac{G \mid \phi, \phi, \Gamma \Rightarrow \Delta}{G \mid \phi, \Gamma \Rightarrow \Delta}$$

$$\mathsf{Bctr}\ \frac{G \mid \langle A, A, \Sigma \rangle, \Gamma \Rightarrow \Delta}{G \mid \langle A, \Sigma \rangle, \Gamma \Rightarrow \Delta}$$

$$\mathsf{Rwk}\ \frac{G \mid \Gamma \Rightarrow \Delta}{G \mid \Gamma \Rightarrow \Delta, A} \qquad \mathsf{Rctr}\ \frac{G \mid \Gamma \Rightarrow \Delta, A, A}{G \mid \Gamma \Rightarrow \Delta, A}$$

$$\mathsf{Ewk}\ \frac{G}{G \mid \Gamma \Rightarrow \Delta} \qquad \mathsf{Ectr}\ \frac{G \mid \Gamma \Rightarrow \Delta \mid \Gamma \Rightarrow \Delta}{G \mid \Gamma \Rightarrow \Delta}$$

Remarquons qu'une conséquence immédiate de l'admissibilité de l'affaiblissement est que toutes les règles sont inversibles : dès que la conclusion d'une règle est dérivable ses prémisses le sont

aussi. Ceci est crucial : lorsqu'une formule est dérivable, nous obtenons ainsi une dérivation dans laquelle l'ordre d'application des règles n'importe pas (voir Sec. 4).

Théorème 2 (Elimination des coupures) *La règle* cut *est admissible dans le calcul* $\mathbf{HS_{ELG}}$.

Ce théorème d'admissibilité des coupures a pour conséquence le théorème de complétude suivant :

Théorème 3 (Complétude axiomatique) *Si A est dérivable dans* **ELG***, alors* $\Rightarrow A$ *est dérivable dans* $\mathbf{HS_{ELG}}$.

Comme remarqué précédemment, les hyperséquents ne sont pas nécessaires dans les dérivations : on peut en particulier prouver qu'un hyperséquent est dérivable dans $\mathbf{HS_{ELG}}$ si et seulement si l'un des séquents qui le composent est dérivable. Cependant, cette utilisation des hyperséquents permet d'obtenir un calcul dans lequel toutes les règles sont inversibles, et donc dans lequel l'ordre d'application des règles n'importe pas : modulo l'ordre d'application des règles, chaque formule a une *unique* dérivation, ou une *unique* preuve qui échoue : il n'est pas nécessaire de faire du *backtracking*. De plus, les hyperséquents sont au cœur du calcul direct de contre-modèles, calcul que l'on peut ainsi effectuer à partir d'un unique hyperséquent non-prouvable apparaissant comme feuille dans une dérivation qui échoue. La prochaine section détaille ces considérations.

4 Recherche de preuves et extraction de contre-modèles

Nous définissons dans cette section une procédure vérifiant la validité (ou la dérivabilité) d'une formule dans la logique d'Elgesem en utilisant le calcul d'hyperséquents introduit dans la section précédente. La procédure utilise une stratégie de recherche de preuve simple, qui commence toujours par la racine. Nous montrons que la stratégie termine toujours, et construisons une dérivation pour toute formule valide. Pour toute formule faisant échouer le processus de recherche, nous montrons qu'il est possible d'extraire directement un contre-modèle pour celle-ci. La stratégie s'appuie sur la notion de saturation suivante. Intuitivement, un hyperséquent est saturé lorsque l'application "à rebours" de toute règle à celui-ci (c'est-à-dire dans laquelle il est en position de conclusion) n'ajoute pas d'information, au sens où l'une des prémisses de la règle est déjà incluse dans l'hyperséquent.

Définition 5 ***(Hyperséquent saturé)*** *Soit* $H = \Gamma_1 \Rightarrow \Delta_1 \mid ... \mid \Gamma_k \Rightarrow \Delta_k$ *un hyperséquent apparaissant dans une preuve de* H'*. Les conditions de saturation associées à chaque application d'une règle de* $\mathbf{HS_{ELG}}$ *sont les suivantes :*

- *Non-prouvabilité :* (init) $\Gamma_n \cap \Delta_n = \emptyset$. ($\bot_{\mathsf{L}}$) $\bot \notin \Gamma_n$. ($\top_{\mathsf{R}}$) $\top \notin \Delta_n$.
- *Règles propositionnelles :* ($\wedge_{\mathsf{L}}$) *Si* $A \wedge B \in \Gamma_n$*, alors* $A \in \Gamma_n$ *et* $B \in \Gamma_n$. ($\wedge_{\mathsf{R}}$) *Si* $A \wedge B \in \Delta_n$*, alors* $A \in \Delta_n$ *ou* $B \in \Delta_n$*. De même pour les règles pour* $\neg$, $\vee$, $\rightarrow$.
- *Règles modales :* ($\mathsf{L}_{\mathbb{E}}$) *Si* $\mathbb{E}_i A \in \Gamma_n$*, alors* $\langle A \rangle_i^{\mathbb{E}} \in \Gamma_n$. ($\mathsf{R}_{\mathbb{E}}$) *Si* $\Gamma, \langle \Sigma \rangle_i^{\mathbb{E}} \Rightarrow \mathbb{E}_i B, \Delta$ *est dans H, alors il y a* $\Gamma', \Sigma \Rightarrow B, \Delta'$ *dans H ou il y a* $\Gamma', B \Rightarrow A, \Delta'$ *dans H pour un* $A \in \Sigma$. ($\mathsf{L}_{\mathbb{C}}$) *et* ($\mathsf{R}_{\mathbb{C}}$) *sont analogues.* ($\mathsf{C}_{\mathbb{E}}$) *Si* $\langle \Sigma \rangle_i^{\mathbb{E}}, \langle \Pi \rangle_i^{\mathbb{E}} \in \Gamma_n$*, alors il y a* $\langle \Omega \rangle_i^{\mathbb{E}} \in \Gamma_n$ *tel que* $\mathsf{set}(\Sigma, \Pi) = \mathsf{set}(\Omega)$. ($\mathsf{T}_{\mathbb{E}}$) *Si* $\langle \Sigma \rangle_i^{\mathbb{E}} \in \Gamma_n$*, alors* $\mathsf{set}(\Sigma) \subseteq \Gamma_n$. ($\mathsf{Q}_{\mathbb{C}}$) *Si* $\Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta$ *est dans H, alors il y a* $\Gamma' \Rightarrow B, \Delta'$ *dans H pour un* $B \in \Sigma$. ($\mathsf{P}_{\mathbb{C}}$) *Si* $\Gamma, \langle \Sigma \rangle_i^{\mathbb{C}} \Rightarrow \Delta$ *est dans H, alors il y a* $\Gamma' \Rightarrow \Delta'$ *dans H tel que* $\mathsf{set}(\Sigma) \subseteq \Gamma'$. *(*$\mathsf{Int}_{\mathbb{E}\mathbb{C}}$*) Si* $\langle \Sigma \rangle_i^{\mathbb{E}} \in \Gamma_n$*, alors il y a* $\langle \Omega \rangle_i^{\mathbb{C}} \in \Gamma_n$ *tel que* $\mathsf{set}(\Sigma) = \mathsf{set}(\Omega)$.

Nous disons que H est saturé relativement à l'application d'une règle R s'il satisfait la condition de saturation (*R*) *correspondant à l'application de cette règle. Un hyperséquent H est saturé relativement à* $\mathbf{HS_{ELG}}$ *s'il est saturé relativement à toute application possible d'une règle du calcul* $\mathbf{HS_{ELG}}$.

La *stratégie de recherche de preuve* est simple : (i) n'appliquer aucune règle aux séquents initiaux, et (ii) ne pas appliquer une règle donnée à un hyperséquent qui est déjà saturé par rapport à l'application de cette règle.

Le coeur de la stratégie est le suivant : éviter les applications de règles qui n'apportent pas d'information additionnelle aux hyperséquents. Nous démontrons que cette stratégie donne un algorithme de recherche de preuve qui termine.

Proposition 4 (Terminaison de la stratégie) *Toute branche d'une preuve d'un hyperséquent*

H obtenue par application de la stratégie définie ci-dessus est finie. Il suit que la procédure de recherche de preuve pour H termine toujours. De plus, toute branche termine par un hyperséquent initial ou par un hyperséquent saturé.

Les hyperséquents apparaissant dans une preuve de H peuvent être exponentiellement plus grands que H. Cela est dû à la présence de la règle $\mathsf{C}_{\mathbb{E}}$ qui, étant donné n formules $\mathbb{E}_i A_1, \ldots, \mathbb{E}_i A_n$, permet de construire un bloc pour chaque sous-ensemble de $\{A_1, ..., A_n\}$. Ainsi, notre procédure de décision n'atteint pas la borne de complexité PSPACE établie pour la logique d'Elgesem par Schröder et Pattinson [10] et Troquard [11].

Un calcul optimal pourrait être obtenu soit en considérant le calcul des séquents de [9], ou (dans l'esprit de [4]) en reformulant les règles de la figure 2 afin que les formules principales ne soient *pas recopiées* dans les prémisses. Ceci aboutirait cependant à la perte de l'inversibilité des règles, et donc à la perte de la possibilité d'extraire directement des contre-modèles à partir de preuves qui échouent. Cette situation est analogue à celle de la logique modale **K** : pour cette logique, on peut obtenir une borne de complexité PSPACE via un calcul de séquents approprié, mais on ne peut pas respecter cette borne avec un calcul dont les règles sont inversibles et permettent à ce titre l'extraction directe et aisée de contre-modèles pour les formules qui ne sont pas valides. Comme pour **K**, nous conjecturons que le calcul d'un contre-modèle peut avoir une complexité exponentielle, car un contre-modèle peut avoir lui-même une taille exponentielle (par rapport à la taille de la formule). Par conséquent, il faut choisir entre un calcul optimal qui vérifie la validité d'une formule et un calcul non-optimal qui permet la construction explicite de contre-modèles.

Nous montrons maintenant comment un contre-modèle peut être construit directement dans la sémantique à voisinages doubles à partir d'un hyperséquent saturé.

Définition 6 ***(Construction de contre-modèle)*** *Soit H un hyperséquent saturé apparaissant dans une preuve de H′. Fixons aussi une énumération e* : $\mathbb{N} \longrightarrow$ *H des composants de H. Pour e fixé, nous écrivons H sous la forme* $\Gamma_1 \Rightarrow \Delta_1 \mid ... \mid \Gamma_k \Rightarrow \Delta_k$. *Le modèle* $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}, \mathcal{V} \rangle$ *est défini comme suit :*

- $\mathcal{W} = \{n \mid \Gamma_n \Rightarrow \Delta_n \in H\}$.
- $\mathcal{V}(p) = \{n \mid p \in \Gamma_n\}$.
- *Pour tout bloc* $\langle\Sigma\rangle_i^{\mathbb{E}}$ *ou* $\langle\Sigma\rangle_i^{\mathbb{C}}$ *apparaissant dans un composant* $\Gamma_m \Rightarrow \Delta_m$ *de H,* $\Sigma^+ = \{n \in \mathcal{W} \mid \mathsf{set}(\Sigma) \subseteq \Gamma_n\}$ *et* $\Sigma^- = \{n \in \mathcal{W} \mid \Sigma \cap \Delta_n \neq \emptyset\}$.
- *Pour tout* $i \in \mathcal{A}$ *et tout* $n \in \mathcal{W}$, $\mathcal{N}_i^{\mathbb{E}}(n) = \{(\Sigma^+, \Sigma^-) \mid \langle\Sigma\rangle_i^{\mathbb{E}} \in \Gamma_n\}$ *and* $\mathcal{N}_i^{\mathbb{C}}(n) = \{(\Sigma^+, \Sigma^-) \mid \langle\Sigma\rangle_i^{\mathbb{C}} \in \Gamma_n\}$.

Lemme 1 *Considérons le modèle* $\mathcal{M}$ *défini comme dans la définition 6. Alors pour tout A,* $\langle\Sigma\rangle_i^{\mathbb{E}}$, $\langle\Pi\rangle_j^{\mathbb{C}}$ *et tout* $n \in \mathcal{W}$, *nous avons : Si* $A \in \Gamma_n$, *alors* $n \Vdash A$ *; si* $\langle\Sigma\rangle_i^{\mathbb{E}} \in \Gamma_n$, *alors* $n \Vdash \mathbb{E}_i \bigwedge \Sigma$ *; si* $\langle\Pi\rangle_j^{\mathbb{C}} \in \Gamma_n$, *alors* $n \Vdash \mathbb{C}_j \bigwedge \Pi$ *; et si* $A \in \Delta_n$, *alors* $n \nVdash A$. *De plus,* $\mathcal{M}$ *est un modèle à voisinages doubles de* **ELG**.

Remarquons que, puisque toutes les règles sont cumulatives, le contre-modèle $\mathcal{M}$ de H est également un contre-modèle de l'hyperséquent H' apparaissant à la racine. Alors, pour tout hyperséquent, nous obtenons soit une dérivation (si l'hyperséquent est valide), soit un contre-modèle. Ceci implique le théorème suivant :

Théorème 4 (Complétude sémantique) *Si H est valide dans tous les modèles à voisinages doubles de* **ELG***, alors il est dérivable dans* $\mathbf{HS}_{\mathbf{ELG}}$.

La procédure de recherche de preuve pour le calcul $\mathbf{HS}_{\mathbf{ELG}}$ peut être employée pour vérifier automatiquement, et constructivement, la validité (ou la dérivabilité) des formules de la logique d'Elgesem. Pour toute formule, cette procédure de recherche de preuve donne soit une dérivation, soit un contre-modèle, selon que la formule est valide ou non.

Exemple 1 ***(Échec de la délégation)*** *Le traitement de la notion de délégation marque une différence majeure entre la modélisation de l'effectivité dans la logique d'Elgesem et d'autres modélisations, telles que celles données par les logiques dites STIT. Elle est explicitement rejetée par Elgesem [6], pour qui un agent n'est pas considéré comme effectuant une action si celle-ci met en jeu l'intervention d'un autre agent. Par exemple, on peut dire que faire réparer une voiture n'est pas équivalent à réparer cette voiture*

soi-même. Représentons Anna par la lettre a, Béatrice par b, et "réparer la voiture" par p. Alors $\mathbb{E}_a\mathbb{E}_b p \rightarrow \mathbb{E}_a p$ *modélise la phrase : "Si Anna fait réparer sa voiture par Béatrice, alors Anna répare la voiture." Notre calcul donne automatiquement un contre-modèle de* $\mathbb{E}_a\mathbb{E}_b p \rightarrow \mathbb{E}_a p$*. Tout d'abord, en figure 4 nous donnons une preuve qui échoue de* $\mathbb{E}_a\mathbb{E}_b p \rightarrow \mathbb{E}_a p$ *dans* $\mathbf{HS}_{\mathbf{ELG}}$*. Nous considérons ensuite l'énumération suivante des composants de l'hyperséquent saturé suivant :* $1 \mapsto \langle\mathbb{E}_b p\rangle^{\mathbb{E}}_a, \langle p\rangle^{\mathbb{E}}_b, \langle\mathbb{E}_b p\rangle^{\mathbb{C}}_a, \langle p\rangle^{\mathbb{C}}_b, p, \mathbb{E}_b p, \mathbb{E}_a\mathbb{E}_b p \Rightarrow \mathbb{E}_a p$*;* $2 \mapsto p \Rightarrow \mathbb{E}_b p$*; et* $3 \mapsto \Rightarrow p$*.*

Nous obtenons les contre-modèles suivants : Contre-modèles à voisinages doubles : En appliquant la construction de la définition 6 nous obtenons le contre-modèle à voisinages doubles $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}^{\mathbb{E}}_i, \mathcal{N}^{\mathbb{C}}_i, \mathcal{V}\rangle$*, où* $\mathcal{W} = \{1, 2, 3\}$*;* $\mathcal{V}(p) = \{1, 2\}$*;* $\mathcal{N}^{\mathbb{E}}_a(1) = \mathcal{N}^{\mathbb{C}}_a(1) = \{(\{1\}, \{2\})\}$ *– puisque* $\mathcal{N}^{\mathbb{E}}_a(1) = \mathcal{N}^{\mathbb{C}}_a(1) = \{(\mathbb{E}_b p^+, \mathbb{E}_b p^-)\}$*,* $\mathbb{E}_b p^+ = \{1\}$*, et* $\mathbb{E}_b p^- = \{2\}$ *–;* $\mathcal{N}^{\mathbb{E}}_b(1) = \mathcal{N}^{\mathbb{C}}_b(1) = \{(\{1, 2\}, \{3\})\}$ *– puisque* $\mathcal{N}^{\mathbb{E}}_b(1) = \mathcal{N}^{\mathbb{C}}_b(1) = \{(p^+, p^-)\}$*,* $p^+ = \{1, 2\}$*, et* $p^- = \{3\}$ *–;* $\mathcal{N}^{\mathbb{E}}_i(n) = \mathcal{N}^{\mathbb{C}}_i(n) = \emptyset$ *pour* $i = a, b$ *et* $n = 2, 3$*.*

Contre-modèle à voisinages simples : La transformation de la proposition 1 donne le contre-modèle à voisinages simples $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}^{\mathbb{E}}_i, \mathcal{N}^{\mathbb{C}}_i, \mathcal{V}\rangle$*, où* $\mathcal{W} = \{1, 2, 3\}$*;* $\mathcal{V}(p) = \{1, 2\}$*;* $\mathcal{N}^{\mathbb{E}}_a(1) = \mathcal{N}^{\mathbb{C}}_a(1) = \{\{1\}, \{1, 3\}\}$*;* $\mathcal{N}^{\mathbb{E}}_b(1) = \mathcal{N}^{\mathbb{C}}_b(1) = \{\{1, 2\}\}$*; et* $\mathcal{N}^{\mathbb{E}}_i(n) = \mathcal{N}^{\mathbb{C}}_i(n) = \emptyset$ *pour* $i = a, b$ *et* $n = 2, 3$*.*

5 Conclusion

Nous avons présenté un calcul d'hyperséquents pour la logique d'effectivité et de capacité d'Elgesem. Ce calcul présente de bonnes propriétés structurelles, parmi lesquelles l'admissibilité syntaxique de la coupure. De plus, nous avons défini une stratégie terminante de recherche de preuve, qui assure le calcul d'une dérivation ou d'un contre-modèle suivant la validité de la formule en entrée. En particulier, on peut extraire directement d'une preuve qui échoue un contre-modèle de la formule considérée dans la sémantique à voisinages doubles, qui peut ensuite aisément être transformé en contre-modèle dans la sémantique de voisinages usuelle.

Le calcul d'hyperséquents présenté dans ce travail se prête bien à l'automatisation, et nous prévoyons de l'implémenter afin d'obtenir le premier prouveur automatique pour la logique d'effectivité et de capacité.

Par ailleurs, la logique d'Elgesem modélise les actions d'un unique agent, qu'il soit un individu ou une institution, ou un groupe pensé comme entité indivisible. Une extension naturelle, proposée par Troquard [11], permet de modéliser des *coalitions*, c'est-à-dire des groupes (divisibles en individus) d'agents. Le travail présenté dans cet article s'étend naturellement à l'extension de Troquard, ceci est présenté par les auteurs de cet article dans [3].

Références

[1] Avron, Arnon: *The Method of Hypersequents in the Proof Theory of Propositional Non-Classical Logics*, page 1–32. Clarendon Press, USA, 1996, ISBN 0198538626.

[2] Belnap, Nuel D, Michael Perloff et Ming Xu: *Facing the future : agents and choices in our indeterminist world*. Oxford University Press on Demand, 2001.

[3] Dalmonte, Tiziano, Charles Grellois et Nicola Olivetti: *Proof systems for the logics of bringing-it-about*. Dans *DEON 2020/2021, the 15th International Conference on Deontic Logic and Normative Systems*, 2021.

[4] Dalmonte, Tiziano, Björn Lellmann, Nicola Olivetti et Elaine Pimentel: *Countermodel Construction via Optimal Hypersequent Calculi for Non-normal Modal Logics*. Dans Artëmov, Sergei N. et Anil Nerode (rédacteurs) : *Logical Foundations of Computer Science - International Symposium, LFCS 2020*, tome 11972 de *Lecture Notes in Computer Science*, pages 27–46. Springer, 2020.

[5] Dalmonte, Tiziano, Nicola Olivetti et Sara Negri: *Non-Normal Modal Logics : Bi-Neighbourhood Semantics and Its Labelled Calculi*. Dans *Advances in Modal Logic 12, proceedings of the 12th conference on "Advances in Modal Logic"*, pages 159–178. College Publications, 2018.

[6] Elgesem, Dag: *The modal logic of agency*. Nordic Journal of Philosophical Logic, 2 :1–46, 1997.

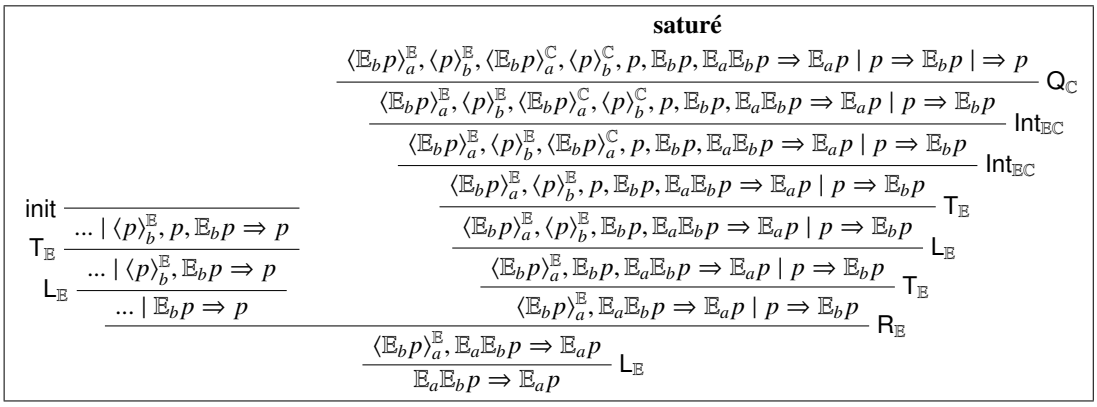


FIGURE 4 – Preuve qui échoue dans $\mathbf{HS_{ELG}}$.

[7] Governatori, Guido et Antonino Rotolo: *On the Axiomatisation of Elgesem's Logic of Agency and Ability*. J. Philosophical Logic, 34(4) :403–431, 2005.

[8] Horty, John F: *Agency and deontic logic*. Oxford University Press, 2001.

[9] Lellmann, Bjoern: *Sequent Calculi with Context Restrictions and Applications to Conditional Logic*. Thèse de doctorat, Imperial College London, 2013.

[10] Schröder, Lutz et Dirk Pattinson: *Shallow Models for Non-iterative Modal Logics*. Dans Dengel, Andreas, Karsten Berns, Thomas M. Breuel, Frank Bomarius et Thomas Roth-Berghofer (rédacteurs) : *KI 2008 : Advances in Artificial Intelligence, 31st Annual German Conference on AI, KI 2008*, tome 5243 de *Lecture Notes in Computer Science*, pages 324–331. Springer, 2008.

[11] Troquard, Nicolas: *Reasoning about coalitional agency and ability in the logics of "bringing-it-about"*. Autonomous Agents and Multi-Agent Systems, 28(3) :381–407, 2014.