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

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

Mémoire présenté en vue de l'obtention de

l'Habilitation à diriger les recherches

spécialité : informatique

par

MARIE-CHRISTINE LAGASQUIE-SCHIEX

# Traitement de l'incohérence dans le raisonnement non monotone : de la restauration de cohérence à l'argumentation

Soutenue le 10 juillet 2009 devant :

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

# Résumé

L'origine de tous mes travaux remonte à ma découverte de la notion de raisonnement non monotone lors de mes études de DEA. Cette notion m'a toujours paru particulièrement intéressante à modéliser dans un système informatique. Elle fait en effet appel à une particularité bien humaine : être capable de bâtir un raisonnement qui a du sens sur des bases qui n'en ont pas. En particulier, il s'agit donc de modéliser la capacité à raisonner à partir d'informations incohérentes, ce raisonnement pouvant être de nature déductive ou autre.

Ce domaine a été très largement étudié en Intelligence Artificielle (IA) depuis maintenant plusieurs dizaines d'années. De nombreux systèmes, le plus souvent ad-hoc, ont été mis en place pour simuler un tel comportement. Les plus fiables du point de vue de la "qualité" du raisonnement sont malheureusement très coûteux en temps et parfois aussi en espace. Cette problématique du coût a été un de mes premiers travaux sur le sujet puisque ma thèse de doctorat a justement porté sur la comparaison de divers systèmes de raisonnement non monotone en particulier du point de vue de la complexité temporelle. Le problème concerné par cette étude était essentiellement un problème de déduction dans un cadre de *restauration de la cohérence* : on part d'une base de formules logiques *inconsistante* représentant les connaissances d'un agent, on sélectionne des sous-bases "préférées" et *consistantes*, on applique la déduction classique et on définit la déduction non monotone en fonction des résultats obtenus avec la déduction classique sur chacune de ces sous-bases préférées.

Les mécanismes mis en œuvre ici sont en général assez simples : on adopte une conclusion si "toutes les sous-bases sont d'accord (pour cette conclusion)", ou bien "si au moins une est d'accord (pour cette conclusion)", ou bien "si au moins une est d'accord (pour cette conclusion) et aucune contre (cette conclusion)". Dans un premier temps, j'ai donc cherché à généraliser cette notion d'inférence non monotone en introduisant un mécanisme de vote pour obtenir des inférences un peu plus riches (plus de formules déduites) tout en restant saines (pas question de conclure une formule et son contraire) ; parmi les relations obtenues on peut par exemple trouver une relation qui conclut positivement si "la majorité des sous-bases est d'accord".

En parallèle, sachant que le problème du raisonnement non monotone tel que je l'avais abordé jusqu'alors peut se ramener à un problème de résolution de conflits, j'ai approfondi les principes de base de ce type de résolution. Ces principes peuvent être de nature très différente : optimalité des solutions du point de vue de l'inclusion, respect des préférences entre les connaissances si elles existent, ….

En général, les solutions à un problème de résolution de conflit sont des ensembles de connaissances permettant de faire de la déduction classique et ces solutions sont en conflit les unes avec les autres. Cela conduit tout naturellement à la notion d'arguments, d'attaque entre les arguments et donc à un type particulier de raisonnement non monotone appelé l'"argumentation" qui est basé sur l'échange et l'évaluation d'arguments interagissant. Ce processus se compose de plusieurs étapes : création d'arguments, identification des interactions entre ces arguments, attribution d'une valeur à chacun de ces arguments, choix des arguments considérés comme les plus acceptables et enfin la détermination des conclusions du système d'argumentation (qui peut correspondre à une phase de déduction non monotone).

Le grand intérêt de l'argumentation est le fait qu'elle permet de modéliser les interactions entre des agents et ainsi de réaliser un raisonnement non monotone sur un groupe d'agents et plus seulement sur un agent unique comme le faisaient les relations d'inférence non monotone que j'avais étudiées jusqu'alors.

Dans ce cadre-là, mon travail a porté sur une généralisation d'un système d'argumentation particulier : le système abstrait proposé par "Dung". J'y ai abordé plusieurs aspects : l'"évaluation" des arguments, leur "acceptabilité", la prise en compte d'interactions "bipolaires", la "fusion" ou la "révision" en argumentation et enfin l'utilisation de l'"argumentation" pour faire du *practical reasoning*.

Cette partie de mon travail s'étendant sur plusieurs années, elle sera sur-représentée dans ce manuscrit.

Et enfin, la théorie des jeux est le dernier point évoqué dans ce manuscrit. Il ne s'agit pas ici de décrire la théorie des jeux dans son intégralité. Je n'ai vraiment travaillé sur ce thème qu'en relation avec les thèmes précédents ; j'ai ainsi pu réutiliser les notions de logique, de conflit, de préférences avec des jeux très spécifiques, les jeux booléens, et cela a débouché sur la mise en évidence de nouveaux liens entre argumentation et jeux.

# Table des matières

**Synthèse des travaux** **1**
Cadre des travaux . . . 1
Évolution thématique de mes travaux . . . 2
Organisation du document . . . 7

**1 Thème 1 : définition et étude de relations d'inférence non monotone** **9**
1.1 Rappel des résultats obtenus en thèse . . . 10
1.1.1 Définition formelle d'une inférence non monotone . . . 10
1.1.2 L'analyse comparative des relations d'inférence non monotone $p$-$m$ . . . 12
1.2 Une étude complémentaire : la généralisation de la définition d'une relation d'inférence non monotone 13
1.2.1 Introduction . . . 13
1.2.2 Processus global de traitement des incohérences . . . 14
1.2.3 Application au raisonnement non monotone . . . 15
1.2.3.1 Où l'on retrouve des relations d'inférence non monotone connues . . . 15
1.2.3.2 Quelques "nouvelles" relations d'inférence non monotone . . . 15
1.2.3.2.1 GRAD-EQUADD-INCL . . . 15
1.2.3.2.2 GRAD-POSMAX-T . . . 16
1.2.3.3 Un exemple illustratif . . . 16
1.2.3.4 Étude théorique . . . 17
1.2.3.4.1 Prudence des relations . . . 17
1.2.3.4.2 Respect des règles de déduction . . . 17
1.2.3.4.3 Quelques remarques sur la complexité . . . 18
1.2.4 Conclusion sur l'étude complémentaire . . . 18
1.3 Conclusion sur le thème 1 . . . 18

**2 Thème 2 : méthodes de résolution de conflit** **21**
2.1 L'approche méthodologique . . . 22
2.1.1 Les définitions et principe de base . . . 22
2.1.2 Utilisation des préférences . . . 23
2.2 Calcul des solutions minimales . . . 25
2.3 Relations avec les méthodes basées sur la sélection directe de sous-ensembles $K$-consistants . . . 26
2.3.1 Dualité entre problème de cohérence et problème de conflits . . . 26
2.3.2 Intégration de préférences . . . 27
2.4 Conclusion sur le thème 2 . . . 29

**3 Thème 3 : l'argumentation** **31**
3.1 Schéma général . . . 32
3.2 Argumentation unipolaire . . . 34
3.2.1 Les bases de l'argumentation unipolaire . . . 34
3.2.2 Évaluation d'arguments . . . 38
3.2.2.1 Approche locale (évaluation générique) . . . 39
3.2.2.2 Approche globale (avec tuples) . . . 43
3.2.2.2.1 Évaluation globale avec tuples pour les graphes acycliques . . . 43
3.2.2.2.2 Étude des circuits . . . 45
3.2.2.2.3 Une évaluation graduelle avec tuples pour des graphes quelconques . . . 48
3.2.2.2.4 Comparaison de valeurs tuplées . . . 50
3.2.2.3 Principales différences entre les évaluations "locale" et "globale" . . . 53
3.2.2.4 Conclusion sur l'évaluation d'arguments . . . 54
3.2.3 Acceptabilité . . . 55
3.2.3.1 Différents niveaux d'acceptabilité collective . . . 55
3.2.3.2 Vers une acceptabilité individuelle graduelle . . . 58
3.2.3.3 Compatibilité entre acceptabilité et évaluation graduelle . . . 58
3.2.3.3.1 Exemples montrant la non-compatibilité dans le cas général . . . 59
3.2.3.3.2 Cas particuliers menant à la compatibilité . . . 61
3.2.3.4 Conclusion sur l'acceptabilité . . . 62
3.2.4 Fusion en argumentation unipolaire . . . 62
3.2.4.1 Quelques exemples . . . 64
3.2.4.1.1 Voter n'est pas suffisant . . . 64
3.2.4.1.2 L'union n'est pas une fusion (en général) . . . 65
3.2.4.2 Systèmes d'argumentation partiels . . . 65
3.2.4.3 Opérateurs de fusion . . . 68
3.2.4.4 Quelques propriétés . . . 72
3.2.4.4.1 Propriétés des PAFs et des expansions consensuelles . . . 72
3.2.4.4.2 Propriétés des opérateurs de fusion . . . 75
3.2.4.5 Acceptabilité pour un profil d'AFs . . . 77
3.2.4.6 Conclusion et perspectives . . . 79
3.2.5 Révision en argumentation unipolaire . . . 80
3.2.5.1 Concepts de base en théorie de la révision . . . 80
3.2.5.2 Révision en argumentation . . . 81
3.2.5.2.1 Définition . . . 81
3.2.5.2.2 Typologie de révisions . . . 82
3.2.5.3 La révision classique en argumentation . . . 84
3.2.5.4 Étude de cas . . . 86
3.2.5.4.1 Révision décisive . . . 86

3.2.5.4.2 Révision expansive . . . 87
3.2.5.5 Discussion et perspectives . . . 88
3.2.6 Practical reasoning et argumentation unipolaire . . . 90
3.2.6.1 Exemple illustratif . . . 91
3.2.6.2 Langage logique utilisé . . . 92
3.2.6.3 Typologie des arguments . . . 93
3.2.6.3.1 Justification de croyances . . . 93
3.2.6.3.2 Justification de désirs . . . 94
3.2.6.4 Interactions entre arguments . . . 95
3.2.6.4.1 Conflits entre arguments épistémiques . . . 95
3.2.6.4.2 Conflits entre arguments explicatifs . . . 96
3.2.6.4.3 Conflits entre arguments instrumentaux . . . 97
3.2.6.4.4 Conflits entre arguments de type différent . . . 98
3.2.6.5 Système d'argumentation pour le PR . . . 99
3.2.6.6 Reprenons l'exemple initial . . . 100
3.2.6.7 Propriétés du système . . . 102
3.2.6.8 Discussion . . . 104
3.2.6.9 Conclusion . . . 105
3.3 Argumentation bipolaire . . . 106
3.3.1 Définition d'un BAF . . . 109
3.3.2 Acceptabilité . . . 111
3.3.2.1 Gestion des conflits . . . 111
3.3.2.2 Nouvelles sémantiques pour l'acceptabilité . . . 112
3.3.3 Coalitions en argumentation bipolaire . . . 114
3.3.3.1 Le concept de coalition . . . 114
3.3.3.2 Un méta-système pour l'argumentation . . . 116
3.3.4 Évaluation d'arguments . . . 117
3.3.4.1 Étiquetages dans HERMES . . . 117
3.3.4.2 Évaluation bipolaire graduelle . . . 118
3.3.5 Conclusion sur la bipolarité . . . 120
3.3.5.1 Acceptabilité et bipolarité . . . 120
3.3.5.2 Coalitions d'arguments . . . 121
3.3.5.3 Evaluation et bipolarité . . . 121
3.4 Conclusion sur le thème 3 . . . 122

**4 Thème 4 : raisonnement à base de jeux** **125**
4.1 Concepts de base de la théorie des jeux . . . 126
4.1.1 Taxonomie des jeux . . . 126
4.1.2 Quelques exemples . . . 127
4.1.3 Représentation des jeux . . . 128
4.1.4 Concepts de solution . . . 129
4.2 Argumentation et théorie des jeux . . . 129
4.2.1 Travaux de Dung . . . 129
4.2.2 Théorie de la preuve dialectique . . . 130
4.3 Jeux booléens . . . 130
4.3.1 Jeux booléens à $n$ joueurs . . . 131
4.3.1.1 Définitions de base . . . 131
4.3.1.2 Dépendance . . . 133
4.3.1.3 Équilibres de Nash . . . 135
4.3.2 Jeux booléens à préférences non-dichotomiques . . . 136
4.3.2.1 Cas des CP-nets . . . 136
4.3.2.2 Cas des buts à priorité . . . 142
4.3.3 Retour sur l'argumentation . . . 145
4.3.3.1 Traduction d'un AF en un CP-jeu booléen . . . 146
4.3.3.2 Calcul des extensions préférées . . . 148
4.3.3.3 Gestion des circuits impairs . . . 149
4.4 Conclusion sur le thème 4 . . . 151

**5 Projet de recherche** **153**
5.1 État de l'art sur le sujet . . . 154
5.2 Originalité du projet . . . 154
5.3 Organisation du projet . . . 155

**Bibliographie** **157**

**A Annexes** **169**
A.1 Répartition des encadrements . . . 169
A.1.1 Encadrements stagiaires Master 2 . . . 169
A.1.2 Encadrements de doctorants . . . 169
A.2 Répartition des publications . . . 170
A.2.1 Par type de publication . . . 170
A.2.2 Par thème de recherche . . . 170
A.3 Participation à des projets et collaborations . . . 170

# Synthèse des travaux

## Cadre des travaux

J'effectue ma recherche dans l'équipe "Raisonnements Plausibles, Décision et Méthodes de Preuve" (RPDMP) de l'IRIT à l'Université Paul Sabatier de Toulouse. Cette équipe est spécialisée dans la formalisation des mécanismes du raisonnement plausible et de la décision notamment en présence d'informations incohérentes et/ou incertaines. Les travaux que j'ai réalisés se situent dans ce cadre et portent principalement sur le raisonnement non monotone et la manière dont on y traite l'incohérence. Présentons ce concept en quelques mots.

Un des buts de l'intelligence artificielle est l'étude des méthodes de manipulation des connaissances pour la réalisation d'un raisonnement (déduction d'informations, explication de faits, prise de décision ...). La logique classique est un des outils formels traditionnels dans ce domaine et elle convient parfaitement tant que les connaissances forment un tout cohérent. Elle ne convient plus dès qu'il s'agit de traiter des connaissances incomplètes, incertaines ou incohérentes (on parle alors de croyances) ; en effet, à partir d'informations contradictoires, la logique classique permet de déduire n'importe quoi. Or la manipulation de telles informations est capitale si on veut pouvoir formaliser et automatiser un raisonnement de "sens commun". Dans ce cas-là, il peut être nécessaire de remettre en question d'anciennes conclusions en présence de nouveaux faits ou de faire un choix entre plusieurs propositions pouvant être antagonistes ; c'est l'aspect non monotone du raisonnement.

Dans ce contexte, il existe au moins trois aspects principaux à prendre en compte et qui permettront de situer mes travaux :

**Comment traiter l'incohérence de la base ? Dispose-t-on d'éléments supplémentaires permettant de "trancher" en cas de conflits ?**

De manière générale, quand on se retrouve en présence d'une incohérence, il y a trois manières différentes de l'aborder :

- soit on *cherche à l'éviter*, ce qui revient à mettre en place des mécanismes qui suppriment réellement les incohérences de l'ensemble des informations utilisé pour raisonner ; cela peut se faire de manière graduelle, au fur et à mesure que les informations arrivent, et c'est ce qu'on appelle la révision des croyances ; cela peut aussi se faire quand les informations proviennent de sources différentes et c'est la fusion de connaissances ;
- soit on *restaure la cohérence*, c'est-à-dire, on identifie des *paquets* d'information qui sont cohérents et on travaille avec ces paquets ; l'idée de base est alors que ces paquets représentent des états possibles du monde ; il existe dans ce cadre de nombreuses méthodes qui dépendent essentiellement de la manière dont on identifie les *paquets cohérents* ; dans ce cas, l'ensemble des informations reste incohérent mais on raisonne à partir de sous-ensembles cohérents issus de cet ensemble ;
- soit on *accepte l'incohérence* et on raisonne avec elle ; pour cela il faut définir de nouveaux langages (logiques non monotones, logiques para-consistantes, ...) ; ici-aussi, l'ensemble des informations reste incohérent et il sert dans sa totalité pour le raisonnement.

Il est évident que la frontière entre ces trois méthodes n'est pas totalement étanche et qu'elle évolue dans le temps. Certains travaux ont par exemple été proposés dans le cadre de l'acceptation de l'incohérence mais peuvent aussi être définis comme des méthodes de restauration de la cohérence si on accepte d'élargir un peu le concept.

L'argumentation en est un exemple type. Initialement, le processus de l'argumentation en IA a été défini comme un moyen de raisonner en acceptant l'incohérence sans avoir à rechercher des paquets cohérents d'information pouvant représenter des états du monde sur lesquels on pourrait ensuite raisonner classiquement ; il s'agissait de mener un raisonnement guidé alors par les conclusions que l'on étudiait ; cela permettait d'identifier des arguments "pour cette conclusion" et des arguments "contre cette conclusion" et de raisonner à partir de ces arguments et de leurs relations. Toutefois, si on considère la restauration de cohérence dans un sens plus large, l'argumentation rentre alors dans ce cadre, car chaque argument peut être vu comme un paquet cohérent sur lequel on peut appliquer un raisonnement classique (même si ce "paquet" ne représente qu'un état très parcellaire du monde).

Tout mon travail s'est fait dans le cadre du **traitement de l'incohérence** en passant de la **restauration de cohérence** à un certain *type* d'**acceptation de l'incohérence**. En effet, mon idée de base est d'*exploiter la décomposition de l'ensemble initial d'informations incohérentes en* ***paquets cohérents*** *mais qui demeurent en relation* (en général, c'est une relation de conflit) en utilisant éventuellement des éléments supplémentaires facilitant la résolution des conflits. Parmi ces éléments, j'ai exploité en particulier la notion de **préférences entre les croyances** qui peut être définie de manière très variée :

- préférences imposées *a priori*,
- préférences déduites à partir de relations explicites données entre ces croyances (relation d'attaque, relation d'appui, ...),
- préférences déduites à partir de relations implicites entre les croyances (dépendance logique, ...),
- préférences devant vérifier certaines contraintes,
- ...

Pour ma part, je me suis surtout intéressée au cas soit de préférences imposées *a priori*, soit de préférences déduites à partir de relations explicites données entre les croyances.

**Quel est le but poursuivi par le raisonnement ?** S'agit-il de déduire de nouvelles informations ? De fournir des explications ? ...

Ici, mon travail a commencé par porter sur l'aspect déductif du raisonnement, donc sur la capacité de déduire de nouvelles informations (ce qui correspond à la définition d'une relation d'inférence non monotone). Puis, il a peu à peu dérivé vers l'aspect explicatif qui est fortement lié au raisonnement abductif, c'est-à-dire la capacité de produire et de gérer des explications pour des faits donnés (ce qui correspond au processus d'argumentation avec création et analyse des arguments). Il est toutefois intéressant de noter que l'argumentation est un outil assez puissant pour pouvoir réaliser n'importe quel type de raisonnement ; de nombreux travaux ont en effet montré que l'argumentation peut être utilisée pour définir des relations d'inférence non monotone, faire de la prise de décision, gérer des dialogues de négociation, ...

**Combien d'agents sont concernés par ce raisonnement :** un seul ? Plusieurs ? S'il n'en existe qu'un, cet agent raisonnera sur un seul ensemble d'informations incohérent, alors que s'il en existe plusieurs, chaque agent peut avoir à sa disposition un ensemble d'informations cohérent ou pas et il faudra alors étudier la manière de bâtir un raisonnement entre tous ces agents.

Sur cet aspect, mon travail a considéré les deux cas mais pas pour tous les thèmes abordés ; par exemple, toute la partie portant sur les relations d'inférence non monotone ne concerne qu'un seul agent, alors que la partie portant sur l'argumentation concerne successivement un seul agent puis plusieurs agents.

# Évolution thématique de mes travaux

Dans cette section, je décris plus précisément les thèmes que j'ai abordés en montrant le passage d'un thème à l'autre, et en mentionnant les collaborations, encadrements et publications associés.

Dans le contexte décrit en section précédente, mon travail de thèse a consisté tout d'abord à classer de nombreuses relations d'inférence non monotone suivant plusieurs points de vue formels :

- la complexité algorithmique, c'est-à-dire la difficulté de résolution dans le cas général et dans des cas plus contraignants ;

– la prudence, c'est-à-dire la capacité de la relation d'inférence à déduire sans risquer d'inférer des conclusions erronées ;
– les propriétés des relations étudiées, c'est-à-dire l'adéquation de la relation au modèle que l'on souhaite formaliser.

Dans une seconde étape, à la fin de ma thèse, j'ai étudié différents algorithmes pour des relations d'inférence non monotone. Cette phase était destinée au passage à une application pratique des résultats théoriques obtenus. Ils ont permis de comparer plusieurs techniques algorithmiques pour une relation d'inférence non monotone sur tout un ensemble de jeux de test, en particulier des tests aléatoires. J'ai pu ainsi identifier un algorithme particulièrement efficace (issu directement de l'étude de complexité théorique) ainsi qu'une technique intéressante dans un cadre de précompilation de la base de croyances.

L'objectif de toutes ces études était de synthétiser les nombreux travaux portant sur ce sujet et ainsi de mieux comprendre les mécanismes entrant en jeu dans la définition de ce type de relations, ainsi que l'utilisation que l'on peut en faire, le but à plus ou moins long terme restant la réalisation d'un système de raisonnement non monotone efficace et réaliste.

Du point de vue "timing", cette phase initiale de mes travaux correspond bien-sûr à ma thèse de doctorat de 1992 à 1995 encadrée par Claudette et Michel CAYROL et ne porte que sur le thème "inférence non monotone"[1]. Tous mes travaux de thèse ont été exposés lors d'une présentation en tant que chercheur invité aux journées du PRC-GDR d'Intelligence Artificielle [LS95c] et ont donné lieu à plusieurs articles ou rapports [CLS93, LS94, CLS94b, CLS94a, CLS95, LS95a, LS95b, CLSS96a]. La partie portant sur les algorithmes pour des relations d'inférence non monotone a été réalisée en 1996 et a donné lieu à trois publications dont une dans une revue internationale [CLSS96a, CLSS96b, CLSS98].

À partir de là, tout mon travail a consisté :
– à rester toujours dans le cadre du traitement de l'incohérence par sélection de paquets cohérents,
– à chercher à généraliser les méthodes de raisonnement existants tout en les enrichissant,
– à établir des liens entre ces diverses méthodes.

Pour cela, depuis la fin de ma thèse, j'ai donc abordé essentiellement les quatre thèmes suivants : l'inférence non monotone, la résolution de conflits, l'argumentation et les jeux.

**Thème 1 : L'inférence non monotone** Il s'agit bien sûr de la suite directe des travaux de ma thèse. J'ai défini un cadre de travail général respectant l'idée stricte de la restauration de cohérence et s'inspirant fortement de l'aspect décisionnel de la déduction en raisonnement non monotone : faut-il, oui ou non, inférer non monotoniquement une formule donnée à partir d'une base de croyances donnée ? Ce cadre permet de retrouver toutes les relations que j'avais étudiées précédemment et permet aussi la définition de nouvelles relations dont les résultats semblent beaucoup plus proches de ceux attendus intuitivement.

L'idée développée ici est d'identifier dans l'ensemble des croyances, à partir duquel on souhaite raisonner, des sous-ensembles susceptibles de représenter des "opinions" (on peut utiliser ici les préférences entre les croyances pour donner une "force" aux différentes "opinions"). Puis, on fait "voter" ces "opinions" en utilisant la logique classique, et c'est en fonction du résultat de ce vote que l'on pourra conclure sur la déduction ou pas d'une formule donnée.

Le but de ce cadre général est à la fois de rassembler les relations existantes et de pouvoir en définir de nouvelles capables de conclure même en présence d'"opinions contre".

Du point de vue "timing", cette période s'étale de 1997 à 2000 et correspond à une collaboration avec Claudette CAYROL. Ces travaux ont débouché sur la rédaction d'un rapport [CLS97] et d'un article [CLS00].

[1] En fait, la source de ces travaux était mon sujet de DEA ; il s'agissait d'étudier un cadre formel unificateur et les outils algorithmiques associés permettant d'utiliser diverses approches définies en Intelligence Artificielle (logique classique, logique des défauts, logiques modales…). J'ai achevé cette étude par un rapport de recherche [LS92]. Bien que ce sujet ne puisse pas déboucher sur une thèse, cette année m'a permis de prendre conscience de la multitude de travaux portant sur le raisonnement non monotone et ainsi est né le désir d'établir une classification des processus déductifs issus de ce contexte.

***Passage de l'inférence non monotone à la résolution de conflits*** Une des question à laquelle je me suis retrouvée confrontée lorsque j'ai étudié la définition de relations d'inférence non monotone dans le cadre strict de la restauration de cohérence était la suivante : pourquoi choisir un certain mécanisme de sélection plutôt qu'un autre ? La partie "comparaison théorique des relations d'inférence non monotone" de ma thèse donnait quelques pistes de réflexion sur ce sujet mais qui mêlaient trop de choses à la fois (le mécanisme de sélection mais aussi le principe d'inférence). J'ai donc voulu répondre à cette question de manière indépendante et c'est ce qui m'a fait m'intéresser aux principes mis en œuvre dans la résolution de conflits.

**Thème 2 : La résolution de conflits** Il s'agit là d'une approche différente permettant aussi de faire du raisonnement non monotone dans le cadre strict de la restauration de cohérence. On identifie dans l'ensemble des croyances les conflits existant entre ces croyances et on cherche à les résoudre pour trouver au final une base de croyances sans conflit (sur laquelle on pourra donc utiliser l'inférence classique si on veut faire de la déduction et ainsi "rejoindre" le thème 1).

L'approche utilisée ici est originale par son aspect méthodologique. En effet, l'accent a été mis ici sur la définition de principes à respecter lors de la résolution de ce type de problème. Ce qui revient à définir plusieurs types de solutions et à chercher à les classer les unes par rapport aux autres.

On peut alors comparer les résultats des diverses solutions proposées avec ceux obtenus par les relations d'inférence non monotone étudiées précédemment.

Ces travaux ont été réalisés de 2000 à 2003 en collaboration avec Claudette CAYROL et Jérôme MENGIN, respectivement professeur et maître de conférences à l'UPS, et ils sont présentés dans un article [CLSM00]. Ils ont aussi permis l'encadrement d'un stagiaire du DESS IRR[2] de l'UPS, Laurent VIDAL, chargé de l'implémentation de certaines des solutions ainsi définies (voir [Vid99]).

Puis, ils ont débouché sur une collaboration avec Hélène FARGIER, directrice de recherche CNRS à l'IRIT, consistant à encadrer conjointement deux stagiaires de DEA[3], Loïc DUSSEUX en 2002 et Pascal LAFOURCADE en 2003, chargés d'étudier les liens entre résolution de conflits et aide à la décision dans le cadre de problèmes réels issus de la gestion de projets ou de produits (voir [Laf03]).

***Passage de l'inférence non monotone à l'argumentation*** Une étude réalisée par Claudette CAYROL (voir [Cay95]) a mis en évidence une équivalence entre certains processus d'argumentation et les relations d'inférence non monotone que j'avais étudiées lors de la thèse. Ce travail m'a donc tout naturellement orienté vers le domaine de l'argumentation, d'autant plus que ce domaine est suffisamment large pour simuler et s'adapter à de nombreux modes de raisonnement non monotone. L'idée ici était à la fois de trouver une méthode de raisonnement permettant de faire de l'inférence non monotone et d'établir une passerelle vers d'autres domaines que je connaissais moins, comme l'étude des interactions entre agents rationnels (sous forme de dialogue – d'argumentation – par exemple), mais qui avaient toujours la même finalité : traiter les incohérences en exhibant des paquets cohérents.

**Thème 3 : L'argumentation** Ici, il s'agit toujours de raisonner à partir d'un ensemble d'informations incohérentes mais en privilégiant l'aspect suivant qui respecte toujours les idées de la restauration de cohérence mais de manière élargie : on peut définir une notion d'"argument" lié à une conclusion donnée (par exemple un sous-ensemble consistant et minimal de la base initiale inférant classiquement la dite-conclusion). On peut alors coupler cette notion d'arguments avec diverses relations explicites entre ces arguments (par exemple, une relation d'attaque : un argument en "contredit" un autre, ou bien une relation d'appui : un argument en "aide" un autre). à partir de ces arguments et de ces relations, on peut construire un système axiomatique suffisamment puissant pour pouvoir prendre en compte non seulement les notions classiques d'arguments "pour" et d'arguments "contre", mais aussi des notions d'évaluation d'arguments, de force d'un argument, d'arguments équivalents, d'arguments "acceptables", ... Ce sont ces arguments acceptables qui seront alors pris en compte pour définir la conclusion du processus d'argumentation. Cette conclusion et sa définition dépendent bien évidemment du but poursuivi qui peut être de nature très variée (déduction, prise de décision, négociation, ...).

[2] DESS IRR : DESS en Intelligence artificielle, Reconnaissance des formes et Robotique.
[3] DEA RCFR (Représentation de la Connaissance et Formalisation du Raisonnement) de l'UPS.

Le lien entre ce thème et les deux précédents est essentiellement du au fait que l'approche dite-argumentative peut servir aussi bien à faire de l'inférence non monotone que de la résolution de conflits.

Il est important de noter que l'argumentation, que j'aborde ici par le biais élargi de la restauration de cohérence, est un processus bien plus général, introduit initialement pour faire de l'acceptation d'incohérence (et pas de la restauration de cohérence, au sens strict du concept) et qui apparaît en tant que thématique dans d'autres disciplines que l'Intelligence Artificielle ; on trouve par exemple des travaux sur l'argumentation dans le cadre de la philosophie mais aussi en psychologie.

Ces travaux ont donné lieu de 2001 à 2009 à de nombreux rapports de recherche [CLS01, CLS02b, CDLSM02a, CLS03a, CLS04, MCLS05, CDLS06b, ALS07], ainsi qu'à plusieurs publications nationales et internationales [CDLSM02b, CLS02a, CLS03d, CLS03b, CLS03c, CLS05b, ACLS04, ACLSL08, CMDK$^+$05], [CLS05a, CLS05c, CDLS06a, CLS07, CDLS07, CMDK$^+$07, ADLS08, ADLS09] (y compris des contributions à des ouvrages de synthèse internationaux, voir [CLS03c, ADLS09, CLS09] et plusieurs articles dans des revues internationales, voir [CLS05b, CMDK$^+$07, ACLSL08]).

Il s'agit d'un des thèmes centraux de mon équipe, ce qui m'a permis de participer à l'encadrement d'une thèse sur ce sujet (Sylvie DOUTRE, voir [Dou02]) ainsi qu'à l'encadrement de plusieurs DEA (Dominique GAY en 2004, Mathieu MARDI en 2005 et Aurore MIQUEL en 2007).

Cette phase de mes travaux a été menée en collaboration avec des membres de mon équipe (en particulier Claudette CAYROL, Leila AMGOUD, Florence DUPIN DE SAINT-CYR et Jérôme MENGIN), mais aussi avec des membres d'autres universités françaises (en particulier avec Sylvie et Pierre MARQUIS et Sébastien KONIECZNY du CRIL à Lens et Caroline DEVRED du LERIA à Angers).

***Passage de l'argumentation aux jeux*** Ici, le lien s'est fait au travers de la notion d'interactions entre agents rationnels. L'argumentation, tout comme les jeux, permet de simuler et de raisonner avec ce type d'interactions. D'autre part, de nombreux liens entre ces deux thèmes ont déjà été mis en évidence.

**Thème 4 : Les jeux** Cet aspect de mon travail est apparu beaucoup plus tard suite à mes travaux sur l'argumentation. En effet, l'argumentation peut être vue comme un processus d'interaction entre agents ; de plus, les jeux, qui servent aussi de modèles d'interactions entre agents rationnels, sont un moyen déjà utilisé pour modéliser des mécanismes d'argumentation (voir [Dun95]). Il m'a donc paru intéressant d'approfondir ce lien entre argumentation et jeux. Après une première étude destinée essentiellement à faire un point sur ce sujet, je me suis intéressée à des jeux très particuliers : les jeux booléens. Leur intérêt était triple :

- utiliser la logique classique pour modéliser des interactions entre agents,
- réutiliser des procédés de représentation de préférence issus de l'inférence non monotone,
- proposer des concepts de solution correspondant à des sous-ensembles consistants.

Cette étude mérite de se poursuivre car ces jeux sont, à mon avis, des outils prometteurs pour simuler un processus d'argumentation.

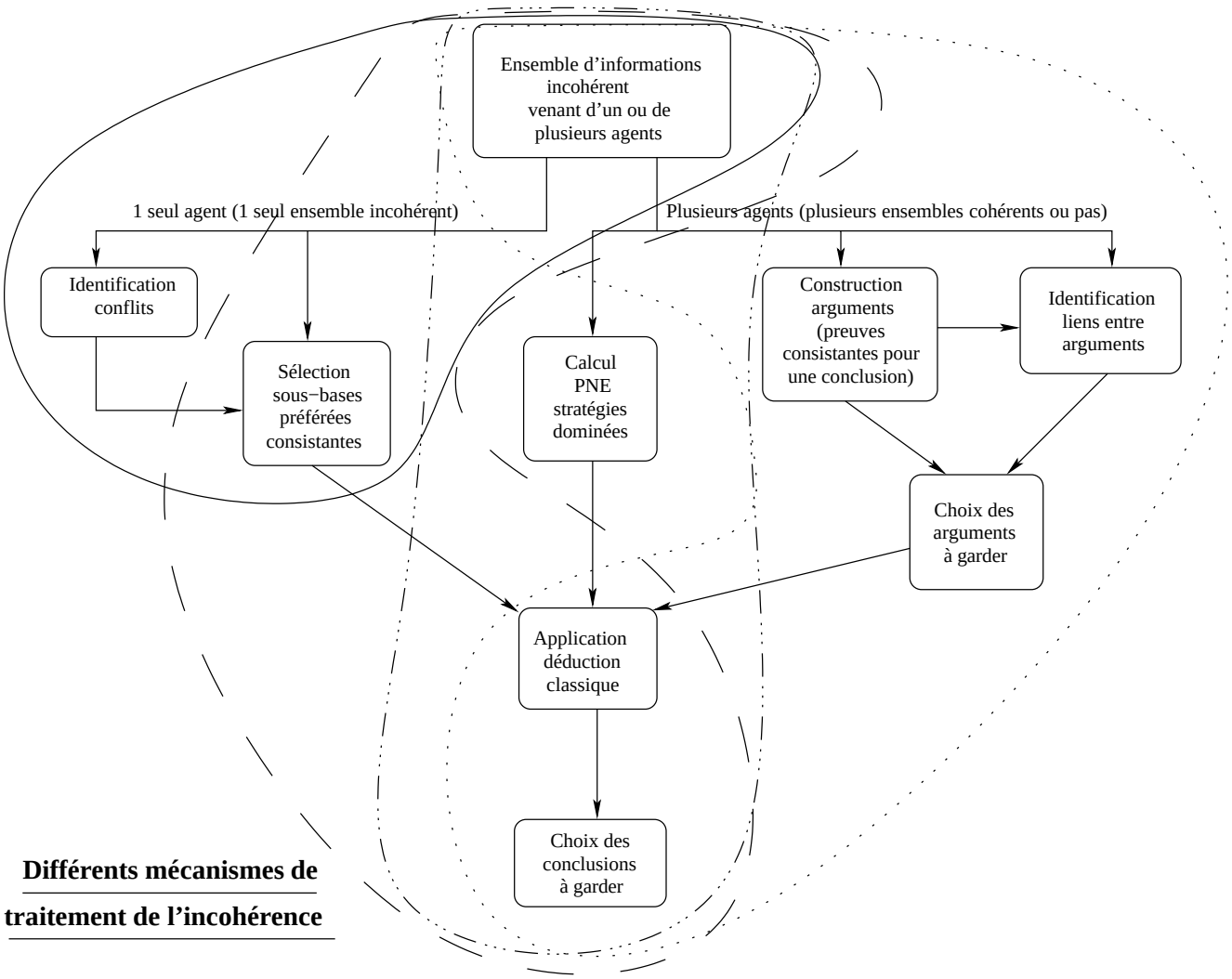
Ces travaux ont donné lieu de 2004 à 2009 à de nombreux rapports de recherche [BLSL06b, BLSL07c], ainsi qu'à plusieurs publications nationales et internationales [BLSL05, BLSL06a, BLSLZ06, BLSL07b, BLSL07d, BLSL07a, BLSL08b, BLSL08a, BLSLZ09] (y compris une contribution à un ouvrage de synthèse international, voir [BLSL08b] et un article dans une revue internationale, voir [BLSLZ09]).

Sur ce thème, j'ai travaillé avec Jérôme LANG, directeur de recherche CNRS à Paris Dauphine (et anciennement chargé de recherche à l'IRIT) et j'ai participé avec lui à l'encadrement d'une thèse (Elise BONZON, voir [Bon07]) ainsi qu'à l'encadrement d'un DEA (Denis SIREYJOL en 2004). De plus, lors de la thèse d'Elise, nous avons mené une collaboration avec Bruno ZANUTTINI (laboratoire du GREYC à Caens).

Tous ces thèmes se retrouvent sur le schéma suivant.

- Entourée d'une ligne avec de larges traits, on y voit la partie "inférence non monotone" qui part de l'ensemble initial, transite par la sélection de sous-bases préférées puis par l'application de la déduction classique et finit par le choix de la conclusion.

– Entourée d'une ligne pleine, on y trouve aussi la "résolution de conflits" qui permet à partir de l'ensemble initial d'identifier les conflits avant de reboucler sur la sélection des sous-bases préférées et de là rejoindre le thème de l'inférence non monotone.
– Entourée d'une ligne pointillée, et partant de l'ensemble initial, le thème "argumentation" produit des arguments et leurs interactions, puis choisit les arguments acceptables avant de reboucler sur l'application de la déduction classique.
– Entourée d'une ligne avec des traits et des points, la partie "jeux" calcule divers concepts de solution (équilibres de Nash – PNE[4] –, stratégies dominées) avant de reboucler elle-aussi sur l'application de la déduction classique.

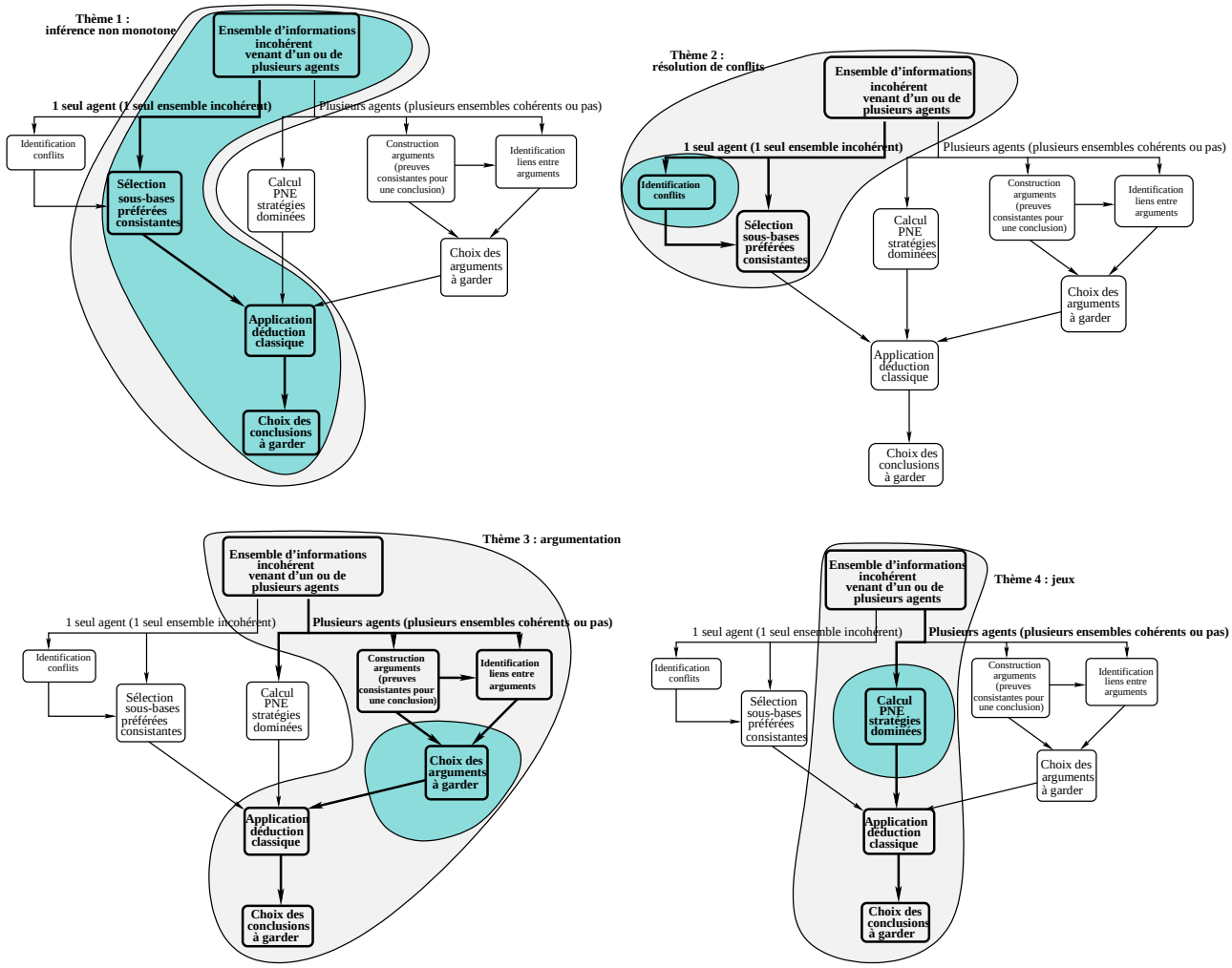



Dans ce schéma on identifie bien le "point d'entrée" : l'ensemble des informations incohérent pouvant être issu d'un ensemble unique (celui d'un seul agent) ou bien construit à partir de l'**union** des informations de chaque agent. Le terme **union** est essentiel : dans tous mes travaux je n'ai pas cherché à étudier la fusion d'information, ni la révision de ces informations *dans le but d'obtenir un ensemble unique qui serait cohérent*. Je me suis toujours placée dans un processus consistant à *extraire des paquets d'information cohérents de l'ensemble initial incohérent* et à raisonner à partir de ces paquets cohérents.

[4] Pure Nash Equilibrium

De même, le “point de sortie” est très clair : il s’agit des informations déduites classiquement par le système quelle que soit la méthode intermédiaire utilisée et quel que soit le type des informations à déduire (nouvelle formule, décision, …).

Sur les schémas suivants, j’ai repris le schéma principal pour mettre en exergue les différents thèmes (parties entourées claires) et les sous-parties de ces thèmes (parties entourées foncées) sur lesquelles j’ai réellement travaillé. Dans le thème 1, j’ai exploré la totalité des étapes, dans le thème 2, je n’ai travaillé que sur la partie d’identification des conflits, sur le thème 3, l’essentiel de mon travail a porté sur le choix des arguments et pour le thème 4, je n’ai abordé que le calcul des concepts de solution :




# Organisation du document

L’organisation est relativement simple, puisqu’il va s’agir de passer en revue les thèmes abordés les uns après les autres.

Le chapitre 1 page 9 est donc consacré aux relations d’inférence non monotone, puis les méthodes de résolution des conflits sont analysées dans le chapitre 2 page 21. Le chapitre 3 page 31 qui est le plus important en taille, eu égard aux nombreuses années passées sur le sujet, porte sur l’argumentation alors que le chapitre 4 page 125 traite des jeux.

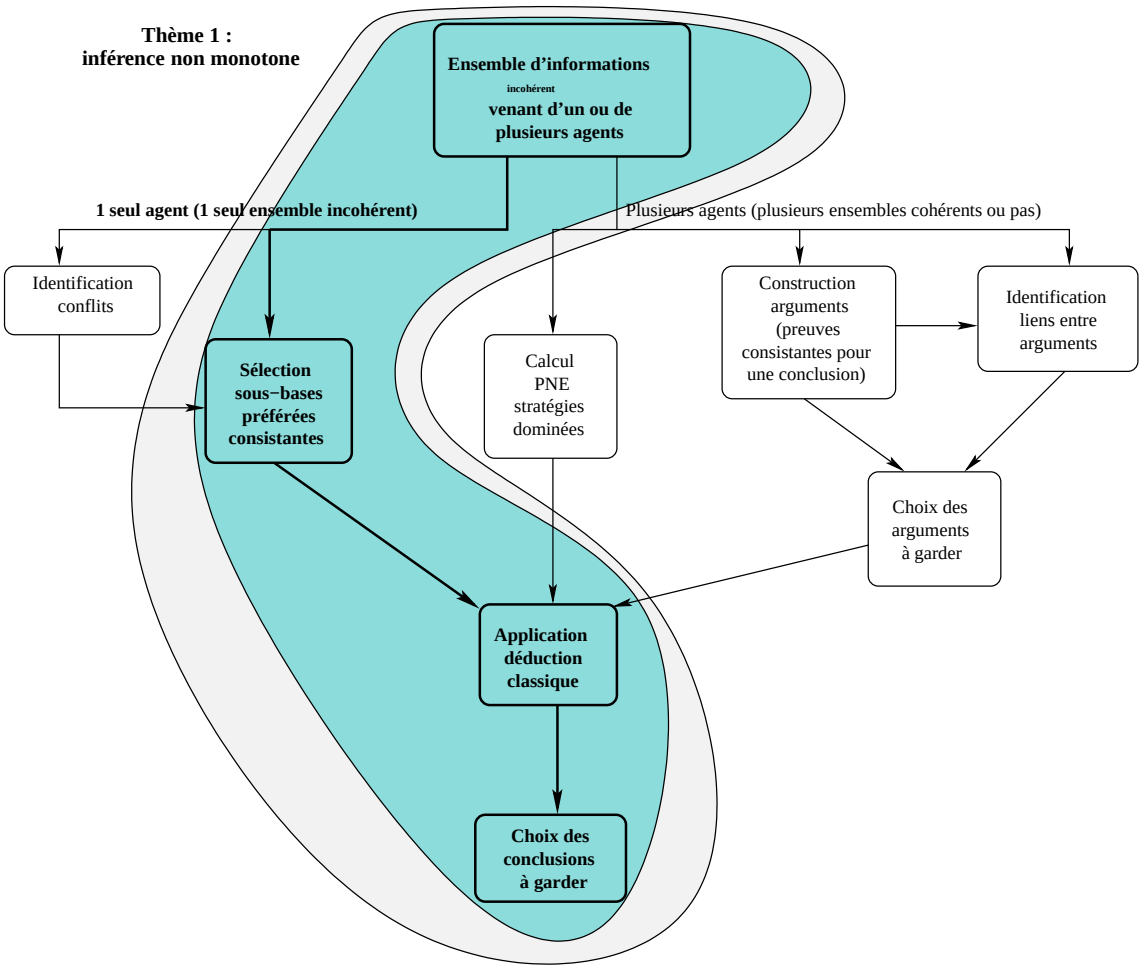
Au début de chacun de ces chapitres le lecteur trouvera une introduction rappelant la nature du thème abordé et comment il se situe par rapport aux autres. Et en fin de chapitre, une conclusion synthétisera les résultats produits pour le thème.

Pour conclure, le chapitre 5 page 153 décrira mon projet de recherche. Et on trouvera en annexe A page 169 diverses informations chiffrées sur mon activité globale de recherche.

# Chapitre 1

# Thème 1 : définition et étude de relations d'inférence non monotone

Le thème abordé ici est représenté par la partie "entourée claire" du schéma suivant et mes travaux sur le sujet présentés dans ce chapitre correspondent à la partie "entourée foncée" :



Il s'agit donc d'extraire d'un ensemble d'informations *incohérent* des sous-ensembles *cohérents* représentant des états possibles du monde et sur lesquels on va pouvoir appliquer la déduction classique ; les relations d'inférence non monotone se définissent alors à partir des résultats de la déduction classique sur ces sous-ensembles. Lors de ma thèse, j'ai étudié les propriétés théoriques et algorithmiques de ces relations d'inférence non monotone. Puis, j'ai poursuivi cette étude en cherchant à généraliser ces relations.

## 1.1 Rappel des résultats obtenus en thèse

Dans cette section, je vais devoir rappeler les définitions et les résultats de ma thèse afin que la lecture des sections et chapitres suivants soit compréhensible.

Au cours de ce processus de définition de relations d'inférence non monotone, il arrive bien souvent que l'on exploite des préférences entre les formules de la base initiale (éventuellement un pré-ordre total sur la base de croyances).

En général, ces relations sont présentées comme étant la combinaison des 2 éléments suivants :

– premièrement un mécanisme de sélection qui fournit un ensemble de sous-bases consistantes issues de la base de croyances initiale ; ce mécanisme peut exploiter ou pas un ordre entre les croyances (voir [Poo88, Bre89, CRS93, DLP91, BCD$^+$93] et [LS95a] pour une synthèse de la plupart de ces travaux) ; on obtient ainsi les "sous-bases" dites "préférées" ;
– puis, on applique un principe d'inférence qui permet de définir la relation d'inférence non monotone à partir de l'inférence classique sur les sous-bases sélectionnées (voir [PL92, LS95a]).

### 1.1.1 Définition formelle d'une inférence non monotone

On part d'une base de croyances stratifiée[1] et considérée de manière syntaxique comme dans [Neb91][2].

Le processus classique de traitement des incohérences peut alors se définir formellement comme suit :

| | |
|---|---|
| $(E, \leq)$ | Base de croyances stratifiée inconsistante |
| $\downarrow$ | *étape 1 : application d'un mécanisme de sélection* $m$ |
| $\{Y_i\}$ | Ensemble de sous-bases $m$-préférées |
| $\downarrow$ | *étape 2 : application d'un principe d'inférence* $p$ |
| $(E, \leq) \mid\!\sim^{p,m} \Phi$ | Conclusion concernant $(E, \leq)$ et $\Phi$ |

avec :

*Les mécanismes de sélection.*

Les méthodes les plus courantes consistent à utiliser des sous-bases de $(E, \leq)$ consistantes maximales (au sens de l'inclusion). D'autre part, il existe différentes manières d'utiliser la relation de priorité $\leq$ sur $E$ pour sélectionner les sous-bases consistantes préférées (voir [CLS94b] pour une synthèse de ces travaux). Ici, je me contenterai d'en rappeler trois : la relation de préférence "Best-Out" issue de la logique possibiliste, une relation de préférence qui combine les priorités et la maximalité pour l'inclusion des sous-bases sélectionnées, et une relation de préférence qui combine les priorités et la maximalité pour la cardinalité.

**Définition 1** *Soit* $X = (X_1 \cup \ldots \cup X_q)$ *et* $Y = (Y_1 \cup \ldots \cup Y_q)$ *deux sous-bases consistantes de* $(E, \leq)$ *(avec* $X_i = (X \cap E_i)$ *et* $Y_i = (Y \cap E_i)$*), on définit :*

[1] Une base stratifiée $E$ est un ensemble de formules partitionné en $q$ strates ($E = E_1 \cup \ldots E_q$) et dans lequel on considère que les formules de la strate $i$ ($i \in [1, q[$) sont plus prioritaires que toutes celles des strates $j > i$.

[2] C'est-à-dire que chaque croyance est un élément d'information bien distinct et que seules les croyances présentes explicitement dans la base sont prises en compte.

Préférence Best-Out *(voir [BCD$^+$93]) : pour $X$ sous-base consistante de $E$, posons $a(X) = \min\{i \mid \exists \Phi \in E_i \setminus X\}$. La préférence Best-Out est le pré-ordre total défini par : $X \ll^{bo} Y$ ssi $a(X) \leq a(Y)$*[3].

Préférence basée sur l'inclusion *(voir [CRS93] et [Gef92] pour des définitions équivalentes) : c'est l'ordre partiel strict défini par : $X \ll^{incl} Y$ ssi $\exists i$ tel que $X_i \subset Y_i$ et $\forall j \mid 1 \leq j < i$, $X_j = Y_j$.*

Préférence basée sur la cardinalité *(voir [Leh95, BCD$^+$93]) : c'est l'ordre strict défini par : $X \ll^{card} Y$ ssi $\exists i$ tel que $|X_i| < |Y_i|$ et $\forall j \mid 1 \leq j < i$, $|X_j| = |Y_j|$ ($|Y|$ dénotant la cardinalité de $Y$).*

Les sous-bases consistantes qui sont maximales pour la préférence Best-Out (resp. basée sur l'inclusion, basée sur la cardinalité) sont appelées les sous-bases bo-(resp. incl, card[4]) préférées de $(E, \leq)$.

La préférence Best-Out n'est pas très sélective puisque certaines sous-bases bo-préférées ne sont même pas maximales pour l'inclusion[5].

La préférence basée sur l'inclusion raffine l'inclusion ensembliste puisque toutes les sous-bases incl-préférées sont maximales pour l'inclusion. D'autre part, elles sont de la forme $(X_1 \cup \ldots \cup X_q)$ telle que l'ensemble $(X_1 \cup \ldots \cup X_i)$ est une sous-base consistante maximale de $(E_1 \cup \ldots \cup E_i)$ pour tout $i = 1 \ldots q$. Cela implique qu'elles sont aussi bo-préférées[6].

La préférence basée sur la cardinalité raffine la préférence basée sur l'inclusion. En effet, toute sous-base card-préférée est aussi incl-préférée, l'inverse étant faux.

Nous noterons T (resp. INCL, CARD, BO) le mécanisme de sélection produisant l'ensemble des sous-bases maximales consistantes (resp. incl-préférées, card-préférées, bo-préférées) de $(E, \leq)$.

*Les principes d'inférence.*

Les plus connus utilisables en présence de sous-bases conflictuelles sont les principes sceptiques et crédules. Il existe toutefois une taxonomie beaucoup plus complète établie par Pinkas et Loui (voir [PL92]) en fonction de la prudence des principes étudiés. Ici, j'en rappellerai trois.

**Définition 2** *Soit $m(E, \leq)$ un ensemble de sous-bases consistantes de $(E, \leq)$ (par exemple, $m(E, \leq)$ peut être obtenu avec l'un des mécanismes* T, INCL, CARD *ou* BO*). Soit $\Phi$ une formule propositionnelle. On définit :*

Le principe UNI *: $\Phi$ est inférée à partir de $m(E, \leq)$ en utilisant le principe d'inférence sceptique (ou universel) si et seulement si $\Phi$ est inférée classiquement par chaque élément de $m(E, \leq)$.*

Le principe EXI *: $\Phi$ est inférée à partir de $m(E, \leq)$ en utilisant le principe d'inférence crédule (ou existentiel) si et seulement si $\Phi$ est inférée classiquement par au moins un des éléments de $m(E, \leq)$*[7].

Le principe ARG[8] *: $\Phi$ est inférée à partir de $m(E, \leq)$ en utilisant le principe d'inférence argumentatif si et seulement si $\Phi$ est inférée classiquement par au moins un élément de $m(E, \leq)$ et qu'aucun élément de $m(E, \leq)$ n'infère classiquement $\neg\Phi$.*

*La relation d'inférence non monotone.*

Nous pouvons donc maintenant donner une définition précise des relations d'inférence non monotone utilisables sur une base de croyances stratifiée. Chacune d'entre elles apparaît au croisement d'un mécanisme de sélection $m$ et d'un principe d'inférence $p$.

**Définition 3** *Soit $(E, \leq)$ une base de croyances stratifiée et $\Phi$ une formule propositionnelle. $(E, \leq) \mid\!\sim^{p,m} \Phi$*[9] *ssi $\Phi$ est inférée à partir de $m(E, \leq)$ en utilisant le principe d'inférence $p$, $m(E, \leq)$ dénotant l'ensemble des sous-bases consistantes de $(E, \leq)$ qui sont préférées au sens du mécanisme $m$.*

[3] Cet ordre dépend seulement de la strate de plus haute priorité dans laquelle au moins une formule a été supprimée pour restaurer la cohérence.

[4] Dans [LS95a], cette relation était notée lex, la notation card étant réservée à une autre relation correspondant à la sélection de sous-bases maximales consistantes et de cardinalité maximale sans tenir compte des préférences.

[5] En effet, soit *amax*$(E) = \mathcal{M}\mathtt{ax}\{i \mid E_1 \cup \ldots \cup E_i$ soit consistant$\}$. Si *amax*$(E) = k$, alors les sous-bases bo-préférées de $E$ sont exactement les sous-bases consistantes de $E$ qui contiennent $(E_1 \cup \ldots \cup E_k)$.

[6] Les sous-bases incl-préférées sont aussi appelées sous-théories dans [Bre89], et correspondent exactement aux sous-bases consistantes fortement maximales de [DLP91].

[7] Il est trivial de montrer que chaque conclusion obtenue à partir de $m(E, \leq)$ par le principe UNI est aussi obtenue par le principe EXI.

[8] Ce principe consiste à ne garder parmi les conséquences crédules que celles dont la négation n'est pas inférée (voir [BDP93] pour une présentation de cette inférence dite argumentative).

[9] On dira que $(E, \leq)$ infère "non monotoniquement" la formule $\Phi$ à l'aide du mécanisme de sélection $m$ et du principe d'inférence $p$.

Par exemple, $m \in$ {T, INCL, CARD, BO} et $p \in$ {UNI, EXI, ARG}.

Ainsi, une relation d'inférence non monotone est représentée par le couple $p$-$m$.

Remarquons qu'une définition plus générale de l'inférence non monotone peut être donnée sous la forme $\Psi \mid\sim^{p,m}_{(E,\leq)} \Phi$. Ce sont les relations $G$-$p$-$m$ évoquées dans [LS95a]. Cette définition plus générale consiste à faire jouer un rôle un peu particulier à la formule $\Psi$ : elle est ajoutée à la base $E$ afin d'en constituer la strate la plus prioritaire. Cela a pour conséquence que ces relations d'inférence non monotone ne sont pas définies pour le mécanisme de sélection T. Dans les autres cas, ces relations donnent les mêmes résultats que les relations $p$-$m$.

**Un exemple illustratif**

Considérons la base de croyances suivante $(E, \leq)$ stratifiée avec 3 niveaux de priorité et représentant le problème très connu des pingouins ($p$ signifie "pingouin", $o$ signifie "oiseau", $v$ signifie "vole", $a$ signifie "a des ailes") :

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

(les formules $p$ et $p \rightarrow o$ sont prioritaires par rapport aux formules $p \rightarrow \neg v$ et $o \rightarrow v$ qui sont elles-mêmes prioritaires par rapport aux formules $o \rightarrow a$ et $a \rightarrow v$).

Parmi toutes les sous-bases consistantes issues de cette base (58 sur cet exemple), nous en isolons cinq :

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

(ce sont les sous-bases T-préférées, les 3 premières étant les sous-bases incl-préférées au sens de Brewka et la troisième étant aussi card-préférée).

Notons que les sous-bases bo-préférées ne sont pas données ici mais elles ont toutes pour caractéristique de contenir les deux formules $p$ et $p \rightarrow o$ et donc d'impliquer $p$ et $o$. Concernant l'inférence des formules réduites aux littéraux (positifs ou négatifs), on a donc :

- avec UNI-T on n'inférera rien,
- avec UNI-BO et UNI-INCL, on inférera $p$ et $o$,
- avec EXI-T, on inférera tous les littéraux,
- avec EXI-BO et EXI-INCL, on inférera aussi bien $v$ que $\neg v$ et $a$ que $\neg a$ tout en inférant $p$ et $o$
- avec ARG-T, on n'inférera rien,
- avec ARG-BO et ARG-INCL, on inférera $p$ et $o$,
- avec EXI-CARD, ARG-CARD et UNI-CARD, on inférera $p$, $o$, $v$ et $a$.

### 1.1.2 L'analyse comparative des relations d'inférence non monotone $p$-$m$

Cette analyse repose sur 4 critères de comparaison :

**La complexité** C'est un critère capital quand on pense efficacité ; que va coûter la réponse à la question "la base $E$ infère-t-elle la formule $\Phi$" ? le coût de cette réponse peut être soit d'ordre temporel, soit d'ordre spatial et l'expérience nous apprend que c'est le coût en temps qui pose problème en premier (voir [GJ79, Neb91, Got92]) ;

**La prudence** C'est un critère subjectif et dépendant du contexte ; malgré cela, elle reste un critère important car elle cherche à exprimer le réalisme d'une relation du point de vue du nombre de conclusions à obtenir (plus une relation est prudente et moins elle produit de conclusions, donc moins elle risque d'inférer des conclusions erronées tout en risquant d'oublier d'en inférer certaines) (voir [PL92]) ;

**Les propriétés de déduction** C'est un critère important car elles définissent une approche axiomatique de l'inférence non monotone et ce faisant elles représentent un "idéal"[10] vers lequel ces relations devraient tendre ; cet idéal a été défini par quelques spécialistes du domaine sous diverses formes qui se recoupent à peu près toutes ; ce sont les propriétés de déduction qui prouvent l'adéquation de la relation au raisonnement que l'on cherche à modéliser au travers des dites relations (voir [KLM90, GM94]).

**Les algorithmes** Du point de vue pratique c'est un critère essentiel.

Notons qu'il existe d'autres critères, que l'on peut étudier, mais que je n'ai pas traités : insensibilité à la redondance, à la forme clausale ... (voir par exemple [BDP94]).

Sachant que toute cette partie concerne mon travail de thèse et que ce n'est pas le sujet de ce document (qui doit traiter du travail réalisé depuis la thèse), je ne m'attarderai donc pas sur les résultats obtenus. Par contre le lecteur intéressé peut trouver tous ces résultats dans les publications suivantes :

- pour la partie complexité [CLS94b, CLSS96a, CLSS98],
- pour la prudence et les postulats de rationalité [CLS95],
- pour l'étude algorithmique [CLSS96a, CLSS98].

## 1.2 Une étude complémentaire : la généralisation de la définition d'une relation d'inférence non monotone

Cette étude a été menée après la thèse et a consisté à définir un cadre plus général pour décrire les relations d'inférence non monotone, dans lequel les relations que j'avais étudiées lors de ma thèse pouvaient s'inscrire sans difficulté.

### 1.2.1 Introduction

L'exemple du pingouin donné en section 1.1.1 page 10 met en évidence certains problèmes liés au mode de définition classique des relations d'inférence non monotone :

- tout d'abord, *comment définir un "bon" mécanisme de sélection, ni trop, ni trop peu sélectif* :
  - si nous sélectionnons toutes les sous-bases consistantes de $(E, \leq)$, nous aurons beaucoup trop de sous-bases à gérer et cela pose un problème d'ordre calculatoire ;
  - si nous ne sélectionnons que trop peu de sous-bases, nous risquons d'"oublier" des croyances intéressantes : sur l'exemple du pingouin, si on sélectionne uniquement la sous-base $Y_3$, on perd une information importante : les pingouins peuvent ne pas voler ;
- d'autre part, *comment définir un "bon" principe d'inférence, ni trop prudent, ni trop permissif* :
  - s'il faut l'unanimité des sous-bases pour pouvoir conclure (principe sceptique), très peu de conclusions seront obtenues ; sur l'exemple du pingouin, si nous raisonnons avec $Y_1$, $Y_2$ et $Y_3$, nous ne pourrons rien conclure à propos des formules $v$ et $a$ ; nous obtenons le même manque de résultat avec le principe suivant : "au moins une sous-base pour et aucune contre" (principe dit argumentatif proposé par [BDP93]) ;
  - par contre, avec le principe "au moins une sous-base pour" (principe crédule), nous tombons dans l'excès inverse en obtenant trop de conclusions (en particulier une formule et son contraire, $v$ et $\neg v$, $a$ et $\neg a$ dans l'exemple du pingouin en raisonnant avec $Y_1$, $Y_2$ et $Y_3$).

Afin de résoudre ces divers problèmes, nous avons exploité les idées suivantes :

[10] Cet "idéal" n'est pas encore parfait, car il existe des relations d'inférence le vérifiant et présentant par ailleurs des problèmes importants, tout comme il existe des relations d'inférence ne le vérifiant pas et qui demeurent pourtant des relations intéressantes par d'autres aspects.

– l'inférence d'une conclusion $\Phi$ est vue comme un problème décisionnel (il faut savoir si "oui ou non" on infère $\Phi$ de notre base de connaissances inconsistante) et on souhaiterait que le mécanisme mis en place pour résoudre ce problème soit sain (si on infère $\Phi$, on ne doit pas inférer $\neg\Phi$) ;
– d'un point de vue plus général, on peut voir les sous-bases sélectionnées comme des "opinions" divergentes (voir sur l'exemple comment les sous-bases $Y_1$ et $Y_2$ qui sont d'accord sur le fait que le pingouin ne vole pas, ne sont plus d'accord dès qu'il s'agit de savoir si le pingouin a des ailes) ;
– donc la sélection des "sous-bases préférées" revient à définir un ensemble d'"opinions" (là-encore avec un *mécanisme de sélection* : la seule contrainte à respecter étant la consistance de ces "opinions") ;
– puis nous déterminons le poids de chacune de ces "opinions" avec un *mécanisme de pondération* qui prend en compte le fait que certaines "opinions" sont plus pertinentes que d'autres ;
– ensuite, le principe d'inférence est remplacé par un *mécanisme de vote* défini en fonction de ce que représentent les "opinions" ; ainsi, nous avons un processus de compensation entre les "opinions pour" et les "opinions contre" qui prend en compte le poids de chaque opinion ; nous serons alors capables de raisonner sur une formule $\Phi$ même s'il existe des "opinions contre" $\Phi$ (donc "pour" $\neg\Phi$) ;
– et enfin, nous concluons après avoir "compté les voix" (*mécanisme de décompte*).

Ainsi, nous nous donnons la possibilité de définir un ensemble réaliste d'"opinions", de mesurer leur "force" (ce qui sous-entend que toutes les "opinions" peuvent ne pas être de la même "force"), de demander leur avis sur une donnée particulière et de faire la synthèse de tous ces avis. Il s'agit là d'une approche très *décisionnelle* qui nous paraît adaptée au fait qu'il existe un conflit à résoudre (les "opinions pour" et les "opinions contre").

### 1.2.2 Processus global de traitement des incohérences

On part encore d'une base de croyances stratifiée inconsistante vue de manière syntaxique et on applique le processus suivant :

| | |
|---|---|
| $(E, \leq)$ | Base de croyances stratifiée inconsistante |
| $\downarrow$ | *étape 1 : mécanisme de sélection* |
| $\{Y_i\}$ | Ensemble d'"opinions" (sous-bases consistantes) |
| $\downarrow$ | *étape 2 : mécanisme de pondération* |
| $\{(Y_i, p(Y_i))\}$ | Ensemble d' "opinions" pondérées |
| $\downarrow$ | *étape 3 : mécanisme de vote* |
| $n(\Phi)$ | Vote des "opinions" pour une formule $\Phi$ |
| $n(\neg\Phi)$ | Vote des "opinions" contre une formule $\Phi$ |
| $n(\text{ni } \Phi, \text{ ni } \neg\Phi)$ | Vote des "opinions" indifférentes à $\Phi$ |
| $\downarrow$ | *étape 4 : mécanisme de décompte des voix* |
| $(E, \leq) \mid\sim \Phi$ | Conclusion concernant $(E, \leq)$ et $\Phi$ |

Aucun processus global de ce type n'a encore été proposé à notre connaissance dans la littérature. Par contre, différentes parties de ce processus ont déjà été étudiées mais, à chaque fois, de manière isolée :
– de nombreux mécanismes de sélection ont été proposés dans [Bre89, CRS93, DLP91, BCD$^+$93] ;
– l'étape 2 a déjà été traitée d'un point de vue possibiliste (voir [DLP91]) ainsi que d'un point de vue probabiliste (voir [Lan94]) ;
– l'étape 3 a été étudiée dans [Lan94] en utilisant des fonctions de croyances (voir [Sme93]) ;
– l'étape 4 a été abordée dans [PL92, Lan94].

En effet, chacun de ces travaux avait été réalisé dans un but bien spécifique :
– la définition d'une taxonomie de relations d'inférence non monotone pour [PL92],
– la révision des croyances pour [Sme93],
– le diagnostic pour [Lan94],
– Le raisonnement non monotone et/ou la logique des défauts pour [Bre89, CRS93, DLP91, BCD$^+$93].

La diversité des thèmes abordés et l'abondance des travaux prouvent l'intérêt de ce processus permettant d'englober dans un même cadre plusieurs contextes de recherche.

D'autre part, ce cadre de travail offre une très grande flexibilité et un pouvoir d'expression important : en effet, à chaque étape, plusieurs choix sont possibles et chaque étape peut être contrôlée de manière indépendante.

Dans la section suivante, nous allons utiliser ce processus dans le contexte du raisonnement non monotone.

### 1.2.3 Application au raisonnement non monotone

#### 1.2.3.1 Où l'on retrouve des relations d'inférence non monotone connues

Les relations d'inférence non monotone $p$-$m$ précédemment décrites peuvent être obtenues à l'aide du processus général décrit dans la section 1.2.2 page précédente :

- pour l'étape 1, nous choisissons le mécanisme de sélection $m$ (avec $m \in$ {T, INCL, CARD, BO}) ;
- pour l'étape 2, nous considérons que chaque sous-base a le même poids 1 ;
- pour l'étape 3, nous définissons :
  - $n(\Phi) =$ somme des poids des sous-bases sélectionnées qui infèrent classiquement la formule $\Phi$[11] ;
  - $n(\neg\Phi) =$ somme des poids des sous-bases sélectionnées qui infèrent classiquement $\neg\Phi$ ;
  - $n(\text{ni } \Phi, \text{ ni } \neg\Phi) =$ somme des poids des sous-bases sélectionnées qui n'infèrent classiquement ni $\Phi$,ni $\neg\Phi$ ;
- pour l'étape 4 :
  - le principe d'inférence UNI est retrouvé en imposant que : $n(\Phi) > 0$, $n(\neg\Phi) = 0$ et $n(\text{ni } \Phi, \text{ ni } \neg\Phi) = 0$ ;
  - le principe d'inférence EXI est retrouvé en imposant que : $n(\Phi) > 0$ ;
  - le principe d'inférence ARG est retrouvé en imposant que : $n(\Phi) > 0$ et $n(\neg\Phi) = 0$.

#### 1.2.3.2 Quelques "nouvelles" relations d'inférence non monotone

Parmi toutes les relations d'inférence non monotone que l'on peut définir à l'aide du processus donné dans ce papier, nous avons choisi d'en présenter deux :

- Le premier exemple est une variante des relations $p$-INCL (avec $p \in$ {UNI, EXI, ARG}) dans laquelle nous avons modifié l'étape 4 (le mécanisme de décompte) ;
- Le second exemple est plus complexe et lié au mode de raisonnement probabiliste.

##### 1.2.3.2.1 GRAD-EQUADD-INCL

Cette relation d'inférence non monotone est définie de la manière suivante :

- pour l'étape 1, nous choisissons le mécanisme de sélection INCL ;
- pour les étapes 2 et 3, nous conservons les choix donnés pour les relations d'inférence non monotone présentées dans la section 1.2.3.1 ;
- pour l'étape 4, nous imposons que $n(\Phi) > n(\neg\Phi)$ et $n(\Phi) \geq n(\text{ni } \Phi, \text{ ni } \neg\Phi)$ : le nombre de sous-bases sélectionnées qui infèrent classiquement $\Phi$ est strictement supérieur à celui des sous-bases sélectionnées qui infèrent classiquement $\neg\Phi$, et supérieur ou égal à celui des sous-bases sélectionnées qui n'infèrent ni $\Phi$, ni $\neg\Phi$.

Nous avons ainsi une relation qui permet de prendre en compte un ensemble d'"opinions" réaliste et qui s'appuie sur un vote démocratique pour décider quelle formule conclure.

[11] Ainsi, étant donné le choix effectué à l'étape 2, $n(\Phi)$ représente le nombre de sous-bases sélectionnées qui infèrent classiquement $\Phi$.

#### 1.2.3.2.2 GRAD-POSMAX-T

Cette relation d'inférence non monotone est définie de la manière suivante :

- pour l'étape 1, nous choisissons le mécanisme de sélection T ;
- pour l'étape 2, nous utilisons les définitions suivantes :
  **Définition 4** *(voir [DLP91]) Soit $Y_i$ une sous-base de $(E, \leq)$, le* poids possibiliste *de $Y_i$ noté avec $p_{po}(Y_i)$ est défini par : si $Y_i \neq E$ alors $p_{po}(Y_i) = x$ ($x$ = numéro de la strate de plus haute priorité dans laquelle on a supprimé une formule afin de restaurer la cohérence) et sinon $p_{po}(E) = q + 1$ ($q$ = nombre max de strates dans $(E, \leq)$).*
- pour l'étape 3, nous définissons :
  - $n(\Phi) =$ max des poids des sous-bases sélectionnées qui infèrent classiquement $\Phi$ ;
  - $n(\neg\Phi) =$ max des poids des sous-bases sélectionnées qui infèrent classiquement $\neg\Phi$ ;
  - $n(\text{ni } \Phi, \text{ ni } \neg\Phi) =$ max des poids des sous-bases sélectionnées qui n'infèrent classiquement ni $\Phi$, ni $\neg\Phi$ ;
- pour l'étape 4, nous imposons que $n(\Phi) > n(\neg\Phi)$ et $n(\Phi) \geq n(\text{ni } \Phi, \text{ ni } \neg\Phi)$.

Ici, nous utilisons le fait que la stratification de la base induit l'aspect probabiliste suivant : les "opinions" ayant le plus fort poids sont celles qui ont la plus grande probabilité d'avoir raison (si on considère que les formules de plus forte priorité sont celles qui ont le plus de chances d'être vraies). Et cette approche est encore renforcée par l'opérateur "max" choisi pour l'étape 3.

Remarquons que, par définition, ces deux "nouvelles" relations d'inférence non monotone sont saines : il n'est pas possible d'inférer à la fois une formule et sa négation.

Étudions maintenant ces "nouvelles" relations, d'abord sur un exemple, puis d'un point de vue théorique, comme ce qui avait été fait dans [LS95a].

### 1.2.3.3 Un exemple illustratif

Considérons la base de croyances suivante composée de 4 strates représentant une nouvelle version du problème des pingouins ($p$ signifie "pingouin", $o$ signifie "oiseau", $v$ signifie "vole", $a$ signifie "a des ailes", $pl$ signifie "a des plumes") :

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

Il est important de noter que la validité des résultats obtenus est fonction de l'intuition que nous avons des résultats attendus. Ici, par exemple, il semble intuitivement correct de conclure $\neg v$ et $pl$, étant donné la base de connaissances et la stratification choisies[12].

Examinons donc les résultats obtenus concernant ces deux formules par les différentes relations d'inférence non monotone ("anciennes" et "nouvelles") :

- ni $\neg v$, ni $pl$ avec UNI-T,
- seulement $\neg v$ avec UNI-BO, UNI-INCL et UNI-CARD,
- $v$, $\neg v$, $pl$ avec EXI-T,
- seulement $pl$ avec ARG-T,
- $\neg v$, $pl$ avec $p$-$m$ ($p \in$ {EXI, ARG} et $m \in$ {BO, INCL, CARD}), et avec GRAD-EQUADD-INCL et GRAD-POSMAX-T.

[12] Il est bien entendu qu'une autre base et une autre stratification induisent un autre ensemble de résultats intuitifs attendus. Le but est donc ici de comparer pour une *même base et une même stratification* les résultats obtenus par différentes relations d'inférence non monotone par rapport aux résultats intuitifs attendus.

Dans cet exemple, nous obtenons de meilleurs résultats avec les “nouvelles” relations d’inférence non monotone qu’avec les relations UNI-$m$ et certaines des relations EXI-$m$ et ARG-$m$.

Dans [CLS97], beaucoup d’autres exemples sont donnés, qui permettent de conclure, qu’en général, on arrive à définir à l’aide du processus présenté dans ce papier des relations fournissant des résultats tout à fait convenables du point de vue intuitif et bien souvent meilleurs que ceux obtenus par des relations d’inférence non monotone bien connues.

#### 1.2.3.4 Étude théorique

Dans cette section, nous comparons les diverses relations décrites précédemment à l’aide des deux points de vue suivants :
– la prudence[13] des relations (voir [PL92]),
– le respect des règles de déduction parfois appelées postulats de rationalité[14] (voir [CLS95]).
Nous donnerons aussi quelques indications concernant la complexité de ces relations.

##### 1.2.3.4.1 Prudence des relations

Nous complétons les résultats déjà connus sur ce sujet (présentés dans [CLS95]) par de nouveaux résultats (dont les preuves sont données dans [CLS97]). Nous obtenons ainsi la taxonomie décrite par la figure 1.1.

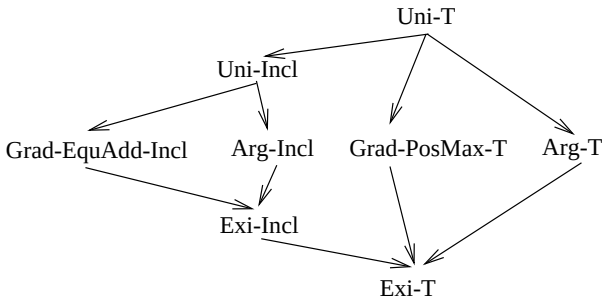



FIG. 1.1 – Comparaison du point de vue de la prudence (les flèches signifient “plus prudent que”)

Ces résultats montrent que les “nouvelles” relations introduites grâce à notre processus ont le même niveau de prudence que les relations du type ARG-$m$ (avec $m \in \{\text{T}, \text{INCL}\}$).

##### 1.2.3.4.2 Respect des règles de déduction

Nous obtenons les résultats suivants pour les “nouvelles” relations d’inférence définies à l’aide de notre processus (les preuves sont dans [CLS97]) :

[13] Soit 2 relations d’inférence $R_1$ et $R_2$, $R_1$ est *plus prudente que* $R_2$ ssi chaque conclusion obtenue par $R_1$ est aussi obtenue par $R_2$.
[14] Les postulats que nous utiliserons ici sont les suivants :
réflexivité : $\alpha \mid\!\sim \alpha$,
équivalence logique gauche : $\dfrac{\models \alpha \leftrightarrow \beta \,;\, \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma}$,
affaiblissement droit (“right weakening”) : $\dfrac{\models \alpha \rightarrow \beta \,;\, \gamma \mid\!\sim \alpha}{\gamma \mid\!\sim \beta}$,
supra-classicité : $\dfrac{\alpha \vdash \beta}{\alpha \mid\!\sim \beta}$,
conditionnalisation faible : $\dfrac{\alpha \mid\!\sim \beta}{\mid\!\sim \alpha \rightarrow \beta}$,
préservation de la consistance : $\dfrac{\alpha \mid\!\sim \bot}{\alpha \vdash \bot}$ (avec $\bot$ représentant la contradiction).

– GRAD-EQUADD-INCL vérifie seulement les propriétés de réflexivité, d'équivalence logique gauche, de supra-classicité et de préservation de la consistance ;
– GRAD-POSMAX-T vérifie seulement les propriétés de réflexivité, d'équivalence logique gauche, de supra-classicité, de préservation de la consistance, d'affaiblissement droit et de conditionnalisation faible.

Ces résultats confirment le fait que les "nouvelles" relations d'inférence ont le même comportement que les relations ARG-$m$.

Rappelons au passage que les résultats de la comparaison théorique semblent privilégier certaines relations (par exemple celles utilisant le principe UNI – voir [CLS97]). Il nous semble que ceci constitue la preuve qu'il est urgent de définir de nouveaux critères de comparaison plus réalistes, ce qui devrait être facilité par le fait que l'on s'inscrit dans un processus général permettant d'établir un lien avec d'autres domaines de recherche.

#### 1.2.3.4.3 Quelques remarques sur la complexité

Nous n'avons pas mené d'étude de complexité. Toutefois, on peut constater facilement que répondre à la question $E \mid\!\sim \Phi$, avec $\mid\!\sim$ une relation non monotone définie à l'aide du processus proposé ici, nécessite (pour la plupart des $\mid\!\sim$ possibles) le calcul de toutes les "opinions" avant de pouvoir implémenter la phase de vote, ce qui correspond en général à une complexité très importante (par exemple $\Delta_3^p$ pour certains cas – voir l'étude de complexité de ARG-$m$ dans [LS95a]). Bien sûr, cette analyse doit être affinée, en particulier pour la recherche de cas particuliers menant à une complexité plus faible (voir par exemple, l'étude de complexité de UNI-CARD toujours dans [LS95a]).

### 1.2.4 Conclusion sur l'étude complémentaire

Dans le contexte du raisonnement non monotone, le processus décrit dans la section 1.2.2 page 14 nous permet :

– de retrouver les relations d'inférence non monotone bien connues et en particulier toutes les $p$-$m$ décrites dans la section 1.1 page 10 et dans [CLS95] ;
– de définir de "nouvelles" relations d'inférence non monotone (par exemple, les deux relations décrites en section 1.2.3.2 page 15) qui paraissent prometteuses du point de vue des résultats obtenus (proches des résultats intuitifs attendus) ;
– de comparer toutes ces relations entre elles avec différents points de vue (par exemple, la prudence et le respect des règles de déduction – voir la section 1.2.3.4 page précédente).

Mais, ce n'est pas le seul domaine d'application. On pourrait par exemple l'appliquer au contexte de la fusion de bases de croyances : chaque ensemble de croyances provenant d'un expert peut être considéré comme une "opinion", auquel cas, le mécanisme de sélection étant inutile, on part directement de l'étape 2 pour effectuer notre raisonnement ; le mécanisme de pondération serait alors lié à une hiérarchie entre les experts (pour un sujet donné, certains experts sont plus compétents que d'autres).

Nous pensons donc que cette approche est tout à fait prometteuse, puisqu'elle permet d'unir dans un même cadre divers modes de raisonnement devant gérer les inconsistances d'un ensemble de croyances à l'aide de sous-bases consistantes de l'ensemble initial et de préférences sur les formules de cet ensemble. Nous pouvons ainsi espérer découvrir de nouveaux liens entre des domaines jusqu'à présent distincts, voire utiliser des outils définis pour un domaine particulier dans un autre domaine. On pourrait, en particulier, exploiter les travaux concernant la théorie de la décision, la théorie du vote et la théorie du choix social afin de définir de nouveaux outils de comparaison pour les relations d'inférence non monotone définies à l'aide du processus présenté ici.

## 1.3 Conclusion sur le thème 1

Dans ce chapitre, j'ai présenté les travaux que j'ai réalisés sur le thème de l'inférence non monotone dans le cadre strict de la restauration de cohérence.

Ces travaux se sont échelonnés sur une période de 8 ans en incluant ma thèse.

Les résultats obtenus peuvent être résumés de la manière suivante :
– pendant la thèse : étude des propriétés théoriques (complexité temporelle dans le pire des cas, prudence, axiomatisation) et algorithmiques des relations d'inférence non monotone issues de la restauration de cohérence,
– après la thèse : définition d'un cadre plus général pour décrire ces relations qui permet la définition de nouvelles relations d'inférence non monotone, et étude de ces nouvelles relations.

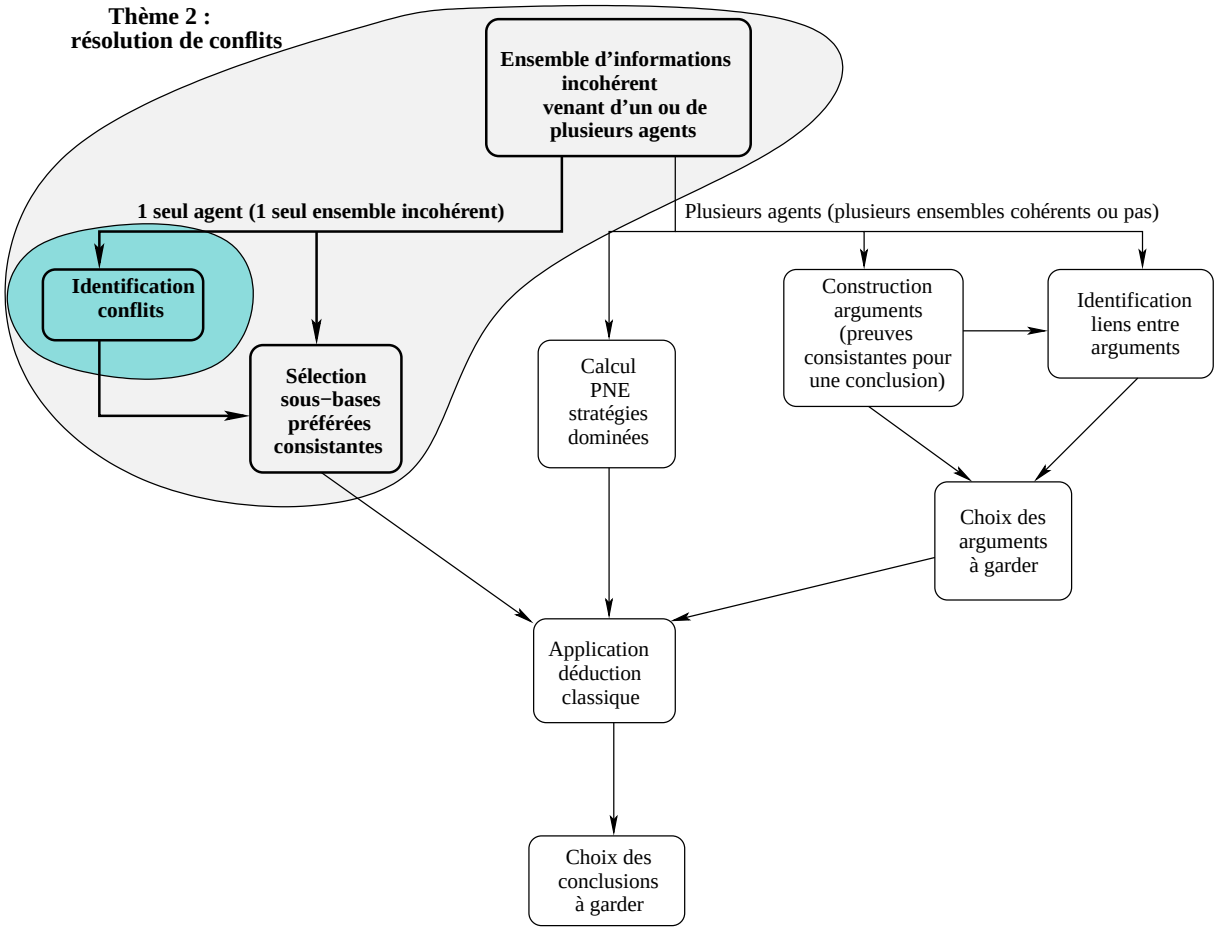
Tous ces travaux ont permis d'identifier une relation d'inférence non monotone particulièrement intéressante du point de vue de ses propriétés. La définition d'un cadre plus général a aussi permis de trouver quelques autres relations qui, bien que moins performantes théoriquement, semblent plus proches de l'intuition.

Cela m'a aussi permis de constater que les mécanismes mis en œuvre pour comparer les relations d'inférence non monotone sont divers et bien souvent en opposition (une relation qui présente des propriétés intéressantes est bien souvent trop prudente, alors qu'une relation peut donner des résultats intuitivement meilleurs mais ne pas respecter des propriétés jugées importantes). Ces remarques sont bien-sûr applicables à toutes tentatives de comparaison de raisonnements non monotones et seront en particulier à prendre en compte lors de mon projet de recherche (voir le chapitre 5 page 153). Elles induisent toute une série de questions parmi lesquelles la question suivante : "pourquoi choisir un mécanisme de sélection plutôt qu'un autre ?" Et c'est cette question qui m'a amenée au thème suivant, la résolution des conflits.

# Chapitre 2

# Thème 2 : méthodes de résolution de conflit

Le thème abordé ici est représenté par la partie “entourée claire” du schéma suivant et mes travaux sur le sujet présentés dans ce chapitre correspondent à la partie “entourée foncée” :




Il s'agit ici de mieux étudier le mécanisme de sélection des sous-ensembles cohérents issus d'un ensemble d'informations incohérent.

Le problème de la restauration de la cohérence d'une base de croyances inconsistante guidée par des relations de préférence entre les croyances est au cœur de nombreuses problématiques en Intelligence Artificielle (raisonnement

non monotone, révision, ...).

Dans ce cadre, il existe deux approches duales :
- Soit on sélectionne directement à partir de la base initiale des sous-ensembles consistants de cette base[1]. Ce principe est à la base du travail décrit dans le chapitre 1 page 9.
- Soit on identifie d'abord les conflits existant dans la base initiale (ce qui correspond à une approche "ATMS" avec calcul des no-goods – voir [Kle86]) puis on cherche à les résoudre en abandonnant des croyances, ce qui nous ramènera, là-encore, à des sous-ensembles consistants de la base initiale.

Dans le cas où il n'existe pas de préférence entre les croyances, ces deux méthodes conduisent aux mêmes résultats[2].

La prise en compte de préférences permet de distinguer les deux approches du point de vue sémantique :

- Dans le cas d'une restauration de cohérence par sélection directe de sous-ensembles consistants, on attribue aux préférences une sémantique que l'on peut qualifier de *globale* (comment se situe une formule donnée par rapport aux autres formules de la base ?).
- Alors que, dans le cas d'une résolution de conflits, on exploite les préférences sous l'angle d'une sémantique dite *locale* (comment se situent les formules d'un conflit donné les unes par rapport aux autres ?).

L'approche locale est essentielle dans le cas où l'on dispose de plusieurs relations de préférence qui peuvent être incompatibles (par exemple, la croyance $a$ sera préférée à la croyance $b$ dans un conflit $c_1$ alors que dans le conflit $c_2$ ce sera l'inverse !). Il s'agit donc d'une approche plus générale que l'approche globale. Notons d'autre part que des préférences locales permettent aussi d'exprimer plus facilement des préférences dépendant d'un contexte ou des préférences conditionnelles (voir [Bre94, PS97, BG98, Gar98]).

Nous allons donc aborder le problème de résolution des conflits dans une base inconsistante équipée d'une relation de préférence, mais sous un angle un peu particulier. En effet, il existe de très nombreux travaux sur ce sujet, mais la plupart du temps, il s'agit de contributions très "expérimentales" et très liées aux applications. Or, nous souhaiterions définir ce problème et ses solutions de manière plus systématique en définissant des principes à respecter et des solutions respectant ces principes.

Ce travail a été réalisé avec Claudette CAYROL et Jérôme MENGIN et ils sont présentés dans un article [CLSM00] (le lecteur y trouvera les preuves des propriétés données ici ainsi que de nombreux autres détails sur ce travail).

## 2.1 L'approche méthodologique

### 2.1.1 Les définitions et principe de base

Soit $L$ un langage logique quelconque. Soit $E$ un ensemble de formules de $L$ représentant des croyances, et $K$ un ensemble consistant de formules de $L$ représentant des connaissances incontestables. L'union $E \cup K$ pourra être inconsistante et $E$ sera appelé une base de croyances.

Le problème de la restauration de cohérence peut être décrit comme étant le suivant :

> *Restaurer la $K$-consistance* consiste à trouver un sous-ensemble $K$-consistant[3] de $E$. En général, on privilégie les solutions maximales pour l'inclusion (voir [Bre89, CRS93, DLP91, BCD+93, LS95a] et le chapitre 1 page 9 de ce document).

Quant au problème de la résolution de conflits, il peut être décrit de la manière suivante :

> $E$ est supposée $K$-inconsistante et elle est donnée sous la forme d'un ensemble de $K$-*conflits* (c'est-à-dire les sous-ensembles de $E$ minimaux pour l'inclusion parmi ceux qui sont $K$-inconsistants).

[1] Ce que l'on peut faire sans calculer nécessairement tous les conflits.

[2] Remarquons que, si on dispose de la base sous la forme d'un ensemble de formules, alors la première méthode sera plus rapide à utiliser que la seconde (en effet, on peut éviter le calcul des conflits). À l'inverse, si la base nous est donnée sous la forme d'un ensemble de conflits, il sera probablement plus judicieux d'appliquer la deuxième méthode.

[3] Un ensemble $Y$ est $K$-consistant ssi $Y \cup K$ est consistant.

On notera $\mathcal{C}(E) = \{c_i\}$ cet ensemble de $K$-conflits issus de $E$ (chaque $c_i$ étant un $K$-conflit), $\mathcal{N}(E) = \bigcup_i c_i$ le noyau de $E$ et $\mathcal{L}(E)$ l'ensemble des formules de $E$ ne participant à aucun conflit (formules *libres* de $E$). On a alors $E = \mathcal{N}(E) \cup \mathcal{L}(E)$[4].

*Résoudre les $K$-conflits* consiste à déterminer un sous-ensemble de $E$ qui ne contienne aucun des $K$-conflits. Il faut trouver, pour chaque conflit, au moins une formule à supprimer et il suffit de travailler sur $\mathcal{N}$ puisque les formules de $\mathcal{L}$ ne sont pas impliquées dans la $K$-inconsistance.

Nous avons donc la définition suivante qui caractérise ce que l'on appelle une solution d'un problème de conflits.

**Définition 5** *Soit $\mathcal{C} = \{c_i\}$ un problème de conflits. $S$ est une* solution[5] *de $\mathcal{C}$ ssi $S$ est un sous-ensemble de $\mathcal{N}$ (union des $c_i$) tel que $c_i \cap S \neq \varnothing, \forall i$.*

**Exemple 1** *Soit $\mathcal{C}$ l'ensemble des conflits suivant : $\mathcal{C} = \{\{a, b\}, \{b, c\}\}$ ($a$, $b$, $c$ étant 3 formules propositionnelles).*

*Si on applique la définition 5, on obtient les 5 solutions suivantes : $S_1 = \{a, b, c\}$, $S_2 = \{a, b\}$, $S_3 = \{b, c\}$, $S_4 = \{a, c\}$, $S_5 = \{b\}$.*

*Remarquons que $S_4$ et $S_5$ semblent plus intéressantes que les 3 autres, puisqu'on ne supprime qu'une seule formule par conflit.*

Par exemple, si on se place dans le cadre du diagnostic basé sur la consistance, chaque conflit est associé à un ensemble de composants ne pouvant avoir un fonctionnement correct en même temps étant donné les observations effectuées ; donc supprimer une formule d'un conflit correspond à supposer un composant en panne. Et on préfère généralement supposer le moins de pannes possible.

Tout conflit doit certes être résolu mais l'élimination d'une formule peut résoudre plusieurs conflits à la fois. Le souci de n'éliminer qu'une seule formule par conflit correspond à ce que nous appelons le *principe de minimalité* (voir section 2.1.2), et nous amène à la définition suivante.

**Définition 6** *Soit $\mathcal{C} = \{c_i\}$ un problème de conflits. $S$ est une* solution minimale *de $\mathcal{C}$ ssi $S$ est un sous-ensemble de $\mathcal{N}$ vérifiant :*
– *$c_i \cap S \neq \varnothing, \forall i$,*
– *$\forall x \in S, \exists i$ tel que $c_i \cap S = \{x\}$.*
*Cette définition caractérise donc les sous-ensembles de $\mathcal{N}$ minimaux pour l'inclusion rencontrant tous les conflits*[6].

**Exemple 1 (suite)** *Quand on applique la définition 6 sur l'exemple 1, on n'obtient alors comme solutions minimales que les sous-ensembles $S_4$ et $S_5$.*

Cette définition 6 permet donc d'obtenir un ensemble de solutions beaucoup plus intéressant que celui obtenu par la définition 5. Toutefois, elle ne tient pas compte du fait que, dans un conflit, on préférera supprimer une formule plutôt qu'une autre.

### 2.1.2 Utilisation des préférences

On introduit donc la prise en compte des préférences dans le problème de résolution de conflits. On considère que chaque $K$-conflit de $E$ (noté $c_i$) est muni d'un pré-ordre total et que $\forall i$, $f(c_i)$ dénote l'ensemble des éléments minimaux de $c_i$ par rapport à ce pré-ordre. On notera $\mathcal{N}_{min} = \bigcup_i f(c_i)$.

À partir de là, on peut expliciter un certain nombre de principes à respecter pour obtenir de "bonnes" solutions et c'est là toute l'originalité de notre approche :

[4] Pour simplifier, on utilisera la notation $\mathcal{C}$, $\mathcal{N}$ et $\mathcal{L}$ à la place de $\mathcal{C}(E)$, $\mathcal{N}(E)$ et $\mathcal{L}(E)$.
[5] Cela correspond à la notion d'*ensemble candidat* ou "*hitting-set*" (voir [Rei87]).
[6] Un ensemble $e_1$ "rencontre" un autre ensemble $e_2$ ssi $e_1 \cap e_2 \neq \varnothing$.

**Le principe de minimalité :**

ne garder que les solutions qui n'en contiennent pas d'autres (minimalité au sens de l'inclusion) ; c'est le principe de base déjà exploité dans la définition 6 page précédente ;

**Le principe de parcimonie :**

éliminer le plus petit nombre possible de croyances (minimalité au sens de la cardinalité) ;

**Le principe du respect des préférences :**

les croyances à éliminer sont choisies autant que possible dans les $f(c_i)$ (l'idée étant de s'assurer que toute formule éliminée l'a été pour une "bonne" raison) ;

**Le principe d'efficacité :**

sélectionner le plus petit nombre possible de solutions (afin de faciliter l'utilisation future de ces solutions – s'il y en a trop, il sera très difficile de les exploiter).

Remarquons que ces principes ne sont pas toujours compatibles.

**Exemple 1 page précédente (suite)** *Soulignons les éléments minimaux de chaque conflit :* $\mathcal{C} = \{\{\underline{a}, b\}, \{b, \underline{c}\}\}$. *Sur cet exemple, parmi les solutions obtenues par la définition 5 page précédente,* $S_4$ *et* $S_5$ *respectent le principe de minimalité,* $S_4$ *respecte les préférences et* $S_5$ *respecte le principe de parcimonie. Et il n'existe pas de solution respectant tous les principes à la fois.*

Le principe de minimalité allié au respect des préférences nous mène à la définition suivante.

**Définition 7** *Soit* $\mathcal{C} = \{c_i\}$ *un problème de conflits équipé d'un pré-ordre total sur chaque* $c_i$. $S$ *est une* solution minimale préférée *ssi* $S$ *est un sous-ensemble de* $\mathcal{N}$ *vérifiant :*
– $c_i \cap S \neq \varnothing, \forall i$,
– $\forall x \in S, \exists i$ *tel que* $c_i \cap S = \{x\}$,
– $\forall x \in S, \exists j$ *tel que* $x \in f(c_j)$.
*Ce qui est équivalent à dire que* $S$ *est un sous-ensemble de* $\mathcal{N}_{min}$ *minimal pour l'inclusion qui rencontre chaque conflit.*

**Exemple 1 page précédente (suite)** *Sur cet exemple, la seule solution minimale préférée est* $S_4$.

**Exemple 2** *Soit le problème de conflits suivant :*
$\mathcal{C} = \{\{a_6, \underline{a_1, a_2}\}, \{a_7, \underline{a_2, a_3}\}, \{a_3, \underline{a_4, a_8}\}, \{a_4, \underline{a_5}\}\}$.
*Sur cet exemple, la définition 7 nous fournit les 5 solutions minimales préférées suivantes :*
$S_1 = \{a_2, a_4\}$, $S_2 = \{a_2, a_8, a_5\}$, $S_3 = \{a_1, a_3, a_5\}$, $S_4 = \{a_2, a_3, a_5\}$, $S_5 = \{a_1, a_3, a_4\}$.

D'autre part, le principe de parcimonie nous conduit à la définition 8.

**Définition 8** *Soit* $\mathcal{C} = \{c_i\}$ *un problème de conflits.* $S$ *est une* solution parcimonieuse *de* $\mathcal{C}$ *ssi* $S$ *est un sous-ensemble de* $\mathcal{N}$ *minimal pour la cardinalité qui rencontre chaque conflit.*

**Exemple 1 page précédente (suite)** *Quand on applique la définition 8, on n'obtient alors comme solution parcimonieuse que le sous-ensemble* $S_5$.

Les solutions parcimonieuses étant toutes minimales[7], on peut donc voir la définition 8 comme un moyen de respecter au mieux le principe d'efficacité.

Par contre, étant donné leur incompatibilité, il est impossible d'associer principe de parcimonie et respect des préférences. La seule chose que l'on puisse faire est d'essayer de respecter le principe d'efficacité en utilisant la cardinalité comme un moyen pour réduire l'ensemble des solutions. Ce qui nous mène à la définition 9 page suivante.

[7] En effet, la définition 8 raffine la définition 6 page précédente.

| Solutions proposées | Principes respectés | | |
|---|---|---|---|
| | Minimalité | Parcimonie | Respect préférences |
| Sol. (de base) (déf. 5 page 23) | | | |
| Sol. minimale (déf. 6 page 23) | X | | |
| Sol. minimale préférée (déf. 7 page ci-contre) | X | | X |
| Sol. parcimonieuse (déf. 8 page précédente) | X | X | |
| Sol. card-min préférée (déf. 9) | X | | X |
| Sol. éli-préférée (déf. 11) | X | | X |

TAB. 2.1 – Tableau récapitulatif des principes et des solutions (le X signifie que la solution proposée respecte le principe donné)

**Définition 9** *Soit $\mathcal{C} = \{c_i\}$ un problème de conflits équipé d'un pré-ordre total sur chaque $c_i$. $S$ est une* solution card-min préférée *ssi : $S$ est un sous-ensemble de $\mathcal{N}_{min}$ minimal pour la cardinalité rencontrant chaque conflit.*

**Exemple 1 page 23 (suite)** *Sur cet exemple, la solution card-min préférée est aussi la solution minimale préférée $S_4$. Notons qu'il ne s'agit pas d'une solution parcimonieuse.*

**Exemple 2 page précédente (suite)** *Si on reprend l'exemple 2, la seule solution card-min préférée est $S_1$.*

Il existe aussi d'autres moyens de réduire l'ensemble des solutions afin de mieux respecter le principe d'efficacité. Nous pouvons par exemple utiliser des notions moins connues comme la préférence dite élitiste issue de [CRS93] (voir les définitions 10 et 11).

**Définition 10** *Soit $S_1$ et $S_2$ deux solutions minimales préférées d'un problème de conflit $\mathcal{C}$. On dira que $S_1 \leq S_2$ ssi pour tout $x \in S_2 \setminus S_1$, il existe $y \in S_1 \setminus S_2$ et un conflit $c_i \in \mathcal{C}$ tels que : $y \in c_i \setminus f(c_i)$ et $x \in f(c_i)$.*

*De manière informelle, cela signifie que $c_i$ est mieux résolu par $x$ que par $y$.*

**Définition 11** *Soit $\mathcal{C} = \{c_i\}$ un problème de conflits équipé d'un pré-ordre total sur chaque $c_i$. $S$ est une* solution éli-préférée *ssi : $S$ est un sous-ensemble de $\mathcal{N}_{min}$ qui rencontre chaque conflit et qui est maximal pour la relation $\leq$ donnée par la définition 10.*

*Donc, on ne conserve que les solutions permettant de résoudre au mieux les conflits (au sens de la définition 10).*

**Exemple 2 page ci-contre (suite)** *Seules $S_1$, $S_2$ et $S_3$ sont éli-préférées.*

Toutes les définitions données ici sont reprises dans le tableau 2.1 avec les principes respectés.

## 2.2 Calcul des solutions minimales

Une propriété intéressante des solutions minimales préférées telles qu'elles ont été définies dans la section précédente est que leur calcul se ramène à un problème bien connu en Intelligence Artificielle : le calcul d'ensembles intersectants d'un ensemble d'ensembles[8].

[8] Un *ensemble intersectant* d'un ensemble d'ensembles $\mathcal{C}$ est une partie de l'union des éléments de $\mathcal{C}$ dont l'intersection avec chaque élément de $\mathcal{C}$ est non vide.

**Proposition 1** *Soit $\mathcal{C} = \{c_i\}$ un ensemble d'ensembles, et soit $f$ une fonction qui associe à tout élément $c_i$ de $\mathcal{C}$ une partie non vide de $c_i$. Soit $\mathcal{N}_{\min}$ l'union des $f(c_i)$, alors $S$ est une solution minimale préférée (au sens de la définition 7 page 24) du problème de conflits défini par $\mathcal{C}$ et $f$ si et seulement si $S$ est un ensemble intersectant minimal (pour l'inclusion) de $\mathcal{C}' = \{c_i \cap \mathcal{N}_{\min}\}$.*

Plusieurs algorithmes de calcul d'ensembles intersectants ont été proposés en Intelligence Artificielle, dans le cadre du diagnostic [Rei87], de la déduction en logique propositionnelle [Cas96, Cas97], ou encore de l'étude de logiques non monotones [JK90, Lev92, Men95, Nie95] entre autres.

Je ne m'étendrai pas davantage sur cette question, mais le lecteur intéressé trouvera dans [CLSM00] des exemples d'algorithmes pour le calcul de ces solutions.

Il est aussi intéressant de noter que, dans le cadre de la révision de croyances, une approche utilisant entre autres le principe de parcimonie a été développée dans [Pap92] pour repérer les sous-ensembles de formules à supprimer. Cette approche a ensuite été utilisée dans le cadre applicatif des systèmes de gestion d'informations géographiques (GIS), voir [WPJ00, BBPW04] à l'aide d'algorithmes eux-aussi définis à partir des ensembles intersectants (*hitting-set* en anglais).

## 2.3 Relations avec les méthodes basées sur la sélection directe de sous-ensembles $K$-consistants

Ici, nous mettons en évidence les liens entre la résolution de conflits et la sélection de sous-bases préférées utilisée au chapitre 1 page 9 pour faire de l'inférence non monotone.

### 2.3.1 Dualité entre problème de cohérence et problème de conflits

Étant donné une base de croyances $E$ $K$-inconsistante, nous considérons les deux problèmes suivants :

- Un problème de cohérence donné sous la forme $(K, E)$, dont les solutions sont les sous-ensembles $K$-consistants de $E$.
- Un problème de résolution de conflits donné sous la forme $(\mathcal{L}, \mathcal{C})$ (où $\mathcal{L}$ est l'ensemble des formules libres de $E$ et $\mathcal{C} = \{c_i\}$ est l'ensemble des $K$-conflits issus de $E$) et dont les solutions sont les sous-ensembles de $E$ qui ne contiennent aucun conflit. Déterminer les sous-ensembles de $E$ qui ne contiennent aucun conflit revient à déterminer les sous-ensembles de $\mathcal{N}$ (union des conflits) qui ne contiennent aucun conflit et donc à considérer le problème de conflits défini en section 2.1.1 page 22, donné sous la forme $\mathcal{C} = \{c_i\}$.

Parmi toutes les solutions d'un problème de cohérence, on considère traditionnellement celles qui sont maximales pour l'inclusion. Elles correspondent à la définition suivante.

**Définition 12** *$F$ sous-ensemble de $E$ est une solution maximale du problème de cohérence ssi $F$ est un sous-ensemble $K$-consistant, maximal pour l'inclusion parmi les sous-ensembles $K$-consistants de $E$.*

La dualité entre problème de cohérence et problème de conflits est exprimée dans la proposition ci-dessous.

**Proposition 2** *Soit $E$ une base de croyances et $K$ un ensemble de connaissances incontestables. Comme il a été indiqué au début de cette section, $E$ et $K$ définissent à la fois un problème de cohérence $(K, E)$ et un problème de conflits $(\mathcal{L}, \mathcal{C})$ (avec $\mathcal{N}$ l'union des éléments de $\mathcal{C}$).*

*On a alors les propriétés suivantes :*

- *$F$ sous-ensemble de $E$ est solution du problème de cohérence ssi $(\mathcal{N} \setminus F)$ est solution (au sens de la déf. 5 page 23) du problème de conflits.*
- *$S$ sous-ensemble de $\mathcal{N}$ est solution minimale (au sens de la déf. 6 page 23) du problème de conflits ssi $(\mathcal{N} \setminus S)$ est un sous-ensemble maximal pour l'inclusion parmi les sous-ensembles $K$-consistants de $\mathcal{N}$.*
- *Les solutions maximales du problème de cohérence sont de la forme $\mathcal{L} \cup (\mathcal{N} \setminus S)$ où $S$ est une solution minimale (au sens de la déf. 6 page 23) du problème de conflits.*

### 2.3.2 Intégration de préférences

Dans le cas où l'on dispose d'une relation de préférence entre les croyances représentée par un pré-ordre partiel $\leq$ sur $E$, certaines solutions maximales du problème de cohérence sont préférées à d'autres. Différents critères de préférence entre sous-ensembles ont été définis à partir d'un pré-ordre sur les croyances. (Voir [LS95a] pour une étude détaillée de ces critères).

Le critère de préférence démocratique par exemple, permet de privilégier les solutions qui contiennent le plus possible (au sens de l'inclusion) de croyances parmi les plus prioritaires.

**Notation 1** *On notera $<$ l'ordre strict associé au pré-ordre $\leq$ : $x < y$ ssi $(x \leq y) \wedge \neg(y \leq x)$. Le problème de cohérence sera maintenant noté $(K, E, <)$.*

**Définition 13** *Soit $F_1$ et $F_2$ deux sous-ensembles $K$-consistants de $E$. $F_2$ est démocratiquement préférée à $F_1$ ssi pour tout $x$ élément de $F_1 \setminus F_2$, il existe un élément $y$ de $F_2 \setminus F_1$ tel que $x < y$.*

*Les solutions du problème de cohérence $(K, E, <)$ qui sont maximales pour la relation de préférence démocratique sont appelées solutions démo-préférées.*

Dans le cas particulier où le pré-ordre $\leq$ sur $E$ est total, il est équivalent de considérer que la base $E$ est stratifiée en $(E_1, E_2, \ldots, E_n)$ où $E_1$ contient les croyances $\leq$-maximales de $E$, et $\forall i$, $E_{i+1}$ contient les croyances $\leq$-maximales de $E \setminus (E_1 \cup E2 \cup \ldots \cup E_i)$. Le problème de cohérence est alors noté $(K, E_1, \ldots, E_n)$. Les solutions démo-préférées du problème de cohérence $(K, E_1, \ldots, E_n)$ correspondent exactement aux sous-bases $K$-consistantes de $E$ préférées au sens de Brewka, définies comme suit. (voir [Bre89, CRS93]).

**Définition 14** *Soit $F$ un sous-ensemble $K$-consistant de $E$ stratifiée en $(E_1, \ldots, E_n)$. Soit $F_i = F \cup E_i$ la projection de $F$ sur la strate $E_i$. $F$ est préférée au sens de Brewka ssi $\forall k$, $1 \leq k \leq n$, $F_1 \cup \ldots \cup F_k$ est un sous-ensemble maximal (pour l'inclusion) parmi les sous-ensembles $K$-consistants de $E_1 \cup \ldots \cup E_k$.*

Tout cela correspond aux sous-bases incl-préférées évoquées au chapitre 1 page 9, quand les préférences sont données sous la forme d'un pré-ordre total.

**Remarque** Dans le cas où la base de croyances est seulement munie d'un pré-ordre partiel, il est commode de considérer la stratification "uniforme" de $E$ construite sur le même principe que ci-dessus (c'est-à-dire que $E_{i+1}$ contient les éléments $\leq$-maximaux de $E \setminus (E_1 \cup E_2 \cup \ldots E_i)$). Les solutions préférées au sens de Brewka du problème de cohérence $(K, E_1, \ldots, E_n)$ ainsi obtenu sont toutes des solutions démo-préférées du problème de cohérence initial $(K, E, <)$. La réciproque est généralement fausse.

#### Caractérisation des solutions minimales préférées d'un problème de conflits

Nous considérons ici un problème de conflits avec expression locale de préférences. Rappelons que $\mathcal{N}_{min}$ désigne l'union des $f(ci)$.

Dans le cas où $\mathcal{N}_{min}$ est strictement inclus dans $\mathcal{N}$, on peut considérer la stratification de $E$ en $(E_1, E_2, E_3)$ où $E_1 = \mathcal{L}$, $E_2 = \mathcal{N} \setminus \mathcal{N}_{min}$ et $E_3 = \mathcal{N}_{min}$.

**Proposition 3** *$S$ sous-ensemble de $\mathcal{N}$ est solution minimale préférée (au sens de la définition 7 page 24) du problème de conflits ssi $E \setminus S$ est une solution démo-préférée du problème de cohérence $(K, E_1, E_2, E_3)$.*

**Exemple 2 page 24 (suite)** *Reprenons ici le problème de conflits défini par $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$ avec :*

$c_1 = \{a_6, \underline{a_1, a_2}\}$, $c_2 = \{a_7, \underline{a_2, a_3}\}$, $c_3 = \{a_3, \underline{a_4, a_8}\}$, $c_4 = \{a_4, \underline{a_5}\}$ *(pour chaque $c_i$, les éléments soulignés sont les éléments minimaux de ce conflit).*

*On a donc :*

$E = \mathcal{N}$ *(rappelons que $\mathcal{N} = \bigcup_i c_i$),*

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

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

*Les solutions démo-préférées (qui sont aussi les solutions préférées au sens de Brewka) du problème de cohérence défini par* $(K, \{a_6, a_7\}, \{a_1, a_2, a_3, a_4, a_5, a_8\})$ *sont :*

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

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

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

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

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

*Ces solutions correspondent respectivement aux solutions minimales préférées du problème de conflits :*

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

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

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

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

*et* $S_5 = \{a_1, a_3, a_4\}$.

**Caractérisation des solutions éli-préférées d'un problème de conflits**

Notre objectif étant toujours d'exploiter la dualité entre problème de conflits et problème de cohérence, nous nous plaçons désormais dans la situation particulière où il existe un ordre sur $E$ qui est compatible avec la stratification locale à chaque conflit. Cet ordre est défini par la construction ci-dessous :

**Définition 15** *Soit* $x$ *et* $y$ *dans* $\mathcal{N}$. *On définit :* $xRy$ *ssi il existe un conflit* $c_i$ *tel que* $x$ *est dans* $f(c_i)$ *et* $y$ *est dans* $c_i \setminus f(c_i)$.

Notons que la fermeture transitive de la relation $R$ ne définit pas toujours un ordre strict, comme l'illustre l'exemple suivant.

**Exemple 3** *Soit le problème de conflits défini par* $\mathcal{C} = \{\{a, \underline{b, e}\}, \{b, f, \underline{c}\}, \{c, g, \underline{a, d}\}\}$.

*On a alors :*

$E = \mathcal{N}$,

$\mathcal{N} \setminus \mathcal{N}_{min} = \{f, g\}$

*et* $\mathcal{N}_{min} = \{a, b, c, d, e\}$.

*On aura alors :*

$bRa$ *: car on préfère abandonner* $b$ *plutôt que* $a$ *dans le conflit* $c_1$.

$aRc$ *: car on préfère abandonner* $a$ *plutôt que* $c$ *dans le conflit* $c_3$.

$cRb$ *: car on préfère abandonner* $c$ *plutôt que* $b$ *dans le conflit* $c_2$.

Nous supposons donc dans la suite que la fermeture transitive de la relation $R$ définit un ordre strict (partiel) sur $\mathcal{N}$, que nous dénoterons par $<$. Nous étendons cet ordre à $E$ par : pour tout $x$ dans $E \setminus \mathcal{N}$, pour tout $y$ dans $\mathcal{N}$, $y < x$.

**Proposition 4** *Soit* $S$ *une solution minimale du problème de conflits. Si* $E \setminus S$ *est une solution démo-préférée du problème de cohérence* $(K, E, <)$ *alors* $S$ *est une solution éli-préférée du problème de conflits.*

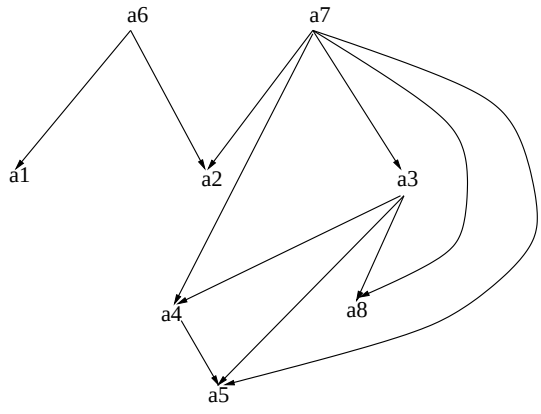



FIG. 2.1 – Graphe de la relation $<$ obtenue à partir de l'exemple 2

**Exemple 2 page 24 (suite)** *À partir du problème de conflit correspondant à cet exemple ($c_1 = \{a_6, \underline{a_1, a_2}\}$, $c_2 = \{a_7, \underline{a_2, a_3}\}$, $c_3 = \{a_3, \underline{a_4, a_8}\}$, $c_4 = \{a_4, \underline{a_5}\}$), on trouve un ordre strict $<$ compatible avec les stratifications locales des conflits. Son graphe est donné par la figure 2.1.*

*Les solutions démo-préférées du problème de cohérence $(K, E, <)$ sont :*

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

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

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

*Remarquons que ce sont aussi les solutions préférées au sens de Brewka du problème de cohérence obtenu par stratification uniforme de $(E, <)$ :*

$E = (\{a_6, a_7\}, \{a_1, a_2, a_3\}, \{a_4, a_8\}, \{a_5\})$.

*Ces 3 solutions fournissent les 3 solutions éli-préférées $S_1$, $S_2$ et $S_3$ du problème de conflits (voir en fin de section 2.1.2 page 23).*

## 2.4 Conclusion sur le thème 2

La définition de solutions d'un problème de résolution de conflits pose une question similaire à celui rencontré lorsqu'on cherche à définir les solutions d'un problème de restauration de cohérence par sélection directe de sous-bases consistantes : il s'agit de la difficulté de concilier deux contraintes dont le respect est intuitivement nécessaire. L'une de ces contraintes impose que les solutions soient les plus petites (respectivement les plus grandes) possibles. L'autre contrainte impose de respecter les préférences éventuellement exprimées entre les croyances.

Afin d'étudier comment ces deux contraintes interagissent lorsqu'on résout des conflits, nous avons été conduits à traduire chacune d'entre elles par un ou plusieurs principes. Ces principes conduisent ensuite à différentes définitions de solutions, et permettent de comparer simplement ces définitions. Nous avons aussi comparé certaines de ces définitions avec les résultats obtenus lorsqu'on aborde le problème, jusqu'ici plus étudié, de la sélection directe de sous-bases consistantes.

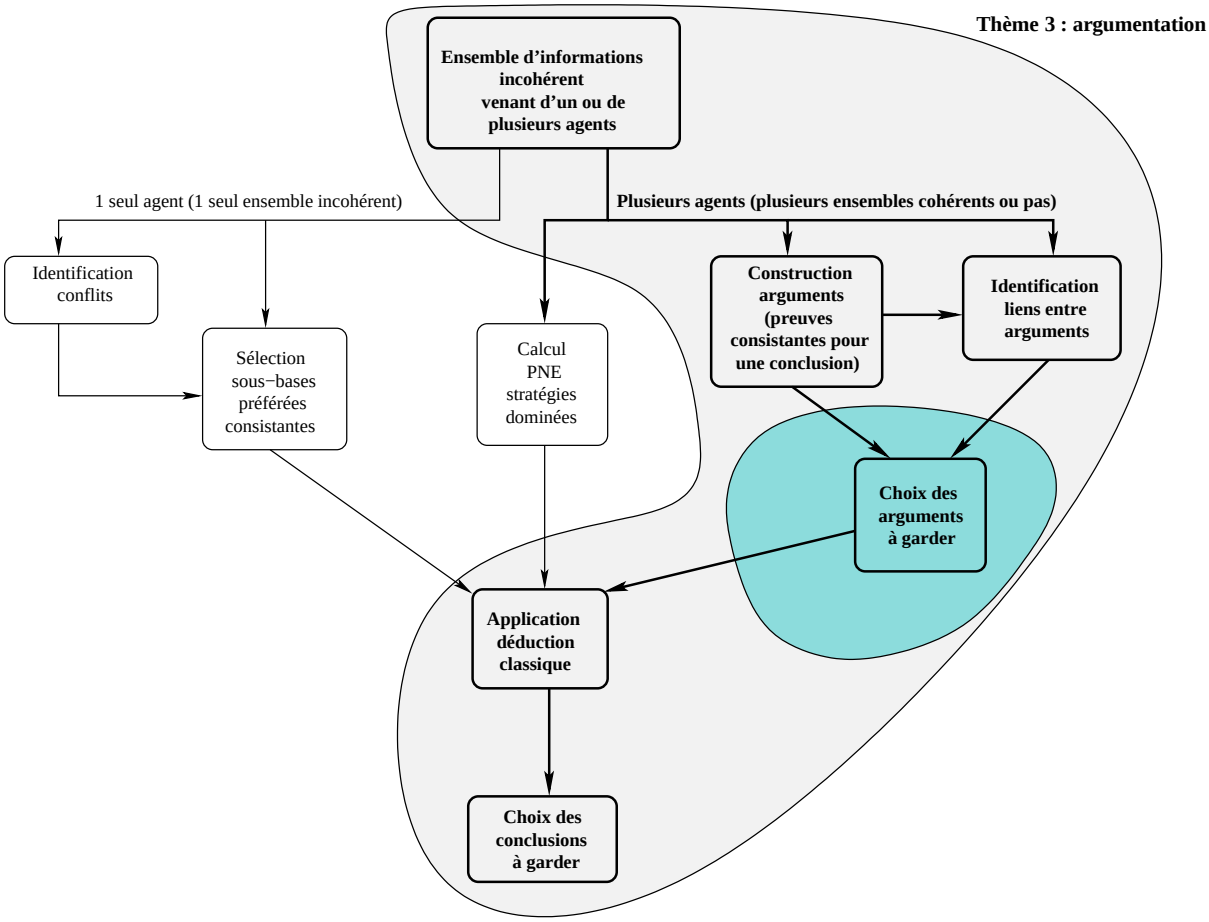
L'une des définitions semble ressortir d'un point de vue calculatoire : calculer les *solutions minimales préférées* revient à calculer les ensembles intersectants minimaux de l'ensemble des parties minimales (au sens des préférences) des conflits. On peut noter que le respect des préférences n'induit pas dans ce cas de difficulté calculatoire supplémentaire (à part le simple calcul des parties minimales des conflits) : au contraire, le calcul des solutions minimales préférées est

plus simple que le calcul des solutions minimales, puisque les ensembles à “intersecter” sont plus petits. Nous avons décrit plusieurs algorithmes qui ont été proposés en Intelligence Artificielle pour effectuer ces calculs. Le lien avec la notion d’impliquants premiers semble particulièrement intéressant, puisqu’il permet d’exploiter des heuristiques mises au point dans le cadre des nombreuses études d’algorithmes permettant de tester la consistance d’un ensemble de formules propositionnelles (voir l’étude algorithmique donnée dans [CLSM00]).

# Chapitre 3

# Thème 3 : l'argumentation

Le thème abordé ici est représenté par la partie "entourée claire" du schéma suivant et mes travaux sur le sujet présentés dans ce chapitre correspondent à la partie "entourée foncée" :




Ce thème rejoint les deux thèmes précédents par le fait qu'il s'agit ici-aussi de traiter l'incohérence en sélectionnant des paquets cohérents (même si ces paquets ne représentent qu'un état très parcellaire du monde, contrairement à ce qui se faisait dans les deux thèmes précédents).

Il est important de noter que l'argumentation ne s'inscrit pas uniquement dans le cadre élargi de la restauration de

cohérence même si c'est sous cet angle que je l'aborde ici. Il s'agit en effet d'une thématique à part entière au même titre que le raisonnement non monotone et qui couvre plusieurs domaines de recherche : l'intelligence artificielle, mais aussi la philosophie et la psychologie cognitive. Les travaux que j'ai réalisés sur ce thème ne concerne bien-sûr que la partie IA.

Il s'agit ici de voir l'argumentation comme un mécanisme complexe, permettant à partir d'un ensemble d'informations incohérent, de *construire* des sous-ensembles cohérents en faveur de ou contre certaines conclusions puis *d'analyser* ces sous-ensembles (les arguments) et leurs interactions afin de conclure. Pour ce faire, je vais m'éloigner nettement de l'aspect logique, puisque je pars avec l'hypothèse que les arguments et leurs interactions existent déjà et donc se réduisent à un simple graphe (les arguments étant des sommets et les interactions étant des arcs), dans lequel je cherche à identifier les arguments à garder. Je ne me suis donc consacrée qu'à la partie *analyse* du processus d'argumentation.

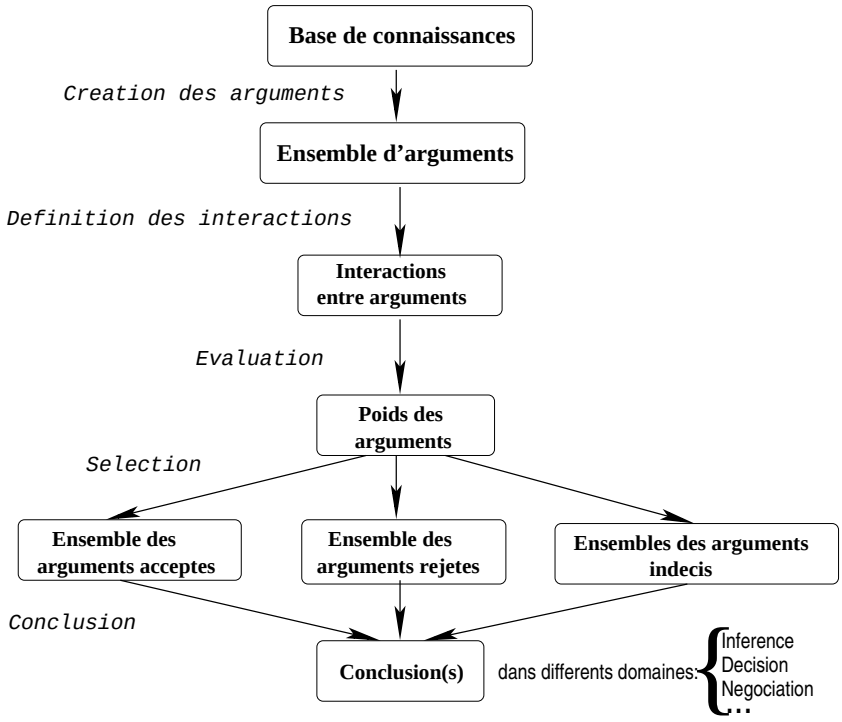
Ce chapitre couvrant de nombreuses années de recherche a une taille conséquente et aborde plusieurs problématiques dans le domaine de l'argumentation.

Le lecteur trouvera d'abord en section 3.1 le schéma général d'un processus d'argumentation qui précise celui donné en début de ce chapitre. Puis, nous aborderons successivement deux types d'argumentation, l'argumentation unipolaire (section 3.2 page 34) et l'argumentation bipolaire (section 3.3 page 106) ; dans le premier cas, les interactions se réduisent à des conflits entre arguments, alors que, dans le second cas, apparaissent aussi des interactions plus "positives" représentant des appuis entre arguments.

Pour chacun de ces types d'argumentation, on présentera les notions d'évaluation (comment affecter un "poids" à un argument – voir sections 3.2.2 page 38 et 3.3.4 page 117) et d'acceptabilité (comment sélectionner un argument – voir sections 3.2.3 page 55 et 3.3.2 page 111). Et dans le cas de l'argumentation unipolaire, on donnera aussi des pistes pour réaliser la fusion ou la révision d'un système d'argument (voir sections 3.2.4 page 62 et 3.2.5 page 80), ainsi que la possibilité d'utiliser un système d'argumentation pour faire du *practical reasoning* (section 3.2.6 page 90).

## 3.1 Schéma général

L'argumentation dans sa globalité peut être vue comme un processus en diverses étapes. Il s'agit donc ici d'un raffinage d'une partie du schéma précédent :




1. Définition des *arguments* : la notion d'argument fait référence en général aux concepts d'explication, de preuve, de justification. Le but d'un argument est de "soutenir" des croyances, ou de critiquer un agent afin d'obtenir un certain comportement. Ces arguments peuvent avoir des formes multiples : texte informel en langage naturel donné dans un dialogue ou bien preuve formelle donnée dans un langage logique parfaitement formalisé.
2. Définition des différentes *interactions entre arguments* : les arguments formés à partir d'une base de connaissances ne peuvent pas être considérés comme indépendants les uns des autres. Il peut exister plusieurs formes d'interactions entre ces arguments. On considère en général deux types principaux d'interactions : les arguments peuvent être en conflit (interaction de conflit) ou bien ils peuvent s'aider (interaction de support). Quand on ne prend en compte que les conflits, on parlera alors de *systèmes d'argumentation unipolaires* et quand conflits et supports apparaissent tous les deux, on parlera alors de *systèmes d'argumentation bipolaires*.
3. *Évaluation des arguments* : ici, l'idée de base est d'attribuer un *poids* à chaque argument. Les différents poids permettent alors de comparer les arguments. Plusieurs critères peuvent être utilisés pour définir ces poids. Par exemple, dans [AC02a], des priorités implicites et explicites portant sur les connaissances ou sur les arguments sont considérées, alors que, dans [CLS03c, ACLS04], les poids sont définis en utilisant les interactions entre les arguments.
4. *Sélection des arguments les plus acceptable* : cette sélection peut être réalisée en utilisant les interactions entre arguments et/ou les évaluations de ces arguments. L'ensemble des arguments acceptables peut être vu comme une "sortie" du processus d'argumentation. Informellement, les arguments considérés comme acceptables sont ceux qui pourraient "gagner" dans un dialogue d'argumentation entre agents. Remarquons que la phase d'évaluation pourrait suffire à elle-seule pour sélectionner des arguments uniquement à partir de leurs poids. On la distingue toutefois de la phase de sélection proprement dite car le mécanisme de sélection peut être bien plus complexe et surtout peut reposer sur des notions totalement différentes de la notion de poids.
5. *Conclure l'argumentation* : cette toute dernière phase consiste à utiliser le statut des arguments pour déterminer celui des "conclusions" du système d'argumentation ; cela dépend du domaine d'application du système d'argumentation (on peut choisir une décision, effectuer une négociation, ...). Il est aussi possible de définir des relations d'inférence non monotone en utilisant les arguments les plus acceptables et ainsi de choisir les formules inférées par le système.

## 3.2 Argumentation unipolaire

Tous les travaux évoqués dans cette section ont donné lieu à de nombreuses publications et sont synthétisés dans quelques articles ([CLS05b, CMDK+07, CdSCLS08, ADLS09, CLS09]) dans lesquels le lecteur pourra trouver les preuves de toutes les propriétés données ici.

Cette section est décomposée de la manière suivante : les bases de l'argumentation unipolaire sont rappelées en section 3.2.1 ; puis j'évoque mes différentes contributions au domaine : en section 3.2.2 page 38 l'évaluation graduelle d'arguments, en section 3.2.3 page 55 l'acceptabilité graduelle d'arguments, en section 3.2.4 page 62 la fusion de systèmes d'argumentation unipolaires, en section 3.2.5 page 80 la révision de systèmes d'argumentation unipolaires et enfin en section 3.2.6 page 90 l'utilisation des systèmes d'argumentation unipolaires pour faire du "practical reasoning".

### 3.2.1 Les bases de l'argumentation unipolaire

En 1995, dans un article très important pour le domaine de l'argumentation en IA ([Dun95]), Dung a proposé un système d'argumentation abstrait dans lequel il se focalisait uniquement sur la manière de sélectionner les arguments. Il ignorait donc totalement les premières étapes du processus d'argumentation, à savoir la création des arguments et le calcul de leurs interactions, ainsi que la valuation de ces arguments. Son hypothèse de départ était donc que l'ensemble des arguments était donné ainsi qu'un ensemble d'interactions représentant uniquement des conflits entre ces arguments.

Je rappelle ici les principales définitions données par Dung :

**Définition 16** *Un système d'argumentation abstrait est un couple $\langle \mathcal{A}, \mathcal{R} \rangle$ constitué d'un ensemble $\mathcal{A}$ d'arguments et d'une relation binaire $\mathcal{R}$ sur $\mathcal{A}$ appelée* = relation d'attaque[1].

*$A_i \mathcal{R} A_j$ signifie que $A_i$* attaque *$A_j$. Un système d'argumentation peut être représenté par un graphe orienté dans lequel les nœuds sont des arguments et les arcs représentent la relation d'attaque.*

*Soit $A \in \mathcal{A}$, $\mathcal{R}^-(A)$ dénote l'ensemble des arguments attaquant $A$ et $\mathcal{R}^+(A)$ dénote l'ensemble des arguments attaqués par $A$.*

Dans ce contexte, Dung propose une notion permettant de "réinstaller" un argument suite à une attaque. Il s'agit de la notion de *défense* définie à partir de la notion d'attaque de la manière suivante :

**Définition 17** *Un argument $A_i$ défend un argument $A_j$ contre l'argument $B$ ssi $B \mathcal{R} A_j$ et $A_i \mathcal{R} B$.*

Dans le système d'argumentation unipolaire proposé par Dung, le mécanisme de sélection des arguments les plus acceptables repose sur une notion d'"acceptabilité collective", *i.e.* sur le fait que ces arguments appartiennent à des ensembles d'arguments vérifiant certaines propriétés. Les principales propriétés sont :

**Définition 18**
- Sans-conflit *: l'ensemble $S \subseteq \mathcal{A}$ est sans-conflit ssi il n'existe pas d'arguments $A_i$ et $A_j$ dans $S$ tels que $A_i \mathcal{R} A_j$.*
- Défense collective *: l'ensemble $S \subseteq \mathcal{A}$ défend collectivement un argument $A_i$ ssi pour chaque argument $B$, tel que $B \mathcal{R} A_i$ il existe $C \in S$ tel que $C \mathcal{R} B$.*

Puis, en utilisant ces propriétés, Dung a proposé dans [Dun95] différentes sémantiques pour l'acceptabilité :

**Définition 19** *Soit $S \subseteq \mathcal{A}$ un ensemble d'arguments,*

[1] Dans ce document, le terme "attaque" est un *terme générique* pouvant recouvrir plusieurs cas de figure :
- les conclusions des deux arguments sont en contradiction,
- la conclusion de l'un des arguments contredit la prémisse de l'autre argument,
- un argument attaque l'autre et lui est préféré,
- …

– Admissible *: $S$ est un ensemble admissible ssi $S$ est sans-conflit et $S$ défend collectivement tous ses éléments.*
– Préférée *: $S$ est une extension préférée ssi $S$ est maximal pour l'inclusion ensembliste parmi les sous-ensembles admissibles de $\mathcal{A}$.*
– Stable *: $S$ est une extension stable ssi $S$ est sans-conflit et $S$ attaque chaque argument qui n'est pas dans $S$.*
– Basique *: $S$ est l'extension basique ssi $S$ est le plus petit point fixe de la fonction caractéristique $\mathcal{F}$ de $\langle \mathcal{A}, \mathcal{R} \rangle$ ($\mathcal{F} : 2^{\langle \mathcal{A}, \mathcal{R} \rangle} \rightarrow 2^{\langle \mathcal{A}, \mathcal{R} \rangle}$ avec $\mathcal{F}(S) = \{A$ tel que $S$ défend collectivement $A\}$).*

Remarquons que Dung considère toujours chaque "attaquant" individuellement (l'"attaque directe" dans sa globalité n'est pas une notion utilisée par Dung).

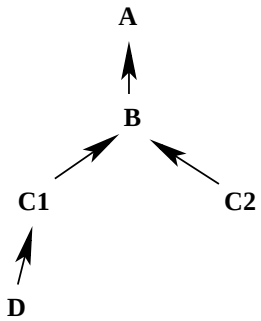
Dans [Dun95], on trouve aussi quelques propriétés intéressantes :

**Proposition 5**
– *Tout ensemble admissible est inclus dans une extension préférée.*
– *Il existe toujours au moins une extension préférée.*
– *Il existe toujours une seule extension basique.*
– *Si le système d'argumentation est bien fondé*[2] *alors il existe une seule extension préférée qui est aussi la seule extension stable et l'extension basique.*
– *Toute extension stable est aussi une extension préférée (la réciproque est fausse).*
– *Il n'existe pas toujours d'extension stable.*
– *L'extension basique est incluse dans chaque extension préférée.*
– *Si $\mathcal{R}$ est finie, l'extension basique se calcule par itération en appliquant la fonction $\mathcal{F}$ à partir de l'ensemble vide.*

**Proposition 6** *L'ensemble des arguments non attaqués (i.e. $\{A | \mathcal{R}^-(A) = \varnothing\}$) est inclus dans l'extension basique, dans chaque extension préférée et dans chaque extension stable.*

Toutes ces notions sont illustrées sur le système d'argumentation suivant :

**Exemple 4**




*Dans ce système, $A$ est défendu par $C_1$ et $C_2$, et il existe une seule extension préférée $\{D, C_2, A\}$ qui est aussi stable et basique.*

Nous avons aussi besoin de préciser certaines notions liées à la représentation graphique d'un système d'argumentation :

**Définition 20** *Soit $\mathcal{G}$ le graphe d'attaque représentant le système d'argumentation $\langle \mathcal{A}, \mathcal{R} \rangle$.*

**Feuille du graphe d'attaque** *Une* feuille de $\mathcal{G}$ *est un argument $A \in \mathcal{A}$ sans attaquant*[3].

**Chemin dans le graphe d'attaque** *Un* chemin de $A$ vers $B$ *est une séquence d'arguments $\mathcal{C} = A_1 - \ldots - A_n$ telle que :*

[2] Un système d'argumentation est dit bien-fondé ssi il n'existe pas de séquence infinie $A_0, A_1, \ldots, A_n, \ldots$ telle que $\forall i, A_i \in \mathcal{A}$ et $A_{i+1} \mathcal{R} A_i$.
[3] $A$ est une feuille ssi $\mathcal{R}^-(A) = \varnothing$.

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

*La* longueur du chemin *est* $n-1$ *(le nombre d'arcs utilisé dans le chemin) et sera notée* $l_{\mathcal{C}}$.

*Un cas particulier est le chemin*[4] *de* $A$ *vers* $A$ *dont la longueur est* $0$.

*L'ensemble des chemins de* $A$ *vers* $B$ *sera noté* $\mathcal{C}(A, B)$.

**Dépendance, indépendance, racine-dépendance d'un chemin**

*Soit deux chemins* $\mathcal{C}_A \in \mathcal{C}(A_1, A_n)$ *et* $\mathcal{C}_B \in \mathcal{C}(B_1, B_m)$.

*Ces deux chemins peuvent être dits* dépendants *ssi* $\exists A_i \in \mathcal{C}_A$, $\exists B_j \in \mathcal{C}_B$ *tels que* $A_i = B_j$. *Sinon ils sont* indépendants.

*Ces deux chemins peuvent être dits* racine-dépendants en $A_n$ *ssi* $A_n = B_m$ *et* $\forall A_i \neq A_n \in \mathcal{C}_A$, $\nexists B_j \in \mathcal{C}_B$ *tel que* $A_i = B_j$.

**Circuits dans le graphe d'attaque** *Un* circuit[5] *est un chemin* $\mathcal{C} = A_1 - \ldots - A_n - A_1$ *tel que* $\forall i, j \in [1, n]$, *si* $i \neq j$, *alors* $A_i \neq A_j$.

*Un circuit* $\mathcal{C}$ *est* isolé *ssi* $\forall A \in \mathcal{C}$, $\nexists B \in \mathcal{A}$ *tel que* $B \mathcal{R} A$ *et* $B \notin \mathcal{C}$.

*Deux circuits* $\mathcal{C}_A = A_1 - \ldots - A_n - A_1$ *et* $\mathcal{C}_B = B_1 - \ldots - B_m - B_1$ *sont* interconnectés *ssi* $\exists i \in [1, n], \exists j \in [1, m]$ *tels que* $A_i = B_j$.

Nous utiliserons aussi les notions d'attaque et de défense indirectes. Ces notions sont inspirées de notions semblables proposées par Dung dans [Dun95] mais ne sont pas strictement équivalentes[6].

**Définition 21 (Attaquant/Défenseur Direct/Indirect d'un argument)** *Soit* $A \in \mathcal{A}$ *:*

- *Les* attaquants directs *de* $A$ *sont les éléments de* $\mathcal{R}^-(A)$.
- *les* défenseurs directs *de* $A$ *sont les attaquants directs des éléments de* $\mathcal{R}^-(A)$.
- *Les* attaquants indirects *de* $A$ *sont les éléments* $A_i$ *définis par :*
$$\exists \mathcal{C} \in \mathcal{C}(A_i, A) \text{ tel que } l_{\mathcal{C}} = 2k+1, \text{ avec } k \geq 1.$$
- *Les* défenseurs indirects *de* $A$ *sont les éléments* $A_i$ *définis par :*
$$\exists \mathcal{C} \in \mathcal{C}(A_i, A) \text{ tel que } l_{\mathcal{C}} = 2k, \text{ avec } k \geq 2.$$

Remarquons qu'un attaquant peut aussi être un défenseur, qu'un attaquant (resp. défenseur) direct peut aussi être un attaquant (resp. défenseur) indirect. Cela nous conduit à la définition d'arguments un peu particuliers : les arguments controversés.

**Définition 22 (Arguments controversés)** *Soit* $A$ *et* $B \in \mathcal{A}$ *deux arguments.* $B$ *est* controversé *par rapport à* $A$ *ssi* $B$ *est à la fois un défenseur direct ou indirect et un attaquant direct ou indirect de* $A$.

**Définition 23 (Branches d'attaque et de défense d'un argument)** *Soit* $A \in \mathcal{A}$, *une* branche d'attaque *(resp.* branche de défense*) pour* $A$ *est un chemin dans* $\mathcal{G}$ *d'une feuille vers* $A$ *dont la longueur est impaire (resp. paire). On dit que* $A$ *est la* racine d'une branche d'attaque *(resp.* branche de défense*).*

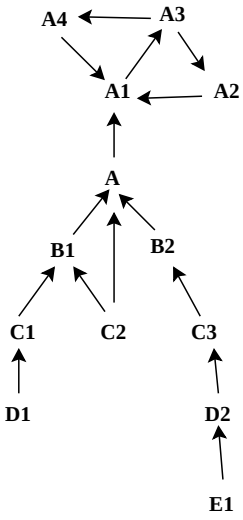
Toutes ces notions sont illustrées sur l'exemple :

**Exemple 5**

[4] Nous supposerons qu'il existe une infinité de chemins de ce type.

[5] Cette définition d'un circuit correspond à la définition d'un circuit élémentaire en théorie des graphes (un circuit élémentaire ne contient pas deux arcs ayant la même extrémité initiale ou la même extrémité finale).

[6] Dans [Dun95], les attaquants (resp. défenseurs) directs sont aussi des attaquants (resp. défenseurs) indirects ce qui n'est pas le cas avec nos définitions.




*Sur ce graphe $\mathcal{G}$, on trouve :*

- *un chemin de $C_2$ vers $A$ dont la longueur est 2 ($C_2 - B_1 - A$) ;*
- *2 circuits $A_1 - A_3 - A_2 - A_1$ et $A_1 - A_3 - A_4 - A_1$, de longueur 3, non isolés (remarquons que $A_1 - A_3 - A_2 - A_1 - A_3 - A_4 - A_1$ n'est pas un circuit avec notre définition) ;*
- *les deux circuits précédents sont interconnectés (en $A_1$ et en $A_3$) ;*
- *les chemins $D_1 - C_1 - B_1$ et $C_3 - B_2 - A$ sont indépendants, les chemins $D_1 - C_1 - B_1 - A$ et $C_3 - B_2 - A$ sont racine-dépendants et les chemins $D_1 - C_1 - B_1 - A$ et $C_2 - B_1 - A$ sont dépendants ;*
- *$D_1$, $C_2$, $E_1$ sont les feuilles de $\mathcal{G}$ ;*
- *$D_1 - C_1 - B_1 - A$ est une branche d'attaque pour $A$ dont la longueur est 3, $C_2 - B_1 - A$ est une branche de défense pour $A$ dont la longueur est 2 ;*
- *$C_2$, $B_1$ et $B_2$ sont les attaquants directs de $A$ ;*
- *$C_1$, $C_2$ (qui est déjà un attaquant direct de $A$) et $C_3$ sont des défenseurs directs de $A$ ;*
- *$D_1$ et $D_2$ sont les deux attaquants indirects de $A$ ;*
- *$E_1$ est le seul défenseur indirect $A$ ;*
- *$C_2$ est controversé par rapport à $A$, et à chacun des $A_i$, tandis que $A_1$, $A_2$, $A_3$, et $A_4$ sont controversés par rapport à eux-mêmes.*

Le système de Dung a été généralisé sous diverses formes dont une qui correspond à un système d'argumentation à contraintes proposé par [CMDM06]. L'idée de base est d'expliciter les *contraintes* que les arguments doivent satisfaire pour pouvoir appartenir aux extensions. Par exemple, on pourrait imposer que deux arguments $A$ et $B$ appartiennent à la même extension. Ces contraintes sont généralement exprimées sous la forme d'une formule logique propositionnelle construite en utilisant $\mathcal{A}$ comme alphabet.

**Définition 24 ([CMDM06] – Contrainte, Complétion)** *Soit $\mathcal{A}$ un ensemble d'arguments et $\mathcal{L}_{\mathcal{A}}$ le langage propositionnel défini en utilisant $\mathcal{A}$ comme l'ensemble des variables propositionnelles.*

– *$C$ est une* contrainte *des arguments de $\mathcal{A}$ si et seulement si $C$ est une formule de $\mathcal{L}_{\mathcal{A}}$.*
– *La* complétion *d'un ensemble $E \subseteq \mathcal{A}$ est : $\widehat{E} = \{A \mid A \in E\} \cup \{\neg A \mid A \in \mathcal{A} \setminus E\}$.*
– *Un ensemble $E \subseteq \mathcal{A}$* satisfait *$C$ si et seulement si $\widehat{E}$ est un modèle de $C$ ($\widehat{E} \vdash C$).*

Un système d'argumentation à contraintes se définit alors par :

**Définition 25 ([CMDM06] – Système d'argumentation à contrainte)** *Un* système d'argumentation à contrainte *est un triplet $\mathsf{CoAF} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ avec $C$ une contrainte d'arguments sur l'ensemble $\mathcal{A}$.*

La notion d'extension de Dung est alors étendue pour prendre en compte la contrainte et on a :

**Définition 26 ([CMDM06] – ensemble $C$-admissible)** *Soit $E \subseteq \mathcal{A}$. $E$ est $C$-*admissible *si et seulement si*

1. *$E$ est admissible,*
2. *$E$ satisfait la contrainte $C$.*

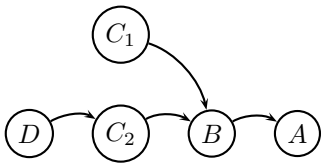
Remarquons que l'ensemble vide est admissible, mais qu'il n'est pas toujours $C$-admissible puisque $\widehat{\varnothing}$ n'implique pas toujours $C$.

**Définition 27 ([CMDM06] – extensions $C$-préférée, $C$-stable)** *Soit $E \subseteq \mathcal{A}$.*

– *$E$ est une* extension $C$-préférée *si et seulement si $E$ est maximal pour l'inclusion ensembliste parmi les ensembles $C$-admissibles.*

– *$E$ est une* extension $C$-stable *si et seulement si $E$ est une extension $C$-préférée qui attaque tous les arguments de $\mathcal{A}\backslash E$.*

**Exemple 6** *Soit* AF *un système d'argumentation défini par $\mathcal{A} = \{A, B, C_1, C_2, D\}$ et $\mathcal{R} = \{(D, C_2), (C_1, B), (C_2, B), (B, A)\}$. Le système* AF *se représente graphiquement par :*




*Ce système possède 4 ensembles admissibles : $E_1 = \varnothing$, $E_2 = \{D, C_1\}$, $E_3 = \{A, C_1\}$ et $E_4 = \{D, C_1, A\}$. Seule $E_4$ est une extension préférée, et dans cet exemple, $E_4$ est aussi une extension stable.*

*Transformons* AF *en un système à contraintes,* CoAF*, en ajoutant par exemple une contrainte qui empêche d'avoir les arguments $A$ et $D$ dans la même extension. Cette contrainte se définit par : $C = D \rightarrow \neg A$. Dans ce cas, $E_4$ = $\{D, C_1, A\}$ n'est pas un ensemble $C$-admissible, puisque l'ensemble $\widehat{E_4}$ = $\{D, C_1, A, \neg B, \neg C_2\}$ n'infère pas la formule $D \rightarrow \neg A$. Par contre, les ensembles admissibles $E_2 = \{D, C_1\}$ et $E_3 = \{A, C_1\}$ sont tous les deux $C$-admissibles et sont aussi des extensions $C$-préférées. Remarquons que sur cet exemple, il n'y a pas d'extensions $C$-stables.*

La dernière notion importante qui découle de la notion d'acceptabilité, quel que soit le type de système d'argumentation utilisé, est la notion de statut d'un argument.

**Définition 28 (Statut d'un argument)** *Soit un système d'argumentation (au sens de Dung, ou généralisé au sens de[CMDM06]), et $E_1, \ldots, E_x$ les extensions de ce système sous une sémantique donnée. Soit un argument $A \in \mathcal{A}$.*

1. *$A$ est* accepté *si et seulement si $A \in E_i$, $\forall E_i$ avec $i = 1, \ldots, x$.*
2. *$A$ est* rejeté *si et seulement si $\nexists E_i$ tel que $A \in E_i$.*
3. *$A$ est* indéfini *si et seulement si $A$ n'est ni accepté, ni rejeté. Cela signifie que $A$ appartient à certaines extensions et pas à d'autres.*

On peut facilement constater qu'un argument rejeté dans un système d'argumentation AF de base à la Dung sera aussi rejeté dans tous les systèmes à contrainte CoAF issus deAF.

**Proposition 7** *Soit $A \in \mathcal{A}$. Sous les sémantiques stable ou préférée, si $A$ est rejeté dans* AF $= \langle \mathcal{A}, \mathcal{R} \rangle$*, alors $A$ est aussi rejeté dans* CoAF $= \langle \mathcal{A}, \mathcal{R}, C \rangle$*, $\forall C$.*

**Exemple 6 (suite)** Dans AF, $D$, $C_1$ et $A$ sont acceptés ; $C_2$ et $B$ sont rejetés.
Dans CoAF, $C_1$ est accepté ; $C_2$ et $B$ sont toujours rejetés ; $D$ et $A$ sont maintenant indéfinis.

### 3.2.2 Évaluation d'arguments

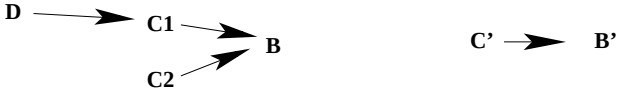
Durant cette phase du processus d'argumentation, nous partons de l'hypothèse suivante : nous avons à notre disposition un ensemble d'arguments et une relation d'attaque et nous allons chercher à attribuer à chaque argument un poids prenant en compte un critère particulier.

Il existe déjà des travaux sur ce sujet suivant le critère utilisé. Par exemple, le poids de l'argument peut être calculé en utilisant une relation de préférences sur les connaissances qui ont servi à créer les arguments. Une autre possibilité peut être d'exploiter la nature du lien qui lie les composants de l'argument (partie prémisse et partie conclusion).

Ici, je me suis attachée à calculer des poids pour les arguments en prenant en compte les interactions dans le cadre d'un système d'argumentation unipolaire, donc les attaques et les défenses que subit l'argument. Il y a deux points de vue possibles :

– soit la valeur d'un argument ne dépend que des valeurs de ses attaquants directs ; dans ce cas, les défenseurs sont pris en compte au travers de ces attaquants directs ; cette approche est dite *locale*.
– soit la valeur d'un argument représente la totalité des branches d'attaque et de défense menant à cet argument ; cette approche est dite *globale*.

La principale différence entre ces deux approches est illustrée par l'exemple suivant :




Avec l'approche locale, $B$ possède deux attaquants directs ($C_2$ et $C_1$) alors que $B'$ n'a en qu'un seul ($C'$). Ainsi $B'$ apparaît meilleur que $B$ (puisque $B'$ subit une attaque alors que $B$ en subit deux).

Avec l'approche globale, deux branches (une d'attaque et une de défense) conduisent à $B$ alors qu'une seule branche d'attaque conduit à $B'$. Ainsi $B$ apparaît meilleur que $B'$ (puisque il existe au moins une défense pour $B$ alors que $B'$ n'en a aucune). Dans ce cas, $C_1$ perd son statut négatif d'attaquant, puisqu'en fait, il est "porteur d'une défense" pour $B$.

Ce travail a été réalisé avec Claudette CAYROL et publié entre autres dans [CLS05b].

#### 3.2.2.1 Approche locale (évaluation générique)

Il existe déjà dans la littérature d'autres approches locales.

Par exemple, dans [JV99b], un étiquetage a été défini permettant d'assigner un statut à chaque argument en utilisant l'ensemble des étiquettes $\{+, -, ?\}$ : $+$ (resp. $-$, $?$) représente le statut "accepté" (resp. "rejeté", "indécis"). Intuitivement, un argument étiqueté avec $?$ est à la fois renforcé et affaibli.

**Définition 29 ([JV99b])** *Soit $\langle \mathcal{A}, \mathcal{R} \rangle$ un système d'argumentation unipolaire. Un étiquetage complet de $\langle \mathcal{A}, \mathcal{R} \rangle$ est une fonction $Lab : \mathcal{A} \rightarrow \{+, ?, -\}$ telle que :*

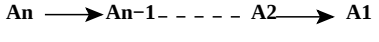
1. *Si $Lab(A) \in \{?, -\}$ alors $\exists B \in \mathcal{R}^-(A)$ tel que $Lab(B) \in \{+, ?\}$*
2. *Si $Lab(A) \in \{+, ?\}$ alors $\forall B \in \mathcal{R}^-(A) \cup \mathcal{R}^+(A)$, $Lab(B) \in \{?, -\}$*

L'intuition sous-jacente est qu'un argument peut être affaibli (étiquette $-$ ou $?$) seulement si un de ses attaquants directs est renforcé (condition 1) ; un argument peut être renforcé seulement si un de ses attaquants directs est affaibli et un argument qui est renforcé (étiquette $+$ ou $?$) affaiblit les arguments qu'il attaque (condition 2). Ainsi :
– si $A$ n'a pas d'attaquant $Lab(A) = +$.
– si $Lab(A) = ?$ alors $\exists B \in \mathcal{R}^-(A)$ tel que $Lab(B) = ?$.
– si ($\forall B \in \mathcal{R}^-(A)$, $Lab(B) = -$) alors $Lab(A) = +$.
– si $Lab(A) = +$ alors $\forall B \in \mathcal{R}^-(A) \cup \mathcal{R}^+(A)$, $Lab(B) = -$.
Tout système d'argumentation unipolaire peut être complètement étiqueté et la sémantique associée dit qu'un ensemble d'arguments $S$ est acceptable ssi il existe un étiquetage complet $Lab$ de $\langle \mathcal{A}, \mathcal{R} \rangle$ tel que $S = \{A | Lab(A) = +\}$.

D'autres types d'étiquetage ont été définis dans [JV99b], dont certains permettent d'établir un lien avec les sémantiques pour l'acceptabilité proposées par [Dun95]. C'est le cas de l'étiquetage enraciné :

**Définition 30 ([JV99b])**
*L'étiquetage complet $Lab$ est enraciné ssi $\forall A \in \mathcal{A}$, si $Lab(A) = -$ alors $\exists B \in \mathcal{R}^-(A)$ tel que $Lab(B) = +$.*

**Exemple 7** *Dans l'exemple suivant :*




*Pour $n$ pair, $Lab(A_n) = Lab(A_{n-2}) = \ldots = Lab(A_2) = +$ et $Lab(A_{n-1}) = Lab(A_{n-3}) = \ldots = Lab(A_1) = -$.*

*Pour $n$ impair, $Lab(A_n) = Lab(A_{n-2}) = \ldots = Lab(A_1) = +$ et $Lab(A_{n-1}) = Lab(A_{n-3}) = \ldots = Lab(A_2) = -$.*

Un autre type d'*évaluation locale* a été défini par [BH01] pour des arguments dits "déductifs". Cette approche se caractérise par les points suivants :
- la structure d'un argument est une paire $\langle$*explication*, *conclusion*$\rangle$ avec *explication* qui est un ensemble consistant de formules permettant d'inférer la formule *conclusion*.
- la relation d'attaque considérée dans [BH01] est stricte et acyclique.
- la notion d'"arbre d'arguments" permet de représenter de manière concise et exhaustive les attaquants et les défenseurs d'un argument donné qui sera la racine de cet arbre.
- une fonction appelée "catégoriseur" assigne à un arbre une valeur qui représente la force relative de la racine de cet arbre en fonction des attaquants et des défenseurs de cette racine.
- une autre fonction appelée "accumulateur" synthétise les valeurs de tous les arbres d'arguments dont les racines sont des arguments pour (resp. contre) une conclusion donnée.

La phase de catégorisation correspond exactement à une évaluation basée sur les interactions. Dans [BH01], on peut trouver une instance possible de la fonction $Cat$ :

- si $\mathcal{R}^-(A) = \varnothing$, alors $Cat(A) = 1$,
- si $\mathcal{R}^-(A) \neq \varnothing$ avec $\mathcal{R}^-(A) = \{A_1, \ldots, A_n\}$, alors $Cat(A) = \frac{1}{1+Cat(A_1)+\ldots+Cat(A_n)}$.

Intuitivement, plus un argument possède d'attaquants directs et plus sa valeur sera faible. De même, plus un argument possède de défenseurs et plus sa valeur sera forte.

**Exemple 7 page précédente (suite)** $Cat(A_n) = 1$, $Cat(A_{n-1}) = 0.5$, $Cat(A_{n-2}) = 0.66$, $Cat(A_{n-3}) = 0.6$, ..., *et* $Cat(A_1) = (\sqrt{5}-1)/2$ *quand* $n \to \infty$ *(cette valeur est l'inverse du nombre d'or*[7]*).*

*Ainsi, on obtient :*

*Si* $n$ *est pair* $Cat(A_{n-1}) \leq \ldots \leq Cat(A_3) \leq Cat(A_1) \leq Cat(A_2) \leq \ldots \leq Cat(A_n) = 1$.

*Si* $n$ *est impair* $Cat(A_{n-1}) \leq \ldots \leq Cat(A_2) \leq Cat(A_1) \leq Cat(A_3) \leq \ldots \leq Cat(A_n) = 1$.

L'approche que nous avons développée pour les *évaluations locales* est une généralisation des deux propositions présentées ci-dessus dans le sens où la fonction $Cat$ de [BH01] et les étiquetages de [JV99b] sont des instances de notre approche.

L'idée principale est que la valeur d'un argument est obtenue par la composition de deux fonctions :
- une pour agréger les valeurs des attaquants directs ; donc, cette fonction calcule la valeur de l'"attaque directe" ;
- et une autre pour calculer les effets de cette "attaque directe" sur la valeur de l'argument : si la valeur de l'"attaque directe" augmente alors la valeur de l'argument décroît et si la valeur de l'"attaque directe" diminue alors la valeur de l'argument augmente.

Soit $(W, \geq)$ un ensemble totalement ordonné possédant un élément minimum ($\mathcal{V}_{\text{min}}$). Soit $V$ un sous-ensemble de $W$, qui contient $\mathcal{V}_{\text{min}}$ et qui possède un élément maximum $\mathcal{V}_{\text{max}}$.

**Définition 31 (Evaluation graduelle générique)** *Soit* $\langle \mathcal{A}, \mathcal{R} \rangle$ *un système d'argumentation unipolaire. Une évaluation graduelle est une fonction* $v : \mathcal{A} \to V$ *telle que :*

1. $\forall A \in \mathcal{A}$, $v(A) \geq \mathcal{V}_{min}$
2. $\forall A \in \mathcal{A}$, *si* $\mathcal{R}^-(A) = \varnothing$, *alors* $v(A) = \mathcal{V}_{max}$
3. $\forall A \in \mathcal{A}$, *si* $\mathcal{R}^-(A) = \{A_1, \ldots, A_n\} \neq \varnothing$, *alors* $v(A) = g(h(v(A_1), \ldots, v(A_n)))$

*avec la fonction* $h : V^* \to W$ *telle que (*$V^*$ *dénote l'ensemble de toutes les séquences finies d'éléments de* $V$*)*
- $h(x) = x$
- $h() = \mathcal{V}_{min}$
- *Pour toute permutation* $(x_{i1}, \ldots, x_{in})$ *de* $(x_1, \ldots, x_n)$, $h(x_{i1}, \ldots, x_{in}) = h(x_1, \ldots, x_n)$
- $h(x_1, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_n)$
- *si* $x_i \geq x'_i$ *alors* $h(x_1, \ldots, x_i, \ldots, x_n) \geq h(x_1, \ldots, x'_i, \ldots, x_n)$

[7] Le nombre d'or est célèbre depuis l'antiquité et on le retrouve un peu partout (en architecture, par exemple).

*et la fonction* $g : W \to V$ *telle que*
– $g(\mathcal{V}_{min}) = \mathcal{V}_{max}$
– $g(\mathcal{V}_{max}) < \mathcal{V}_{max}$
– $g$ *est décroissante (si* $x \leq y$ *alors* $g(x) \geq g(y)$*)*

Remarquons que $h(x_1, \ldots, x_n) \geq \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)$ est une conséquence logique des propriétés imposées à la fonction $h$.

Une première propriété de la fonction $g$ explique le comportement de l'évaluation locale dans le cas d'un argument racine d'une seule branche (comme dans l'exemple 7 page 39) :

**Proposition 8** *La fonction* $g$ *vérifie pour tout* $n \geq 1$ *:*

$$g(\mathcal{V}_{max}) \leq g^3(\mathcal{V}_{max}) \leq \ldots \leq g^{2n+1}(\mathcal{V}_{max}) \leq g^{2n}(\mathcal{V}_{max}) \leq \ldots \leq g^2(\mathcal{V}_{max}) \leq \mathcal{V}_{max}$$

*De plus, si* $g$ *est strictement décroissante et que* $g(\mathcal{V}_{max}) > \mathcal{V}_{min}$*, l'inégalité précédente devient stricte.*

Une seconde propriété montre que l'évaluation locale induit une relation d'ordre sur les arguments :

**Proposition 9 (Préordre complet)** *Soit* $v$ *une valuation au sens de la définition 31 page précédente.* $v$ *induit un préordre complet*[8] $\succeq$ *sur l'ensemble des arguments* $\mathcal{A}$ *défini par :* $A \succeq B$ *ssi* $v(A) \geq v(B)$*.*

Une troisième propriété permet de gérer les circuits :

**Proposition 10 (Valeur dans un circuit)** *Soit* $\mathcal{C}$ *un circuit isolé d'un graphe d'attaque, dont la longueur est* $n$*. Si* $n$ *est impair, tous les arguments du circuit ont la même valeur et cette valeur est un point fixe de la fonction* $g$*. Si* $n$ *est pair, la valeur de chaque argument du circuit est un point fixe de la fonction* $g^n$*.*

La propriété suivante montre les principes sous-jacents satisfaits par les évaluations locales :

**Proposition 11 (Principes sous-jacents)** *Les évaluations graduelles locales données par la définition 31 page ci-contre respectent les principes suivants :*

**P1** *La valuation est maximale pour un argument sans attaquant et non maximale pour un argument attaqué et non défendu.*

**P2** *La valuation d'un argument est une fonction de la valuation de ses attaquants directs (l'"attaque directe").*

**P3** *La valuation d'un argument est une fonction décroissante de la valuation de l'"attaque directe".*

**P4** *Chaque attaquant de l'argument contribue à l'augmentation de la valuation de l'"attaque directe" pour cet argument.*

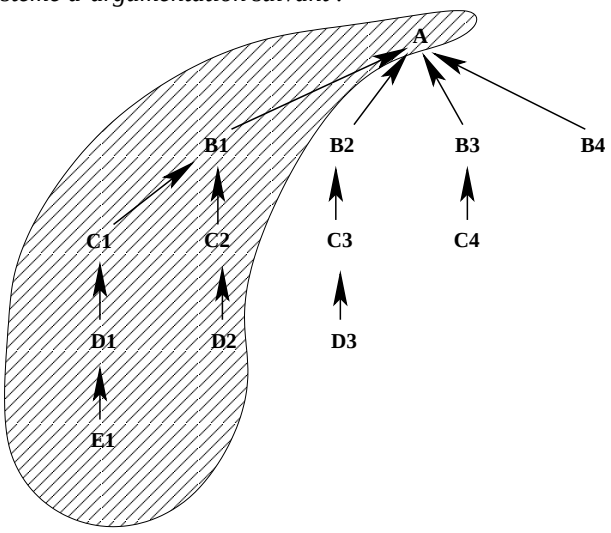
Les dernières propriétés expliquent pourquoi les évaluations de [JV99b] et de [BH01] sont des instances de l'évaluation locale décrite dans la définition 31 page précédente :

**Proposition 12 (Lien avec [JV99b])**
*Chaque étiquetage enraciné au sens de [JV99b] peut être défini comme une instance de l'évaluation graduelle générique telle que :*
– $V = W = \{-, ?, +\}$ *avec* $- < ? < +$*,*
– $\mathcal{V}_{min} = -$*,*
– $\mathcal{V}_{max} = +$*,*
– $g$ *définie par* $g(-) = +$*,* $g(+) = -$*,* $g(?) = ?$
– *et* $h$ *étant la fonction* $\mathcal{M}\mathtt{ax}$*.*

[8] Un préordre complet sur $\mathcal{A}$ signifie que tous les éléments de $\mathcal{A}$ sont comparables deux à deux.

**Proposition 13 (Lien avec [BH01])** *L'évaluation graduelle proposée par [BH01] peut être définie comme une instance de l'évaluation graduelle générique telle que :*
- $V = [0, 1]$,
- $W = [0, \infty[$,
- $\mathcal{V}_{min} = 0$,
- $\mathcal{V}_{max} = 1$,
- $g : W \rightarrow V$ *définie par* $g(x) = \frac{1}{1+x}$
- *et* $h$ *définie par* $h(x_1, \ldots, x_n) = x_1 + \ldots + x_n$.

**Exemple 8** *Considérons le système d'argumentation suivant :*




*Dans cet exemple, avec l'évaluation générique, on obtient :*
- $v(E_1) = v(D_2) = v(D_3) = v(C_4) = v(B_4) = \mathcal{V}_{max}$
- $v(D_1) = v(C_2) = v(C_3) = v(B_3) = g(\mathcal{V}_{max})$
- $v(C_1) = v(B_2) = g^2(\mathcal{V}_{max})$
- $v(B_1) = g(h(g^2(\mathcal{V}_{max}), g(\mathcal{V}_{max})))$
- $v(A) = g(h(g(h(g^2(\mathcal{V}_{max}), g(\mathcal{V}_{max}))), g^2(\mathcal{V}_{max}), g(\mathcal{V}_{max}), \mathcal{V}_{max}))$

*Et on a :*

$$E_1, D_2, D_3, C_4, B_4$$
$$\succeq$$
$$C_1, B_2$$
$$\succeq$$
$$D_1, C_2, C_3, B_3$$

*Toutefois, on peut remarquer que les contraintes sur* $v(A)$ *et* $v(B_1)$ *sont insuffisantes pour comparer* $A$ *et* $B_1$ *avec les autres arguments.*

*Le même problème apparaît si on réduit l'exemple à la partie hachurée du graphe ; on obtient* $E_1, D_2 \succeq C_1 \succeq D_1, C_2$*, mais* $A$ *et* $B_1$ *ne peuvent toujours pas être comparés aux autres arguments*[9].

*Si on utilise l'instance de l'évaluation graduelle proposée par [BH01], on obtient :*
- $v(E_1) = v(D_2) = v(D_3) = v(C_4) = v(B_4) = 1$,
- $v(D_1) = v(C_2) = v(C_3) = v(B_3) = \frac{1}{2}$,
- $v(C_1) = v(B_2) = \frac{2}{3}$,
- $v(B_1) = \frac{6}{13}$,
- $v(A) = \frac{78}{283}$.

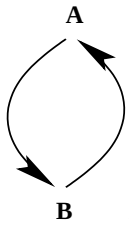
[9] $v(A) = g^2(h(g^2(\mathcal{V}_{\max}, g(\mathcal{V}_{\max}))))$ et $v(B_1) = g(h(g^2(\mathcal{V}_{\max}), g(\mathcal{V}_{\max})))$.

*et on a :*

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

*Toutefois, si on réduit l'exemple à sa partie hachurée, alors la valeur de $A$ est $\frac{13}{19}$. On a alors $v(A)$ meilleure que $v(B_1)$ et $v(D_1)$, mais aussi meilleure que $v(C_1)$ ($A$ devient meilleur que son défenseur).*

**Exemple 9 (Circuit isolé)** *Considérons maintenant un système d'argumentation réduit à un circuit isolé :*




.

*L'évaluation générique donne $v(A) = v(B) =$ le point fixe de la fonction $g^2$.*

*Et avec l'instance proposée par [BH01], $v(A) = v(B) = \frac{-1+\sqrt{5}}{2} \approx 0.618$ (l'inverse du nombre d'or).*

#### 3.2.2.2 Approche globale (avec tuples)

Considérons désormais la seconde approche dite "globale". Ici, l'idée clé est que la valeur de $A$ doit décrire la valeur du sous-graphe ayant $A$ pour racine. Nous voulons donc mémoriser la longueur de chaque branche conduisant à $A$ dans un tuple (pour une branche d'attaque, il y aura un nombre impair dans le tuple et pour une branche de défense, il y aura un nombre pair dans le tuple).

La contrainte principale consiste donc à identifier chaque branche menant à l'argument que l'on veut évaluer ainsi que le "type" de cette branche. C'est très simple à faire quand le graphe est acyclique et beaucoup plus complexe quand il existe des circuits. Je vais donc présenter d'abord une évaluation globale pour les cas sans circuit puis je donnerai quelques pistes pour traiter les circuits.

##### 3.2.2.2.1 Évaluation globale avec tuples pour les graphes acycliques

La mémorisation des longueurs des branches menant à un argument donné va se faire dans un tuple d'entiers. Pour manipuler ces tuples, on donne les définitions suivantes :

**Définition 32 (Tuple)** *Un tuple est une séquence d'entiers. Le tuple $\underbrace{(0, \ldots, 0, \ldots)}_{\infty}$ sera noté $0^\infty$. Le tuple $\underbrace{(1, \ldots, 1, \ldots)}_{\infty}$ sera noté $1^\infty$.*

**Notation 2** *$\mathcal{T}$ dénote l'ensemble de tous les tuples pouvant être construits à partir de valeurs positives entières.*

**Définition 33 (Opérations sur les tuples)** *Il existe deux types d'opérations sur les tuples :*

– *la concaténation de deux tuples définie par la fonction* $\star : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ *telle que*

$$\begin{aligned} 0^\infty \star t = t \star 0^\infty &= t \text{ pour } t \neq () \\ (x_1, \ldots, x_n, \ldots) \star (x'_1, \ldots, x'_n, \ldots) &= \texttt{Tri}(x_1, \ldots, x_n, \ldots, x'_1, \ldots, x'_n, \ldots) \end{aligned}$$

`Tri` *étant la fonction qui trie un tuple par ordre croissant de valeurs.*
– *l'addition d'un tuple et d'un entier définie par la fonction* $\oplus : \mathcal{T} \times \mathbb{N} \rightarrow \mathcal{T}$ *telle que*

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

Remarquons que les tuples infinis sont autorisés puisque, entre autres, ils permettront de prendre en compte les circuits (voir la section 3.2.2.2.4 page 50).

Ces opérations sur les tuples vérifient les propriétés suivantes :

**Proposition 14 (Propriétés de $\star$ et de $\oplus$)**
*La concaténation $\star$ est commutative et associative.*

*Pour tout tuple $t$ et pour tous entiers $k$ et $k'$, $(t \oplus k) \oplus k' = t \oplus (k + k')$.*

*Pour tout entier $k$ et pour tous tuples $t$ et $t'$ différents de $0^\infty$*[10]*, $(t \star t') \oplus k = (t \oplus k) \star (t' \oplus k)$.*

Afin d'évaluer les arguments, l'ensemble des longueurs des branches menant à l'argument est divisé en deux sous-ensembles, celui des valeurs paires (longueurs des branches de défense)) et celui des valeurs impaires (longueurs des branches d'attaque). Ceci nous conduit à la notion de valeurs tuplées :

**Définition 34 (Valeurs tuplées)** *Une* valeur tuplée *est une paire de tuples $vt = [vt_p, vt_i]$ avec :*
– *$vt_p$ un tuple d'entiers pairs trié par ordre croissant ; ce tuple est le composant pair de $vt$ ;*
– *$vt_i$ un tuple d'entiers impairs trié par ordre croissant ; ce tuple est le composant impair de $vt$.*

**Notation 3** *$\mathcal{V}$ dénote le sous-ensemble de $\mathcal{T} \times \mathcal{T}$ contenant toutes les valeurs tuplées (donc, $\forall vt \in \mathcal{V}$, $vt$ est une paire de tuples vérifiant la définition 34).*

En utilisant ces valeurs tuplées, on peut donner un *processus de calcul* de l'évaluation globale en cas de graphes acycliques.

**Définition 35 (Evaluation avec tuples pour des graphes acyclique)** *Soit $\langle \mathcal{A}, \mathcal{R} \rangle$ un système d'argumentation unipolaire sans circuits. Une évaluation avec tuples est une fonction $v : \mathcal{A} \rightarrow \mathcal{V}$ telle que :*

*Si $A \in \mathcal{A}$ est une feuille du graphe alors*

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

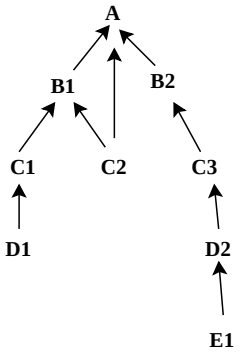
*Si $A \in \mathcal{A}$ a des attaquants directs notés $B_1$, …, $B_n$, … alors*

$$v(A) = [v_p(A), v_i(A)] \text{ avec : } \begin{vmatrix} v_p(A) = (v_i(B_1) \oplus 1) \star \ldots \star (v_i(B_n) \oplus 1) \star \ldots \\ v_i(A) = (v_p(B_1) \oplus 1) \star \ldots \star (v_p(B_n) \oplus 1) \star \ldots \end{vmatrix}$$

Remarquons que le choix de la valeur $[0^\infty, ()]$ pour les feuilles est justifié par le fait que la valeur d'un argument mémorise toutes les longueurs des branches y menant et il est impossible d'avoir pour un argument donné $v_p(A) = v_i(A) = ()$.

**Exemple 10** *Sur ce graphe, l'évaluation avec tuples donne les résultats suivants :*

[10] Sinon c'est faux : $(0^\infty \star (p)) \oplus k = (p + k)$, alors que $(0^\infty \oplus k) \star ((p) \oplus k) = (k) \star (p + k) = (k, p + k)$.



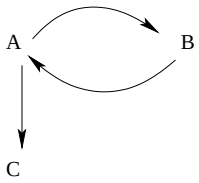

*On obtient :*
- $v(D_1) = v(C_2) = v(E_1) = [0^\infty, ()]$,
- $v(C_1) = v(D_2) = [(), (1)]$,
- $v(C_3) = [(2), ()]$,
- $v(B_1) = [(2), (1)]$,
- $v(B_2) = [(), (3)]$,
- $v(A) = [(2, 4), (1, 3)]$.

#### 3.2.2.2.2 Étude des circuits

Quand le graphe d'argumentation contient des circuits, la notion de branche n'est parfois plus vraiment significative (par exemple, dans un circuit non attaqué comme ceux des exemples 9 page 43 et 11), et même quand cette notion garde son sens, la longueur d'une branche doit être définie différemment.

Ces remarques sont facilement illustrées par différents exemples :

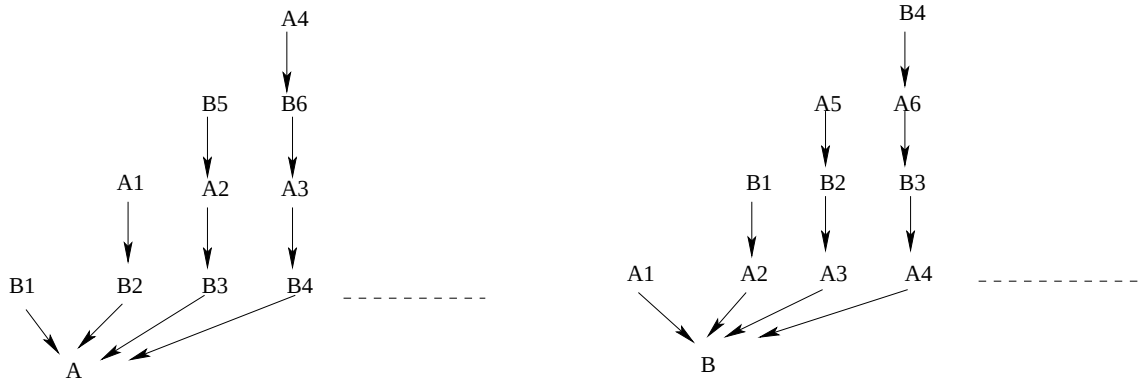
**Exemple 11 (Circuit non attaqué)** *Le graphe se réduit au circuit non attaqué $A - B - A$ qui attaque l'argument $C$ :*




*La notion de branche est inutilisable ici, puisqu'il n'existe pas de feuille.*

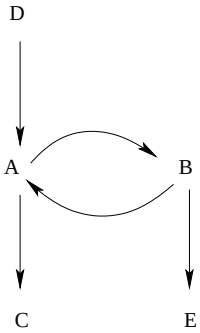
*Il nous reste alors deux possibilités :*
- *soit on considère que le circuit correspond à une branche infinie ; mais alors on ne peut connaître sa parité.*
- *soit on considère que le circuit correspond à une infinité de branches ; donc $A$ (resp. $B$) est la racine d'une infinité de branches d'attaque et d'une infinité de branches de défense dont les longueurs sont connues et finies.*

*La seconde possibilité signifie que le circuit peut avoir deux représentations sous la forme d'un graphe acyclique et infini (un avec $A$ en racine et un autre avec $B$ en racine). Il s'agit donc d'un processus de réécriture du circuit :*



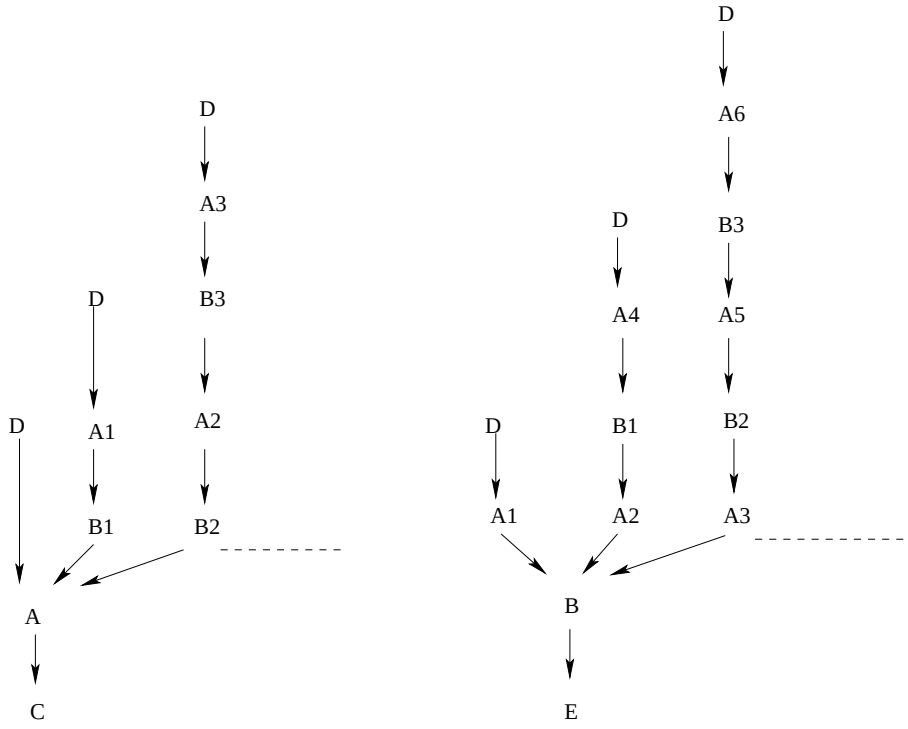

*Les $A_i$ et $B_i$ sont de nouveaux arguments créés lors du processus de réécriture du circuit.*

**Exemple 12 (Circuit attaqué)** *Le circuit $A - B - A$ est attaqué par au moins un argument qui n'appartient pas au circuit (ici, l'attaquant est l'argument $D$ qui est non attaqué) :*




*Dans ce cas, la notion de branche a un sens parce qu'il existe bien une feuille dans le graphe, mais la difficulté consiste ici à calculer la longueur de ces branches. Comme dans l'exemple 11 page précédente, nous pouvons soit considérer qu'il existe une seule branche infinie (et du coup, on ne peut savoir si c'est une branche d'attaque ou une branche de défense), soit considérer qu'il existe une infinité de branches dont les longueurs sont connues et finies.*

*En appliquant la seconde possibilité, le graphe peut être réécrit en utilisant les structures suivantes :*

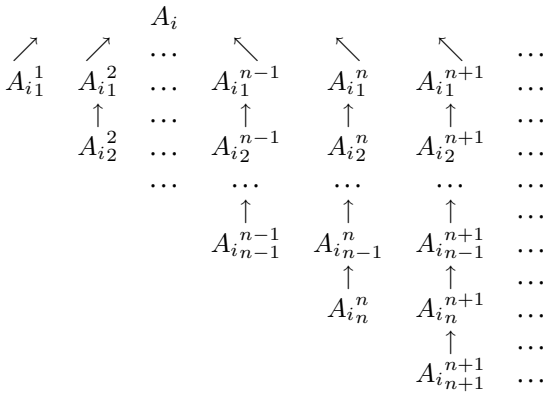



*Les $A_i$ et $B_i$ sont de nouveaux arguments créés lors du processus de réécriture du circuit.*

En partant des exemples précédents, j'ai choisi de gérer un circuit sous la forme d'une infinité de branches d'attaque et de défense ; cela me permet d'appliquer la définition 35 page 44 pour l'évaluation par tuples à tout type de graphes (cyclique ou pas). Par contre, il me faut disposer d'un processus de réécriture automatique des circuits. Ce processus dépend des attaques que subit le circuit :

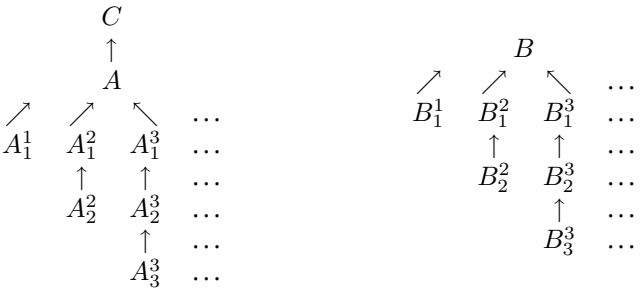
**Définition 36 (Réécriture d'un circuit non attaqué)** *Soit $\mathcal{C} = A_0 - A_1 - \ldots - A_{n-1} - A_0$ un circuit non attaqué. Le graphe $\mathcal{G}$ qui contient $\mathcal{C}$ est réécrit comme suit :*

1. *le circuit $\mathcal{C}$ est supprimé,*
2. *et il est remplacé par les graphes acycliques infinis suivants (un pour chaque $A_i$, $i = 0 \ldots n-1$) :*



3. *les arcs entre les $A_i$ et les arguments n'appartenant pas à $\mathcal{C}$ sont conservés.*

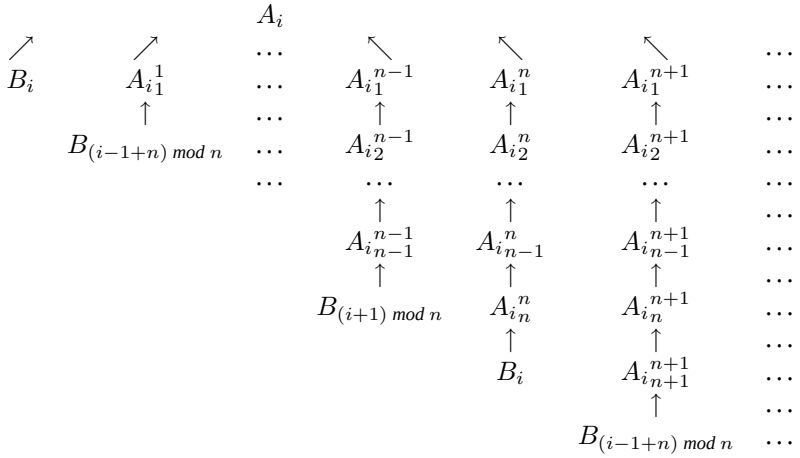
**Exemple 11 page 45 (suite)** *Le graphe $\mathcal{G}$ contenant le circuit non attaqué $A - B - A$ et l'argument $C$, attaqué par $A$, est réécrit sous la forme :*



*avec $A_k^l$ et $B_k^l$ de nouveaux arguments.*

**Définition 37 (Réécriture d'un circuit attaqué)** *Soit $\mathcal{C} = A_0 - A_1 - \ldots - A_{n-1} - A_0$ un circuit attaqué, l'attaquant direct de chaque $A_i$ extérieur à $\mathcal{C}$ est noté $B_i$, s'il existe. Le graphe $\mathcal{G}$ contenant $\mathcal{C}$ est réécrit comme suit :*

1. *le circuit $\mathcal{C}$ est supprimé,*
2. *et il est remplacé par les graphes acycliques infinis suivants (un pour chaque $A_i$ $i = 0 \ldots n-1$) :*



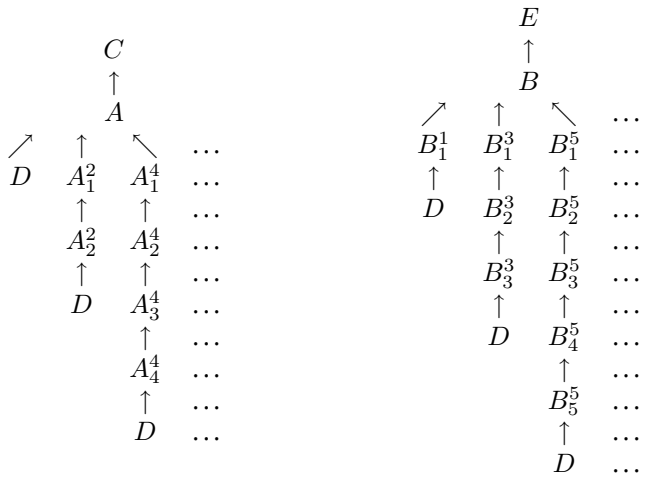
*(les branches conduisant à $B_k$ existent si et seulement si $B_k$ existe[11]).*

3. *les arcs entre les $A_i$ et les arguments n'appartenant pas à $\mathcal{C}$ sont conservés.*

[11] L'opérateur mod correspond à la fonction modulo.

4. *les arcs entre les $B_i$ et les arguments n'appartenant pas à $\mathcal{C}$ sont conservés.*

**Exemple 12 page 45 (suite)** *Le graphe $\mathcal{G}$ contenant le circuit $A - B - A$ attaqué en $A$ par l'argument $D$ et contenant aussi l'argument $C$ (resp. $E$) attaqué par $A$ (resp. $B$) est réécrit sous la forme :*



*avec $A_k^l$ et $B_k^l$ de nouveaux arguments.*

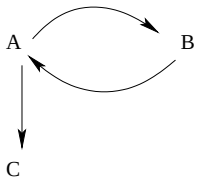
S'il existe plusieurs circuits dans le graphe, deux cas sont alors possibles :
– Si ces circuits ne sont pas interconnectés, on peut réécrire chaque circuit séparément et le résultat ne dépend pas de l'ordre de réécriture.
– S'ils sont interconnectés, ils sont considérés comme un métacyle pouvant être soit attaqué soit non attaqué et on peut ici aussi réécrire le graphe (il s'agit d'un processus plus complexe décrit dans [CLS05b]).

#### 3.2.2.2.3 Une évaluation graduelle avec tuples pour des graphes quelconques

En utilisant les définitions données en sections 3.2.2.2.1 page 43 et 3.2.2.2.2 page 45, l'évaluation graduelle avec tuples donnée dans la définition 35 page 44 est applicable pour des graphes quelconques *après l'exécution du processus de réécriture*.

Voyons sur quelques exemples ce que l'on obtient.

**Exemple 11 page 45 (suite)** *Considérons le graphe suivant :*




*La réécriture de ce graphe est donnée en section 3.2.2.2.2 page 45.*

*La définition 35 page 44 fournit alors les résultats suivants :*

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

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

*Ce qui nous donne :*
– $v(A_n^n) = [0^\infty, ()]$ *pour chaque* $n \geq 1$

– $v(A^n_{n-1}) = [(), (1)]$ *pour chaque* $n \geq 2$
– $v(A^m_n) = [v_p(A^m_{n+2}) \oplus 2, v_i(A^m_{n+2}) \oplus 2]$ *pour chaque* $n \geq 1$ *et* $m \geq n+2$
*Et en combinant avec* $v_p(A)$ *et* $v_i(A)$*, on obtient deux séquences de tuples : une séquence (*$x_k$*,* $k \geq 1$*) de tuples infinis d'entiers pairs, et une séquence (*$y_k$*,* $k \geq 1$*) de tuples infinis d'entiers impairs :*

$$x_k = (2) \star (v_i(A^{2k+1}_{2k-1}) \oplus 1) \star \ldots \star (v_i(A^n_{2k-1}) \oplus 1) \star \ldots$$

$$y_k = (1) \star (v_p(A^{2k+1}_{2k-1}) \oplus 1) \star \ldots \star (v_p(A^n_{2k-1}) \oplus 1) \star \ldots$$

*D'après la propriété 14 page 44, on peut facilement montrer que* $v_p(A) = x_1$ *et que pour chaque* $k \geq 1$, $x_k = (2) \star (x_{k+1} \oplus 2)$.

*De la même manière,* $v_i(A) = y_1$ *et pour chaque* $k \geq 1$, $y_k = (1) \star (y_{k+1} \oplus 2)$.

*Ces équations prouvent que :*

*Pour chaque entier pair* $p$, $p > 0$, $p$ *appartient à chaque tuple* $x_i, i \geq 1$.
*Pour chaque entier impair* $p$, $p$ *appartient à chaque tuple* $y_i$, $i \geq 1$.

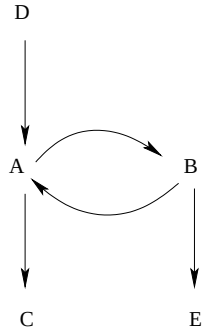
*La preuve se fait par induction sur* $p$.

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

Remarquons que tous les résultats évoqués ci-dessus peuvent sans problème être étendus à des circuits de longueur $n$, $n \geq 2$.

**Proposition 15 (Propriétés d'un circuit non attaqué)**
*Pour chaque circuit non attaqué, pour chaque argument* $A$ *du circuit,* $v(A) = [(2, 4, 6, \ldots), (1, 3, 5, \ldots)]$.

**Exemple 12 page 45 (suite)** *Considérons le graphe suivant :*




*La réécriture de ce graphe a été donnée en section 3.2.2.2.2 page 45.*

*La définition 35 page 44 fournit alors les résultats suivants :*

$$v_p(A) = (v_i(D) \oplus 1) \star (v_i(A^2_1) \oplus 1) \star \ldots \star (v_i(A^{2n}_1) \oplus 1) \star \ldots$$

$$v_i(A) = (v_p(D) \oplus 1) \star (v_p(A^2_1) \oplus 1) \star \ldots \star (v_p(A^{2n}_1) \oplus 1) \star \ldots$$

*et aussi*
$v(D) = [0^\infty, ()]$
$v(A^n_n) = [(), (1)]$ *pour chaque* $n \geq 2$

*Exactement comme dans l'exemple 11 page 45, les formules $v_p(A)$ et $v_i(A)$ peuvent être réécrites pour mettre en lumière certaines séquences de tuples.*

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

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

*On peut alors montrer facilement que $v_p(A) = x'_1$ et que pour chaque $k \geq 1$, $x'_k = (x'_{k+1} \oplus 2)$.*

*De même, $v_i(A) = y'_1$ et pour chaque $k \geq 1$, $y'_k = (1) \star (y'_{k+1} \oplus 2)$.*

*Cela permet de prouver que $x'_1$ est le tuple vide et que $y'_1 = (1, 3, 5, \ldots)$.*

*Ainsi, $v(A) = [(), (1, 3, 5, \ldots)]$. Pour $B$, le raisonnement est le même que pour $A$, et on obtient $v(B) = [(2, 4, 6, \ldots), ()]$. On a alors $v(C) = [(2, 4, 6, \ldots), ()]$ et $v(E) = [(), (3, 5, 7 \ldots)]$.*

**Conclusion à propos des circuits** Le traitement des circuits est coûteux puisque les valeurs obtenues sont infinies.

Dans [CLS05b, CLS05a], des algorithmes ont été proposés pour calculer ces valeurs ; ce sont des algorithmes qui travaillent par propagation et qui sont paramétrés par un nombre limite de passage dans un circuit ; ils fournissent donc en sortie des valeurs tuplées finies et donc incomplètes.

#### 3.2.2.2.4 Comparaison de valeurs tuplées

Dans cette section, je donne la définition de la relation de comparaison entre valeurs tuplées. Cette comparaison utilise l'idée suivante : un argument $A$ est meilleur qu'un argument $B$ ssi $A$ possède une meilleure défense et subit une attaque plus faible.

La première possibilité semble être d'utiliser un ordre lexicographique sur les tuples. Cet ordre noté par $\leq_{lex\infty}$ sur $\mathcal{T}$ peut se définir de la manière suivante :

**Définition 38 (Ordre lexicographique sur les tuples)** *Soit $(x_1, \ldots, x_n, \ldots)$ et $(y_1, \ldots, y_m, \ldots)$ deux tuples $\in \mathcal{T}$.*

*$(x_1, \ldots, x_n, \ldots) <_{lex\infty} (y_1, \ldots, y_m, \ldots)$ ssi $\exists i \geq 1$ tel que :*
- *$\forall j < i$, $x_j = y_j$ et*
- *$y_i$ existe et :*
  - *soit le tuple $(x_1, \ldots, x_n, \ldots)$ est fini avec un nombre d'éléments égal à $i - 1$ (donc, $x_i$ n'existe pas),*
  - *soit $x_i$ existe et $x_i < y_i$.*

*$(x_1, \ldots, x_n, \ldots) =_{lex\infty} (y_1, \ldots, y_m, \ldots)$ ssi les tuples contiennent le même nombre $p \in \mathbb{N} \cup \{\infty\}$ d'éléments et $\forall i$, $1 \leq i \leq p$, $x_i = y_i$.*

*Ainsi, on a : $(x_1, \ldots, x_n, \ldots) \leq_{lex\infty} (y_1, \ldots, y_m, \ldots)$ ssi*
*$(x_1, \ldots, x_n, \ldots) =_{lex\infty} (y_1, \ldots, y_m, \ldots)$ ou $(x_1, \ldots, x_n, \ldots) <_{lex\infty} (y_1, \ldots, y_m, \ldots)$.*

La relation d'ordre $<_{lex\infty}$ est une généralisation de l'ordre lexicographique classique (voir [Xuo92]) au cas des tuples infinis. Cette relation est complète mais n'est pas bien-fondée (il existe des séquences infinies qui sont strictement décroissantes : $(0) <_{lex\infty} (0, 0) <_{lex\infty} \ldots <_{lex\infty} (0, \ldots, 0, \ldots) <_{lex\infty} \ldots <_{lex\infty} (0, 1)$).

Par contre, les valeurs paires et les valeurs impaires ne jouant pas le même rôle, une comparaison reposant sur un ordre lexicographique classique ne suffit pas. L'idée est donc de faire la comparaison de deux arguments en suivant en deux étapes bien distinctes :

– La “première étape” *compare le nombre de branches d'attaque et le nombre de branches de défense* de chaque argument. Nous avons ainsi deux critères (un pour la défense et un pour l'attaque) qu'il nous faut agréger suivant une *méthode prudente* : nous ne pourrons conclure que si un des deux arguments possède plus de branches de défense (il est meilleur du point de vue de la défense) et moins de branches d'attaque (il est aussi meilleur du point de vue de l'attaque) que l'autre. Il faut que les deux points de vue s'accordent : en effet, si un des deux arguments possède plus de branches de défense (il est meilleur du point de vue de la défense) et plus de branches d'attaque (il est moins bon du point de vue de l'attaque) que l'autre, il est impossible de conclure.
– Dans le cas où les arguments possèdent les mêmes nombres de branches de défense et de branches d'attaque, une “seconde étape” *compare la qualité des attaques et la qualité des défenses* en utilisant la longueur de chaque branche. Cette comparaison est réalisée en respectant un principe lexicographique (voir La définition 38 page précédente) et fournit deux critères qui doivent à nouveau être agrégés de manière prudente. Ici aussi, en cas de désaccord entre les deux critères, les arguments seront considérés comme incomparables.

Voyons sur quelques exemples ce que cela donne :

– $[(2),(1)]$ est meilleure que $[(2),(1,1)]$ parce qu'il y a moins de branches d'attaque dans la première valeur tuplée que dans la seconde, le nombre de branches de défense étant le même (première étape).
– $[(2),(1)]$ est incomparable avec $[(2,2),(1,1)]$ parce qu'il y a moins de branches de défense et moins de branches d'attaque dans la première valeur tuplée que dans la seconde (première étape).
– $[(2),(3)]$ est meilleure que $[(2),(1)]$ parce que les branches d'attaque dans la première valeur tuplée sont plus “faibles” que celles de la seconde valeur tuplée (la branche d'attaque de la première valeur est plus longue que celle de la seconde valeur), les branches de défense étant identiques (seconde étape, en utilisant la comparaison lexicographique sur les parties paires et sur les parties impaires).
– $[(2),(3)]$ est meilleure que $[(4),(3)]$ parce que les branches de défense sont plus “fortes” dans la première valeur que dans la seconde valeur (la branche de défense est plus courte dans la première valeur que dans la seconde), les branches d'attaque étant identiques (seconde étape).
– $[(2),(1)]$ est incomparable avec $[(4),(3)]$ parce que, dans la première valeur, les branches d'attaque sont moins bonnes alors que celles de défense sont meilleures que leurs équivalents dans la seconde valeur (seconde étape).

L'algorithme 1 page suivante permet d'implémenter cette comparaison d'arguments en utilisant le principe d'une double comparaison (d'abord quantitative, puis ensuite qualitative) avec deux critères (un portant sur l'attaque et l'autre portant sur la défense) que l'on agrège de manière prudente.

L'algorithme 1 page suivante définit un préordre partiel sur l'ensemble $v(\mathcal{A})$ :

**Proposition 16 (Préordre partiel)** *L'algorithme 1 page suivante définit sur l'ensemble $v(\mathcal{A})$ un préordre partiel $\succeq$, dont la seule valeur maximale est $[0^\infty,()]$ et dont la seule valeur minimale est $[(),1^\infty]$.*

**Notation 4** *le préordre partiel $\succeq$ sur l'ensemble $v(\mathcal{A})$ induit un préordre partiel sur les arguments (le préordre partiel sur $\mathcal{A}$ sera noté comme le préordre partiel sur $v(\mathcal{A})$) : $A \succeq B$ si et seulement si $v(A) \succeq v(B)$*[12].

Afin d'identifier précisément les principes sous-jacents satisfaits par l'évaluation globale avec tuples, nous considérons d'abord les différentes manières de modifier les parties attaque ou défense d'un argument :

**Définition 39 (Ajout/suppression d'une branche d'un argument)**
*Soit $A$ un argument dont la valeur tuplée est $v(A) = [v_p(A), v_i(A)]$ avec $v_p(A) = (x_1^p, \ldots, x_n^p)$ et $v_i(A) = (x_1^i, \ldots, x_m^i)$ sachant que $v_p(A)$ ou $v_i(A)$ peuvent être vides mais pas en même temps.*

*L'ajout (resp. la suppression) d'une branche de défense de $A$ est définie par :*

*$v_p(A)$ devient $\texttt{Tri}(x_1^p, \ldots, x_n^p, x_{n+1}^p)$ où $x_{n+1}^p$ est la longueur de la branche ajoutée (resp. $\exists j \in [1..n]$ tel que $v_p(A)$ devient $(x_1^p, \ldots, x_{j-1}^p, x_{j+1}^p, \ldots, x_n^p)$).*

*Et la même chose sur $v_i(A)$ pour l'ajout (resp. la suppression) d'une branche d'attaque de $A$.*

[12] Nous utiliserons aussi la notation $B \preceq A$ définie par : $B \preceq A$ ssi $A \succeq B$.

**Algorithme 1:** Comparaison de deux valeurs tuplées

```
% Description des paramètres :                                                    %
% v, w : 2 valeurs tuplées                                                       %
% Notations :                                                                    %
%     |v_p| (resp. |w_p|) : nombre d'éléments dans le composant pair de v (resp. w)   %
%         si v_p (resp. w_p) est infini alors |v_p| (resp. |w_p|) est considéré égal à ∞  %
%     |v_i| (resp. |w_i|) : nombre d'éléments dans le composant impair de v (resp. w)  %
%         si v_i (resp. w_i) est infini alors |v_i| (resp. |w_i|) est considéré égal à ∞  %
%     ≻ dénotera la relation stricte associée à ⪰ et sera définie par : v ≻ w ssi v ⪰ w et non(w ⪰ v).  %
début
1   si v = w alors v ⪰ w ET w ⪰ v                                          % Cas 1 %
2   sinon
3       si |v_i| = |w_i| ET |v_p| = |w_p| alors
                                  % comparaisons lexicographiques entre v_p et w_p et entre v_i et w_i %
4           si v_p ≤_lex∞ w_p ET v_i ≥_lex∞ w_i alors v ≻ w                  % cas 2 %
5           sinon
6               si v_p ≥_lex∞ w_p ET v_i ≤_lex∞ w_i alors v ≺ w              % cas 3 %
7               sinon v ⋡ w ET v ⋠ w          % valeurs tuplées incomparables. cas 4 %
8       sinon
9           si |v_i| ≥ |w_i| ET |v_p| ≤ |w_p| alors v ≺ w                   % cas 5 %
10          sinon
11              si |v_i| ≤ |w_i| ET |v_p| ≥ |w_p| alors v ≻ w               % cas 6 %
12              sinon v ⋡ w ET v ⋠ w          % valeurs tuplées incomparables. cas 7 %
fin
```

**Définition 40 (Augmentation/diminution de la longueur d'une branche d'un argument)**
*Soit $A$ un argument dont la valeur tuplée est $v(A) = [v_p(A), v_i(A)]$ avec $v_p(A) = (x_1^p, \ldots, x_n^p)$ et $v_i(A) = (x_1^i, \ldots, x_m^i)$ sachant que $v_p(A)$ ou $v_i(A)$ peuvent être vides mais pas en même temps.*

*L'augmentation (resp. la diminution) de la longueur d'une branche de défense de $A$ est définie par :*

*$\exists j \in [1..n]$ tel que $v_p(A)$ devient $(x_1^p, \ldots, x_{j-1}^p, x_j'^p, x_{j+1}^p, \ldots, x_n^p)$ où $x_j'^p > x_j^p$ (resp. $x_j'^p < x_j^p$) et la parité de $x_j'^p$ est égale à la parité de $x_j^p$.*

*Et la même chose sur $v_i(A)$ pour l'augmentation (resp. la diminution) d'une branche d'attaque de $A$.*

**Définition 41 (Amélioration/dégradation des défenses/attaques)**
*Soit $A$ un argument dont la valeur tuplée est $v(A) = [v_p(A), v_i(A)]$ sachant que $v_p(A)$ ou $v_i(A)$ peuvent être vides mais pas en même temps. On définit :*

**Une amélioration (resp. dégradation) de la défense** *consiste à*
- *ajouter une branche de défense à $A$ si initialement $v_p(A) \neq 0^\infty$ (resp. supprimer une branche de défense de $A$) ;*
- *ou diminuer (resp. augmenter) la longueur d'une branche de défense de $A$ ;*
- *ou supprimer la seule branche de défense menant à $A$ (resp. ajouter une branche de défense menant $A$ si initialement $v_p(A) = 0^\infty$) ;*

**Une amélioration (resp. dégradation) de l'attaque** *consiste à*
- *ajouter (resp. supprimer) une branche d'attaque de $A$ ;*
- *ou diminuer (resp. augmenter) la longueur d'une branche d'attaque de $A$.*

**Proposition 17 (Principes sous-jacents)** *Soit $v$ une évaluation avec tuples correspondant à la définition 35 page 44) et associée à l'algorithme 1, $v$ respecte les principes suivants :*

**P1′** *L'évaluation est maximale pour un argument sans attaquant et non maximale pour un argument attaqué (qu'il soit défendu ou pas).*

**P2′** *L'évaluation d'un argument prend en compte toutes les branches menant à cet argument.*

**P3′** *L'amélioration de la défense ou la dégradation de l'attaque d'un argument augmente la valeur de cet argument.*

**P4′** *L'amélioration de l'attaque ou la dégradation de la défense d'un argument diminue la valeur de cet argument.*

**Exemple 8 page 42 (suite)** *Avec l'évaluation avec tuples, on obtient :*

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

*Donc, on a :*

$$\begin{array}{c} E_1, D_2, D_3, C_4, B_4 \\ \succ \\ C_1, B_2 \\ \succ \\ B_1 \\ \succ \\ D_1, C_2, C_3, B_3 \end{array} \qquad \textit{mais aussi} \qquad \begin{array}{c} E_1, D_2, D_3, C_4, B_4 \\ \succ \\ A \end{array}$$

*$A$ est incomparable avec presque tous les autres arguments (excepté avec les feuilles du graphe).*

*De la même manière, si on se restreint à la partie hachurée du graphe, on obtient :*

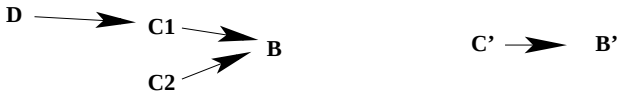
$$E_1, D_2 \succ C_1 \succ B_1 \succ A \succ D_1, C_2$$

*$A$ est maintenant comparable avec tous les autres arguments (en particulier, $A$ est "pire" que son défenseur $C_1$ et que son attaquant direct $B_1$).*

#### 3.2.2.3 Principales différences entre les évaluations "locale" et "globale"

Dans [CLS03c], on trouvera une comparaison de ces approches avec quelques approches existantes ([Dun95, JV99b, BH01]), ainsi qu'une comparaison entre approches "locale" et "globale" (remarquons que l'évolution de l'approche globale proposée dans [CLS05b] ne remet pas en cause ces résultats).

Rappelons juste ici le point essentiel qui les distingue l'une de l'autre (l'exemple utilisé ici a déjà été présenté au début de cette section) :




Dans l'approche locale, $B'$ est meilleur que $B$ (puisque $B'$ ne subit qu'une attaque alors que $B$ en subit deux).

Dans l'approche globale, $B$ est meilleur que $B'$ (puisqu'il possède au moins une défense alors que $B'$ n'en a aucune). Dans ce cas, $C_1$ perd son statut négatif d'attaquant, puisqu'il est "porteur d'une défense" pour $B$.

Le tableau suivant synthétise les résultats concernant les différentes évaluations proposées :

*approche globale*

| | | | | | | |
|---|---|---|---|---|---|---|
| arguments ayant seulement des branches d'attaque | $\preceq$ | arguments ayant des branches d'attaque et de défense | $\preceq$ | arguments ayant seulement des branches de défense | $\preceq$ | arguments non attaqués |

*approche locale*

| arguments ayant plusieurs attaquants directs non attaqués | $\succeq$ | arguments ayant un seul attaquant direct non attaqué | $\succeq$ | arguments ayant un seul attaquant direct attaqué (éventuellement défendu) | $\succeq$ | arguments non attaqués |
|---|---|---|---|---|---|---|
| arguments ayant plusieurs attaquants directs attaqués (éventuellement défendus) | | | | | | |

La différence entre approches locale et globale est aussi illustrée par la propriété suivante :

**Proposition 18 (Indépendance des branches dans l'approche globale)**
*Soit $A$ un argument ayant les attaquants directs suivants :*
– $A_1$ *dont la valeur est* $v(A_1) = [(a^1_{p_1}, \ldots, a^1_{p_{m_1}}), (a^1_{i_1}, \ldots, a^1_{i_{m_1}})]$,
– …,
– $A_n$ *dont la valeur est* $v(A_n) = [(a^n_{p_1}, \ldots, a^n_{p_{m_n}}), (a^n_{i_1}, \ldots, a^n_{i_{m_n}})]$.
*Soit $A'$ un argument ayant les attaquants directs suivants :*
– $A^1_{p_1}$ *dont la valeur est* $v(A^1_{p_1}) = [(a^1_{p_1}), ()]$,
– …,
– $A^1_{p_{m_1}}$ *dont la valeur est* $v(A^1_{p_{m_1}}) = [(a^1_{p_{m_1}}), ()]$,
– $A^1_{i_1}$ *dont la valeur est* $v(A^1_{i_1}) = [(), (a^1_{i_1})]$,
– …,
– $A^1_{i_{m_1}}$ *dont la valeur est* $v(A^1_{i_{m_1}}) = [(), (a^1_{i_{m_1}})]$,
– …,
– $A^n_{p_1}$ *dont la valeur est* $v(A^n_{p_1}) = [(a^n_{p_1}), ()]$,
– …,
– $A^n_{p_{m_n}}$ *dont la valeur est* $v(A^n_{p_{m_n}}) = [(a^n_{p_{m_n}}), ()]$,
– $A^n_{i_1}$ *dont la valeur est* $v(A^n_{i_1}) = [(), (a^n_{i_1})]$,
– …,
– $A^n_{i_{m_n}}$ *dont la valeur est* $v(A^n_{i_{m_n}}) = [(), (a^n_{i_{m_n}})]$.
*Alors* $v(A) = v(A')$.

Cette propriété montre l'"indépendance" des branches lors du calcul des valeurs dans l'approche globale, même si ces branches ne sont pas graphiquement indépendantes. Dans l'exemple suivant, $A$ et $A'$ ont la même valeur $[(2, 2), ()]$ bien qu'ils soient la racine de deux graphes différents :

Cette propriété n'est pas satisfaite par l'approche locale (voir les principes sous-jacents à ce type d'approche, Propriété 11 page 41), la valeur de l'argument $A$ devant être au moins aussi bonne que (et quelquefois meilleure que[13]) celle de l'argument $A'$ ($A$ ayant un attaquant direct alors que $A'$ en a deux).

#### 3.2.2.4 Conclusion sur l'évaluation d'arguments

Dans le domaine de l'argumentation unipolaire, une partie essentielle de mon travail a consisté à introduire de la gradualité dans deux des étapes du processus d'argumentation :

– l'évaluation des arguments,
– l'acceptabilité des arguments.

[13] Avec l'évaluation proposée par [BH01], on aurait : $v(A) = \frac{3}{4}$ et $v(A') = \frac{1}{2}$.

La section courante concernait le premier point. Mon travail a essentiellement consisté à définir et à comparer deux types d'évaluations basées sur les interactions (donc sur les attaques puisque c'est dans le cadre d'un système unipolaire).
– Le premier type est une évaluation locale générique recouvrant des propositions déjà existantes (par exemple [BH01] et [JV99b]). Cette approche est dite "locale" puisqu'elle permet le calcul d'une valeur pour un argument uniquement en fonction des valeurs des attaquants directs de cet argument.
– Le second type est basé sur un étiquetage prenant la forme de paires de tuples ; cet étiquetage mémorise la structure du sous-graphe des interactions menant à l'argument, en associant à chaque branche sa longueur (nombre des arcs depuis les feuilles jusqu'à l'argument ; si la longueur est paire alors il s'agit d'une branche de défense et si elle est impaire alors il s'agit d'une branche d'attaque). Cette approche est dite "globale" puisque la valeur d'un argument se calcule en prenant en compte la totalité du sous-graphe menant à cet argument.

Chacun de ces types d'évaluation induit un préordre sur l'ensemble des arguments.

Ces évaluations pourront alors être utilisées afin de faire de la sélection d'arguments (voir la section 3.2.3).

### 3.2.3 Acceptabilité

Dans le contexte de l'argumentation unipolaire et de l'acceptabilité d'arguments, mon travail a consisté à décliner à nouveau la notion de gradualité afin d'obtenir une acceptabilité graduelle.

L'idée de base est de sélectionner un argument en fonction de la non-sélection de ses attaquants directs. Pour cela, on peut utiliser deux méthodes différentes :
– La première méthode consiste à raffiner la partition issue de l'acceptabilité collective proposée par Dung ; Ce raffinage peut se faire en exploitant les évaluations graduelles définies dans la section 3.2.2 page 38.
– La seconde méthode se place plutôt dans le cadre de l'acceptabilité individuelle et consiste à définir une nouvelle acceptabilité en utilisant uniquement les évaluations graduelles de la section 3.2.2 page 38.

Ce travail a été réalisé avec Claudette CAYROL et publié entre autres dans [CLS05b].

#### 3.2.3.1 Différents niveaux d'acceptabilité collective

Sous une sémantique donnée, et en suivant les idées de Dung, l'acceptabilité d'un argument dépend de son appartenance à une extension de cette sémantique. Trois cas sont possibles[14] :

– l'argument peut être *uni-accepté*, quand il appartient à toutes les extensions de cette sémantique,
– ou l'argument peut être *exi-accepté*, quand il appartient au moins à une extension de cette sémantique,
– ou l'argument peut être *non-accepté* quand il n'appartient à aucune extension de cette sémantique.

Il me semble toutefois que ces trois niveaux sont insuffisants pour pouvoir parler vraiment de gradualité. Par exemple, que conclure dans le cas de deux arguments $A$ et $B$ exi-acceptés et tels que $A\mathcal{R}B$ ou $B\mathcal{R}A$ ?

Ainsi, nous proposons une nouvelle définition qui tient compte de la situation d'un argument par rapport à ses attaquants directs pour raffiner la classe des arguments exi-acceptés pour une sémantique $S$ donnée.

**Définition 42 (Argument proprement-accepté)** *Soit $A \in \mathcal{A}$, $A$ est* proprement-accepté *ssi $A$ appartient au moins à une extension de $S$ et $\forall B \in \mathcal{A}$ tel que $B\mathcal{R}A$, $B$ n'appartient à aucune extension de $S$.*

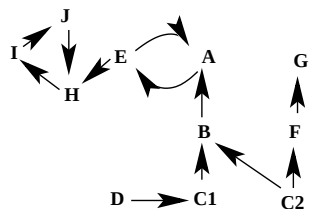
Ainsi, on capture l'idée qu'un argument est mieux accepté si ses attaquants ne le sont pas.

**Proposition 19** *Soit $A \in \mathcal{A}$ et une sémantique $S$ telle que chaque extension de $S$ est sans-conflit. Si $A$ est uni-accepté alors $A$ est proprement-accepté. La réciproque est fausse.*

[14] La terminologie utilisée dans cette section est la même que celle utilisée dans le domaine du raisonnement non monotone (voir [PL92] et le chapitre 1 page 9 de ce document) : le mot *uni* provient du mot *universel* lui-même "synonyme" du mot *sceptique*, et le mot *exi* provient du mot *existentiel*, "synonyme" du mot *crédule*. J'ai choisi d'utiliser les mots *uni* et *exi* parce qu'ils font allusion aux quantificateurs logiques $\forall$ (*quel que soit*) et $\exists$ (*il existe au moins un*).

La notion d'argument proprement-accepté raffine la classe des arguments exi-acceptés. Pour une sémantique $S$ et un argument $A$, il existe alors 4 états possibles pour $A$ :
– $A$ peut être *uni-accepté*, si $A$ appartient à toutes les extensions pour $S$ (donc, il est aussi proprement-accepté) ;
– ou $A$ peut être *proprement-accepté* (donc, il est aussi exi-accepté) ;
– ou $A$ peut être *seulement-exi-accepté*, si $A$ n'est pas proprement-accepté, mais est quand même exi-accepté ;
– ou $A$ est *non-accepté* si $A$ n'est même pas exi-accepté.

**Exemple 13** *Considérons le système d'argumentation suivant.*




*Il existe deux extensions préférées $\{D, C_2, A, G\}$ et $\{D, C_2, E, G, I\}$. Donc, pour la sémantique préférée, les niveaux d'acceptabilité sont les suivants :*
– *$D$, $C_2$ et $G$ sont uni-acceptés,*
– *$I$ est proprement-accepté mais pas uni-accepté,*
– *$A$ et $E$ sont seulement-exi-acceptés,*
– *$B$, $C_1$, $F$, $H$ et $J$ sont non-acceptés.*

Remarquons que dans les cas où il n'existe qu'une seule extension, les trois premiers niveaux d'acceptabilité coïncident. C'est le cas :
– Sous la sémantique préférée, quand il n'y a pas de circuit pair (voir [Dou02]).
– Sous la sémantique basique (qui a la particularité de n'avoir qu'une seule extension).
On a aussi la propriété suivante :

**Proposition 20** *Sous la sémantique stable, la classe des arguments uni-acceptés coïncide avec celle des arguments proprement-acceptés.*

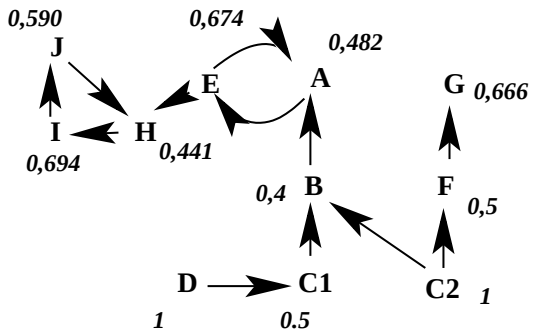
Et si on exploite en plus un résultat donné dans [DBC01, DBC02], et réutilisé par [Dou02], qui montre qu'en l'absence de circuits impairs les extensions préférées sont aussi stables, nous pouvons alors appliquer la propriété 20 et nous obtenons ainsi le résultat suivant :

**Consequence 1** *Sous la sémantique préférée, quand il n'existe pas de circuit impair, la classe des arguments uni-acceptés coïncide avec celle des arguments proprement-acceptés.*

Une autre possibilité pour introduire de la gradualité consiste à utiliser les évaluations graduelles basées sur les interactions définies en section 3.2.2 page 38.

Soit $v$ une évaluation graduelle et $\succeq$ le préordre associé (partiel ou complet) sur $\mathcal{A}$. Ce préordre peut être utilisé à l'intérieur de chaque niveau d'acceptabilité (par exemple, à l'intérieur du niveau des arguments exi-acceptés) pour identifier les arguments qui seraient plus intéressants à accepter que d'autres.

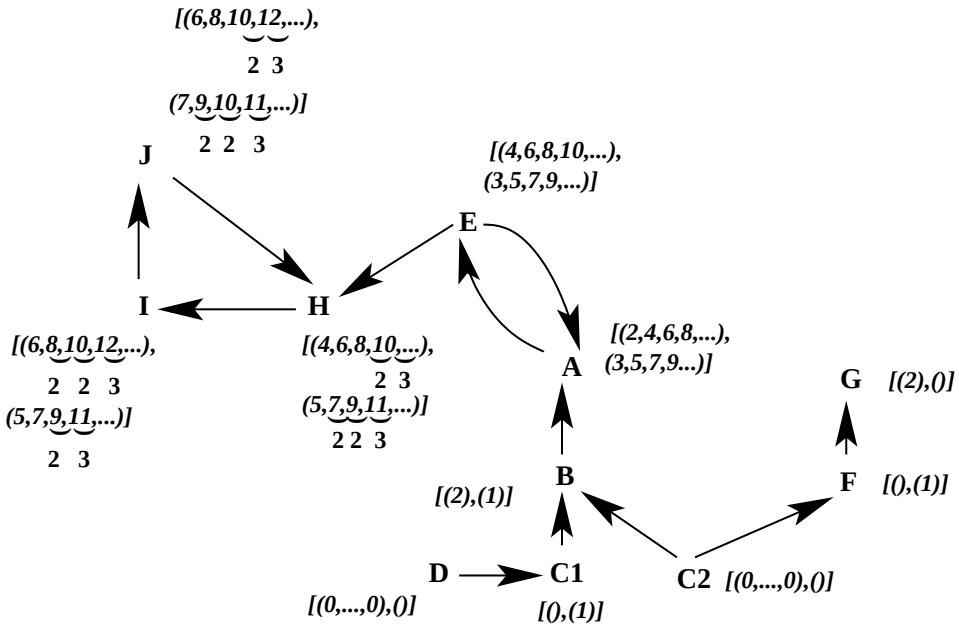
**Exemple 13 (suite)** *Deux évaluations graduelles différentes sont appliquées sur le même graphe :*




**Évaluation selon (Besnard & Hunter, 2001)**

*Avec l'instance de l'évaluation générique proposée par [BH01] (voir section 3.2.2.1 page 39), nous obtenons les résultats de comparaison suivants :*

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




**Évaluation par tuples**

*Avec l'évaluation globale avec tuples présentée en section 3.2.2.2 page 43, nous obtenons les résultats de comparaison suivants*[15] *:*

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

*Donc, tous les arguments appartenant à un circuit sont incomparables avec $G$, $B$, $F$, $C_1$ et, même entre eux, il y a peu de comparaisons possibles.*

Si nous appliquons le préordre induit par une évaluation sans respecter les niveaux d'acceptabilité, nous obtenons alors des résultats contre-intuitifs. Dans l'exemple 13 page ci-contre, on obtient :

– Avec l'évaluation de [BH01] et sous la sémantique préférée, $E \succ G$ en dépit du fait que $G$ soit uni-accepté et que $E$ ne soit que seulement-exi-accepté.
– Avec l'évaluation avec tuples et sous la sémantique préférée, $H \succ E$ en dépit du fait que $E$ soit seulement-exi-accepté et que $H$ soit non-accepté.

Ces situations contre-intuitives illustrent la différence importante entre la définition d'une acceptabilité collective et la définition d'une évaluation qui repose sur des bases individuelles, alors même que toutes les deux exploitent les interactions entre les arguments.

[15] Afin de simplifier l'écriture, on ne répète pas les valeurs à l'intérieur des tuples (on indique juste combien de fois cette valeur apparaît). Par exemple : $[(2,4,4,6,6,6,8,8,8,8\ldots),(3,5,5,7,7,7,9,9,9,9\ldots)]$ sera notée $[(2,\underbrace{4}_{2},\underbrace{6}_{3},\underbrace{8}_{4},\ldots),(3,\underbrace{5}_{2},\underbrace{7}_{3},\underbrace{9}_{4},\ldots)]$

#### 3.2.3.2 Vers une acceptabilité individuelle graduelle

L'acceptabilité individuelle est basée sur la comparaison d'un argument avec ses attaquants.

La toute première proposition faite en ce sens a été de sélectionner seulement les arguments qui n'ont pas d'attaquants (voir [EGFK93b]).

Plus tard, ce travail a été étendu par [AC98] : en utilisant une relation de préférence entre arguments (correspondant à une évaluation intrinsèque), un argument a été considéré accepté ssi il était préféré à chacun de ses attaquants.

Suivant cette idée, nous avons proposé un mécanisme semblable mais qui exploite une *évaluation basée sur les interactions*.

Soit $v$ une évaluation graduelle, le préordre induit par $v$ peut être directement utilisé afin de comparer, du point de vue de l'acceptabilité, un argument et ses attaquants[16]. Cela permet de définir une nouvelle classe d'arguments : les arguments bien-défendus.

**Définition 43 (Argument bien-défendu)** *Soit $A \in \mathcal{A}$, $A$ est* bien-défendu (pour $v$) *ssi $\forall B \in \mathcal{A}$ tel que $B\mathcal{R}A$, $B \not> A$.*

Ainsi, on capture l'idée qu'un argument sera mieux accepté s'il est au moins aussi bon que ses attaquants directs (ou incomparable avec eux dans le cas d'un préordre partiel). L'ensemble des arguments bien-défendus dépendra bien-sûr de la fonction d'évaluation choisie.

En utilisant cette notion, l'ensemble de tous les arguments peut être partitionné de la manière suivante :

– La première classe contient les arguments non attaqués,
– la seconde classe contient les arguments attaqués mais bien-défendus,
– la troisième classe contient les autres arguments (attaqués et pas bien-défendus).

Remarquons que l'ensemble des arguments bien-défendus correspond à l'union des deux premières classes. Un raffinage supplémentaire peut être obtenu en appliquant l'évaluation à l'intérieur de chaque classe comme dans la section 3.2.3.1 page 55.

Dans l'exemple 13 page 56 présenté en section 3.2.3.1 page 55, les arguments bien-défendus sont :

– $D$, $C_2$, $G$, $H$ et $A$ ($A$ est incomparable avec $B$ mais meilleur que $E$) pour l'évaluation avec tuples,
– et avec l'évaluation de [BH01] les arguments bien-défendus sont $D$, $C_2$, $G$, $I$ et $E$ ($E$ est meilleur que $A$).

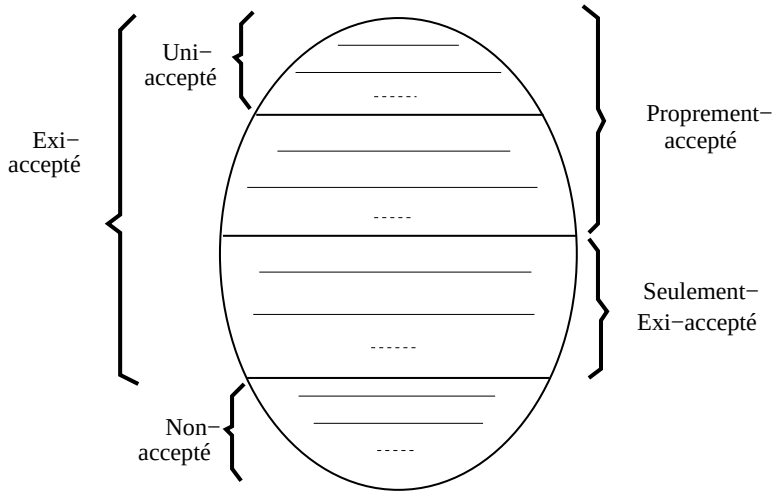
Remarquons aussi que, comme dans les sémantiques proposées dans [Dun95], la définition 43 considère les attaquants un par un. Cela n'est bien-sûr pas applicable pour une évaluation qui gère l'"attaque directe" comme un tout (comme l'évaluation de [BH01] – voir les contre-exemples présentés en section 3.2.3.3).

#### 3.2.3.3 Compatibilité entre acceptabilité et évaluation graduelle

En suivant ce qui a été dit dans les sections précédentes, nous avons donc à notre disposition deux moyens pour partitionner l'ensemble des arguments :

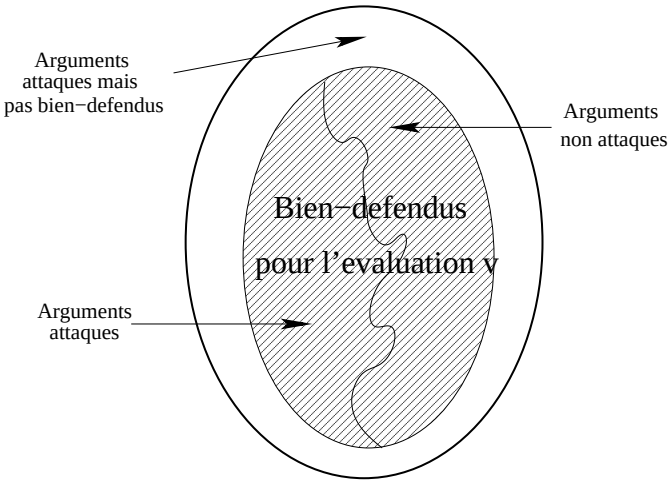
– Soit $S$ une sémantique et $v$ une évaluation graduelle, on peut utiliser la partition issue de [Dun95] et la raffiner :

[16] Cette idée est aussi utilisée dans la notion de "defeat" proposée par [BC02]. Remarquons que, dans le travail de [BC02], l'évaluation est une information supplémentaire ajoutée au système d'argumentation. Par opposition, dans notre travail, la $v$-préférence est extraite du graphe d'attaque.




*Raffinement de chaque niveau avec l'évaluation graduelle v*

– Soit $v$ une évaluation graduelle, on peut utiliser la notion d'argument bien-défendu pour partitionner l'ensemble des arguments :




Une question très naturelle se pose alors : est-il possible de trouver une sémantique $S$ et une évaluation graduelle $v$ telles que les partitions associées soient compatibles ?

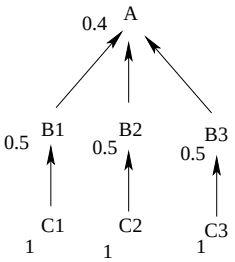
Les exemples suivants montrent que la classe des arguments bien-défendus ne correspond pas à la classe des arguments proprement-acceptés (dans certains cas, des arguments uni-acceptés ne sont même pas bien-défendus).

#### 3.2.3.3.1 Exemples montrant la non-compatibilité dans le cas général

Nous allons donner des exemples pour chaque évaluation usuelle (l'évaluation globale avec tuples et 2 instances de l'évaluation locale générique : [BH01, JV99b]) et pour les sémantiques les plus classiques pour l'acceptabilité (les sémantiques préférée et stable de [Dun95]).

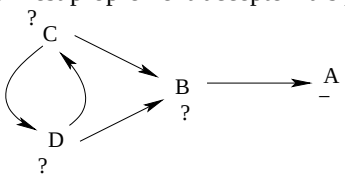
*Argument proprement-accepté mais pas bien-défendu.* Sont donnés 3 exemples (chacun pour une évaluation différente) :

– l'argument $A$ est proprement-accepté mais pas bien-défendu :



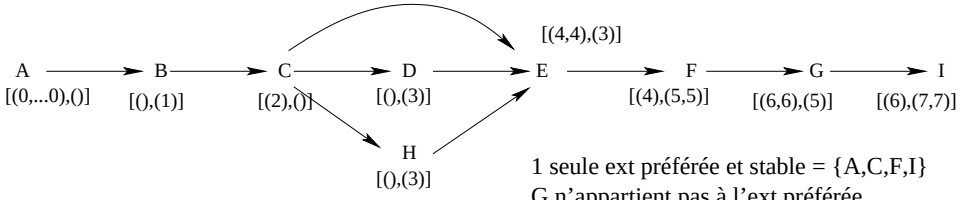

1 seule ext préférée et stable = { C1, C2, C3, A}
B1, B2, B3 n'appartiennent à aucune ext préférée
Bi $\succ$ A quel que soit i = 1, 2, 3

– l'argument $A$ est proprement-accepté mais pas bien-défendu :




Ici, 2 ext préférées et stables : {C,A} et {D,A}
B n'appartient à aucune ext préférée
B $\succ$ A

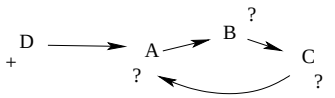
– l'argument $I$ est proprement-accepté mais pas bien-défendu :




1 seule ext préférée et stable = {A,C,F,I}
G n'appartient pas à l'ext préférée
G $\succ$ I

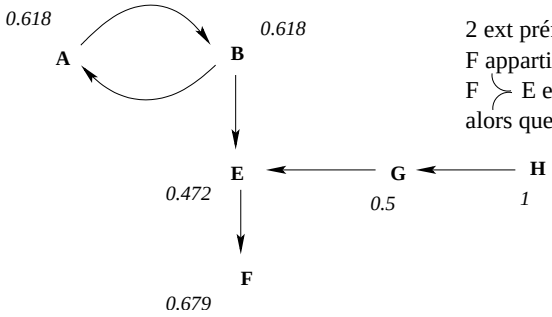
*Argument bien-défendu mais pas proprement-accepté.* Pour les mêmes trois évaluations, on obtient :

– l'argument $C$ est bien-défendu mais il n'est pas proprement-accepté :



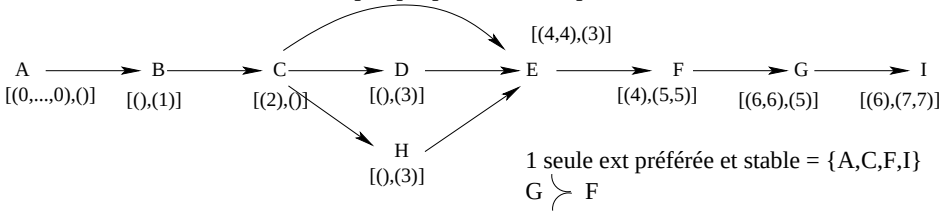

1 seule ext préférée et stable : {D, B}
C $\succ$ B
C n'appartient pas à l'ext préférée et stable

– l'argument $F$ est bien-défendu mais il n'est pas proprement-accepté :




2 ext préférées et stables = {A,H,E} et {B,H,F}
F appartient à 1 ext préférée et stable
F $\succ$ E et E appartient aussi à 1 ext préférée et stable alors que E attaque F

– l'argument $G$ est bien-défendu mais il n'est pas proprement-accepté :




1 seule ext préférée et stable = {A,C,F,I}
G $\succ$ F
G n'appartient pas à l'ext préférée

#### 3.2.3.3.2 Cas particuliers menant à la compatibilité

Dans le contexte d'un système d'argumentation avec une relation $\mathcal{R}$ finie et sans circuit[17], les sémantiques stable et préférée ne produisent qu'une seule extension et les 3 premiers niveaux d'acceptabilité (uni-acceptés, exi-acceptés, proprement-acceptés) coïncident.

Dans ce cas-là, il existe au moins deux cas menant à la compatibilité.

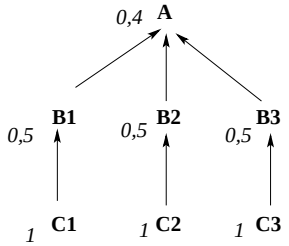
*Premier cas :* Il concerne l'évaluation globale avec tuples.

**Théorème 1** *Soit $\mathcal{G}$ le graphe associé à $\langle \mathcal{A}, \mathcal{R} \rangle$, $\langle \mathcal{A}, \mathcal{R} \rangle$ étant un système d'argumentation avec une relation $\mathcal{R}$ finie, acyclique et satisfaisant la condition suivante : $\exists\ A \in \mathcal{A}$ tel que :*
- *$\forall X_i$, feuille de $\mathcal{G}$, $\exists$ seulement un chemin de $X_i$ vers $A$, $X_i^1 - \ldots - X_i^{l_i} - A$ avec $X_i^1 = X_i$ et $l_i$ la longueur de ce chemin (si $l_i$ est paire, ce chemin est une branche de défense pour $A$, sinon c'est une branche d'attaque),*
- *tous les chemins de $X_i$ vers $A$ sont racine-dépendants en $A$,*
- *$\forall A_i \in \mathcal{A}$, $\exists X_j$ une feuille de $\mathcal{G}$ telle que $A_i$ appartient à un chemin de $X_j$ vers $A$.*

*Soit $v$ une évaluation globale avec tuples. Soit $S$ une sémantique $\in$ {préférée, stable}.*

1. *$\forall B \in \mathcal{A}$, $B \neq A$, $B$ (exi, uni, proprement)-accepté pour $S$ ssi $B$ bien-défendu pour $v$.*
2. *Si $A$ est (exi, uni, proprement)-accepté pour $S$ alors $A$ est bien-défendu pour $v$ (la réciproque est fausse).*
3. *Si $A$ est bien-défendu pour $v$ et si toutes les branches conduisant à $A$ sont des branches de défense pour $A$ alors $A$ est (exi, uni, proprement)-accepté pour $S$.*

Remarquons que le théorème 1 est en général faux pour une évaluation locale. Voir le contre-exemple suivant pour l'évaluation de [BH01] :




Le graphe satisfait les conditions du théorème 1. L'ensemble des arguments bien-défendus est $\{C_1, C_2, C_3\}$ (donc, $A$ n'est pas bien-défendu). Néanmoins, $\{C_1, C_2, C_3, A\}$ est l'extension préférée.

*Second cas :* Il concerne l'évaluation locale générique :

**Théorème 2** *Soit $\langle \mathcal{A}, \mathcal{R} \rangle$ un système d'argumentation avec une relation $\mathcal{R}$ finie et acyclique. Soit $S$ une sémantique $\in$ {préférée, stable}. Soit $v$ une évaluation locale générique satisfaisant la condition suivante $(*)$ :*

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

*$\forall A \in \mathcal{A}$, $A$ (exi, uni, proprement)-accepté pour $S$ ssi $A$ est bien-défendu pour $v$.*

Ce théorème est une conséquence directe du lemme suivant :

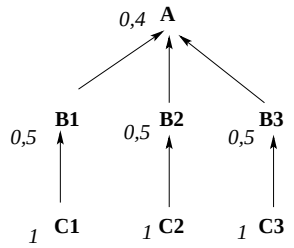
**Lemme 1** *Soit $\langle \mathcal{A}, \mathcal{R} \rangle$ un système d'argumentation avec une relation $\mathcal{R}$ finie et acyclique. Soit $S$ une sémantique $\in$ {préférée, stable}. Soit $v$ une évaluation locale générique satisfaisant la condition $(*)$.*

*$(i)$ Si $A$ est exi-accepté et possède un seul attaquant direct $B$ alors $A \succeq B$.*

*$(ii)$ Si $B$ est non-accepté et possède un seul attaquant direct $C$ alors $C \succeq B$.*

[17] Donc, $\langle \mathcal{A}, \mathcal{R} \rangle$ est bien-fondé.

Remarquons que la condition $(*)$ utilisée dans le théorème 2 page précédente est :
– fausse pour l'évaluation locale proposée par [BH01] :




$g(x) = \frac{1}{1+x}$ et $h(x_1, \ldots, x_n) = \Sigma_{i=1}^{n} x_i$ (voir la propriété 13 page 42). On obtient :
– $\forall i = 1 \ldots 3$, $x_i = v(B_i) = 0.5$,
– $\forall i = 1 \ldots 3$, $g(x_i) = 0.66$, donc $g(x_i) \geq x_i$,
– et néanmoins $g(h(x_1, x_2, x_3)) = v(A) = 0.4 \ngeq h(x_1, x_2, x_3) = 1.5$.
– fausse pour les évaluations locales définies en utilisant une fonction $h$ telle que $\exists n > 1$ avec $h(x_1, \ldots, x_n) > \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)$ (pour toutes les fonctions $g$ strictement décroissantes) : voir le graphe précédent où $h(x_1, x_2, x_3) = 1.5$ et $\mathcal{M}\mathtt{ax}(x_1, x_2, x_3) = 0.5$.
– vraie pour les évaluations locales définies en utilisant une fonction $h = \mathcal{M}\mathtt{ax}$ (pour toutes les fonctions $g$) : si $h = \mathcal{M}\mathtt{ax}$ alors $g(h(x_1, \ldots, x_n)) = g(\mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)) = g(x_j)$, $x_j$ étant le maximum des $x_i$ ; et, par hypothèse, $g(x_i) \geq x_i$, $\forall x_i$, donc en particulier pour $x_j$ ; ainsi, on a :

$$g(h(x_1, \ldots, x_n)) = g(x_j) \geq x_j = \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n) = h(x_1, \ldots, x_n).$$

#### 3.2.3.4 Conclusion sur l'acceptabilité

Dans le domaine de l'argumentation unipolaire, une partie essentielle de mon travail a consisté à introduire de la gradualité dans deux des étapes du processus d'argumentation :

– l'évaluation des arguments,
– l'acceptabilité des arguments.

En ce qui concerne le premier point, j'ai présenté dans la section 3.2.2 page 38, deux cas possibles d'évaluation.

Pour le second point, dans la section courante, deux approches différentes ont aussi été étudiées :
– Tout d'abord, dans le contexte de l'acceptabilité collective de [Dun95] : les trois niveaux d'acceptabilité de base (uni-acceptés, exi-acceptés, non-acceptés) ont été raffinés en utilisant la notion d'arguments *proprement-acceptés* (ceux dont les attaquants directs sont non-acceptés).
– Puis, dans le contexte de l'acceptabilité individuelle : en utilisant les évaluations graduelles introduites précédemment, j'ai défini la nouvelle notion d'arguments *bien-défendus* (ceux qui sont préférés à leurs attaquants directs d'après une évaluation graduelle $v$ donnée).

Le premier concept permet de raffiner la classe des exi-acceptés en deux sous-classes (les arguments proprement-acceptés et les arguments seulement-exi-acceptés). Les évaluations graduelles peuvent alors être utilisées en interne pour chaque niveau de cette acceptabilité collective.

Le second concept introduit deux nouveaux niveaux d'acceptabilité (les arguments bien-défendus et ceux qui ne le sont pas). Et là-aussi, les évaluations graduelles permettent d'introduire de la gradualité à l'intérieur de chaque niveau de cette acceptabilité individuelle.

Dans tous les cas, nous avons appliqué deux principes de base :
– l'acceptabilité est fortement liée aux interactions entre arguments,
– et un argument est toujours plus acceptable s'il est préféré à ses attaquants directs.

### 3.2.4 Fusion en argumentation unipolaire

Dans un cadre multi-agents, l'argumentation peut servir à représenter le processus d'échange d'informations, comme par exemple en négociation ou en persuasion (voir [Mac79, WK95, Gor95, PJ96, AMP00, AP02, AP04]). Cela peut

se faire sous la forme d'un jeu à deux joueurs, chaque joueur donnant un argument attaquant un argument de l'autre joueur.

Dans cette section, je considère l'argumentation dans un cadre multi-agents mais dans une perspective différente : ici, mon but n'est pas de caractériser l'échange d'informations entre agents, mais de caractériser l'ensemble des arguments acceptables par un groupe d'agents quand les données fournies par chaque agent constituent un système d'argumentation abstrait au sens de Dung. Il n'y a donc pas d'échange entre agents : chaque agent donne ses informations sous la forme d'un système d'argumentation puis on cherche à en extraire une (ou des) conclusion(s) sous la forme d'un ensemble d'arguments acceptables pour le groupe d'agents.

Dans ce cas-là, on pourrait choisir une méthode très simple consistant à faire "voter" les agents : un ensemble d'arguments serait considéré comme acceptable pour le groupe si et seulement si il est acceptable pour "suffisamment" d'agents dans le groupe (le terme "suffisamment" faisant référence aux diverses types de vote possibles). Cette méthode ne nécessite aucune fusion d'information. Toutefois, certains exemples montrent qu'elle peut conduire à des résultats contre-intuitifs.

J'ai donc proposé, en collaboration avec des collègues du CRIL de Lens, une approche plus sophistiquée sous la forme d'un processus en trois étapes : premièrement, chaque système d'argumentation est étendu en un système d'argumentation partiel prenant en compte l'ensemble complet des arguments (les siens et ceux des autres agents) ainsi que les interactions entre arguments (les siennes et celles proposées par les autres agents qui ne sont pas en contradiction avec les siennes) ; puis on effectue une fusion de tous ces systèmes d'argumentation partiels afin de résoudre les conflits pouvant exister et ainsi de trouver le (ou les) système(s) d'argumentation le(s) plus proche(s) de chaque système d'argumentation partiel ; et enfin, on calcule les arguments acceptables pour l'ensemble des systèmes d'argumentation fourni par la fusion.

Pour cela, j'introduis d'abord la notion de *système d'argumentation partiel (PAF)*, qui étend la notion de système d'argumentation proposée par Dung puisqu'elle prend en compte un nouveau type d'interaction, *l'ignorance concernant la relation d'attaque*. Cela permet d'exprimer le fait que tous les agents ne sont pas supposés partager le même ensemble d'arguments. Cela va donc permettre d'étendre le système d'argumentation de chaque agent en lui rajoutant tous les arguments qui lui sont inconnus mais qui sont connus par au moins un autre agent. Bien-sûr, cet ajout d'argument peut se faire de diverses manières, chaque agent pouvant avoir sa propre méthode d'expansion. J'en présenterai quelques-unes mais je me focaliserai ensuite sur une méthode particulière, *l'expansion consensuelle* : quand on incorpore un nouvel argument dans un système d'argumentation, on ajoute une interaction le concernant seulement si tous les agents connaissant cet argument ET celui qui lui est lié par l'interaction sont d'accord sur cette interaction ; sinon, l'interaction est remplacée par une ignorance.

Une fois que toutes les expansions des systèmes d'argumentation initiaux sont calculées, on peut effectuer la fusion ; cela consiste à calculer tous les systèmes d'argumentation pouvant être construits à partir de l'ensemble complet d'argument qui soient les "plus proche possibles" des systèmes d'argumentation partiels. La proximité est caractérisée par la notion de distance entre un système d'argumentation et un profil (*i.e.*, un vecteur) de systèmes d'argumentation partiels, induite par une distance primitive entre systèmes d'argumentation partiels et une fonction d'agrégation. Il existe de nombreuses possibilités tant pour les distances primitives que pour les fonctions d'agrégation ; nous nous sommes principalement intéressés à la distance d'édition (en gros, le nombre d'ajouts et de suppressions nécessaires pour passer d'un système à l'autre), et nous avons considéré comme fonctions d'agrégation la somme, le max et le leximax.

Le processus que nous proposons part donc d'un ensemble de systèmes d'argumentation pour fournir un nouvel ensemble de systèmes d'argumentation. Toutefois, alors que l'ensemble initial reflète différents points de vue (chaque système donnant le point de vue d'un agent), l'ensemble résultant de la fusion représente l'incertitude liée à la fusion d'informations conflictuelles. La dernière étape consiste donc à définir les arguments acceptables pour le groupe d'agents en tenant compte de cette incertitude (et, une fois de plus, il existe plusieurs manières de faire cela). Après cette étape, nous montrons que les arguments choisis sont différents de ceux qui l'auraient été avant la fusion et sont plus en accord avec l'intuition.

Tous les travaux présentés dans cette section ont donné lieu à coopération avec des chercheurs du CRIL (Lens) et du LERIA (Angers). Le lecteur trouvera dans [CMDK$^+$05, CMDK$^+$07] l'ensemble des preuves des propriétés obtenues dans ce domaine.

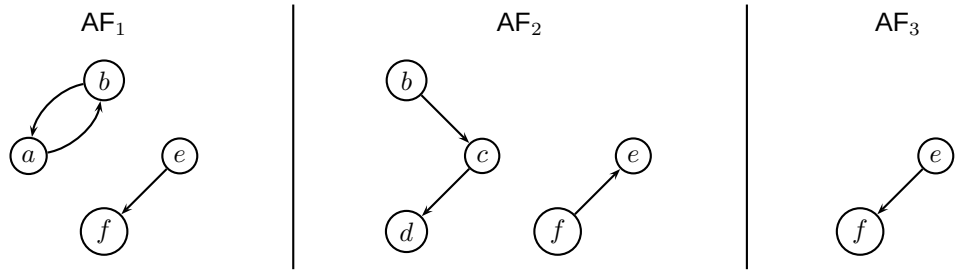
#### 3.2.4.1 Quelques exemples

Soit un profil de systèmes d'argumentation $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ de $n$ AFs (avec $n \geq 1$) où chaque $\mathsf{AF}_i = \langle A_i, R_i \rangle$ représente les données fournies par l'agent $i$, notre but est de trouver les sous-ensembles de $\bigcup_i A_i$ qui sont acceptables par le groupe des $n$ agents. Le vote est une des méthodes possibles.

##### 3.2.4.1.1 Voter n'est pas suffisant

En effet, on pourrait considérer qu'un ensemble d'arguments est acceptable pour le groupe s'il est acceptable pour "suffisamment" d'agents du groupe. La méthode de vote choisie permet d'expliciter ce que "suffisamment" signifie : cela peut être par exemple la majorité simple. Voyons ce que cela donne sur un exemple :

**Exemple 14** *Soit les 3 systèmes d'argumentation suivants :*
- $\mathsf{AF}_1 = \langle \{a, b, e, f\}, \{(a, b), (b, a), (e, f)\} \rangle$,
- $\mathsf{AF}_2 = \langle \{b, c, d, e, f\}, \{(b, c), (c, d), (f, e)\} \rangle$,
- $\mathsf{AF}_3 = \langle \{e, f\}, \{(e, f)\} \rangle$.




*Quelle que soit la sémantique choisie (parmi celles de Dung), $c$ n'appartiendra pas à une extension de $\mathsf{AF}_2$. Comme $c$ n'est pas connu des deux autres agents, il ne peut pas être considéré comme acceptable par le groupe quelle que soit la méthode de vote retenue (en supposant qu'il s'agisse d'une fonction de choix basée sur les extensions, i.e., que seuls des sous-ensembles d'extensions d'un $\mathsf{AF}_i$ sont éligibles pour être des ensembles acceptables). Toutefois, puisque $c$ (resp. $a$) ne fait pas partie des arguments connus par le premier (resp. le second ou le troisième) agent, on pourrait considérer que les trois agents sont d'accord sur le fait que $a$ attaque $b$, $b$ attaque $a$ et $b$ attaque $c$. En effet, cette hypothèse est compatible avec les systèmes d'argumentation proposés par chaque agent. Sous cette hypothèse, il est alors correct de considérer que $\{c\}$ est crédulement acceptable pour le groupe sachant que $c$ est considéré comme défendu par $a$ contre $b$ par l'agent 1 et qu'il n'y a pas de preuve du contraire fournie dans les systèmes d'argumentation des deux autres agents.*

Cet exemple illustre donc, qu'en général, une procédure de vote n'est pas un moyen satisfaisant pour agréger les données fournies par différents agents sous la forme de systèmes d'argumentation. Les deux principaux problèmes apparaissant lors d'un vote sont :

**Problème 1** Un vote n'a de sens que si tous les agents considèrent le même ensemble d'alternatives (donc le même ensemble d'arguments), ce qui n'est pas le cas en général.

**Problème 2** Un vote ne concernerait que les extensions choisies : les relations d'attaque, à partir desquelles les extensions ont été construites, ne seraient plus prises en considération, une fois les extensions obtenues. Cela nous conduirait à laisser de côté beaucoup d'informations significatives qui pourraient être utilisées pour déterminer les ensembles d'arguments acceptables pour le groupe d'agents.

#### 3.2.4.1.2 L'union n'est pas une fusion (en général)

Pour résoudre les deux problèmes identifiés précédemment, une approche simple paraît être de faire l'union des systèmes d'argumentation $\mathsf{AF}_1$, ..., $\mathsf{AF}_n$, *i.e.*, de considérer le système d'argumentation noté $\mathsf{AF} = \bigcup_{i=1}^{n} \langle A_i, R_i \rangle$ et défini par $\mathsf{AF} = \langle \bigcup_{i=1}^{n} A_i, \bigcup_{i=1}^{n} R_i \rangle$. Malheureusement, cette approche présente certains inconvénients majeurs illustrés à l'aide de l'exemple précédent :

**Exemple 14 page précédente (suite)** *Le résultat de l'union est* $\bigcup_{i=1}^{3} \mathsf{AF}_i = \langle \{a, b, c, d, e, f\}, \{(a, b), (b, a), (b, c), (c, d), (e, f), (f, e)\} \rangle$.

Cet exemple met en évidence le fait suivant : la résolution des conflits se fait en introduisant une prédominance de la relation d'attaque explicite par rapport à la relation de non-attaque implicite. Ainsi, quand une paire d'arguments n'appartient pas à la relation d'attaque d'un agent (comme par exemple $(f, e)$ pour l'agent 1) alors que les deux arguments sont connus de cet agent, cela signifie qu'il n'existe pas d'attaque (l'argument $f$ *n'attaque pas* l'argument $e$). Imaginons alors que 999 agents soient en accord avec l'agent 1, et que le $1000^{ime}$ agent soit l'agent 2 qui croit à l'attaque de $e$ par $f$. Si on fait l'union, le système d'argumentation résultat contiendra l'attaque de $e$ par $f$ alors il y a 999 agents sur 1000 qui croient le contraire !

### 3.2.4.2 Systèmes d'argumentation partiels

L'exemple introduit dans la section précédente a illustré les différents cas qu'il va falloir prendre en compte :
– un argument existe dans le système d'argumentation $\mathsf{AF}_1$ de l'un des agents et n'existe pas dans le système d'argumentation $\mathsf{AF}_2$ d'un autre agent ;
– une interaction entre deux arguments existe dans le système d'argumentation $\mathsf{AF}_1$ de l'un des agents et n'existe pas dans le système d'argumentation $\mathsf{AF}_2$ d'un autre agent.
Dans le premier cas, le nouvel argument peut être ajouté à $\mathsf{AF}_2$ mais la question ouverte est de savoir comment le lier aux autres arguments de $\mathsf{AF}_2$.

Dans le second cas, les choses sont un peu plus compliquées : si une interaction entre deux arguments $a$ et $b$ existe dans le système $\mathsf{AF}_1$ et pas dans le système $\mathsf{AF}_2$, même quand $a$ et $b$ sont dans $\mathsf{AF}_2$, nous ne pouvons pas ajouter cette interaction dans $\mathsf{AF}_2$ (on considère que c'est volontairement que l'agent 2 n'a pas inclus cette attaque dans $\mathsf{AF}_2$). En effet, si une interaction est absente d'un AF, cela signifie que cette interaction *n'existe pas* pour l'agent correspondant. La conséquence de cette approche est la nécessité d'établir une discrimination entre plusieurs cas de figure correspondant tous à l'ajout d'un argument $a$ dans un AF donné. Soit $b$ un argument de cet AF, trois cas sont à considérer :
– l'agent croit que l'interaction $(a, b)$ existe (il y a donc une attaque) ;
– l'agent croit que l'interaction $(a, b)$ n'existe pas (il y a donc une non-attaque) ;
– l'agent ne sait pas si l'interaction $(a, b)$ existe ou pas (il y a donc ignorance).
Les deux premiers cas expriment le fait que les connaissances de l'agent sont suffisantes pour lui permettre de calculer les interactions liant le nouvel argument $a$ aux autres arguments de son système. Le troisième cas exprime le fait que l'agent n'est pas capable de calculer les interactions liant $a$ aux arguments de son système (il peut y avoir plusieurs raisons à cela : manque d'information ou ressources de calcul insuffisantes).

Pour prendre en compte tous ces cas, nous avons donc défini une extension du cadre argumentatif proposé par Dung en introduisant une nouvelle interaction entre les arguments appelée l'*ignorance*[18] ; cela nous permet de définir des *systèmes d'argumentation partiels*.

**Définition 44 (système d'argumentation partiel (PAF))** *Un* système d'argumentation partiel (fini) *sur $\mathcal{A}$ est un quadruplet* $\mathsf{PAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{I}, \mathcal{N} \rangle$ *avec*
– *$\mathcal{A}$ un ensemble fini d'arguments,*

[18] Dans [CLS05a], un autre type d'extension du cadre de Dung avait été proposé : l'argumentation bipolaire ; cela correspond à l'introduction d'une autre relation binaire entre arguments représentant une notion de support entre arguments. Cette extension rappelée en Section 3.3 page 106 est bien-sûr totalement différente de celle présentée dans la section courante.

– $\mathcal{R}, \mathcal{I}, \mathcal{N}$ *des relations binaires sur* $\mathcal{A}$ *:*
  – $\mathcal{R}$ *est la* relation d'attaque,
  – $\mathcal{I}$ *est la* relation d'ignorance *et est telle que* $\mathcal{R} \cap \mathcal{I} = \varnothing$,
  – *et* $\mathcal{N} = (\mathcal{A} \times \mathcal{A}) \setminus (\mathcal{R} \cup \mathcal{I})$ *est la* relation de non-attaque.

$\mathcal{N}$ *se déduit de* $\mathcal{A}$*,* $\mathcal{R}$ *et* $\mathcal{I}$*, ainsi un système d'argumentation partiel est entièrement défini par* $\langle \mathcal{A}, \mathcal{R}, \mathcal{I} \rangle$ *(les deux notations seront utilisées dans la suite).*

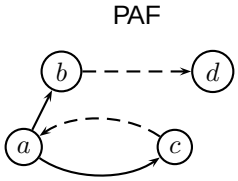
Chaque AF est un PAF particulier dans lequel l'ensemble $\mathcal{I}$ est vide (nous dirons alors que cet AF est équivalent au PAF associé). Dans un AF, la relation $\mathcal{N}$ existe déjà mais elle n'est pas donnée de manière explicite ($\mathcal{I} = \varnothing$ et $\mathcal{N} = \mathcal{A} \times \mathcal{A} \setminus \mathcal{R}$). Donc un AF peut aussi être dénoté par $\langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$.

Chaque PAF sur $\mathcal{A}$ peut être vu comme une représentation compacte de plusieurs AFs sur $\mathcal{A}$, appelés les *complétions* du PAF :

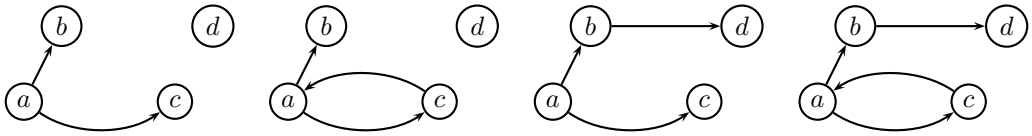
**Définition 45 (Complétion d'un PAF)** *Soit* $\mathsf{PAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{I} \rangle$*. Soit* $\mathsf{AF} = \langle \mathcal{A}, S \rangle$*.* $\mathsf{AF}$ *est la* complétion *de* $\mathsf{PAF}$ *ssi* $\mathcal{R} \subseteq S \subseteq \mathcal{R} \cup \mathcal{I}$*.*

*L'ensemble de toutes les complétions de* $\mathsf{PAF}$ *sera noté* $\mathcal{C}(\mathsf{PAF})$*.*

**Exemple 15** *Le système d'argumentation partiel* $\mathsf{PAF} = \langle \mathcal{A} = \{a, b, c, d\}, \mathcal{R} = \{(a, b), (a, c)\}, \mathcal{I} = \{(c, a), (b, d)\}, \mathcal{N} = \{(a, a), (b, b), (c, c), (d, d), (b, a), (b, c), (c, b), (a, d), (d, a), (d, b), (c, d), (d, c)\}\rangle$ *est représenté par la figure suivante (les flèches pleines donnent la relation d'attaque et les flèches en pointillé donnent la relation d'ignorance ; la relation de non-attaque ne sera pas représentée de manière explicite comme c'est déjà le cas quand on représente graphiquement un AF) :*




*Les complétions de ce PAF sont :*




Désormais, le **problème 1** peut être traité en associant à chaque système d'argumentation $\mathsf{AF}_i$ le système d'argumentation partiel correspondant $\mathsf{PAF}_i$ de telle sorte que tous les $\mathsf{PAF}_i$ travaillent sur le même ensemble d'arguments $\bigcup_{i=1}^{n} A_i$. Dans ce but, nous introduisons la notion *d'expansion* d'un AF :

**Définition 46 (Expansion d'un AF)** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil de* $n$ *AFs tels que* $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i, \mathcal{N}_i \rangle$*. Soit* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ *un système d'argumentation. Une* expansion *de* $\mathsf{AF}$ *suivant* $\mathcal{P}$ *est un PAF dénoté par* $\mathsf{exp}(\mathsf{AF}, \mathcal{P})$ *et défini par* $\langle \mathcal{A} \cup \bigcup_i \mathcal{A}_i, \mathcal{R}', \mathcal{I}', \mathcal{N}' \rangle$ *tel que* $\mathcal{R} \subseteq \mathcal{R}'$ *et* $(\mathcal{A} \times \mathcal{A}) \setminus \mathcal{R} \subseteq \mathcal{N}'$*.* $\mathsf{exp}$ *sera appelée la* fonction d'expansion*.*

Afin d'être la plus générale possible, cette définition n'impose que très peu de contraintes au PAF résultat : la seule contrainte essentielle est la préservation des relations d'attaque et de non-attaque de l'AF initial au moment où on étend l'ensemble des arguments de cet AF. Il existe beaucoup de manières d'implémenter cette définition en imposant davantage de contraintes, reflétant diverses attitudes face à l'ajout de nouveaux arguments ; par exemple, si $a$ est un des arguments initiaux de l'agent $i$ et qu'il arrive un "nouvel" argument $b$, l'agent $i$ peut (entre autres choses) :
– rejeter systématiquement $b$ (e.g., en ajoutant $(b, b)$ à sa relation d'attaque $\mathcal{R}'_i$),
– accepter systématiquement $b$ (en ajoutant $(a, b)$, $(b, a)$ et $(b, b)$ à sa relation de non-attaque $\mathcal{N}'_i$),

– exprimer son ignorance à propos de $b$ (en ajoutant $(a, b)$, $(b, a)$ et $(b, b)$ à sa relation d'ignorance $\mathcal{I}'_i$).
Chaque agent peut aussi calculer l'interaction exacte entre $a$ et $b$ lorsque la relation d'attaque n'est pas primitive mais qu'elle est donnée à partir de notions plus basiques (comme par exemple dans le travail de Elvang-Gøransson et al., voir e.g., [EGFK93a, EGFK93b, EGH95]). Remarquons que si l'agent dispose de capacités de calcul limitées, il peut calculer le plus possible d'interactions puis décider que celles qu'il n'a pu calculer sont des ignorances.

Dans la suite, nous allons nous focaliser sur une expansion particulière : *l'expansion consensuelle*. Intuitivement, l'expansion consensuelle d'un système d'argumentation $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ suivant un profil donné d'AFs s'obtient en ajoutant une attaque (resp. une non-attaque) $(a, b)$ (avec au moins un des $a$ ou $b$ qui n'appartenaient pas à $\mathcal{A}$) si cette interaction *existe pour tous les autres agents du profil connaissant ces deux arguments*; sinon, le couple $(a, b)$ est ajouté à la relation d'ignorance.

Cette manière d'implémenter la notion d'expansion signifie que chaque agent fait confiance aux autres agents en ce qui concerne les arguments qui lui étaient inconnus avant l'expansion : si une pièce d'information proposée par un agent n'est pas en conflit avec les informations données par les autres agents, chaque agent du groupe l'acceptera.

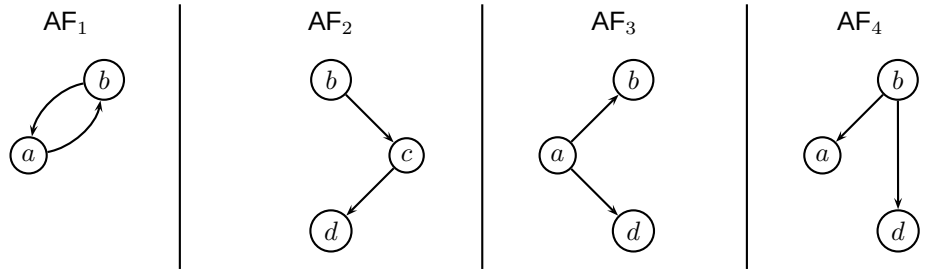
**Définition 47 (Expansion consensuelle)** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil de $n$ AFs tels que $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle$. Soit $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$ un système d'argumentation. Soit $conf(\mathcal{P}) = (\bigcup_i \mathcal{R}_i) \cap (\bigcup_i \mathcal{N}_i)$ un ensemble d'interactions pour lesquelles un conflit existe parmi les agents du profil.* L'expansion consensuelle *de* $\mathsf{AF}$ *suivant $\mathcal{P}$ est le tuple noté* $\mathsf{exp}_C = \langle \mathcal{A}', \mathcal{R}', \mathcal{I}', \mathcal{N}' \rangle$ *avec :*
– $\mathcal{A}' = \mathcal{A} \cup \bigcup_i \mathcal{A}_i$,
– $\mathcal{R}' = \mathcal{R} \cup ((\bigcup_i \mathcal{R}_i \setminus conf(\mathcal{P})) \setminus \mathcal{N})$,
– $\mathcal{I}' = conf(\mathcal{P}) \setminus (\mathcal{R} \cup \mathcal{N})$,
– $\mathcal{N}' = (\mathcal{A}' \times \mathcal{A}') \setminus (\mathcal{R}' \cup \mathcal{I}')$.

La proposition suivante confirme le fait que l'expansion consensuelle d'un AF suivant un profil est bien une expansion :

**Proposition 21** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil de $n$ AFs tels que $\mathsf{AF}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle$. Soit $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle$ un système d'argumentation. L'expansion consensuelle $\mathsf{exp}_C$ de $\mathsf{AF}$ suivant $\mathcal{P}$ est une expansion de $\mathsf{AF}$ suivant $\mathcal{P}$ dans le sens de la définition 46 page précédente.*

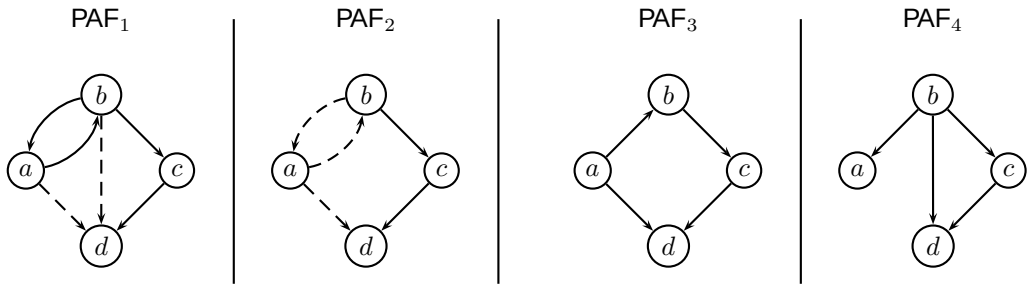
L'expansion consensuelle est une des expansions les plus prudentes puisqu'on ajoute une interaction du type attaque ou non-attaque seulement si tous les agents sont d'accord avec cette interaction.

**Exemple 16** *Considérons le profil contenant les quatre systèmes d'argumentation suivants :*
– $\mathsf{AF}_1 = \langle \mathcal{A}_1 = \{a, b\}, \mathcal{R}_1 = \{(a, b), (b, a)\} \rangle$,
– $\mathsf{AF}_2 = \langle \mathcal{A}_2 = \{b, c, d\}, \mathcal{R}_2 = \{(b, c), (c, d)\} \rangle$,
– $\mathsf{AF}_3 = \langle \mathcal{A}_3 = \{a, b, d\}, \mathcal{R}_3 = \{(a, b), (a, d)\} \rangle$,
– $\mathsf{AF}_4 = \langle \mathcal{A}_4 = \{a, b, d\}, \mathcal{R}_4 = \{(b, d), (b, a)\} \rangle$.




*Pour chaque $i$, l'expansion consensuelle $\mathsf{PAF}_i$ de $\mathsf{AF}_i$ est :*
– $\mathsf{PAF}_1 = \langle \quad \{a, b, c, d\}, \quad \{(a, b), (b, a), (b, c), (c, d)\}, \quad \{(a, d), (b, d)\} \quad \rangle$,
– $\mathsf{PAF}_2 = \langle \quad \{a, b, c, d\}, \quad \{(b, c), (c, d)\}, \quad \{(a, b), (b, a), (a, d)\} \quad \rangle$,
– $\mathsf{PAF}_3 = \langle \quad \{a, b, c, d\}, \quad \{(a, b), (a, d), (b, c), (c, d)\}, \quad \varnothing \quad \rangle$,

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




Quand les expansions considérées par chaque agent sont obtenues par la même fonction d'expansion $\mathsf{exp}$, alors, pour un profil $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ donné, on notera $\mathsf{exp}(\mathcal{P})$ le profil des PAFs $\langle \mathsf{exp}(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathsf{exp}(\mathsf{AF}_n, \mathcal{P}) \rangle$.

#### 3.2.4.3 Opérateurs de fusion

Afin de traiter le **problème 2**, nous proposons de fusionner les AFs au niveau des interactions plutôt qu'au niveau des ensembles d'arguments acceptables. Ainsi nous pourrons caractériser les systèmes d'argumentation qui sont les plus proches possible du profil d'AFs donné.

Une manière d'arriver à cela consiste à définir une notion de "distance" entre un AF et un profil d'AFs, ou plus généralement entre un PAF et un profil de PAFs. Pour cela, il va nous falloir une notion de pseudo-distance entre deux PAFs, et une manière de combiner de telles pseudo-distances :

**Définition 48 (Pseudo-distance)** *Une* pseudo-distance $d$ *entre PAFs sur $\mathcal{A}$ est une fonction qui associe un nombre réel non négatif à chaque paire de PAFs sur $\mathcal{A}$ et qui satisfait les propriétés de symétrie ($d(x,y) = d(y,x)$) et de minimalité ($d(x,y) = 0$ si et seulement si $x = y$).*

*$d$ est une distance si elle satisfait aussi l'inégalité triangulaire ($d(x,z) \leq d(x,y) + d(y,z)$).*

**Définition 49 (Fonction d'agrégation)** *Une* fonction d'agrégation *est une fonction*[19] $\otimes$ *de* $(\mathbb{R}+)^n$ *vers* $(\mathbb{R}+)$*, qui satisfait*
– *si* $x_i \geq x'_i$*, alors* $\otimes(x_1, \ldots, x_i, \ldots, x_n) \geq \otimes(x_1, \ldots, x'_i, \ldots, x_n)$ **(fonction croissante)**
– $\otimes(x_1, \ldots, x_n) = 0$ *si* $\forall i$, $x_i = 0$ **(minimalité)**
– $\otimes(x) = x$ **(identité)**

La fusion d'un profil d'AFs est définie comme un ensemble d'AFs :

**Définition 50 (Fusion de $n$ AFs)** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil de $n$ AFs. Soit $d$ une pseudo-distance quelconque entre PAFs, soit $\otimes$ une fonction d'agrégation, et soit $\mathsf{exp}_1, \ldots, \mathsf{exp}_n$ $n$ fonctions d'expansion. La* fusion *de $\mathcal{P}$ est l'ensemble d'AFs défini par :*

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

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

*Afin d'alléger un peu les notations, l'ensemble des AFs résultant de cette fusion $\{\mathsf{AF}'_1, \ldots, \mathsf{AF}'_k\}$ sera noté avec le profil $\langle \mathsf{AF}'_1, \ldots, \mathsf{AF}'_k \rangle$ (ou toute permutation de ce profil).*

Ainsi, fusionner un profil d'AFs $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ est un processus en deux étapes :

[19] En fait, strictement parlant, c'est une famille de fonctions, une pour chaque $n$.

**expansion :** Tout d'abord, on calcule l'expansion de chaque $\mathsf{AF}_i$ de $\mathcal{P}$. Remarquons que l'on peut choisir une fonction d'expansion propre à chaque agent. La seule chose vraiment importante est que $\mathtt{exp}_i(\mathsf{AF}_i, \mathcal{P})$ soit un PAF sur $\mathcal{A} = \bigcup_i \mathcal{A}_i$.

**fusion :** Les AFs sur $\mathcal{A}$ représentant le résultat de la fusion seront ceux qui sont les plus proches possibles de l'expansion de $\mathcal{P}$ étant donné les distances agrégées.

Nous supposerons dans la suite que chaque agent utilise l'expansion consensuelle. Ce qui nous permettra de simplifier les notations en supprimant $\langle \mathtt{exp}_1, \ldots, \mathtt{exp}_n \rangle$ de la liste des paramètres des opérateurs de fusion.

Remarquons qu'il serait possible de réécrire la définition 50 page ci-contre de manière à inclure certaines contraintes d'intégrité. Cela peut être en particulier utile s'il existe des connaissances (imposées) concernant le résultat attendu de la fusion (on impose par exemple que certaines attaques doivent être vérifiées pour tout le groupe). Ainsi, on ne tiendrait compte que des AFs vérifiant ces contraintes ; un travail similaire existe en fusion de bases de connaissances propositionnelles (voir par exemple, [KP02]). Par rapport à ce qui peut se faire en fusion de bases de connaissances, on pourrait trouver ici des contraintes un peu particulières ; par exemple, si on s'intéresse à des contraintes *structurelles*, on pourrait imposer que les AFs candidats soient acycliques, afin de préserver des propriétés intéressantes comme le fait d'avoir des AFs (1) bien-fondés (ce qui implique l'existence d'une extension unique quelle que soit la sémantique choisie parmi celles proposées par Dung), et que (2) cette extension (qui est alors l'extension basique, voir [Dun95]) soit calculable en un temps polynomial dans la taille de l'AF (alors qu'en général ce calcul est exponentiel, voir [DBC02]).

Il existe beaucoup de possibilités, tant pour les pseudo-distances entre PAFs que pour les fonctions d'agrégation. Parmi les fonctions d'agrégation les plus classiques, on trouve la somme $\Sigma$, le max $\mathcal{M}\mathtt{ax}$ et le leximax $\mathcal{L}\mathtt{eximax}$[20] mais on peut aussi utiliser des fonctions non symétriques (en particulier si certains agents sont considérés comme plus importants que d'autres). Certaines des fonctions d'agrégation (comme la somme) permettent de prendre en compte le nombre d'agents croyant ou pas en une attaque donnée :

**Exemple 14 page 64 (suite)** *Deux agents sur trois sont d'accord avec l'existence de l'attaque de $e$ sur $f$ et la non-existence de l'attaque de $f$ sur $e$. Si on utilise la règle de la majorité, le groupe devrait être du même avis que ces deux agents.*

Le choix de la fonction d'agrégation est très important et lié au comportement attendu lors de la fusion. Par exemple, la somme est un choix possible si on veut résoudre les conflits entre agents en utilisant la règle de la majorité. D'un autre côté, si on veut trouver une manière de conclure plus consensuelle, alors la fonction leximax sera mieux adaptée car elle fournira un résultat plus proche de chaque agent. Dans le cadre de la fusion de bases de connaissances propositionnelles, il existe déjà une étude explicitant les conséquences du choix entre majorité et arbitrage (voir [KP02]) et cette étude peut aussi s'appliquer au problème de la fusion d'AFs.

Dans la suite, nous nous intéresserons à la *distance d'édition* entre PAFs :

**Définition 51 (Distance d'édition)** *Soit $\mathsf{PAF}_1 = \langle \mathcal{A}, \mathcal{R}_1, \mathcal{I}_1, \mathcal{N}_1 \rangle$ et $\mathsf{PAF}_2 = \langle \mathcal{A}, \mathcal{R}_2, \mathcal{I}_2, \mathcal{N}_2 \rangle$ deux PAFs sur $\mathcal{A}$.*
- *Soit $a$, $b$ deux arguments $\in \mathcal{A}$. La* distance d'édition entre $\mathsf{PAF}_1$ et $\mathsf{PAF}_2$ sur $a, b$ *est la fonction $de_{a,b}$ telle que :*
  - $de_{a,b}(\mathsf{PAF}1, \mathsf{PAF}2) = 0$ *si et seulement si* $(a, b) \in \mathcal{R}_1 \cap \mathcal{R}_2$ *ou* $\mathcal{I}_1 \cap \mathcal{I}_2$ *ou* $\mathcal{N}_1 \cap \mathcal{N}_2$,
  - $de_{a,b}(\mathsf{PAF}1, \mathsf{PAF}2) = 1$ *si et seulement si* $(a, b) \in \mathcal{R}_1 \cap \mathcal{N}_2$ *ou* $\mathcal{N}_1 \cap \mathcal{R}_2$,
  - $de_{a,b}(\mathsf{PAF}1, \mathsf{PAF}2) = 0.5$ *sinon.*
- *La* distance d'édition entre $\mathsf{PAF}_1$ et $\mathsf{PAF}_2$ *est donnée par*

$$de(\mathsf{PAF}1, \mathsf{PAF}2) = \Sigma_{(a,b)\in\mathcal{A}\times\mathcal{A}} de_{a,b}(\mathsf{PAF}1, \mathsf{PAF}2).$$

La distance d'édition entre deux PAFs est le nombre (minimum) d'additions/suppressions à réaliser pour rendre les deux PAFs identiques. L'ignorance est traitée comme un moyen terme entre attaque et non-attaque.

Il est alors facile de montrer que :

[20] Appliquée à un vecteur de $n$ nombres réels, la fonction $\mathcal{L}\mathtt{eximax}$ renvoie la liste triée par ordre décroissant. Ces listes peuvent alors être comparées en suivant l'ordre lexicographique induit par l'ordre classique sur les nombres réels.

**Proposition 22** *La distance d'édition $de$ entre PAFs est une distance.*

Cette notion de distance d'édition ainsi que quelques opérateurs de fusion sont illustrés sur l'exemple 16 page 67.

**Exemple 16 page 67 (suite)** *Considérons l'AF défini par* $\mathsf{AF}'_1 = \langle\{a,b,c,d\}, \{(a,b),(b,a),(b,c),(c,d)\}\rangle$.

*La distance d'édition entre* $\mathsf{AF}'_1$ *et chacun des PAFs* $\mathsf{PAF}_1$, $\mathsf{PAF}_2$, $\mathsf{PAF}_3$, $\mathsf{PAF}_4$ *obtenus par expansion consensuelle à partir du profil* $\langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4\rangle$ *est la suivante :*

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

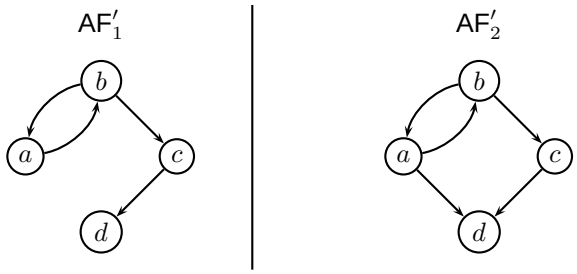
*En prenant la somme comme opérateur de fusion, on obtient :*
$\Sigma_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 6.5$.

*En prenant le max, on obtient :* $\mathcal{M}\mathtt{ax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 2$.

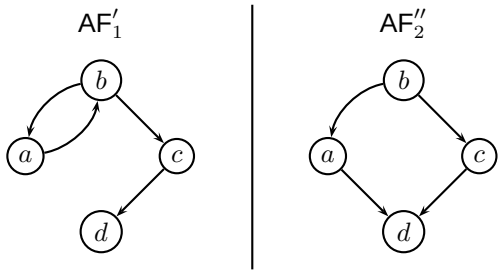
*En prenant le leximax, on obtient :* $\mathcal{L}\mathtt{eximax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = (2, 2, 1.5, 1)$.

*Si on effectue ce calcul de distance pour tous les AFs candidats (i.e., tous les AFs sur* $\{a,b,c,d\}$*), on peut faire le calcul complet de la fusion :*

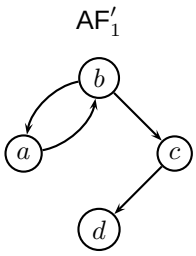
$\Delta_{de}^{\Sigma}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_4\rangle)$ *est l'ensemble contenant les deux AFs suivants :*
- $\mathsf{AF}'_1 = \langle\{a,b,c,d\}, \{(a,b),(b,a),(b,c),(c,d)\}\rangle$,
- $\mathsf{AF}'_2 = \langle\{a,b,c,d\}, \{(a,b),(b,a),(b,c),(a,d),(c,d)\}\rangle$.




$\Delta_{de}^{\mathcal{M}\mathtt{ax}}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_4\rangle)$ *est l'ensemble contenant* $\mathsf{AF}'_1$ *et* $\mathsf{AF}''_2 = \langle\{a,b,c,d\}, \{(b,a),(b,c),(a,d),(c,d)\}\rangle$.




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



*Les différences de résultats suivant les opérateurs utilisés s'expliquent aisément car :*

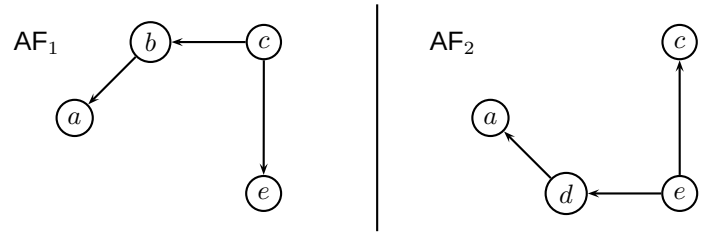
- $\mathsf{AF}'_1$ *est l'AF le plus consensuel car il est équidistant à chaque PAF ;*
- $\mathsf{AF}'_2$ *est plus proche de* $\mathsf{PAF}_1$, $\mathsf{PAF}_2$ *et* $\mathsf{PAF}_3$ *que de* $\mathsf{PAF}_4$*, il peut alors être sélectionné en utilisant la somme mais pas en utilisant le* $\mathcal{M}\mathtt{ax}$ *ou le* $\mathcal{L}\mathtt{eximax}$ *car il est trop loin de* $\mathsf{PAF}_4$*.*
- $\mathsf{AF}''_2$ *est presque équidistant de chacun des quatre PAFs du profil mais moins consensuel que* $\mathsf{AF}'_1$*, il sera alors sélectionné ni avec* $\Sigma$*, ni avec* $\mathcal{L}\mathtt{eximax}$ *mais seulement avec* $\mathcal{M}\mathtt{ax}$ *étant donné qu'il n'est pas très éloigné des PAFs donnés.*

*Remarquons que le fait que* $\mathsf{AF}'_1$ *apparaisse à chaque fois, quelle que soit la fonction d'agrégation choisie, semble très intuitif. En effet, toutes les interactions (attaque ou non-attaque) qui existent pour une majorité (faible) appartiennent aussi à* $\mathsf{AF}'_1$*.*

Quelques autres exemples illustrent aussi la démarche suivie jusqu'alors :

**Exemple 17** *Considérons les AFs définis par :*

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




*Remarquons que l'attaque de* $c$ *sur* $e$ *est connue de l'agent 1 mais pas de l'agent 2 et que l'attaque de* $e$ *sur* $c$ *est connue de l'agent 2 mais pas de l'agent 1. Cela illustre bien le fait que les agents peuvent ne pas partager les mêmes connaissances concernant les attaques et donc avoir des relations d'attaque différentes.*

$\mathsf{AF}_1$ *possède une unique extension préférée :* $\{c, a\}$*.* $\mathsf{AF}_2$ *possède une unique extension préférée :* $\{e, a\}$*.*

*Les expansions consensuelles de* $\mathsf{AF}_1$ *et de* $\mathsf{AF}_2$ *sont respectivement :*

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

*Le résultat de la fusion du profil* $\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle$ *avec* $de$ *et* $\otimes = \mathcal{M}\mathtt{ax}$ *(ou* $\otimes = \mathcal{L}\mathtt{eximax}$*) est :*

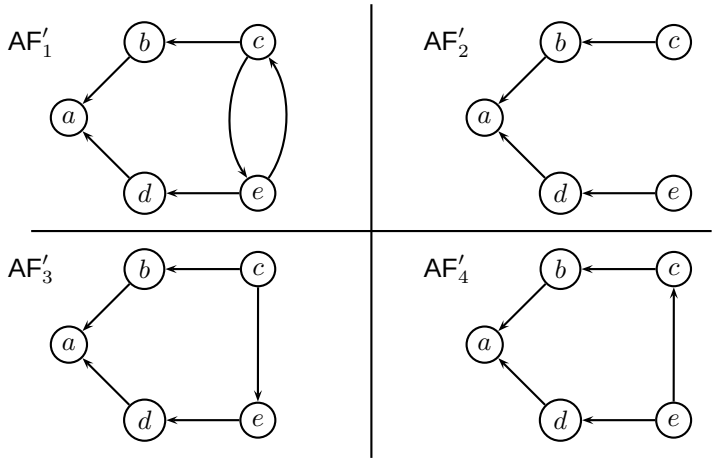
$$\Delta_{de}^{\mathcal{M}\mathtt{ax}}(\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle) = \Delta_{de}^{\mathcal{L}\mathtt{eximax}}(\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle) = \{\mathsf{AF}'_1, \mathsf{AF}'_2\} \text{ avec}$$

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

*Et si on utilise la somme comme fonction d'agrégation, deux AFs supplémentaires sont générés :*

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

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




*Chacun des résultats de la fusion contient un AF permettant de dériver l'argument $a$ (c'est le cas dans $\mathsf{AF}_1$ et $\mathsf{AF}_2$). L'utilisation de la somme comme fonction d'agrégation conduit au résultat le plus consensuel puisqu'il conserve les AFs initiaux de chaque agent ($\mathsf{AF}'_3$ est équivalent à $\mathsf{PAF}_1$ et $\mathsf{AF}'_4$ est équivalent à $\mathsf{PAF}_2$).*

#### 3.2.4.4 Quelques propriétés

Dans cette section, nous donnons quelques-unes des propriétés les plus importantes obtenues grâce à l'expansion consensuelle et aux opérateurs de fusion utilisant la distance d'édition. Ces propriétés illustrent bien le fait que ces choix d'expansion et de distance sont des choix intéressants.

##### 3.2.4.4.1 Propriétés des PAFs et des expansions consensuelles

De manière intuitive, on attend du résultat de la fusion qu'il préserve toutes les informations sur lesquelles les agents étaient déjà en accord, et même, plus généralement, celles sur lesquelles ils n'étaient pas en désaccord.

Cette propriété intuitive repose sur les notions de "partie commune" et "partie en accord" d'un profil de PAFs :

**Définition 52 (Partie en accord d'un profil de PAFs)**
*Soit $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ un profil de PAFs. La* partie en accord *de $\mathcal{P}$ est notée $CFP(\mathcal{P})$ et est définie par :*

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

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

La partie en accord d'un profil de PAFs représente les informations (attaque ou non-attaque) qui ne sont remises en question par aucun agent. Comme elles ne sont pas source de conflit, on s'attend à les retrouver telles quelles dans le résultat de la fusion.

**Exemple 16 page 67 (suite)** *Avec $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$,*

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

Remarquons qu'avec $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathsf{exp}_C(\mathsf{AF}_4, \mathcal{P}) \rangle$, $CFP(\mathsf{exp}_C(\mathcal{P})) = \langle \{a,b,c,d\}, \{(b,c), (c,d)\}, \{(a,b),(b,a),(a,d),(b,d)\} \rangle$ (maintenant $(a,c)$ et $(c,a)$ sont des non-attaques); on a donc $CFP(\mathcal{P}) \neq CFP(\mathsf{exp}_C(\mathcal{P}))$.

**Définition 53 (Partie commune d'un profil de PAFs)**
*Soit* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *un profil de PAFs. La* partie commune *de* $\mathcal{P}$ *est notée* $CP(\mathcal{P})$ *et est définie par :* $CP(\mathcal{P}) = \langle \bigcap_i \mathcal{A}_i, \bigcap_i \mathcal{R}_i, \bigcap_i \mathcal{I}_i, \bigcap_i \mathcal{N}_i \rangle$.

La notion de partie commune d'un profil de PAFs est une notion plus stricte que celle de partie en accord. Elle représente les informations pour lesquelles tous les agents sont d'accord. Ce sont donc des informations qui sont vérifiées quelle que soit la vision consensuelle du groupe et donc elles doivent apparaître dans tous les résultats de la fusion.

**Exemple 16 page 67 (suite)** *Avec* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$, $CP(\mathcal{P}) = \langle \{b\}, \varnothing, \varnothing, \{(b,b)\} \rangle$.

Bien-sûr il existe des liens entre ces deux notions :

**Proposition 23** *Soit* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *un profil de PAFs. La* partie commune *de* $\mathcal{P}$ *est "incluse" dans la partie en accord de* $\mathcal{P}$*, i.e. :*
– $\bigcap_i \mathcal{R}_i \subseteq \bigcup_i \mathcal{R}_i \setminus \bigcup_i \mathcal{N}_i$ ;
– $\bigcap_i \mathcal{I}_i \subseteq \mathcal{I}_{CFP}$ ;
– $\bigcap_i \mathcal{N}_i \subseteq \bigcup_i \mathcal{N}_i \setminus \bigcup_i \mathcal{R}_i$.

La partie commune d'un profil de $n$ PAFs (resp. AFs) n'est pas toujours un PAF (resp. un AF). Par contre, la partie en accord d'un profil de $n$ PAFs est aussi un PAF (toutefois, la partie en accord d'un profil de $n$ AFs n'est pas toujours un AF).

Il existe un cas particulier intéressant : si les différents PAFs d'un profil sont basés sur le même ensemble d'arguments et si pour chaque paire d'arguments $(a, b)$ telle que $(a, b)$ appartient à la relation d'ignorance d'un des PAFs, cette paire appartient aussi à la relation d'attaque d'un autre PAF et à la relation de non-attaque d'un troisième PAF alors les parties communes et en accord sont identiques :

**Proposition 24** *Soit* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *un profil de* $n$ *PAFs sur le même ensemble d'arguments* $\mathcal{A}$*. Soit la partie en accord de* $\mathcal{P}$ *dénotée par* $CFP(\mathcal{P}) = \langle \mathcal{A}_{CFP}, \mathcal{R}_{CFP}, \mathcal{I}_{CFP}, \mathcal{N}_{CFP} \rangle$ *et la partie commune de* $\mathcal{P}$ *dénotée par* $CP(\mathcal{P}) = \langle \mathcal{A}_{CP}, \mathcal{R}_{CP}, \mathcal{I}_{CP}, \mathcal{N}_{CP} \rangle$*. Si* $\bigcup_i \mathcal{I}_i \subseteq conf(\mathcal{P}) = (\bigcup_i \mathcal{R}_i) \cap (\bigcup_i \mathcal{N}_i)$*, nous avons :*
– $\mathcal{A}_{CFP} = \mathcal{A}_{CP}$,
– $\mathcal{R}_{CFP} = \mathcal{R}_{CP}$,
– $\mathcal{N}_{CFP} = \mathcal{N}_{CP}$.

Ce résultat est intéressant pour nous car c'est exactement ce qui se produit si tous les agents utilisent l'expansion consensuelle.

**Exemple 16 page 67 (suite)** *Avec* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$ *et* $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_2, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_3, \mathcal{P}), \mathsf{exp}_C(\mathsf{AF}_4, \mathcal{P}) \rangle$*, on a :*
– $CFP(\mathsf{exp}_C(\mathcal{P})) =$
$\langle \{a,b,c,d\}, \quad \{(b,c),(c,d)\}, \quad \{(a,b),(b,a),(a,d),(b,d)\}, \quad \{(a,a), (b,b), (c,c), (d,d), (a,c), (c,a), (d,a), (d,b), (d,c), (c,b)\} \rangle$
– $CP(\mathsf{exp}_C(\mathcal{P})) =$
$\langle \{a,b,c,d\}, \quad \{(b,c),(c,d)\}, \quad \varnothing, \quad \{(a,a),(b,b),(c,c),(d,d),(a,c),(c,a),(d,a),(d,b),(d,c),(c,b)\} \rangle$.

Une autre propriété intéressante est le fait que l'expansion consensuelle préserve la partie en accord du profil :

**Proposition 25** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs. Pour chaque* $i$*, on a :*
– $\mathcal{A}_{CFP(\mathcal{P})} = \mathcal{A}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$,
– $\mathcal{R}_{CFP(\mathcal{P})} \subseteq \mathcal{R}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$,
– $\mathcal{N}_{CFP(\mathcal{P})} \subseteq \mathcal{N}_{\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})}$.

On peut aussi introduire la notion de concordance entre AFs :

**Définition 54 (Concordance)** *Soit* $\mathsf{AF}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$, $\mathsf{AF}_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$ *deux AFs.* $\mathsf{AF}_1$, $\mathsf{AF}_2$ *sont dits* concordants *si et seulement si* $\forall (a,b) \in (\mathcal{A}_1 \cap \mathcal{A}_2) \times (\mathcal{A}_1 \cap \mathcal{A}_2)$, $(a,b) \in \mathcal{R}_1$ *si et seulement si* $(a,b) \in \mathcal{R}_2$. *Sinon, on dit qu'ils sont* discordants.
*Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs.* $\mathcal{P}$ *est dit* concordant *si et seulement si tous ses AFs sont concordants deux à deux. Sinon, il est dit* discordant.

Bien-sûr, la concordance est liée à l'ensemble $conf(\mathcal{P})$ représentant les conflits entre attaque et non-attaque entre les différents AFs du profil :

**Proposition 26** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs.* $\mathcal{P}$ *est concordant si et seulement si* $conf(\mathcal{P}) = \bigcup_i \mathcal{R}_i \cap \bigcup_i \mathcal{N}_i$ *est vide.*

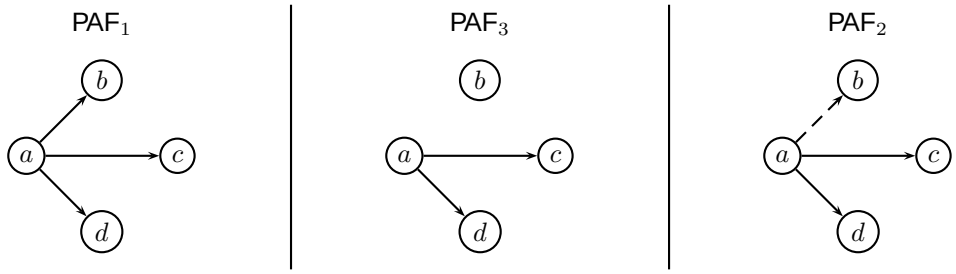
Quand un profil d'AFs est concordant, sa partie en accord est l'union de ses éléments et vice versa :

**Proposition 27** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs.* $\mathcal{P}$ *est concordant si et seulement si* $CFP(\mathcal{P}) = \bigcup_i \mathsf{AF}_i$.

**Proposition 28** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs.*
$\mathcal{P}$ *est concordant si et seulement si* $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathsf{exp}_C(\mathsf{AF}_n, \mathcal{P}) \rangle$ *se réduit à* $\langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle$ *(i.e., chacun des* $n$ *éléments du vecteur est égal à* $\bigcup_i \mathsf{AF}_i$*).*

Remarquons que $\bigcup_i \mathsf{AF}_i$ peut aussi apparaître dans $\mathsf{exp}_C(\mathcal{P})$, même si $\mathcal{P}$ est discordant, comme le montre l'exemple suivant :

**Exemple 18** *Considérons le profil* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *composé des trois AFs suivants :*
– $\mathsf{AF}_1 = \langle \{a,b,c\}, \{(a,b),(a,c)\} \rangle$,
– $\mathsf{AF}_2 = \langle \{a,b,c\}, \{(a,c)\} \rangle$,
– $\mathsf{AF}_3 = \langle \{a,d\}, \{(a,d)\} \rangle$.
*Le profil* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *est discordant et* $\mathsf{exp}_C(\mathcal{P}) = \langle \mathsf{PAF}_1, \mathsf{PAF}_2, \mathsf{PAF}_3 \rangle$ *avec :*
– $\mathsf{PAF}_1 = \langle \{a,b,c,d\}, \{(a,b),(a,c),(a,d)\}, \varnothing \rangle$ *(*$= \bigcup_i \mathsf{AF}_i$*)*,
– $\mathsf{PAF}_2 = \langle \{a,b,c,d\}, \{(a,c),(a,d)\}, \varnothing \rangle$,
– $\mathsf{PAF}_3 = \langle \{a,b,c,d\}, \{(a,c),(a,d)\}, \{(a,b)\} \rangle$.




La proposition suivante montre que si une attaque $(a,b)$ n'est pas en conflit avec un profil d'AFs, alors cette attaque sera présente dans tous les PAFs obtenus avec l'expansion consensuelle si et seulement si elle est présente dans au moins un AF du profil.

**Proposition 29** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil d'AFs. Soit* $(a,b)$ *une paire d'arguments telle que* $a, b \in \bigcup_i \mathcal{A}_i$ *et* $\nexists \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$ *tels que* $(a,b) \in (\mathcal{R}_i \setminus \mathcal{R}_j) \cup (\mathcal{R}_j \setminus \mathcal{R}_i)$.

$\exists \mathsf{AF}_l \in \mathcal{P}$ *tel que* $(a,b) \in \mathcal{R}_l$ *si et seulement si* $\forall \mathsf{AF}_k \in \mathcal{P}$, $(a,b) \in R'_k$ *avec* $R'_k$ *dénotant la relation d'attaque du PAF* $\mathsf{exp}_C(\mathsf{AF}_k, \mathcal{P})$.

Une notion de compatibilité d'un profil de PAFs sur un même ensemble d'arguments peut alors être définie :

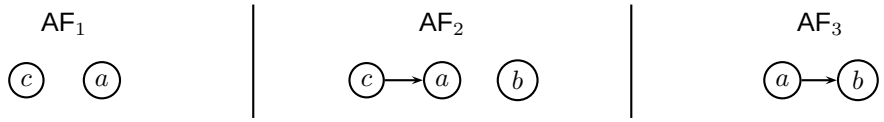
**Définition 55 (Compatibilité)** *Soit $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ un profil de PAFs sur un ensemble d'arguments $\mathcal{A}$.* $\mathsf{PAF}_1, \ldots, \mathsf{PAF}_n$ *sont dits* compatibles *si et seulement si ils ont au moins en commun une complétion. Sinon, ils sont dits* incompatibles.

*Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs. Soit* $\mathsf{exp}$ *une fonction d'expansion.* $\mathsf{AF}_1, \ldots, \mathsf{AF}_n$ *sont dits* compatibles *étant donné* $\mathsf{exp}$ *si et seulement si, $\forall i = 1 \ldots n$, les $\mathsf{exp}(\mathsf{AF}_i, \mathcal{P})$ sont compatibles. Sinon, les $\mathsf{AF}_1, \ldots, \mathsf{AF}_n$ sont dits* incompatibles.

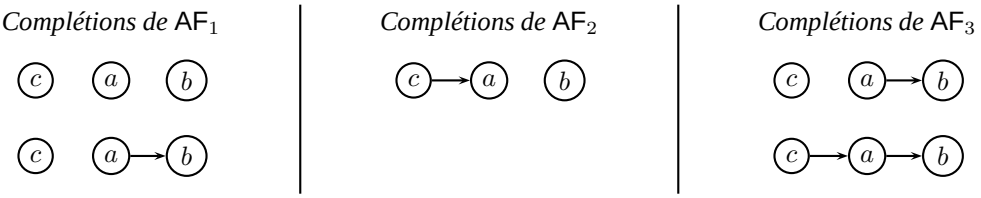
Il existe un lien entre compatibilité et concordance lorsqu'on utilise l'expansion consensuelle :

**Proposition 30** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs. $\mathcal{P}$ est concordant si et seulement si $\mathsf{exp}_C(\mathsf{AF}_1, \mathcal{P})$, ..., $\mathsf{exp}_C(\mathsf{AF}_n, \mathcal{P})$ sont compatibles.*

**Exemple 19** *Considérons* $\mathsf{AF}_1$, $\mathsf{AF}_2$ *et* $\mathsf{AF}_3$.




*Les complétions de leurs expansions consensuelles respectives* $\mathsf{PAF}_1$, $\mathsf{PAF}_2$ *et* $\mathsf{PAF}_3$ *sont :*




$\mathsf{AF}_1$ *et* $\mathsf{AF}_2$ *sont discordants et incompatibles étant donné* $\mathsf{exp}_C$. $\mathsf{AF}_3$ *et* $\mathsf{AF}_1$ *sont concordants et compatibles étant donné* $\mathsf{exp}_C$.

#### 3.2.4.4.2 Propriétés des opérateurs de fusion

Donnons maintenant quelques propriétés concernant les opérateurs de fusion en se focalisant en particulier sur l'utilisation de la distance d'édition.

**Proposition 31** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs. Supposons que la fonction d'expansion utilisée par chaque agent soit l'expansion consensuelle. Si $\mathcal{P}$ est concordant alors $\Delta^{\otimes}_{de}(\mathcal{P}) = \{\bigcup_i \mathsf{AF}_i\}$.*

La propriété suivante est une propriété attendue : la partie en accord d'un profil $\mathcal{P}$ est incluse dans chaque AF résultat de la fusion de $\mathcal{P}$ quand la distance d'édition est utilisée.

**Proposition 32** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs. Supposons que chaque agent utilise l'expansion consensuelle. Quelle que soit la fonction d'agrégation choisie $\otimes$, on a : $\forall \mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle)$ :*
– $\mathcal{A}_{CFP(\mathcal{P})} \subseteq \mathcal{A}$,
– $\mathcal{R}_{CFP(\mathcal{P})} \subseteq \mathcal{R}$,
– $\mathcal{N}_{CFP(\mathcal{P})} \subseteq \mathcal{N}$.

Un corollaire direct des propositions 23 page 73 et 32 est le suivant :

**Corollaire 1** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs. Supposons que chaque agent utilise l'expansion consensuelle. Quelle que soit la fonction d'agrégation choisie $\otimes$, on a : $\forall \mathsf{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{N} \rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle)$ :*
– $\mathcal{A}_{CP(\mathcal{P})} \subseteq \mathcal{A}$,

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

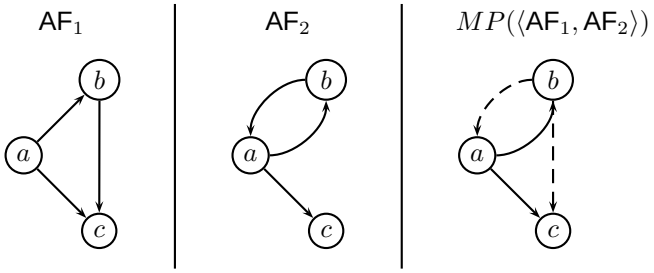
Quand la somme est utilisée comme fonction d'agrégation et que tous les AFs d'un profil utilisent le même ensemble d'arguments, la fusion de ce profil peut être caractérisée de manière beaucoup plus concise grâce à la notion de graphe de majorité. Intuitivement, un graphe de majorité issu d'un profil d'AFs sur un même ensemble d'arguments est le PAF obtenu en appliquant la règle de la majorité absolue pour décider si $a$ attaque $b$ ou pas, pour toute paire $(a, b)$ d'arguments. Et si aucune majorité ne se dessine, alors on considère qu'il y a ignorance.

**Définition 56 (PAF de majorité)** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs sur le même ensemble $\mathcal{A}$ d'arguments. Le* PAF de majorité *$MP(\mathcal{P})$ de $\mathcal{P}$ est le quadruplet $\langle \mathcal{A}, \mathcal{R}, \mathcal{N}, \mathcal{I} \rangle$ tel que $\forall a, b \in \mathcal{A}$ :*[21]
– *$(a, b) \in \mathcal{R}$ si et seulement si $\#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{R}_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{N}_i\})$ ;*
– *$(a, b) \in \mathcal{N}$ si et seulement si $\#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{N}_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in \mathcal{R}_i\})$ ;*
– *$(a, b) \in \mathcal{I}$ sinon.*

La proposition suivante établit que le PAF de majorité d'un profil d'AFs sur un même ensemble d'arguments est un PAF :

**Proposition 33** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs sur le même ensemble $\mathcal{A}$ d'arguments. Le PAF de majorité $MP(\mathcal{P})$ de $\mathcal{P}$ est un PAF.*

**Exemple 20** *Considérons $\mathsf{AF}_1 = \langle \{a, b, c\}, \{(a, b), (b, c), (a, c)\} \rangle$, et $\mathsf{AF}_2 = \langle \{a, b, c\}, \{(a, b), (b, a), (a, c)\} \rangle$.*

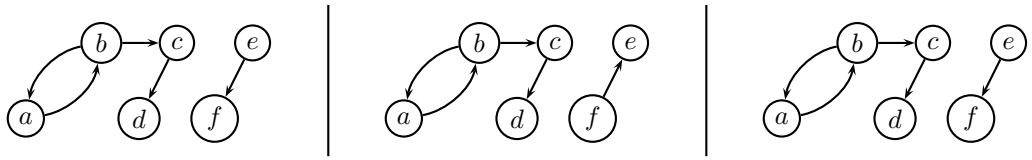



*On a $MP(\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle) = \langle \{a, b, c\}, \{(a, b), (a, c)\}, \{(b, c), (b, a)\}, \{(a, a), (b, b), (c, c), (c, a), (c, b)\} \rangle$.*

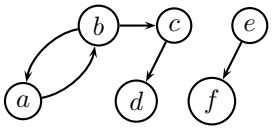
**Proposition 34** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs sur le même ensemble $\mathcal{A}$ d'arguments. $\Delta^{\Sigma}_{de}(\mathcal{P}) = \mathcal{C}(MP(\mathcal{P}))$.*

La proposition précédente est illustrée par l'exemple 14 page 64 :

**Exemple 14 page 64 (suite)** *Les expansions consensuelles de $\mathsf{AF}_1$, $\mathsf{AF}_2$ et $\mathsf{AF}_3$ sont respectivement :*




*Donc, le PAF de majorité de $\langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ est :*




[21] Pour tout ensemble $S$, $\#(S)$ dénote la cardinalité de $S$.

*Si on utilise la distance d'édition et la somme comme fonction d'agrégation, ce PAF représente aussi le résultat de la fusion dans le sens où cette fusion fournit l'ensemble des complétions de ce PAF.*

Le calcul du PAF de majorité d'un profil d'AFs sur un même ensemble d'arguments consiste donc à faire *voter sur les relations d'attaque* associées à chaque AF. Rappelons que nous avons déjà signalé, en section 3.2.4.1 page 64, que cette approche nous paraissait mieux adaptée qu'un vote effectué directement sur les ensembles d'arguments acceptables de chaque agent.

#### 3.2.4.5 Acceptabilité pour un profil d'AFs

À partir d'un profil d'AFs (éventuellement sur des ensembles d'arguments différents), un opérateur de fusion produit un nouvel ensemble d'AFs (mais cette fois, sur le même ensemble d'arguments) qui sont les meilleurs candidats possibles pour représenter les AFs du groupe (ils symbolisent donc une sorte de "consensus").

Il y a une différence importante du point de vue épistémique entre ces deux ensembles d'AFs : le premier reflète les différents points de vue donnés par les agents (et il peut arriver que deux agents aient le même point de vue), alors que le second exprime une certaine incertitude sur la fusion due à la présence de conflits.

Rappelons que ce travail porte principalement sur la caractérisation des ensembles d'arguments acceptables par le groupe d'agents. Pour cela, il reste donc à définir les mécanismes permettant d'exploiter le résultat de la fusion. Il faut donc définir une notion d'acceptabilité collective issue d'un ensemble d'AFs.

**Définition 57 (Acceptabilité collective issue d'un ensemble d'AFs)** *Une* relation d'acceptabilité collective pour un profil $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *d'AFs, dénotée par* $Acc_{\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle}$*, est une fonction de* $2^{\bigcup_i \mathcal{A}_i}$ *dans* $\{vrai, faux\}$ *qui associe à chaque sous-ensemble* $E$ *de* $\bigcup_i \mathcal{A}_i$ *la valeur vrai si* $E$ *est collectivement acceptable pour* $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *et la valeur faux sinon.*

Par exemple, une relation d'acceptabilité collective pour un profil $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ peut être définie par les relations d'acceptabilité *locales* $Acc_{\mathsf{AF}_i}$ (basées elles-mêmes sur certaines sémantiques ou certains principes de sélection), qui peuvent coïncider ou pas pour certains $\mathsf{AF}_i$, et une méthode de vote $V : \{vrai, faux\}^n \mapsto \{vrai, faux\}$ :

$$Acc_{\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle}(E) = V(Acc_{\mathsf{AF}_1}(E), \ldots, Acc_{\mathsf{AF}_n}(E)).$$

Voici quelques exemples possibles de la définition 57 basés sur des méthodes de vote classiques :

**Définition 58 (Acceptabilités pour des profils d'AFs)** *Soit* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *un profil de* $n$ *AFs sur le même ensemble d'arguments* $\mathcal{A}$*. Soit* $Acc_{\mathsf{AF}_i}$ *la relation d'acceptabilité associée à* $\mathsf{AF}_i$*. Si* $n = 1$*, alors* $Acc_{\langle \mathsf{AF}_1 \rangle} = Acc_{\mathsf{AF}_1}$*. Sinon, pour tout sous-ensemble* $S$ *de* $\mathcal{A}$*, on a :*

- $S$ *est* sceptiquement collectivement acceptable pour $\mathcal{P}$ *si et seulement si* $S$ *est inclus dans au moins un ensemble acceptable pour chaque* $\mathsf{AF}_i$ *:*
$$\forall \mathsf{AF}_i \in \mathcal{P}, \exists E_i \text{ tel que } Acc_{\mathsf{AF}_i}(E_i) = vrai \text{ et } S \subseteq E_i.$$
- $S$ *est* crédulement collectivement acceptable pour $\mathcal{P}$ *si et seulement si* $S$ *est inclus dans au moins un ensemble acceptable pour au moins un* $\mathsf{AF}_i$ *:*
$$\exists \mathsf{AF}_i \in \mathcal{P}, \exists E_i \text{ tel que } Acc_{\mathsf{AF}_i}(E_i) = vrai \text{ et } S \subseteq E_i.$$
- $S$ *est* collectivement acceptable à la majorité pour $\mathcal{P}$ *si et seulement si* $S$ *est inclus dans au moins un ensemble acceptable pour au moins une majorité faible d'*$\mathsf{AF}_i$ *:*
$$\#(\{\mathsf{AF}_i \mid \exists E_i \text{ tel que } Acc_{\mathsf{AF}_i}(E_i) = vrai \text{ et } S \subseteq E_i\}) \geq \tfrac{n}{2}.$$

Évidemment quand les relations d'acceptabilité locales $Acc_{\mathsf{AF}_i}$ ne sont pas triviales (*i.e.*, équivalentes à la fonction constante *faux*, on peut dire que tout ensemble d'arguments accepté sceptiquement collectivement est aussi accepté collectivement à la majorité et que tout ensemble accepté collectivement à la majorité est aussi accepté crédulement collectivement.

Remarquons aussi que le fait de choisir un principe (sceptique, crédule, à la majorité) pour définir l'acceptabilité collective d'un profil n'implique pas qu'il faille utiliser ce même principe pour définir les relations d'acceptabilité locales $Acc_{\mathsf{AF}_i}$.

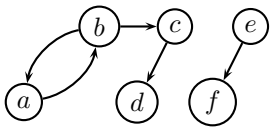
Nous allons maintenant illustrer ces idées en reprenant quelques-uns des exemples déjà utilisés ici sur lesquels on utilisera la sémantique préférée de Dung et le principe crédule pour définir les acceptabilités locales :

**Exemple 17 page 71 (suite)** *Avec la distance d'édition et* $\otimes = \mathcal{L}\texttt{eximax}$ *(ou* $\mathcal{M}\texttt{ax}$*) comme fonction d'agrégation, la fusion produit les deux AFs* $\mathsf{AF}'_1$ *et* $\mathsf{AF}'_2$*.*

*Si les relations d'acceptabilité locales sont basées sur l'inférence crédule à partir de la sémantique préférée, on a :*
– $Acc_{\mathsf{AF}'_1}(E) =$ *vrai si et seulement si* $E \subseteq \{c, d\}$ *ou* $E \subseteq \{b, e\}$ *;*
– $Acc_{\mathsf{AF}'_2}(E) =$ *vrai si et seulement si* $E \subseteq \{a, c, e\}$*.*
$\{c\}$ *et* $\{e\}$ *sont sceptiquement collectivement acceptables et* $\{b, e\}$,$\{c, d\}$ *et* $\{a, c, e\}$ *(et leurs sous-ensembles) sont crédulement (et à la majorité) collectivement acceptables.*

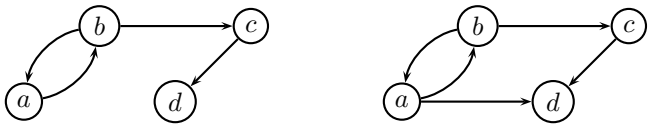
*Avec cette méthode, l'argument* $a$ *peut encore être déduit crédulement, alors que cela n'est plus plus possible si on fait simplement l'union de* $\mathsf{AF}_1$ *et* $\mathsf{AF}_2$*.*

**Exemple 14 page 64 (suite)** *Avec la distance d'édition et la somme comme fonction d'agrégation, la fusion produit un seul AF noté* $\mathsf{AF}$ *:*




$\mathsf{AF}$ *possède deux extensions préférées :* $\{a, c, e\}$ *et* $\{b, d, e\}$*. Donc,* $Acc_{\mathsf{AF}}(E) =$ *vrai si et seulement si* $E \subseteq \{a, c, e\}$ *ou* $E \subseteq \{b, d, e\}$*. Les trois relations d'acceptabilité collectives coïncident ici (puisqu'il n'y a qu'un seul AF dans le résultat de la fusion). Les ensembles* $\{a, c, e\}$ *et* $\{b, d, e\}$ *(et leurs sous-ensembles) sont crédulement, sceptiquement et à la majorité, collectivement acceptables ; ce résultat est plus intéressant que celui qu'on aurait obtenu en appliquant une méthode de vote sur les arguments déduits par chaque agent (voir section 3.2.4.1 page 64).*

**Exemple 16 page 67 (suite)** *Avec la distance d'édition et la somme comme fonction d'agrégation, la fusion produit les deux AFs suivants :*




*Les extensions préférées pour ces 2 AFs coïncident (*$\{a, c\}$ *et* $\{b, d\}$*). Du coup, les trois relations d'acceptabilité collectives coïncident aussi. Ainsi, les ensembles* $\{a, c\}$ *et* $\{b, d\}$ *(et leurs sous-ensembles) sont sceptiquement, crédulement et à la majorité collectivement acceptables.*

Il est intéressant de comparer la relation d'acceptabilité collective pour le profil initial $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ avec celle obtenue pour le résultat de la fusion $\Delta_d^{\otimes}(\mathcal{P})$. Il n'est pas surprenant de constater que ces deux relations sont différentes et n'ont pas de connexion logique (*i.e.,* aucune implique l'autre), même dans le cas où ces deux relations sont basées sur les mêmes notions d'acceptabilité locale (par exemple, en considérant qu'un ensemble d'arguments $E$ est acceptable pour un AF s'il est inclus dans au moins une des extensions préférées) et la même méthode de vote (par exemple, la règle de la majorité simple).

Ainsi, il peut arriver que certains ensembles soient acceptables après la fusion, alors qu'ils ne l'étaient pas au départ :

**Proposition 35** *Soit $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ un profil d'AFs sur un même ensemble d'arguments $\mathcal{A}$. L'ensemble de tous les ensembles collectivement acceptables pour le profil $\mathcal{P}$ n'est pas nécessairement égal à l'ensemble de tous les ensembles acceptables pour la fusion de $\mathcal{P}$.*

Un contre-exemple est donné par l'exemple 16 page 67.

Quand chaque relation d'acceptabilité locale correspond exactement à l'acceptabilité collective proposée par Dung (pour une sémantique donnée et $\forall \mathsf{AF}_i$, $Acc_{\mathsf{AF}_i}(E) = vrai$ si et seulement si $E$ est une extension de $\mathsf{AF}_i$ pour cette sémantique), on peut faire les remarques suivantes :

– Si un ensemble d'arguments est inclus dans *un* des ensembles acceptables d'un agent, il ne sera pas nécessairement inclus dans un des ensembles acceptables d'un des AFs issus de la fusion (et ceci est aussi vérifié pour les singletons). L'inverse est aussi vérifié.
– D'autre part, et c'est plus surprenant, même si un ensemble d'arguments est inclus dans *chacun* des ensembles acceptables pour un agent, cela ne garantit pas qu'il soit inclus dans un ensemble acceptable de l'un des AFs issus de la fusion. De même, si un ensemble d'arguments est inclus dans chaque ensemble acceptable des AFs issus de la fusion, cela ne garantit pas qu'il ait été inclus dans un ensemble acceptable pour un des agents. Intuitivement, cela s'explique par le fait que si un argument est accepté par tous les agents *pour de mauvaises raisons* (par exemple, par manque d'information pour calculer les attaques le concernant), il pourra être rejeté par le groupe après la fusion. Plus formellement, cela provient du fait que rien n'assure qu'un argument donné par un agent sera accepté par le groupe et du fait que le processus d'acceptabilité est non monotone (dans le sens où l'ajout d'une simple attaque $(a, b)$ dans un AF peut modifier complètement les extensions de cet AF).

#### 3.2.4.6 Conclusion et perspectives

Nous avons présenté un cadre pour inférer des informations à partir d'une collection de systèmes d'argumentation *à la* Dung. Cette approche consiste à fusionner ces AFs. Le cadre proposé est suffisamment général pour autoriser plusieurs scénarios. On ne suppose pas que les agents utilisent le même ensemble d'arguments, ni le même type de relation d'attaque (la notion d'attaque peut être différente d'un agent à l'autre, soit parce que les agents n'ont pas le même point de vue sur la manière dont des arguments donnés interagissent, soit parce qu'ils ne sont pas d'accord sur ce que peuvent être ces interactions). À chaque agent, on associe une fonction d'expansion spécifique qui lui permet de prendre en compte les nouveaux arguments. Il existe aussi plusieurs notions de distances entre PAFs et différentes fonctions d'agrégation permettant de définir les AFs qui représentent le mieux le point de vue du groupe.

Nous avons illustré cette approche sur plusieurs exemples en montrant en particulier qu'elle fournit des résultats paraissant intuitivement meilleurs que si l'on faisait soit l'union des AFs initiaux, soit un vote direct sur ces AFs. Nous avons aussi donné quelques propriétés intéressantes de ces opérateurs de fusion (en particulier de ceux reposant sur la distance d'édition). En enfin, nous avons montré que les résultats fournis par la fusion peuvent être totalement différents de ceux qu'on aurait pu déduire des AFs initiaux.

Ce travail pourrait se poursuivre suivant différents axes :

**Fusion de PAFs.** Notre cadre pourrait être généralisé pour permettre la fusion de PAFs et plus seulement d'AFs. Cela nous permettrait de prendre en compte des agents avec des états de croyance incomplets du point de vue du calcul de la relation d'attaque. Les expansions des PAFs peuvent être définies de manière similaire à celles des AFs (ce qui change c'est la gestion de l'ignorance). Comme les PAFs sont plus expressifs que les AFs, il serait alors intéressant de définir une notion d'acceptabilité directement sur un PAF.

**Notion de force d'une attaque.** Supposons qu'à chaque attaque connue par l'agent $i$ soit associée une valeur numérique qui reflète la force de cette attaque d'après l'agent, *i.e.*, le degré avec lequel l'agent $i$ croit que $a$ attaque $b$. Cette nouvelle information pourra facilement être prise en compte en modifiant légèrement la notion de distance d'édition. Une autre possibilité concernant la force d'une attaque serait de générer une telle force à partir d'un degré d'accord (ou de désaccord) dans le groupe. Par exemple, chaque attaque $(a, b)$ dans le PAF de majorité d'un profil $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ pourrait être étiquetée par le ratio $\frac{\#(\{i \in 1 \ldots n | (a,b) \in \mathcal{R}_i\})}{n}$ et idem pour la relation de non-attaque (cela revient à considérer que ces deux relations d'attaque et de non-attaque sont des relations floues). Il resterait alors à définir une notion

d'acceptabilité pour ce type de PAF. Cette notion d'attaque "graduelle" existe déjà dans un autre contexte (voir les travaux [RMF+08]).

**Fusion d'auditoires.** Dans [BCDD07], une extension de la notion d'AF, appelée AF valué (VAF), a été proposée pour prendre en compte les valeurs représentant les préférences d'un agent pour un auditoire donné. Il serait intéressant d'étudier la fusion de ces VAFs.

### 3.2.5 Révision en argumentation unipolaire

Quand un agent reçoit une nouvelle information, il doit adapter ses croyances ; cette adaptation n'est pas toujours facile car elle peut impliquer la remise en cause de connaissances considérées jusqu'alors comme acquises. Ce problème est un problème classique qui a déjà été beaucoup étudié en IA en particulier en raisonnement non monotone. Parmi les premiers travaux sur ce sujet, on trouve le travail de Alchourrón, Gärdenfors and Makinson (AGM) [AGM85] qui avait proposé un cadre formel pour raisonner sur le changement de croyances et qui avait introduit la notion de "révision de croyances". La révision de croyances consiste à définir comment choisir les croyances à conserver quand arrive une nouvelle croyance. Ici, nous avons étendu cette notion au cadre de l'argumentation en cherchant à savoir comment gérer l'arrivée d'un nouvel argument.

Quand un nouvel argument apparaît, avec éventuellement de nouvelles interactions entre cet argument et les arguments déjà dans le système, les "sorties" du système, c'est-à-dire les ensembles d'arguments acceptables (extensions), peuvent changer du tout au tout. Le but de cette étude est donc d'analyser l'impact de ces ajouts (argument et interaction) sur l'ensemble initial d'extensions. Nous avons ainsi caractériser divers types de révision. Cette étude a aussi au moins deux applications possibles, une pour l'aspect calculatoire des extensions (quels sont les changements sur les extensions à effectuer suivant les modifications du système d'argumentation ? Peut-on éviter ces changements ?) et l'autre pour l'aspect stratégique des dialogues utilisant l'argumentation (comment rendre un dialogue plus "ouvert" ?).

La sous-section 3.2.5.1 rappelle très brièvement la notion de révision des croyances, puis on donne la définition d'une révision en argumentation (section 3.2.5.2 page suivante) sachant que l'on se restreint ici à un cas simple : ajout d'un *seul* argument et d'une *seule* interaction. Dans ce cadre, une typologie de différents types de révision est proposée ainsi que quelques propriétés intéressantes (voir section 3.2.5.3 page 84). Je développe alors deux de ces types de révision en section 3.2.5.4 page 86, les révisions "décisive" et "expansive".

Les résultats présentés dans cette section ont été obtenus en collaboration avec Claudette CAYROL et Florence DUPIN DE SAINT-CYR et publiés dans [CdSCLS08] (le lecteur y trouvera les preuves des propriétés évoquées ici).

#### 3.2.5.1 Concepts de base en théorie de la révision

La notion de "révision de croyances" a été introduite dans le papier de AGM [AGM85]. Il s'agit de savoir quoi garder parmi les anciennes croyances quand arrive une nouvelle croyance, bien-sûr contradictoire avec les anciennes. Les croyances sont représentées par des formules. AGM considère qu'il existe trois types de changement, la contraction, l'expansion et la révision. L'expansion consiste juste à ajouter la nouvelle information sans vérifier sa consistance avec les anciennes croyances. La contraction permet de supprimer une information. La révision consiste à ajouter la nouvelle information en préservant la consistance. Le principal intérêt du travail d'AGM est la définition d'un ensemble de postulats décrivant ce que doit être un opérateur de révision "rationnel". Dans [Som94] une analyse de ces principes a montré qu'ils reposaient sur trois principes :
– un principe de consistance (le résultat doit être consistant),
– un principe de changement minimal (il faut modifier le moins possible de croyances),
– la priorité à la nouvelle information (cette nouvelle information doit être vérifiée après la révision).
Formellement, un opérateur de révision associe à un ensemble de formules clos pour la déduction $K$ (encodant les croyances initiales[22]) et à une formule $p$ (encodant la nouvelle information), un autre ensemble de croyances noté $K * p$. Pour que cet opérateur noté $*$ soit "rationnel", il devra satisfaire les postulats d'AGM :

**K* 1** $K * p = Th(K * p)$.

[22] Si $BC$ est un ensemble de formules encodant ces croyances alors $K = Th(BC)$ avec $Th$ l'opérateur de fermeture pour la déduction logique.

**K* 2** $p \in K * p$.

**K* 3** $K * p \subseteq Th(K \cup \{p\})$.

**K* 4** Si $\neg p \notin K$, alors $Th(K \cup \{p\}) \subseteq K * p$.

**K* 5** $\bot \in K * p$ si et seulement si $p \leftrightarrow \bot$.

**K* 6** Si $p \leftrightarrow q$ alors $K * p = K * q$.

**K* 7** $K * (p \wedge q) \subseteq Th((K * p) \cup \{q\})$.

**K* 8** Si $\neg q \notin K * p$ alors $Th((K * p) \cup \{q\}) \subseteq K * (p \wedge q)$.

**K* 1** impose que le résultat de la révision soit clos pour la déduction logique. **K* 2** assure que la nouvelle information appartiendra aux croyances révisées. **K* 3** implique que les croyances après révision ne sont que celles déductibles de l'union de $K$ avec $p$. **K* 4** associé à **K* 3** signifie que si la nouvelle information n'est pas contradictoire avec les anciennes croyances alors la révision n'est rien d'autre d'une expansion. **K* 5** dit que le résultat de la révision n'est inconsistant que si la nouvelle information était elle-même inconsistante. **K* 6** exprime le fait que la révision est indépendante de la syntaxe. Ces six premiers postulats sont appelés postulats de base alors que les deux derniers expriment la minimalité du changement. **K* 7** implique que réviser par une conjonction $p \wedge q$ n'apporte pas plus d'information que de réviser d'abord $K$ par $p$, puis de réviser le résultat par $q$. **K* 8** signifie que, quand on révise $K$ par $p \wedge q$, chaque déduction logique de $q$ et $K * p$ est conservée à condition que $q$ ne soit pas contradictoire avec $K * p$.

Remarquons que dans la suite, nous limiterons notre étude au postulat **K* 2**. Et nous appellerons "classique" un opérateur qui satisfera **K*2**. On verra quand même dans la suite que ce postulat n'est pas toujours très bien adapté au cas de la révision en argumentation (voir section 3.2.5.3 page 84).

Un dernier point concernant la théorie de la révision est la différence entre "révision de croyances" et "mise à jour de croyances" (cette dernière notion a été introduite par [Win88]). Cette distinction repose sur la nature de la nouvelle information : soit elle complète la connaissance de l'agent concernant du monde, soit elle informa cet agent d'une modification de ce monde. Plus précisément, la mise à jour est un processus qui rend compte d'une évolution physique du système alors que la révision rend compte d'une évolution épistémique (c'est la connaissance que l'on a du monde qui change). Dans le cadre de l'argumentation, nous pensons être plutôt dans le cas d'une révision : l'agent n'était pas conscient d'un argument qui apparaît soudainement ; c'est plutôt la marque que le monde n'a pas changé mais que c'est la conscience que l'agent a de ce monde qui a évoluée.

#### 3.2.5.2 Révision en argumentation

Tout d'abord, nous allons donner une définition formelle de la notion de révision en argumentation en se basant sur les "sorties" du système d'argumentation (les extensions pour une sémantique donnée). Nous proposerons alors une typologie des différents types de révision.

##### 3.2.5.2.1 Définition

Informellement, une révision se produit lorsqu'un nouvel argument apparaît. Si cet argument n'est pas connecté aux autre arguments de $\langle \mathcal{A}, \mathcal{R} \rangle$, la révision est alors triviale : cet argument est ajouté à chaque extension. Par contre, le processus de révision est beaucoup plus intéressant lorsque cet argument est ajouté à $\langle \mathcal{A}, \mathcal{R} \rangle$ avec au moins une interaction le liant aux autres arguments. Dans ce travail, nous nous sommes donc intéressées à l'ajout d'un *seul* argument $Z$ et d'une *seule* interaction, $Z\mathcal{R}X$ ou $X\mathcal{R}Z$, avec $X$ appartenant à $\mathcal{A}$.

Dans la suite, nous identifierons un système d'argumentation $\langle \mathcal{A}, \mathcal{R} \rangle$ à son graphe d'attaque $\mathcal{G}$ et nous écrirons $X \in \mathcal{G}$ au lieu de "$X$ est un argument représenté par un nœud de $\mathcal{G}$". L'ensemble des extensions de $\langle \mathcal{A}, \mathcal{R} \rangle$ est noté $\mathcal{E}$ (avec $E_1$, …, $E_n$ dénotant les extensions).

Réviser $\langle \mathcal{A}, \mathcal{R} \rangle$ consiste à ajouter un argument $Z$ qui attaque (ou est attaqué par) un argument $X$ de $\mathcal{A}$. La révision produit un nouveau système d'argumentation représenté par le graphe $\mathcal{G}'$ et un nouvel ensemble d'extensions $\mathcal{E}'$ (avec $E'_1$, …, $E'_p$ dénotant les nouvelles extensions).

$$(\mathcal{G},\mathcal{E}) \xrightarrow[i=(Z,X) \text{ ou } i=(X,Z)]{\text{révision avec } Z \text{ et } i} (\mathcal{G}',\mathcal{E}')$$

**Définition 59** *Soit $\mathcal{G}$ un graphe d'attaque. Soit $s$ une sémantique. Soit $X \in \mathcal{G}$, $Z \notin \mathcal{G}$ et $i$ une paire d'arguments composée de $Z$ et de $X$ (soit $(X,Z)$, soit $(Z,X)$). Soit $\mathcal{G}'$ le graphe obtenu à partir de $\mathcal{G}$ en ajoutant le sommet $Z$ et l'arc $i$. L'*opérateur de révision $\Theta$ *associe à $(Z,i,\mathcal{G},s)$ l'ensemble $\mathcal{E}'$ qui est l'ensemble des extensions de $\mathcal{G}'$ pour la sémantique $s$.*

Dans le contexte de la définition 59, la propriété suivante énumère quelques résultats de base utiles pour la suite :

**Proposition 36**
– *Si la nouvelle interaction est $(Z,X)$, alors $Z$ n'est pas attaqué dans $\mathcal{G}'$.*
– *Si la nouvelle interaction est $(X,Z)$, alors $Z$ n'attaque aucun argument de $\mathcal{G}'$.*
– *Le processus de révision n'introduit aucun circuit dans $\mathcal{G}'$.*

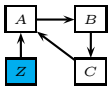
La révision d'un AF peut modifier l'ensemble des extensions. Suivant la sémantique choisie pour l'acceptabilité cette modification peut être plus ou moins importante. Cela dépend aussi du type de l'interaction qui a été ajoutée ainsi que du statut de l'argument $X$ concerné par cette interaction. Dans la section suivante, nous proposons une typologie de divers types de révision suivant la manière dont l'ensemble des extensions est modifié. L'étape suivante consistera à caractériser chaque type de révision en imposant des conditions sur l'interaction $i$.

#### 3.2.5.2.2 Typologie de révisions

Soit $\langle \mathcal{A},\mathcal{R}\rangle$ un système d'argumentation et $\mathcal{E}$ l'ensemble des extensions de $\langle \mathcal{A},\mathcal{R}\rangle$ sous une sémantique donnée $s$. Différentes situations peuvent se rencontrer dans le cas général. $\mathcal{E}$ peut être vide (ce qui se produit quand $s$ est la sémantique stable), peut se réduire à un singleton $\{E_1\}$ (avec $E_1$ pouvant être vide), ou peut contenir plus d'une extension $\{E_1,\ldots,E_n\}$. La situation correspondant au cas d'une seule extension non vide est particulièrement intéressante pour le statut d'un argument. Par opposition, quand il existe plusieurs extensions, il existe aussi plusieurs choix possibles. Nous allons tout d'abord considérer les révisions telles que $\mathcal{G}'$ possède une extension unique non vide, alors que cela n'est pas le cas pour $\mathcal{G}$. Une telle révision est dite **décisive**.

**Exemple 21**

1. *Sous la sémantique stable, avec $i = (Z,A)$*

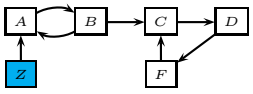



*Avant la révision $\mathcal{E} = \varnothing$,*
*après la révision $\mathcal{E}' = \{\{Z,B\}\}$*

2. *Sous la sémantique basique, avec $i = (Z,A)$*

$\mathcal{E} = \{\{\}\}$, $\mathcal{E}' = \{\{Z,B\}\}$

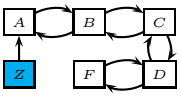
3. *Sous la sémantique préférée, avec $i = (Z,A)$*




$\mathcal{E} = \{\{A\},\{B,D\}\}$,
$\mathcal{E}' = \{\{Z,B,D\}\}$

Une contrainte plus faible serait une diminution des choix. Une révision permettant à $\mathcal{G}'$ de posséder moins d'extensions que $\mathcal{G}$, mais au moins deux, est dite **sélective**. Remarquons que la révision sélective n'a pas de sens sous la sémantique basique, puisqu'il n'existe qu'une seule extension basique.

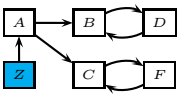
**Exemple 22** *Sous la sémantique préférée (ou stable), avec $i = (Z,A)$*




$\mathcal{E} = \{\{A, C, F\}, \{A, D\}, \{B, D\}, \{B, F\}\}$,
$\mathcal{E}' = \{\{Z, C, F\}, \{Z, B, D\}, \{Z, B, F\}\}$

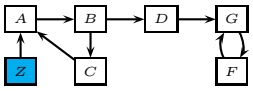
Un point de vue opposé serait de considérer les révisions qui augmentent l'ambiguïté, en augmentant le nombre d'extensions. C'est par exemple le cas quand $\mathcal{G}$ possède une extension non vide et que $\mathcal{G}'$ possède strictement plus d'extensions que $\mathcal{G}$. Une situation assez proche correspond au cas où $\mathcal{G}$ n'a pas d'extension ou bien une extension vide, alors que $\mathcal{G}'$ possède plus d'une extension. Dans ce cas, la révision apporte quelque information, mais pas de manière décisive. De telles révisions sont dites **ambiguës**. Tout comme la révision sélective, la révision ambiguë n'a pas de sens sous la sémantique basique.

**Exemple 23** *Sous la sémantique préférée (ou stable), avec* $i = (Z, A)$




$\mathcal{E} = \{\{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, B, C\}, \{Z, B, F\}, \{Z, D, C\}, \{Z, D, F\}\}$

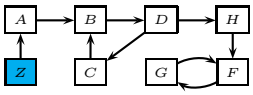
*Sous la sémantique stable, avec* $i = (Z, A)$




$\mathcal{E} = \varnothing$,
$\mathcal{E}' = \{\{Z, B, F\}, \{Z, B, G\}\}$

Dans la même optique, nous considérons les révisions supprimant chaque extension, ce qui nous amène à une sorte d'impasse décisionnelle. Une révision telle que $\mathcal{G}'$ ne possède aucune extension, alors que $\mathcal{G}$ en avait au moins une, est dite **destructrice**. Remarquons qu'une révision destructrice n'a que sens que sous la sémantique stable.

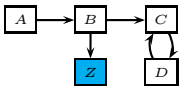
**Exemple 24** *Sous la sémantique stable, avec* $i = (Z, A)$




$\mathcal{E} = \{\{A, D, F\}, \{A, D, G\}\}$,
$\mathcal{E}' = \varnothing$

Jusqu'alors nous n'avons considéré que des révisions qui ont un impact sur le nombre d'extensions. On peut aussi s'intéresser aux révisions qui modifient non plus le nombre d'extensions mais le contenu des extensions. La situation la plus intéressante correspond à des révisions par lesquelles les extensions résultat de $\mathcal{G}'$ contiennent les extensions initiales de $\mathcal{G}$, le nombre d'extensions restant le même. De telles révisions sont dites **expansives**.

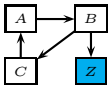
**Exemple 25** *Sous la sémantique préférée (ou stable), avec* $i = (B, Z)$




$\mathcal{E} = \{\{A, C\}, \{A, D\}\}$,
$\mathcal{E}' = \{\{Z, A, C\}, \{Z, A, D\}\}$

Quand rien n'est modifié, c'est-à-dire que $\mathcal{E} = \mathcal{E}'$, la révision est dite **conservative**.

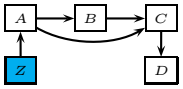
**Exemple 26** *Sous la sémantique préférée, avec* $i = (B, Z)$




$\mathcal{E} = \{\{\}\}, \mathcal{E}' = \{\{\}\}$

Et sinon, il peut arriver que certaines extensions (et parfois toutes) sont modifiées. Ce type de révision est dit **modifiante**. Cela est aussi le cas quand chaque extension de $\mathcal{G}'$ possède une intersection non vide avec une extension de $\mathcal{G}$ mais qu'elle ne la contient pas.

**Exemple 27** *Sous la sémantique basique, avec* $i = (Z, A)$




$\mathcal{E} = \{\{A, D\}, \mathcal{E}' = \{\{Z, B, D\}\}$

Toutes ces révisions sont récapitulées dans le tableau suivant.

| $\mathcal{E}' =$ <br> $\mathcal{E} =$ | $\varnothing$ | $\{\{\}\}$ | $\{E'_1\}$ | $\{E'_1, \ldots, E'_p\}$ <br> $p \geq 2$ |
|---|---|---|---|---|
| $\varnothing$ | conservative | #1 | décisive | ambiguë |
| $\{\{\}\}$ | #2 | conservative | | |
| $\{E_1\}$ | destructrice | #3 | conservative <br> expansive <br> modifiante | ambiguë |
| $\{E_1,$ <br> $\ldots,$ <br> $E_n\}$ <br> $n \geq 2$ | | #4 | décisive | $n < p$ : <br> ambiguë <br> $n > p$ : <br> sélective <br> $n = p$ : <br> conservative <br> expansive <br> modifiante |

Avec $E_i \neq \varnothing$ et $E'_i \neq \varnothing$. Chaque case de la table donne le nom de la révision correspondante. Les cases contenant #i correspondent à des situations impossibles :

**#1 et #2** La seule sémantique pour l'acceptabilité avec laquelle un AF pourrait ne pas avoir d'extension est la sémantique stable. Toutefois, avec la sémantique stable, un AF ne peut pas avoir une extension vide quand l'ensemble des arguments n'est pas vide. Et, par hypothèse, les cases #1 et #2 correspondent à des AFs non vides (parce qu'il existe au moins un $X$ appartenant à $\mathcal{G}$ et $X$ et $Z$ appartiennent à $\mathcal{G}'$). Donc ces cas ne peuvent pas se produire dans le cadre des sémantiques pour l'acceptabilité choisies dans ce travail.

**#3** Sous la sémantique stable, ce cas ne peut se produire pour les mêmes raisons que les cas #1 et #2.

Sous la sémantique basique, comme $\mathcal{G}$ possède une seule extension, il existe au moins un argument non-attaqué $W$ ; donc, si l'interaction ajoutée est $(X, Z)$, $W$ est toujours non-attaqué et $\mathcal{G}'$ possède toujours une extension non vide ; et, si l'interaction ajoutée est $(Z, X)$, alors $Z$ est non-attaqué et appartient à l'extension basique de $\mathcal{G}'$ ; donc, $\mathcal{G}'$ ne peut pas avoir une extension vide.

Sous la sémantique préférée, si l'interaction ajoutée est $(Z, X)$, $Z$ est non-attaqué et appartient aux extensions préférées de $\mathcal{G}'$ ; donc, ces extensions ne sont pas vides. Et, si l'interaction ajoutée est $(X, Z)$, alors $Z$ n'attaque aucun des arguments de $E_1$ ; donc ces arguments appartiennent toujours à une extension préférée de $\mathcal{G}'$ et les extensions préférées ne sont pas vides.

En conclusion, ce cas ne peut ne produire avec aucune des sémantiques de l'acceptabilité choisie dans ce travail.

**#4** Ce cas ne pourrait apparaître qu'avec la sémantique préférée (parce qu'avec la sémantique basique, il n'existe qu'une seule extension et avec la stable une extension ne peut être vide que si l'ensemble des arguments est vide). Si l'interaction ajoutée est $(Z, X)$, $Z$ peut "supprimer" ou "créer" des extensions, mais il appartient à chacune d'entre elles (parce qu'il est non-attaqué), donc $\mathcal{G}'$ ne peut pas avoir une extension vide. Et si l'interaction ajoutée est $(X, Z)$, $Z$ n'attaque aucun des arguments de $E_i, \forall i$, donc ces arguments appartiennent aux extensions préférées de $\mathcal{G}'$ et $\mathcal{G}'$ ne peut pas avoir une extension vide. Ce cas est donc impossible.

#### 3.2.5.3 La révision classique en argumentation

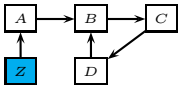
Réviser une base de connaissances consiste à changer les croyances de manière minimale tout en considérant la nouvelle information comme "prioritaire" (suivant en cela le postulat AGM **K* 2**). Toutefois, les opérateurs de révision définis dans la section précédente n'assurent nullement que le nouvel argument sera accepté dans les nouvelles extensions du graphe. Dans cette section, nous allons donc caractériser cette propriété (dite "classique") pour le cas de la révision d'un système d'argumentation.

**Définition 60** *La révision $\Theta$ est* classique *si et seulement si $\mathcal{G}'$ possède au moins une extension et l'argument ajouté $Z$ appartient à chaque extension de $\mathcal{G}'$.*

**Proposition 37** *Si l'interaction ajoutée est $(Z, X)$, alors la révision sera classique sous la sémantique basique et la sémantique préférée.*

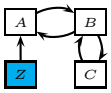
*De plus, si $\mathcal{G}$ ne possède aucun circuit de longueur impaire alors la révision est aussi classique pour la sémantique stable.*

**Exemple 28** *Sous la sémantique basique :*

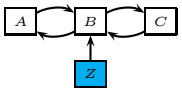



$\mathcal{E} = \{\{A, C\}\}, \mathcal{E}' = \{\{Z\}\}$

**Exemple 29** *Calculons les extensions préférées des deux graphes suivants :*

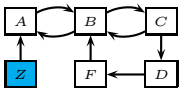



$\mathcal{E} = \{\{A, C\}, \{B\}\}$,
$\mathcal{E}' = \{\{Z, B\}, \{Z, C\}\}$




$\mathcal{E} = \{\{A, C\}, \{B\}\}$,
$\mathcal{E}' = \{\{Z, A, C\}\}$

**Exemple 30** *Sous la sémantique stable :*




$\mathcal{E} = \{\{A, C, F\}, \{B, D\}\}$,
$\mathcal{E}' = \{\{Z, C, F\}, \{Z, B, D\}\}$

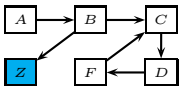
La condition disant que $\mathcal{G}$ ne doit pas avoir un circuit de longueur impaire assure l'existence d'au moins une extension stable après l'ajout de $Z$. C'est une condition suffisante mais pas nécessaire.

**Exemple 28 (suite)** *Avant la révision, l'extension stable est $\{A, C\}$, et après la révision il n'existe plus d'extension stable.*

**Proposition 38** *Si l'interaction ajoutée est $(X, Z)$ telle que $X$ soit attaquée par chaque extension de $\mathcal{G}$ alors la révision est classique sous les sémantiques basique et préférée.*

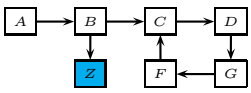
*De plus, si $\mathcal{G}$ possède au moins une extension stable alors la révision est aussi classique pour la sémantique stable.*

**Exemple 31** *Sous la sémantique basique :*




$\mathcal{E} = \{\{A\}\}, \mathcal{E}' = \{\{Z, A\}\}$

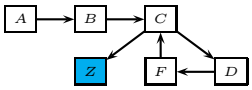
**Exemple 32** *Sous les sémantiques stable ou préférée :*




$\mathcal{E} = \{\{A, C, G\}, \{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, A, C, G\}, \{Z, A, D, F\}\}$

Remarquons que, sous la sémantique basique, $X$ doit être attaqué et que le fait que $X$ n'appartienne pas à la seule extension $E$ de $\mathcal{E}$ n'est pas suffisant pour assurer que la révision soit classique :

**Exemple 33** *Sous la sémantique basique :*




$E = E' = \{A\}$.
*Donc,* $\mathcal{E} = \mathcal{E}' = \{\{A\}\}$.

$C$ *n'appartient pas à* $E$*, et néanmoins* $Z$ *n'appartient pas à* $E'$.

Ainsi que cela a déjà été signalé, réviser un graphe par un argument (et une seule interaction) ne consiste pas toujours à accepter cet argument. La "classicalité" n'est donc pas la seule propriété intéressante à étudier pour les opérateurs de révision dans le cadre de l'argumentation.

#### 3.2.5.4 Étude de cas

Dans cette section, nous allons étudier deux cas : la révision décisive et la révision expansive. Dans le premier cas, après la révision, il ne reste qu'une seule extension (et il est donc facile de prendre une décision) ; dans le second cas, le nombre d'extensions est inchangé mais chaque nouvelle extension (après la révision) contient une des extensions initiales (avant la révision).

##### 3.2.5.4.1 Révision décisive

La révision décisive facilite la prise de décision : avant cette révision, dans $\mathcal{G}$, soit il n'y avait pas d'ensemble d'arguments acceptables, soit il y en avait trop (plus d'un), et après la révision il n'y a plus qu'un seul ensemble d'arguments acceptables dans $\mathcal{G}'$.

**Définition 61** *La révision* $\Theta$ *est* décisive *si et seulement si* $\Theta$ *appliquée à* $\mathcal{G}$*, avec* $\mathcal{E} = \varnothing$*, ou* $\mathcal{E} = \{\{\}\}$*, ou* $\mathcal{E} = \{E_1, \ldots, E_n\}$, $n \geq 2$*, fournit un nouveau graphe* $\mathcal{G}'$ *avec* $\mathcal{E}' = \{E'\}$, $E' \neq \varnothing$.

**Proposition 39** *Si une révision est décisive alors l'interaction ajoutée est de la forme* $(Z, X)$*. Une révision décisive est classique.*

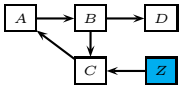
**Exemple 21 page 82 (suite)**

1. *Sous la sémantique stable, l'exemple 21. 1 page 82 illustre la révision décisive avec* $\mathcal{E} = \varnothing$ *et* $\mathcal{E}' = \{\{Z, B\}\}$.
2. *Sous la sémantique basique, l'exemple 21. 2 page 82 illustre la révision décisive avec* $\mathcal{E} = \{\{\}\}$ *et* $\mathcal{E}' = \{\{Z, B\}\}$.
3. *Sous la sémantique préférée, l'exemple 21. 3 page 82 illustre la révision décisive avec* $\mathcal{E} = \{\{A\}, \{B, D\}\}$ *et* $\mathcal{E}' = \{\{Z, B, D\}\}$.

**Théorème 3** *Sous la sémantique basique, si l'interaction ajoutée est* $(Z, X)$ *et* $\mathcal{E} = \{\{\}\}$*, alors la révision est décisive.*

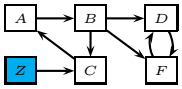
**Théorème 4** *Sous la sémantique préférée, si l'interaction ajoutée est* $(Z, X)$*, si* $\mathcal{E} = \{\{\}\}$ *et s'il n'existe pas de circuit de longueur paire dans* $\mathcal{G}$*, alors la révision est décisive.*

**Exemple 34** *Sous la sémantique préférée :*




$\mathcal{E} = \{\{\}\}, \mathcal{E}' = \{\{Z, A, D\}\}$

Remarquons que si des circuits de longueur paire existent dans le graphe, la révision peut produire plusieurs extensions ; ce sera donc une révision ambiguë :




$\mathcal{E} = \{\{\}\}$,
$\mathcal{E}' = \{\{Z, A, D\}, \{Z, A, F\}\}$

C'est pour cette raison que nous n'avons considéré que des graphes sans circuit de longueur paire dans le théorème 4 page précédente.

Remarquons aussi que nous n'avons pas trouvé de théorème de caractérisation dans le cas de la sémantique stable pour la révision décisive.

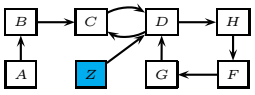
#### 3.2.5.4.2 Révision expansive

Une révision est dite "expansive" quand elle ne fait rien d'autre qu'ajouter des arguments dans les extensions déjà existantes.

**Définition 62** *La révision $\Theta$ est* expansive *si et seulement si $\mathcal{G}$ et $\mathcal{G}'$ possèdent le même nombre non nul d'extensions non vides et chaque extension de $\mathcal{G}'$ inclut strictement une extension de $\mathcal{G}$.*

**Proposition 40** *La révision expansive est classique.*

**Exemple 35** *Sous la sémantique basique :*

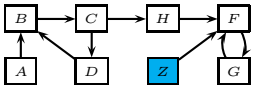



$\mathcal{E} = \{\{A\}\}$,
$\mathcal{E}' = \{\{Z, A, C, H, G\}\}$

L'exemple 25 page 83 illustre lui-aussi la révision expansive sous les sémantiques préférée et stable.

**Théorème 5** *Sous la sémantique basique avec $\mathcal{E} = \{E\}$, si l'interaction ajoutée est $(Z, X)$, si $X \notin E$ et si $E \neq \varnothing$, alors la révision est expansive.*

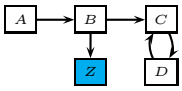
**Exemple 36** *Sous la sémantique basique :*




$\mathcal{E} = \{\{A, C\}\}$,
$\mathcal{E}' = \{\{Z, A, C, G\}\}$

**Théorème 6** *Sous la sémantique basique, si l'interaction ajoutée est $(X, Z)$, si $X \notin E$ et si $E$ attaque $X$, alors la révision est expansive et $\mathcal{E}' = \{E \cup \{Z\}\}$.*

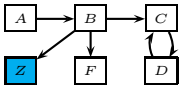
**Exemple 37** *Sous la sémantique basique :*




$\mathcal{E} = \{\{A\}\}, \mathcal{E}' = \{\{Z, A\}\}$

**Théorème 7** *Sous la sémantique stable, si l'interaction ajoutée est $(X, Z)$, si $\mathcal{E} \neq \varnothing$, et si $\forall i \geq 1$, $X \notin E_i$, alors la révision est expansive et $\forall i$, $E'_i = E_i \cup \{Z\}$.*

**Exemple 38** *Sous la sémantique stable :*




$\mathcal{E} = \{\{A, C, F\}, \{A, D, F\}\}$,
$\mathcal{E}' = \{\{Z, A, C, F\}, \{Z, A, D, F\}\}$

Remarquons que, pour un graphe acyclique, le théorème 6 peut aussi s'appliquer à la sémantique stable. C'est alors un cas particulier du théorème 7.

**Théorème 8** *Sous la sémantique préférée, si l'interaction ajoutée est $(X, Z)$, et si $\forall i \geq 1$, $E_i$ attaque $X$, alors la révision est expansive et $\forall i$, $E'_i = E_i \cup \{Z\}$.*

**Exemple 37 page précédente (suite)** *Sous la sémantique préférée, $\mathcal{E} = \{\{A, C\}, \{A, D\}\}$ et $\mathcal{E}' = \{\{Z, A, C\}, \{Z, A, D\}\}$*

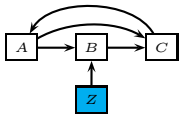
Remarquons là-aussi que, pour un graphe acyclique, le théorème 6 page précédente peut s'appliquer à la sémantique préférée. Et c'est donc un cas particulier du théorème 8.

Si l'interaction ajoutée est $(Z, X)$, des résultats plus faibles peuvent être obtenus. Dans ce cas, la révision n'est pas expansive dans le sens où $\mathcal{G}'$ peut avoir plus d'extensions que $\mathcal{G}$ ; toutefois, l'union de $Z$ à une extension de $\mathcal{G}$ donne une extension de $\mathcal{G}'$.

**Proposition 41** *Sous la sémantique stable, si l'interaction ajoutée est $(Z, X)$, et si $\forall i \geq 1$, $X \notin E_i$, alors $\forall i$, $E_i \cup \{Z\}$ est une extension stable de $\mathcal{G}'$.*

Toutefois, d'autres extensions stables peuvent apparaître dans $\mathcal{G}'$ (voir l'exemple 39). Donc la révision n'est pas expansive.

**Exemple 39** *Sous la sémantique stable :*




$\mathcal{E} = \{\{A\}\}$,
$\mathcal{E}' = \{\{Z, A\}, \{Z, C\}\}$

Remarquons que, dans le cas où le graphe est acyclique, le théorème 5 page précédente peut être utilisé sous la sémantique stable. Et le résultat obtenu est plus fort que celui donné dans la propriété 41.

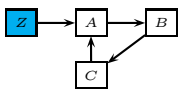
On a aussi une autre propriété intéressante :

**Proposition 42** *Sous la sémantique préférée, si l'interaction ajoutée est $(Z, X)$ et si $\forall i \geq 1$, $X \notin E_i$, alors $\forall i$, $E'_i = E_i \cup \{Z\}$ est admissible dans $\mathcal{G}'$.*

*De plus, s'il existe aucun circuit de longueur impaire dans $\mathcal{G}$, $\forall i$, $E'_i = E_i \cup \{Z\}$ est une extension préférée de $\mathcal{G}'$.*

L'exemple 39 montre aussi que d'autres extensions préférées peuvent apparaître dans $\mathcal{G}'$. Et l'exemple suivant illustre la première partie de la propriété 42.

**Exemple 40** *Sous la sémantique préférée :*




$\mathcal{E} = \{\{\}\}$
$\{\} \cup \{Z\}$ *est admissible dans* $\mathcal{G}'$
*mais* $\mathcal{E}' = \{\{Z, B\}\}$.

Remarquons que si le graphe initial est acyclique, le théorème 5 page précédente peut aussi être appliqué à la sémantique préférée et il donne alors un résultat plus fort que celui issu de la propriété 42.

#### 3.2.5.5 Discussion et perspectives

Dans cette section, nous avons étudié le concept de révision d'un système d'argumentation. Cette révision est pour l'instant très sommaire puisqu'il s'agit juste de traiter l'introduction d'un nouvel argument et d'une nouvelle interaction le concernant dans un AF. Cette révision est étudiée d'après l'impact qu'elle produit sur l'ensemble des ensembles d'arguments acceptables de l'AF résultat par rapport à celui de l'AF initial. Le mot "révision" est utilisé ici par analogie avec la notion traditionnelle de révision de bases de connaissances. Il existe toutefois deux différences importantes.

– Le formalisme sous-jacent est différent : en révision des croyances, les formules logiques sont utilisées pour représenter les croyances alors qu'en argumentation on utilise une représentation à base d'arguments et de relations binaires entre ces arguments (les attaques). Dans le premier cas, la "sortie" du système est un nouvel ensemble de formules logiques, alors que, dans le second cas, la "sortie" sera un nouvel ensemble d'arguments acceptables (ou acceptés).
– La révision est un processus en représentation de connaissance qui est fortement lié à la notion d'inférence et de consistance. C'est la raison pour laquelle les postulats proposés par AGM pour la révision de croyances utilisent la notion de consistance (l'ajout d'une nouvelle information doit se faire en préservant la consistance de la base), même quand cette "révision" s'effectue dans un contexte non monotone (voir les travaux définissant la notion de révision pour des théories non monotones – [WvdH97]). La théorie de l'argumentation est intimement liée à la notion de non-monotonie, mais sans s'appuyer obligatoirement sur les notions d'inférence et de consistance , du coup les postulats proposés pour les théories non monotones ne s'appliquent pas vraiment à la révision en argumentation. Seuls certains d'entre eux seront exploitables pour nous, comme celui sur la classicalité, mais pour le reste il faudra trouver d'autres principes.

Ce travail est donc un travail préliminaire qui se distingue nettement des quelques travaux existant sur le sujet. En effet, nous avons choisi de rester au niveau abstrait sans nous préoccuper de la manière dont les arguments et leurs interactions ont été obtenus. Et c'est ce qui nous différencie des approches existantes traitant de l'ajout d'une nouvelle information dans un cadre argumentatif. Par exemple, Wassermann [Was99], tout comme [FGS04, PC05], définissent sous quelles conditions, en termes d'arguments, des croyances non justifiées peuvent être ajoutées. L'approche de [PG00] étudie les propriétés de la révision de connaissances sous le point de vue argumentatif, *i.e.*, le problème consistant à générer une base de connaissances dans laquelle chaque information est justifiée par de "bons" arguments.

Plus récemment, [RMF+08] a proposé une opération de révision basée sur une notion de "garantie", qui consiste à n'ajouter un argument que s'il peut ensuite être garanti. Même si les idées sous-jacentes sont similaires, ce travail diffère du nôtre sur au moins deux points :

– Premièrement, dans [RMF+08], les arguments ne sont pas totalement abstraits puisqu'ils sont construits à l'aide d'une structure qui permet d'identifier une relation entre argument et sous-argument, ainsi que des propriétés de minimalité, consistance et atomicité. La définition des arguments garantis s'effectue par évaluation des lignes d'argumentation. Par opposition, notre approche reste au niveau abstrait et l'ensemble des arguments acceptables ou acceptés se fait par l'emploi des sémantiques basées sur la notion d'extensions.
– D'autre part, la révision basée sur la notion de garantie vise à satisfaire le postulat d'AGM **K*2**, puisque l'argument ajouté doit être garanti. Notre travail suit une toute autre direction. Nous proposons une étude théorique la plus complète possible de l'impact d'un ajout sur les sorties d'un AF abstrait, ce qui nous permet d'identifier plusieurs types de révision.

Rappelons que ce n'est pas la première fois que des processus définis dans le cadre de la représentation de connaissances sont retranscrits dans le cadre de l'argumentation (voir par exemple des travaux sur la fusion d'AFs présentés dans [CMDK+07] et dans la section 3.2.4 page 62).

Il existe plusieurs applications possibles de notre travail. Une des plus intéressantes consiste à l'utiliser pour concevoir des stratégies de dialogue. La plupart des travaux sur ce sujet considèrent qu'un dialogue se définit par un protocole donnant l'ensemble de mouvements autorisés et qu'une stratégie consiste alors à choisir le prochain mouvement à effectuer. Par exemple, [BC98] propose une stratégie de sélection conduisant à des dialogues plus coopératifs. D'autres approches proposent des jeux de dialogue permettant de répondre à diverses questions : comment faire pour qu'un argument donné appartienne à une extension ? Dans ce cas, une stratégie efficace consistera à identifier quel argument mettre en échec pour que l'argument recherché soit accepté. [AM02] ont proposé des heuristiques pour choisir les arguments les moins attaquables dans un dialogue de persuasion. De la même manière, [RPRS08] ont proposé une stratégie optimale afin de gagner un débat basée sur la probabilité de succès d'un argument et le coût de cet argument pour un agent. [Hun04], avec une approche plus globale, a défini une stratégie permettant de construire un sous-arbre d'arguments optimisant la réussite des buts des agents et minimisant le coût.

Notre approche adopte un autre point de vue. Nous ne définissons aucun protocole et nous ne nous restreignons pas à un type de dialogue. Soit un ensemble d'arguments qui interagissent, nous nous intéressons aux extensions du système pour une sémantique donnée. Nous étudions donc l'impact de l'ajout d'un argument et d'une interaction sur la structure de l'ensemble des extensions. Nous ne nous focalisons pas sur un argument en particulier, mais nous voulons

juste voir comment modifier la structure des extensions. Ce travail nous permet donc de choisir la bonne manière de réviser suivant la structure recherchée après la révision (quel argument sera affecté par la révision et avec quel type d'interaction). C'est pourquoi nous pensons que ce travail peut être utilisé pour définir des stratégies pour diriger le dialogue plutôt que pour participer au dialogue lui-même. Par exemple, si un arbitre du dialogue veut que le dialogue reste le plus ouvert possible, il imposera aux participants d'utiliser des arguments appropriés à une révision expansive. Par contre, si l'arbitre veut que le dialogue soit le plus pointu possible, il favorisera une révision sélective (ou décisive).

Ce travail peut se poursuivre suivant différents axes :

1. généraliser notre option de révision afin de permettre l'ajout d'un argument avec plusieurs interactions ou même carrément l'ajout d'un sous-graphe d'argumentation ;
2. étudier les effets des autres postulats de la révision de croyances ainsi que de ceux définis pour la révision de systèmes non monotones, dans le cas où les arguments sont construits à partir d'une base de connaissances et où les sorties du système d'argumentation sont des formules logiques ;
3. puisque la révision décisive paraît être une "bonne" révision, il pourrait être intéressant de répondre à la question suivante : "Comment identifier le *changement minimal*[23] à effectuer dans un AF donné afin qu'il ne possède plus qu'une seule extension non vide ?".

### 3.2.6 Practical reasoning et argumentation unipolaire

Le *Practical reasoning* (PR) [Raz78], est un mode de raisonnement qui traite le problème suivant "Quelle est la bonne action à faire pour un agent donné dans une situation donnée". Dans [Woo00], le PR est présenté comme un processus en deux étapes. La première étape, souvent appelée *délibération*, consiste à identifier les désirs d'un agent. Durant la seconde étape, appelée *means-end reasoning*, on cherche les moyens de réaliser ces désirs, c'est-à-dire les actions ou les plans à effectuer.

Un désir est *justifié* s'il est vrai dans l'état courant du monde et il est *réalisable* s'il existe un plan permettant de l'obtenir. Les *intentions* de l'agent sont les désirs qu'un agent décide d'obtenir et doivent être à la fois justifiées et réalisables.

Dans la plupart des travaux sur le PR, les arguments sont utilisés pour fournir des raisons permettant de dire si certains désirs peuvent ou pas être considérés comme des intentions. Dans la littérature, on trouve ainsi certains systèmes pour le PR basés sur l'argumentation (voir [AK05, HvdT03, RA06]).

Toutefois, dans la plupart des cas, ce problème du PR est modélisé à l'aide de deux systèmes séparés, chacun d'entre eux capturant une des étapes du processus. Une telle approche présente un inconvénient majeur. En effet, certains désirs qui ne seront pas réalisables pourraient être quand même acceptés lors de la phase de délibération au détriment d'autres désirs qui auraient été réalisables. Une autre limitation de ces systèmes est le fait que leurs propriétés théoriques n'ont pas été étudiées.

Dans cette section, nous proposons un système d'argumentation pour le PR qui permet le calcul des intentions en un seule étape. Ce système est basé sur un système d'argumentation un peu particulier : un AF *à contraintes* [CMDM06] ; ce type d'AF étend le système général de Dung [Dun95] en ajoutant des contraintes que les arguments appartenant aux extensions doivent vérifier (voir les définitions 24 page 37 à 27 page 37).

Le système proposé ici a en entrée :

1. trois types d'arguments : les arguments *épistémiques* qui donnent les raisons de croire en certaines croyances, les arguments *explicatifs* qui montrent qu'un désir peut être vrai dans l'état courant du monde, et les arguments *instrumentaux* qui montrent qu'un désir peut être réalisable,
2. différents conflits entre tous ces arguments, et
3. une contrainte particulière sur les arguments capturant l'idée qu'un désir sera une intention si et seulement s'il est justifié et réalisable. Cela se traduit par le fait que, dans une extension donnée, chaque argument instrumental pour un désir doit être accompagné par au moins un argument explicatif pour ce même désir, et vice-versa.

[23] En nombre d'arcs à ajouter ou à supprimer et en nombre d'arguments à ajouter ou à supprimer.

Les sorties du système sont les différents ensembles d'arguments acceptés (les extensions) ainsi que les différents ensembles d'intentions. L'utilisation de la contrainte permet de calculer directement les intentions à partir des extensions et de ne retourner ainsi que les informations véritablement utiles. En effet, chaque extension correspond à un ensemble de croyances "garanties" et de désirs à la fois justifiés et réalisables. Nous avons alors étudié les propriétés théoriques de ce système et, en particulier, vérifié que ce système est consistant et complet au sens proposé par [CA05].

Cette section est organisée de la manière suivante : la sous-section 3.2.6.1 donne un exemple de PR et la sous-section 3.2.6.2 page suivante présente le langage logique utilisé. La sous-section 3.2.6.3 page 93 décrit les différents types d'arguments à utiliser dans un problème de PR, et la sous-section 3.2.6.4 page 95 donne les conflits pouvant exister entre tous ces arguments. La sous-section 3.2.6.5 page 99 présente notre système d'argumentation à contraintes pour le PR. Ce système est illustré sur l'exemple introductif dans la sous-section 3.2.6.6 page 100. Les propriétés du système sont données en sous-section 3.2.6.7 page 102. Et enfin, la sous-section 3.2.6.8 page 104 compare notre approche aux systèmes existants.

Le travail décrit dans cette section a été réalisé en collaboration avec Leila AMGOUD (IRIT) et Caroline DEVRED (LERIA). Tous les résultats donnés dans cette section ont fait l'objet de publication (voir [ADLS08, ADLS09]) et le lecteur trouvera dans les dites-publications les preuves de toutes les propriétés citées ici.

#### 3.2.6.1 Exemple illustratif

Dans cet exemple, Paula est une étudiante en doctorat. Elle a quatre désirs différents et voudrait savoir lesquels sont réalisables et comment. Les désirs en question sont les suivants :

1. Faire un voyage en Afrique Centrale ($jca$)
2. Finir une soumission avant de partir en voyage ($fp$)
3. Obtenir un poste d'enseignant-chercheur ($lec$)
4. Faire une visite à son amie Carla ($vc$) si Carla est rentrée de son voyage ($cb$)

Les trois premiers désirs sont inconditionnels alors le dernier dépend du retour de l'amie que Paula veut visiter.

En plus de ses désirs, Paula a quelques croyances sur la manière de les satisfaire :

1. Pour pouvoir aller en Afrique Centrale, il faut avoir des billets d'avion ($t$) et être vaccinée ($vac$)
2. Pour pouvoir avoir des billets d'avion, il faut soit se rendre dans une agence de voyage ($ag$) soit demander à un ami de nous les apporter ($fr$)
3. Pour être vaccinée, il faut soit aller à l'hôpital ($hop$) soit aller chez le médecin ($dr$)
4. Pour finir la soumission, il faut travailler ($w$)
5. Pour rendre visite à son amie, il faut prendre la voiture ($gc$) si elle est en état de marche ($gs$)
6. Pour avoir un poste d'enseignant-chercheur, il faut finir sa thèse ($ft$)

Paula a aussi quelques croyances sur l'état du monde :

1. Si on travaille, on ne peut pas aller à l'agence de voyage, ni chez le médecin
2. la voiture est en état de marche
3. Carla n'est pas revenue de son voyage ($\neg cb$)
4. La thèse de Paula n'est pas finie ($\neg ft$)
5. Paula n'est pas vaccinée et n'a pas de billets d'avion

D'après les informations précédentes, le désir de Paula de devenir enseignant-chercheur n'est pas réalisable. Celui de faire une visite à Carla est réalisable puisqu'il existe une manière d'y arriver ; par contre, étant donné l'état courant du monde, ce désir n'est pas justifié (puisque Carla n'est pas revenue de voyage). En ce qui concerne les deux premiers désirs ($jca$ et $fp$), les choses sont différentes. Ils sont tous les deux justifiés et réalisables mais pas en même temps car les manières de les réaliser sont en conflit.

Dans l'idéal, tous les désirs devraient être des intentions, mais on voit sur l'exemple que cela ne pourra pas être le cas. Nous allons donc chercher à savoir quels désirs deviendront des *intentions* et avec *quels plans*.

#### 3.2.6.2 Langage logique utilisé

Dans cette section, nous donnons le langage logique utilisé pour modéliser les croyances, les désirs et les plans. Soit $\mathcal{L}$ un *langage propositionnel*, et $\equiv$ la relation classique d'équivalence.

Dans $\mathcal{L}$, nous distinguons un sous-ensemble $\mathcal{D}$ utilisé pour encoder les *désirs*. Pour nous, un désir est un état que l'agent veut atteindre. Les éléments de $\mathcal{D}$ sont des *littéraux*. $d_1, \ldots, d_n$ qui dénoteront les désirs et de manière générale les lettres minuscules dénoteront les formules de $\mathcal{L}$.

À partir des ensembles décrits précédemment, on peut définir des *règles de génération de désir*. Une telle règle exprime sous quelles conditions un désir peut être "atteint". Par exemple, un désir peut être issu de croyances : "si le temps est ensoleillé, alors je désire aller au parc". Dans ce cas, le désir d'aller au parc dépend de ma croyance au sujet du temps. On peut aussi avoir des désirs qui dépendent d'autres désirs. Par exemple, s'il y a une conférence en Inde et si je désire m'y rendre, alors je désire aussi assister aux "tutoriels". Dans cet exemple, le désir d'assister aux "tutoriels" dépend de l'existence de la conférence en Inde et du désir que j'ai de m'y rendre. Et enfin, un désir peut être inconditionnel, ce qui signifie qu'il ne dépend ni de croyances, ni d'autres désirs. Ces trois cas sont décrits formellement dans la définition suivante.

**Définition 63 (Règles de génération de désirs)** *Une* règle de génération de désirs *(ou* règle de désir*) est une expression de la forme*

$$b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow^{24} d_m$$

*où $b$ est une formule propositionnelle de $\mathcal{L}$, et chaque $d_i$ est un élément de l'ensemble $\mathcal{D}$. De plus, $\nexists d_i, d_j$ avec $i, j \leq m$ tels que $d_i \equiv d_j$.*

*$b \wedge d_1 \wedge \ldots \wedge d_{m-1}$ est appelé le* corps *de la règle (ce corps peut être vide ; cela correspond alors au cas d'un désir inconditionnel), et $d_m$ est appelé le* conséquent *de la règle.*

La signification de la règle est la suivante : "si l'agent *croit* $b$ et *désire* $d_1, \ldots, d_{m-1}$, alors l'agent *désirera* aussi $d_m$". Remarquons qu'un même désir $d_i$ peut apparaître dans le conséquent de plusieurs règles ; cela signifie qu'il existe plusieurs "raisons" d'avoir le même désir.

Une règle de désir est consistante si elle dépend de croyances consistantes et de désirs non contradictoires.

**Définition 64 (Règle de désir consistante)** *Une règle de désir $b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow d_m$ est* consistante *si et seulement si*

- *$b \nvdash \bot$,*
- *$\nexists d_i, d_j$ avec $i, j \leq m$ tels que $d_i \equiv \neg d_j$.*

*Sinon, la règle est dite* inconsistante.

Un agent possède aussi différents *plans* fournis par un système de planification. Un plan est un moyen de réaliser un désir. La génération de ces plans n'est pas le sujet de ce travail. Nous prenons donc comme hypothèse qu'il existe un système de planification capable de nous les fournir sous la forme d'un triplet : i) un ensemble de préconditions qui doivent être satisfaites avant de démarrer le plan, ii) un ensemble de postconditions qui seront vérifiées après la réalisation du plan, et iii) le désir réalisé par le plan. Formellement, on a :

**Définition 65 (Plan)** *Un plan est un triplet $\langle S, T, x \rangle$ tel que*

- *$S$ et $T$ sont des ensembles de formules propositionnelles* consistants *issus de $\mathcal{L}$,*
- *$x \in \mathcal{D}$,*
- *$T \vdash x$ et $S \nvdash x$.*

[24] Le symbole $\hookrightarrow$ ne correspond pas à l'implication matérielle.

Naturellement, il existe un lien entre $S$ et $T$. Mais ce lien n'est pas défini de manière explicite ici puisqu'il dépend essentiellement du système de planification choisi (voir quelques exemple de systèmes de planification sains et complets dans [GNT04, RN95]).

Remarquons que l'ensemble des préconditions peut être vide ($S = \varnothing$), par exemple quand le désir est déjà réalisé dans l'état courant du monde. Dans ce cas, le plan est toujours activable.

Nous considérerons qu'un agent est équipé avec trois *bases finies* :

1. une base $\mathcal{K} \neq \varnothing$ et $\mathcal{K} \neq \{\bot\}$ contenant ses *croyances de base* au sujet de son environnement (les éléments de $\mathcal{K}$ sont des formules propositionnelles issues du langage $\mathcal{L}$),
2. une base $\mathcal{B}_d$ contenant ses règles de désir "consistantes",
3. une base $\mathcal{P}$ contenant ses plans.

À partir de $\mathcal{B}_d$, on peut caractériser les *désirs potentiels* d'un agent :

**Définition 66 (Désirs potentiels)** *L'ensemble des* désirs potentiels *d'un agent est* $\mathcal{PD} = \{d_m | \exists b \wedge d_1 \wedge \ldots \wedge d_{m-1} \hookrightarrow d_m \in \mathcal{B}_d\}$.

Ces désirs sont dits "potentiels" car, quand le corps de la règle n'est pas vide, l'agent ne peut pas savoir *a priori* si tous les antécédents seront vrais ou pas.

**Exemple 41** *La phrase suivante peut être mise sous la forme d'une règle de désir consistante : "Si le temps est beau et si je désire me relaxer alors je désire aller au parc". La traduction se fait en utilisant le vocabulaire suivant :*
*– $wb$ signifie "le temps est beau" (c'est une croyance),*
*– $r$ signifie "Je me relaxe" (c'est un désir),*
*– $bp$ signifie "Je suis dans le parc" (c'est un désir).*
*Nous avons alors la règle suivante : $wb \wedge r \hookrightarrow bp$ (et le désir $bp$ est un désir potentiel).*

*Si nous supposons avoir les croyances suivantes :*
*– le parc est proche de ma maison si j'utilise le métro, ligne 3 (croyance dénotée par $b_1$),*
*– le métro fonctionne et j'ai des tickets valides ($b_2$),*
*– le parc est ouvert et l'entrée est libre et gratuite ($b_3$)*
*et s'il existe une manière d'obtenir $bp$ en utilisant $b_1$, $b_2$ et $b_3$ alors le plan suivant peut être utilisé : $\langle \{b_1, b_2, b_3\}, \{bp\}, bp \rangle$.*

#### 3.2.6.3 Typologie des arguments

Nous allons maintenant présenter les différents types d'arguments utilisés dans un problème de PR. Il y a trois catégories différentes. La première permet de justifier ou d'attaquer des croyances de la base $\mathcal{K}$, alors que les deux autres sont destinées à justifier l'adoption des désirs potentiels de $\mathcal{PD}$. Les arguments seront ici dénotés avec des lettres grecques minuscules.

##### 3.2.6.3.1 Justification de croyances

La première catégorie d'arguments est celle qui a déjà été très largement étudiée dans la littérature, en particulier pour gérer l'inconsistante dans des bases de connaissances. De tels arguments permettent en effet de supporter ou d'attaquer des conclusions potentielles ou des inférences. Ils sont appelés arguments *épistémiques* dans [Har04a]. Dans notre cas, ils seront construits à partir de la base $\mathcal{K}$. Nous utiliserons ici la définition proposée par [SL92].

**Définition 67 (Argument Épistémique )** *Soit $\mathcal{K}$ une base de connaissances. Un* argument épistémique $\alpha$ *est une paire $\alpha = \langle H, h \rangle$ telle que :*

1. $H \subseteq \mathcal{K}$,

2. *$H$ est consistant,*
3. *$H \vdash h$ et*
4. *$H$ est minimal (pour l'inclusion ensembliste $\subseteq$) parmi les ensembles satisfaisant les conditions 1, 2, 3.*

*Le* support *de l'argument est donné par la fonction* $\texttt{SUPP}(\alpha) = H$*, alors que sa* conclusion *est renvoyée par* $\texttt{CONC}(\alpha) = h$*. $\mathcal{A}_b$ dénote l'ensemble de tous les arguments épistémiques pouvant être construits à partir de $\mathcal{K}$.*

**Exemple 41 page précédente (suite)** *Supposons que $\mathcal{K}_{41}$ contient les croyances suivantes :*
– *le temps est beau ($wb$),*
– *si le temps est beau alors je ne prends pas mon parapluie ($wb \rightarrow \neg u$),*
– *si le temps est beau alors je prends un chapeau ($wb \rightarrow h$).*
*Avec $\mathcal{K}_{41}$, on peut construire au moins les arguments suivants :*
– $\alpha_1 = \langle \{wb, wb \rightarrow \neg u\}, \neg u \rangle$,
– $\alpha_2 = \langle \{wb, wb \rightarrow h\}, h \rangle$.
*Remarquons que $\langle \{wb, wb \rightarrow \neg u, wb \rightarrow h\}, h \rangle$ n'est pas un argument épistémique puisqu'il n'est pas minimal.*

#### 3.2.6.3.2 Justification de désirs

Un désir pourra être atteint pour un agent seulement si ce désir est *justifié* et *réalisable*. Il existe donc deux *raisons* pour adopter un désir :

– les conditions soutenant ce désir sont vraies dans l'état courant du monde. De telles raisons conduisent à la notion d'*arguments explicatifs*.
– il existe un plan permettant de réaliser ce désir. De telles raisons conduisent à la notion d'*arguments instrumentaux*.

La définition des arguments explicatifs se fait en utilisant deux bases : la base de croyances $\mathcal{K}$ et la base de règles de désirs $\mathcal{B}_d$. Nous utiliserons une définition arborescente de ces arguments issue de [Vre97]. Nous utiliserons aussi quelques fonctions particulières : les fonctions $\texttt{BELIEFS}(\delta)$, $\texttt{DESIRS}(\delta)$, $\texttt{CONC}(\delta)$ et $\texttt{SUB}(\delta)$ qui retournent respectivement, pour un argument explicatif $\delta$, les croyances utilisées dans $\delta$, les désirs supportés par $\delta$, la conclusion et l'ensemble des sous-arguments de l'argument $\delta$.

**Définition 68 (Argument explicatif)** *Soit $\langle \mathcal{K}, \mathcal{B}_d \rangle$ deux bases.*
– *Si $\exists \hookrightarrow d \in \mathcal{B}_d$ alors $\longrightarrow d$ est un* argument explicatif[25] *($\delta$) avec*
  – $\texttt{BELIEFS}(\delta) = \varnothing$,
  – $\texttt{DESIRS}(\delta) = \{d\}$,
  – $\texttt{CONC}(\delta) = d$,
  – $\texttt{SUB}(\delta) = \{\delta\}$.
– *Si $\alpha$ est un argument épistémique, et $\delta_1, \ldots, \delta_m$ sont des arguments explicatifs, et $\exists\ \texttt{CONC}(\alpha) \wedge \texttt{CONC}(\delta_1) \wedge \ldots \wedge \texttt{CONC}(\delta_m) \hookrightarrow d \in \mathcal{B}_d$ alors $\alpha, \delta_1, \ldots, \delta_m \longrightarrow d$ est un* argument explicatif *($\delta$) avec*
  – $\texttt{BELIEFS}(\delta) = \texttt{SUPP}(\alpha) \cup \texttt{BELIEFS}(\delta_1) \cup \ldots \cup \texttt{BELIEFS}(\delta_m)$,
  – $\texttt{DESIRS}(\delta) = \texttt{DESIRS}(\delta_1) \cup \ldots \cup \texttt{DESIRS}(\delta_m) \cup \{d\}$,
  – $\texttt{CONC}(\delta) = d$,
  – $\texttt{SUB}(\delta) = \{\alpha\} \cup \texttt{SUB}(\delta_1) \cup \ldots \cup \texttt{SUB}(\delta_m) \cup \{\delta\}$.
*$\mathcal{A}_d$ dénote l'ensemble de tous les arguments explicatifs que l'on peut construire à partir de $\langle \mathcal{K}, \mathcal{B}_d \rangle$ en respectant le fait que leurs ensembles* DESIRS *soient consistants*[26].

Cette définition arborescente explique pourquoi l'ensemble $\texttt{DESIRS}(\delta)$ contient le désir $d$, conclusion de $\delta$, mais aussi, dans le cas d'un désir conditionnel, *tous les désirs* utilisés pour justifier $d$ (*i.e.* les désirs utilisés par $\delta_1$, …, les désirs utilisés par $\delta_m$).

[25] La flèche longue $\longrightarrow$ représente un argument explicatif alors que la flèche courte $\rightarrow$ représente l'implication logique classique.
[26] Nous n'acceptons pas les cas où des désirs contradictoires sont nécessaires pour justifier un autre désir.

Remarquons que le fait que les règles dans $\mathcal{B}_d$ soient consistantes n'est pas suffisant pour assurer la consistance de l'ensemble DESIRS d'un argument explicatif.

On peut facilement montrer que cet ensemble DESIRS d'un argument explicatif est un sous-ensemble de $\mathcal{PD}$ et que l'ensemble BELIEFS est un sous-ensemble de la base $\mathcal{K}$.

**Proposition 43** *Soit $\delta \in \mathcal{A}_d$.*
– $\texttt{BELIEFS}(\delta) \subseteq \mathcal{K}$.
– $\texttt{DESIRS}(\delta) \subseteq \mathcal{PD}$.

Remarquons aussi que le même désir peut être supporté par plusieurs arguments explicatifs puisqu'un désir peut être le conséquent de différentes règles de désir.

La dernière catégorie d'arguments que nous allons utiliser ici correspond à l'obligation pour "un désir d'avoir un plan pour le réaliser". La définition de ce type d'arguments utilisera la base de croyances $\mathcal{K}$, la base des plans $\mathcal{P}$, et l'ensemble $\mathcal{PD}$.

**Définition 69 (Argument instrumental)** *Soit $\langle \mathcal{K}, \mathcal{P}, \mathcal{PD} \rangle$ trois bases, et $d \in \mathcal{PD}$. Un* argument instrumental *est une paire $\pi = \langle \langle S, T, x \rangle, d \rangle$ où*
– $\langle S, T, x \rangle \in \mathcal{P}$,
– $S \subseteq \mathcal{K}$,
– $x \equiv d$.
*$\mathcal{A}_p$ dénote l'ensemble de tous les arguments instrumentaux pouvant être construits à partir de $\langle \mathcal{K}, \mathcal{P}, \mathcal{PD} \rangle$. La fonction* CONC *retournera pour un argument $\pi$ le désir $d$. De même, les fonctions* PLAN, Prec *et* Postc *retourneront respectivement le plan $\langle S, T, x \rangle$ de l'argument, les préconditions $S$ du plan, et ses postconditions $T$.*

La seconde condition de la définition précédente dit que les préconditions du plan sont vraies dans l'état courant du monde, donc que le plan peut être exécuté. En effet, il pourrait arriver qu'il existe dans la base de plans $\mathcal{P}$ des plans dont les préconditions ne sont pas vraies. De tels plans ne peuvent donc pas être exécutés et ne donnent donc pas lieu à des arguments instrumentaux.

**Exemple 41 page 93 (suite)** *En utilisant la croyance $wb$, les règles de désir consistantes $\hookrightarrow r$ et $wb \wedge r \hookrightarrow bp$ et le plan $\langle \{b_1, b_2, b_3\}, \{bp\}, bp \rangle$, les arguments suivants peuvent être définis :*
– $\alpha_3 = \langle \{wb\}, wb \rangle$ *(argument épistémique),*
– $\delta_1 = \longrightarrow r$ *(argument explicatif),*
– $\delta_2 = \alpha_3, \delta_1 \longrightarrow bp$ *(argument explicatif),*
– $\pi_1 = \langle \langle \{b_1, b_2, b_3\}, \{bp\}, bp \rangle, bp \rangle$ *(argument instrumental).*
Pour la suite, nous posons $\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p$. Remarquons que $\mathcal{A}$ est *fini* puisque les trois bases initiales ($\mathcal{K}$, $\mathcal{B}_d$ et $\mathcal{P}$) sont aussi finies.

#### 3.2.6.4 Interactions entre arguments

Les arguments construits à partir d'une base de connaissances ne peuvent pas être considérés indépendamment les uns des autres. En effet, un argument constitue une raison de croire ou d'adopter un désir. Mais, ce n'est pas une preuve que cette croyance est vraie, ou que ce désir doive être adopté. En effet, il existe des conflits entre les arguments (certains en attaquant d'autres). Dans cette section, nous allons donc identifier les conflits pouvant apparaître entre les différents types d'arguments décrits dans la section précédente.

##### 3.2.6.4.1 Conflits entre arguments épistémiques

Un argument épistémique peut être attaqué par un autre argument épistémique pour trois sortes de raisons : i) ils ont des conclusions contradictoires (cela correspond à la notion de "*rebuttal*"), ii) la conclusion d'un argument contredit une prémisse d'un autre argument (cela correspond à la notion d'"*assumption attack*"), iii) la conclusion d'un

argument contredit une règle d'inférence utilisée pour construire un autre argument (cela correspond à la notion d'"*undercutting*").

Puisque la base $\mathcal{K}$ est construite en utilisant un langage logique propositionnel, on peut donc appliquer un résultat de [AC02a] qui dit que la notion d'"assumption attack" est suffisante pour capturer tous les conflits entre arguments épistémiques.

**Définition 70** *Soit $\alpha_1, \alpha_2 \in \mathcal{A}_b$. $\alpha_1\ \mathcal{R}_b\ \alpha_2$ si et seulement si $\exists h \in \texttt{SUPP}(\alpha_2)$ tel que $\texttt{CONC}(\alpha_1) \equiv \neg h$.*

**Exemple 41 page 93 (suite)** *Supposons que l'on ajoute les croyances suivantes à $\mathcal{K}_{41}$ :*
– *le vent est très fort ($wvs$),*
– *si le vent est très fort alors mon chapeau risque de s'envoler ($wvs \rightarrow hf$),*
– *si mon chapeau risque de s'envoler alors je ne le prends pas ($hf \rightarrow \neg h$).*
*Un nouvel argument épistémique peut être construit : $\alpha_4 = \langle\{wvs, wvs \rightarrow hf, hf \rightarrow \neg h, wb\}, \neg(wb \rightarrow h)\rangle$.*

*Puis, en utilisant $\alpha_2 = \langle\{wb, wb \rightarrow h\}, h\rangle$, le conflit suivant apparaît : $\alpha_4 \mathcal{R}_b \alpha_2$.*

Remarquons que ce type d'attaque ***n'est pas symétrique***. De plus, on peut montrer que les arguments qui s'auto-attaquent ne peuvent pas exister.

**Proposition 44** *$\nexists \alpha \in \mathcal{A}_b$ tel que $\alpha\ \mathcal{R}_b\ \alpha$.*

Dans [Cay95], le système d'argumentation $\langle\mathcal{A}_b, \mathcal{R}_b\rangle$ a été utilisé pour gérer l'inconsistance dans les bases de connaissances. On peut donc l'appliquer à $\mathcal{K}$. Dans ce cas particulier, on a alors une correspondance parfaite entre les extensions stables du système et les sous-ensembles maximaux consistants de la base $\mathcal{K}$. Avant de rappeler formellement ce résultat, nous introduisons quelques notations importantes :
– Soit $E \subseteq \mathcal{A}_b$, $\texttt{Base}(E) = \bigcup H_i$ tel que $\langle H_i, h_i\rangle \in E$.
– Soit $T \subseteq \mathcal{K}$, $\texttt{Arg}(T) = \{\langle H_i, h_i\rangle$ est un argument épistémique $|H_i \subseteq T\}$.

**Proposition 45 ([Cay95])** *Soit $E$ extension stable de $\langle\mathcal{A}_b, \mathcal{R}_b\rangle$.*
– *$\texttt{Base}(E)$ est un sous-ensemble maximal (pour l'inclusion ensembliste) consistant de $\mathcal{K}$.*
– *$\texttt{Arg}(\texttt{Base}(E)) = E$.*

**Proposition 46 ([Cay95])** *Soit $T$ un sous-ensemble maximal (pour l'inclusion ensembliste) consistant de $\mathcal{K}$.*
– *$\texttt{Arg}(T)$ est une extension stable de $\langle\mathcal{A}_b, \mathcal{R}_b\rangle$.*
– *$\texttt{Base}(\texttt{Arg}(T)) = T$.*

Une conséquence directe de ce résultat est que, si la base $\mathcal{K}$ n'est pas réduite à $\bot$, alors le système $\langle\mathcal{A}_b, \mathcal{R}_b\rangle$ possède au moins une extension stable non vide.

**Proposition 47** *Le système d'argumentation $\langle\mathcal{A}_b, \mathcal{R}_b\rangle$ possède des extensions stables non vides.*

#### 3.2.6.4.2 Conflits entre arguments explicatifs

Les arguments explicatifs peuvent aussi être en conflit. En effet, il suffit qu'ils soient basés sur des désirs contradictoires. Ce type de conflit est capturé par la relation suivante :

**Définition 71** *Soit $\delta_1, \delta_2 \in \mathcal{A}_d$. $\delta_1\ \mathcal{R}_d\ \delta_2$ si et seulement si $\exists d_1 \in \texttt{DESIRS}(\delta_1)$, $d_2 \in \texttt{DESIRS}(\delta_2)$ tels que $d_1 \equiv \neg d_2$.*

**Exemple 41 page 93 (suite)** *En ajoutant la croyance "Je suis en retard dans mon travail" (dénotée par $l$) à la base $\mathcal{K}_{41}$ et en utilisant la nouvelle règle de désir $l \hookrightarrow \neg r$, on peut construire les arguments suivants :*
– *$\alpha_5 = \langle\{l\}, l\rangle$ (argument épistémique),*
– *$\delta_3 = \alpha_5 \longrightarrow \neg r$ (argument explicatif).*

*Puis en utilisant $\delta_1 = \longrightarrow r$, les conflits suivants apparaissent : $\delta_1 \mathcal{R}_d \delta_3$ et $\delta_3 \mathcal{R}_d \delta_1$.*

**Proposition 48** *La relation $\mathcal{R}_d$ est symétrique et irréflexive.*

Il est aussi facile de vérifier que deux arguments explicatifs ayant des désirs conflictuels sont aussi en conflit au sens de la relation $\mathcal{R}_d$ :

**Proposition 49** *Soit $d_1, d_2 \in \mathcal{PD}$. Si $d_1 \equiv \neg d_2$, alors $\forall \delta_1, \delta_2 \in \mathcal{A}_d$ tels que (1) $\exists \delta'_1 \in \texttt{SUB}(\delta_1)$ avec $\texttt{CONC}(\delta'_1) = d_1$, et (2) $\exists \delta'_2 \in \texttt{SUB}(\delta_2)$ avec $\texttt{CONC}(\delta'_2) = d_2$, alors $\delta_1 \mathcal{R}_d \delta_2$.*

Remarquons enfin que, d'après la définition d'un argument explicatif, ses ensembles DESIRS ne peuvent pas être inconsistants. Ce qui n'est pas le cas de ses ensembles BELIEFS qui, eux, peuvent l'être. L'union des croyances de deux arguments explicatifs peut aussi être inconsistante. Nous montrerons un peu plus loin qu'il est inutile de détecter ce type de conflit, puisqu'ils sont capturés par les conflits entre arguments explicatif et épistémique (voir les propriétés 52 page suivante et 53 page suivante).

#### 3.2.6.4.3 Conflits entre arguments instrumentaux

Deux plans peuvent être en conflit pour 4 raisons :

1. incompatibilité de leurs préconditions (ce qui signifie que les deux plans ne peuvent pas être exécutés en même temps),
2. incompatibilité de leurs postconditions (l'exécution des deux plans conduit à des états du monde contradictoires),
3. incompatibilité entre les postconditions d'un plan et les préconditions de l'autre (cela signifie que l'exécution d'un plan empêche l'exécution de l'autre),
4. incompatibilité des désirs supportés (là-aussi, cela revient à dire que ces deux plans conduisent à des états du monde contradictoires).

Toutes ces raisons sont capturées dans la définition suivante.

**Définition 72** *Soit $\pi_1, \pi_2 \in \mathcal{A}_p$ et $\pi_1 \neq \pi_2$. $\pi_1 \mathcal{R}_p \pi_2$ si et seulement si :*
– $\texttt{Prec}(\pi_1) \wedge \texttt{Prec}(\pi_2) \models \bot$, *ou*
– $\texttt{Postc}(\pi_1) \wedge \texttt{Postc}(\pi_2) \models \bot$, *ou*
– $\texttt{Postc}(\pi_1) \wedge \texttt{Prec}(\pi_2) \models \bot$ *ou* $\texttt{Prec}(\pi_1) \wedge \texttt{Postc}(\pi_2) \models \bot$

D'après cette définition, il est évident que $\mathcal{R}_p$ est symétrique et irréflexive[27].

**Proposition 50** *La relation $\mathcal{R}_p$ est symétrique et irréflexive.*

**Exemple 41 page 93 (suite)** *En utilisant l'argument instrumental $\pi_1 = \langle \langle \{b_1, b_2, b_3\}, \{bp\}, bp \rangle, bp \rangle$ ainsi que le nouvel argument instrumental $\pi_2 = \langle \langle \{b_4, b_5, b_6\}, \{st, nl\}, nl \rangle, nl \rangle$ défini avec :*
– *$b_4$ = J'ai une voiture, $b_5$ = Le métro est en panne, $b_6$ = Je dois aller à la réunion,*
– *$st$ = Je suis très stressée et fatiguée, $nl$ = Je ne veux pas être en retard à la réunion (c'est un désir),*
*Il existe un conflit entre $\pi_1$ et $\pi_2$ car $b_2$ (le métro fonctionne et j'ai des tickets valides) et $b_5$ (le métro est en panne) sont en contradiction. Donc $\pi_1 \mathcal{R}_p \pi_2$ et $\pi_2 \mathcal{R}_p \pi_1$.*

En utilisant la définition précédente, on peut montrer que si deux plans réalisent des désirs conflictuels, alors les arguments instrumentaux correspondants sont aussi en conflit.

[27] Le fait que les postconditions d'un plan soient inconsistantes avec les préconditions de ce même plan n'est pas considéré comme un conflit. Dans ce cas, après l'exécution du plan, nous aurons un mécanisme de mise à jour des croyances. C'est aussi pour cette raison qu'il n'existe pas de conflit entre arguments épistémique et instrumental concernant les postconditions (voir la définition 73 page suivante).

**Proposition 51** *Soit $d_1, d_2 \in \mathcal{PD}$. Si $d_1 \equiv \neg d_2$, alors $\forall \pi_1, \pi_2 \in \mathcal{A}_p$ tels que $\texttt{CONC}(\pi_1) = d_1$ et $\texttt{CONC}(\pi_2) = d_2$, alors $\pi_1 \mathcal{R}_p \pi_2$.*

Dans toute cette section, nous n'avons considéré que des *conflits binaires* entre plans, et donc entre les arguments instrumentaux correspondants. Toutefois, dans la vie de tous les jours, il arrive qu'on puisse avoir trois plans qui tous les trois ensembles sont en conflit alors qu'ils ne le sont pas pris deux à deux. Pour des raisons de simplicité, nous n'avons pas traité ce type de conflit.

#### 3.2.6.4.4 Conflits entre arguments de type différent

Dans cette section, nous allons décrire les conflits pouvant exister entre arguments de catégories différentes. Tout d'abord, on considère que les arguments épistémiques jouent un rôle clé pour assurer l'acceptabilité des arguments explicatifs ou instrumentaux. Du coup, un argument épistémique pourra attaquer un argument explicatif ou instrumental, ce qui revient à dire que nous permettons l'invalidation d'une des croyances utilisées par ces arguments explicatifs ou instrumentaux. De même, un argument explicatif peut aussi être en conflit avec un argument instrumental, en particulier quand les désirs qu'ils supportent ou réalisent sont contradictoires.

**Définition 73** *Soit $\alpha \in \mathcal{A}_b$, $\delta \in \mathcal{A}_d$, $\pi \in \mathcal{A}_p$.*
- *$\alpha \mathcal{R}_{bd} \delta$ si et seulement si $\exists h \in \texttt{BELIEFS}(\delta)$ tel que $h \equiv \neg\texttt{CONC}(\alpha)$.*
- *$\alpha \mathcal{R}_{bp} \pi$ si et seulement si $\exists h \in \texttt{Prec}(\pi)$, tel que $h \equiv \neg\texttt{CONC}(\alpha)$.*
- *$\delta \mathcal{R}_{pdp} \pi$ et $\pi \mathcal{R}_{pdp} \delta$ si et seulement si $\texttt{CONC}(\pi) \equiv \neg d$ avec $d \in \texttt{DESIRS}(\delta)$*[28].

Une conséquence triviale de cette définition est l'existence d'un lien entre $\mathcal{R}_b$ et $\mathcal{R}_{bd}$ :

**Consequence 2** *Soit $\alpha_1 \in \mathcal{A}_b$, $\alpha_2 \in \mathcal{A}_b$ et $\delta \in \mathcal{A}_d$ tels que $\alpha_1 \in \texttt{SUB}(\delta)$. Si $\alpha_2 \mathcal{R}_b \alpha_1$ alors $\alpha_2 \mathcal{R}_{bd} \delta$.*

**Exemple 41 page 93 (suite)** *Dans $\mathcal{K}_{41}$, on rajoute les croyances suivantes :*
- *$b_5 \rightarrow \neg b_2$ (si le métro tombe en panne alors il ne fonctionne pas !),*
- *$ca$ (il y a des nuages),*
- *$ca \rightarrow \neg wb$ (s'il y a des nuages alors le temps n'est pas beau),*

*et dans $\mathcal{B}_d$, on rajoute la règle suivante : $l \hookrightarrow \neg bp$ (si je suis en retard dans mon travail alors je désire ne pas aller au parc).*

*On peut alors créer de nouveaux arguments (2 épistémiques et 1 explicatif) :*
- *$\alpha_6 = \langle \{b_5, b_5 \rightarrow \neg b_2\}, \neg b_2 \rangle$,*
- *$\alpha_7 = \langle \{ca, ca \rightarrow \neg wb\}, \neg wb \rangle$,*
- *$\delta_4 = \alpha_5 \longrightarrow \neg bp$.*

*Il existe alors les interactions suivantes entre ces nouveaux arguments :*
- *$\alpha_6 \mathcal{R}_{bp} \pi_1$*
- *$\alpha_7 \mathcal{R}_{bd} \delta_2$ (remarquons que $\alpha_7 \mathcal{R}_b \alpha_3$ et $\alpha_3 \in \texttt{SUB}(\delta_2)$ ce qui illustre la conséquence 2),*
- *$\delta_4 \mathcal{R}_{pdp} \pi_1$ et $\pi_1 \mathcal{R}_{pdp} \delta_4$.*

D'autre part, ainsi que cela a déjà été signalé, l'ensemble des croyances d'un argument explicatif peut être inconsistant. Dans ce cas, l'argument explicatif est attaqué (au sens de $\mathcal{R}_{bd}$) par un argument épistémique :

**Proposition 52** *Soit $\delta \in \mathcal{A}_d$. Si $\texttt{BELIEFS}(\delta) \vdash \bot$, alors $\exists \alpha \in \mathcal{A}_b$ tel que $\alpha \mathcal{R}_{bd} \delta$.*

De la même manière, quand les croyances de deux arguments explicatifs sont inconsistantes, il existe alors un argument épistémique qui attaque au moins l'un des deux arguments explicatifs :

**Proposition 53** *Soit $\delta_1, \delta_2 \in \mathcal{A}_d$ avec $\texttt{BELIEFS}(\delta_1) \nvdash \bot$ et $\texttt{BELIEFS}(\delta_2) \nvdash \bot$. Si $\texttt{BELIEFS}(\delta_1) \cup \texttt{BELIEFS}(\delta_2) \vdash \bot$, alors $\exists \alpha \in \mathcal{A}_b$ tel que $\alpha \mathcal{R}_{bd} \delta_1$, ou $\alpha \mathcal{R}_{bd} \delta_2$.*

[28] Remarquons que si $\delta_1 \mathcal{R}_{pdp} \pi_2$ et qu'il existe $\delta_2$ tel que $\texttt{CONC}(\delta_2) = \texttt{CONC}(\pi_2)$ alors $\delta_1 \mathcal{R}_d \delta_2$.

Les conflits peuvent aussi exister entre un argument instrumental et un argument explicatif puisque les croyances de l'argument explicatif peuvent être conflictuelles avec les préconditions de l'argument instrumental. Et ici-aussi, ce conflit est "couvert" par celui qui existe entre un argument épistémique et l'un des deux autres arguments :

**Proposition 54** *Soit $\delta \in \mathcal{A}_d$ et $\pi \in \mathcal{A}_p$ avec* $\texttt{BELIEFS}(\delta) \not\vdash \bot$. *Si* $\texttt{BELIEFS}(\delta) \cup \texttt{Prec}(\pi) \vdash \bot$ *alors* $\exists \alpha \in \mathcal{A}_b$ *tel que* $\alpha \mathcal{R}_{bd} \delta$, *ou* $\alpha \mathcal{R}_{bp} \pi$.

Les trois propositions précédentes sont suffisantes pour ignorer ces conflits particuliers (ceux entre deux arguments explicatifs, et ceux entre un argument explicatif et un argument instrumental). Rappelons d'autre part qu'il est interdit à des arguments explicatifs ou instrumentaux d'attaquer des arguments épistémiques : un désir ne peut invalider une croyance. Cette interdiction est illustrée par un exemple issu de [Tho00]. Un agent pense qu'il va pleuvoir et que s'il pleut il va se mouiller. Il est évident que cet agent ne désire pas être mouillé, par contre il n'est pas question que le désir de ne pas être mouillé change la croyance de l'agent dans le fait qu'il va pleuvoir (il ne faut pas prendre ses désirs pour des réalités).

#### 3.2.6.5 Système d'argumentation pour le PR

Ici, nous allons utiliser la notion de système d'argumentation à contrainte introduit par [CMDM06] et rappelé dans la section 3.2.1 page 34. Dans le contexte du PR, la notion de contrainte permettra d'encoder le lien entre la justification d'un désir et le plan qui le réalise. L'idée de base est donc ici de n'accepter un argument explicatif pour un désir donné que si est accepté en même temps un argument instrumental permettant de le réaliser, et vice-versa. Cette contrainte va donc être formalisée de la manière suivante :

**Définition 74 (Contrainte pour le PR)** *Soit $\mathcal{A} = \mathcal{A}_d \cup \mathcal{A}_p$ un ensemble d'arguments et $\mathcal{L}_\mathcal{A}$ le langage propositionnel défini en prenant $\mathcal{A}$ comme ensemble de variables propositionnelles. Une* contrainte pour le PR *est une contrainte $C$ sur les arguments de $\mathcal{A}$ telle que :*

$$\begin{aligned} C \quad = \quad & \Big( \bigwedge_{\pi_i \in \mathcal{A}_p} (\pi_i \Rightarrow \Big( \bigvee_{\delta_j \in \{\delta \in \mathcal{A}_d | \texttt{CONC}(\pi_i) \equiv \texttt{CONC}(\delta)\}} \delta_j \Big)\Big) \\ & \wedge \\ & \Big( \bigwedge_{\delta_k \in \mathcal{A}_d} (\delta_k \Rightarrow \Big( \bigvee_{\pi_l \in \{\pi \in \mathcal{A}_p | \texttt{CONC}(\delta_k) \equiv \texttt{CONC}(\pi)\}} \pi_l \Big)\Big) \end{aligned}$$

*avec la convention :* $(\bigvee_{x \in X} x) = \bot$ *si* $X = \varnothing$.

On peut alors définir un système d'argumentation à contrainte pour le PR :

**Définition 75 (Système d'argumentation à contrainte pour le PR)** *Un* système d'argumentation à contrainte pour le PR *est le triplet* $\mathsf{CoAF}_{\mathsf{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ *avec :*
– $\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p$,
– $\mathcal{R} = \mathcal{R}_b \cup \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{bd} \cup \mathcal{R}_{bp} \cup \mathcal{R}_{pdp}$,
– $C$ *une contrainte sur les arguments définie sur $\mathcal{A}$ au sens de la définition 74.*

Dans ce qui suit, $E_1, \ldots, E_n$ dénoteront les $C$-extensions de $\mathsf{CoAF}_{\mathsf{PR}}$ sous une sémantique donnée. À cause de la contrainte $C$, chaque extension $E_i$ ne contient, parmi les arguments instrumentaux, que ceux pour lesquels il existe dans la même extension au moins un argument explicatif ayant la même conclusion, et vice-versa (une extension ne contient, parmi les arguments explicatifs, que ceux pour lesquels il existe dans la même extension au moins un argument instrumental ayant la même conclusion). L'emploi de la contrainte permet ainsi de filtrer le contenu des extensions et de ne garder que les informations utiles.

**Remarque 1** *Une remarque importante concerne la notion de défense : cette notion a deux sémantiques totalement différentes dans le contexte du PR. Quand on ne considère que les arguments épistémiques ou explicatifs, la défense correspond exactement à la notion définie par Dung et utilisée dans les systèmes d'argumentation à contrainte de [CMDM06] : un argument $\alpha$ attaque l'attaquant d'un autre argument $\beta$ ; donc $\alpha$ "réinstalle" $\beta$ ; sans cette défense, nous ne pourrions pas garder $\beta$ dans un ensemble admissible. Les choses sont différentes avec les arguments instrumentaux : quand un argument instrumental attaque un autre argument, cette attaque est toujours symétrique (donc, chaque argument instrumental se défend lui-même contre un autre argument instrumental). Dans ce cas, il pourrait être suffisant de ne prendre en compte que l'absence de conflit pour choisir les arguments instrumentaux appartenant à un ensemble admissible* [29]*. Toutefois, afin de garder une définition homogène de l'admissibilité, la notion de défense sera aussi utilisée pour les arguments instrumentaux, tout en sachant qu'elle n'a aucun impact sur les conflits concernant un argument instrumental.*

Une autre remarque concerne l'existence des $C$-extensions, en particulier les préférées.

**Remarque 2** *L'ensemble vide est toujours un ensemble $C$-admissible pour* $\mathsf{CoAF_{PR}}$*. En effet, $\varnothing$ est admissible (voir [Dun95]) et toutes les variables $\pi_i$ et $\delta_k$ sont fausses dans $\widehat{\varnothing}$, donc $\widehat{\varnothing} \vdash C$)*[30]*. Ainsi, le système d'argumentation* $\mathsf{CoAF_{PR}}$ *possède toujours au moins une extension $C$-préférée.*

Rappelons que le but d'un problème de PR est de calculer les intentions d'un agent, *i.e.* les désirs à la fois justifiés et réalisables. Cette notion d'intention est définie de la manière suivante :

**Définition 76 (Ensemble d'intentions)** *Soit $\langle \mathcal{K}, \mathcal{B}_d, \mathcal{P} \rangle$ les trois bases d'un agent et* $\mathsf{CoAF_{PR}}$ *le système d'argumentation à contrainte pour le PR correspondant. Soit $E_1, \ldots, E_n$ les $C$-extensions de* $\mathsf{CoAF_{PR}}$ *pour une sémantique donnée.*

*Un ensemble $\mathcal{I} \subseteq \mathcal{PD}$ est un* d'ensemble d'intentions *de* $\mathsf{CoAF_{PR}}$ *sous la sémantique donnée si et seulement si il existe une $C$-extension $E_i$ telle que pour chaque $d \in \mathcal{I}$,*

1. *il existe $\pi \in \mathcal{A}_p \cap E_i$ tel que $d = \mathtt{CONC}(\pi)$*
2. *et $\bigcup_{\alpha_j \in E_i \cap \mathcal{A}_b} \mathtt{SUPP}(\alpha_j) \not\vdash d$.*

La première condition correspond au fait qu'une intention doit être un désir justifié *et* réalisable. Et l'idée derrière la seconde condition est que les désirs déjà vrais dans l'état courant du monde ne sont pas des intentions. En effet, une intention est un désir qui ne doit pas être déjà réalisé. Notons toutefois, que le système que nous proposons renverra non seulement les intentions de l'agent, mais aussi les désirs déjà réalisés.

Différents ensembles d'intentions peuvent être retournés par $\mathsf{CoAF_{PR}}$, puisque chaque extension donne à la fois un ensemble d'intentions, l'état du monde justifiant ces intentions et les plans permettant de les réaliser. Il restera donc à l'agent à choisir entre ces divers ensembles ; ceci est alors un problème de décision qui sort de la problématique attribuée au PR (voir [AP07]) et que nous ne traiterons donc pas ici.

#### 3.2.6.6 Reprenons l'exemple initial

Dans cette section, nous allons encoder et résoudre l'exemple de la section 3.2.6.1 page 91 en utilisant notre système d'argumentation à contrainte pour le PR. Voyons tout d'abord les trois bases. La première base $\mathcal{B}_d$ contient les règles de génération de désir.

$$\mathcal{B}_d = \{\hookrightarrow jca, \hookrightarrow fp, \hookrightarrow lec, cb \hookrightarrow vc\}$$

Rappelons la signification des variables : $jca$ pour "faire un voyage en Afrique Centrale", $fp$ pour "finir un papier", $lec$ pour "devenir maître de conférences", $cb$ pour "Carla est de retour", et $vc$ pour "visiter Carla". Dans ce cas, l'ensemble des désirs potentiels est $\mathcal{PD} = \{jca, fp, lec, vc\}$.

Paula a quatre plans différents pour réaliser son désir d'aller en Afrique Centrale :

[29] Cette propriété peut être étendue à $\langle \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{pdp} \rangle$ puisque cette partie de $\mathsf{AF_{PR}}$ est symétrique au sens de [CMDM05] ; dans ce cas, l'admissibilité est équivalente à l'absence de conflit.

[30] Cela est du à la forme particulière de la contrainte dans le cas du PR. Cela n'est pas vrai pour n'importe quelle contrainte (voir [CMDM06] et la section 3.2.1 page 34).

– $p_1 = \langle\{\neg t, \neg vac\}, \{t, vac, ag, dr, jca, \neg w\}, jca\rangle$,
– $p_2 = \langle\{\neg t, \neg vac\}, \{t, vac, fr, dr, jca, \neg w\}, jca\rangle$,
– $p_3 = \langle\{\neg t, \neg vac\}, \{t, vac, ag, hop, jca, \neg w\}, jca\rangle$, et
– $p_4 = \langle\{\neg t, \neg vac\}, \{t, vac, fr, hop, jca\}, jca\rangle$.
Elle a aussi un plan pour chacun des autres désirs ($fp$, $lec$ et $vc$).
– $p_5 = \langle\{\}, \{fp, w\}, fp\rangle$,
– $p_6 = \langle\{ft\}, \{lec\}, lec\rangle$, et
– $p_7 = \langle\{gs\}, \{vc\}, vc\rangle$.
Ainsi, la base des plans $\mathcal{P}$ sera la suivante :

$$\mathcal{P} = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}$$

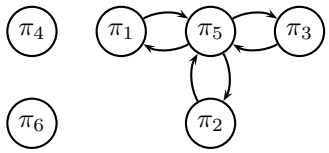
En plus de ces deux bases, Paula possède les croyances suivantes sur l'état du monde :

$$\mathcal{K} = \{w \rightarrow (\neg ag \wedge \neg dr), gs, \neg cb, \neg ft, \neg vac, \neg t\}$$

Remarquons que cette base est consistante dans le cas de cet exemple. Il n'existera donc aucun conflit entre arguments épistémiques. Donc les relations $\mathcal{R}_b$, $\mathcal{R}_{bd}$ et $\mathcal{R}_{bp}$ sont vides et nous ne donnerons pas le contenu de $\mathcal{A}_b$.

Il existe trois arguments explicatifs :
– $\delta_1 : \longrightarrow jca$
– $\delta_2 : \longrightarrow fp$
– $\delta_3 : \longrightarrow lec$
Le désir $vc$, *i.e.* rendre visite à Carla, n'est pas justifié puisqu'il n'existe pas d'argument explicatif en sa faveur (la condition $cb$ n'est pas satisfaite).On a donc $\mathcal{A}_d = \{\delta_1, \delta_2, \delta_3\}$ et il n'existe pas de conflit entre ces arguments, donc $\mathcal{R}_d = \varnothing$.

En ce qui concerne la réalisation des désirs, il est évident que le désir $lec$ n'est pas réalisable puisque le plan correspondant exige que la thèse soit finie, ce qui n'est pas le cas. Les trois autres désirs sont par contre réalisables, ce qui donne les arguments instrumentaux suivants :
– $\pi_1 : \langle p_1, jca\rangle$
– $\pi_2 : \langle p_2, jca\rangle$
– $\pi_3 : \langle p_3, jca\rangle$
– $\pi_4 : \langle p_4, jca\rangle$
– $\pi_5 : \langle p_5, fp\rangle$
– $\pi_6 : \langle p_7, vc\rangle$
Ainsi, $\mathcal{A}_p = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$. Il existe des conflits entre ces arguments, donc la relation $\mathcal{R}_p$ n'est pas vide et se décrit graphiquement par :




Sur cet exemple simple, la relation $\mathcal{R}_{pdp}$ est vide puisqu'il n'existe pas de désirs contradictoires. On a donc $\mathsf{CoAF}_{\mathsf{PR}} = \langle\mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_p, C\rangle$ où $C$ est la contrainte suivante :

$$\begin{aligned} C = \quad ( \quad & (\pi_1 \Rightarrow \delta_1) \\ & \wedge(\pi_2 \Rightarrow \delta_1) \\ & \wedge(\pi_3 \Rightarrow \delta_1) \\ & \wedge(\pi_4 \Rightarrow \delta_1) \\ & \wedge(\pi_5 \Rightarrow \delta_2) \end{aligned}$$

$$\begin{array}{ll} & \wedge(\pi_6 \Rightarrow \bot)) \\ \wedge( & (\delta_1 \Rightarrow (\pi_1 \vee \pi_2 \vee \pi_3 \vee \pi_4) \\ & \wedge(\delta_2 \Rightarrow \pi_5) \\ & \wedge(\delta_3 \Rightarrow \bot)) \end{array}$$

Remarquons les cas particuliers de $\delta_3$ et $\pi_6$ : pour $\delta_3$ (resp. $\pi_6$) il n'existe pas d'argument instrumental (resp. explicatif) correspondant.

Le système $\mathsf{AF_{PR}}$ possède deux extensions stables et préférées :
– $E_1 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \delta_3, \pi_1, \pi_2, \pi_3, \pi_4, \pi_6\}$
– $E_2 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \delta_3, \pi_4, \pi_5, \pi_6\}$
Remarquons que ces extensions contiennent l'argument explicatif $\delta_3$ en faveur du désir $lec$ même si ce désir n'est pas réalisable. De même, elles contiennent l'argument instrumental $\pi_6$ alors que le désir $vc$ n'est pas justifié. Si, maintenant, nous utilisons le système $\mathsf{CoAF_{PR}}$, nous obtenons aussi deux extensions $C$-stables et $C$-préférées :
– $E'_1 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \pi_1, \pi_2, \pi_3, \pi_4\}$
– et $E'_2 = \mathcal{A}_b \cup \{\delta_1, \delta_2, \pi_4, \pi_5\}$.
Remarquons que les extensions $C$-stables ne contiennent que les informations utiles.

En utilisant les $C$-extensions, nous pouvons maintenant donner les ensembles d'intentions de Paula. Il y en a deux, quelle que soit la sémantique choisie :
– $\mathcal{I}_1 = \{jca\}$
– $\mathcal{I}_2 = \{jca, fp\}$
Le choix entre ces deux ensembles n'est pas notre propos ici. On pourrait toutefois supposer que Paula choisit le second ensemble puisqu'il permet de satisfaire deux désirs au lieu d'un seul (en utilisant les plans correspondant à $\pi_4$ et $\pi_5$). Si par contre, elle est très prudente et qu'elle tient absolument à aller en Afrique Centrale, elle choisira plutôt l'ensemble $I_1$ qui donne 4 plans différents pour réaliser son désir le plus cher (sur les quatre plans, elle peut espérer qu'il en aura bien un qui fonctionnera).

#### 3.2.6.7 Propriétés du système

Le but de cette section est d'étudier les propriétés du système d'argumentation que nous proposons pour faire du PR : $\mathsf{CoAF_{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. Nous noterons par $\mathsf{AF_{PR}}$ le système d'argumentation basique correspondant à $\mathsf{CoAF_{PR}}$ (c'est-à-dire le système sans la contrainte $C$ : $\langle \mathcal{A}, \mathcal{R} \rangle$).

Les premières propriétés concernent les extensions du système, et sont des conséquences directes des résultats obtenus par [CMDM06]. La première propriété établit un lien entre les ensembles $C$-admissibles et les extensions $C$-préférées, et montre l'impact de la contrainte sur la notion d'admissibilité.

**Proposition 55** *Soit $\mathsf{CoAF_{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. Soit $\Omega$ l'ensemble des ensembles $C$-admissibles de $\mathsf{CoAF_{PR}}$.*

1. *Soit $E \in \Omega$. Il existe une extension $C$-préférée $E'$ de $\mathsf{CoAF_{PR}}$ telle que $E \subseteq E'$.*
2. *Soit $\mathsf{CoAF_{PR}}' = \langle \mathcal{A}, \mathcal{R}, C' \rangle$ telle que $C' \models C$. Soit $\Omega'$ l'ensemble des ensembles $C'$-admissibles de $\mathsf{CoAF_{PR}}'$. L'inclusion $\Omega' \subseteq \Omega$ est vérifiée.*

Les deux propriétés suivantes montrent que le système à contrainte est plus général que le système classique, mais qu'ils peuvent coïncider dans certains cas.

**Proposition 56** *Soit $\mathsf{CoAF_{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$. Pour chaque extension $C$-préférée $E$ de $\mathsf{CoAF_{PR}}$, il existe une extension préférée $E'$ de $\mathsf{AF_{PR}}$ telle que $E \subseteq E'$.*

Cette propriété est illustrée par l'exemple de la section 3.2.6.6 page 100 : $E_1 \subseteq E'_1$ et $E_2 \subseteq E'_2$, avec $E_1$, $E_2$ les extensions $C$-préférées et $E'_1$, $E'_2$ les extensions préférées.

**Proposition 57** *Soit $\mathsf{CoAF_{PR}} = \langle \mathcal{A}, \mathcal{R}, C \rangle$ tel que $C$ est une formule valide sur $\mathcal{A}$. Alors les extensions préférées de $\mathsf{AF_{PR}}$ sont les extensions $C$-préférées de $\mathsf{CoAF_{PR}}$.*

Rappelons que $\mathsf{AF_{PR}} = \langle \mathcal{A}_b \cup \mathcal{A}_d \cup \mathcal{A}_p, \mathcal{R}_b \cup \mathcal{R}_d \cup \mathcal{R}_p \cup \mathcal{R}_{bd} \cup \mathcal{R}_{bp} \cup \mathcal{R}_{pdp} \rangle$, on a donc une importante propriété qui montre que le sous-ensemble des arguments épistémiques d'une extension stable donnée de $\mathsf{AF_{PR}}$ est aussi une extension stable du système $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$. Sachant que le système $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$ est destiné à gérer l'inconsistance de la base $\mathcal{K}$, ce résultat montre que les extensions stables de $\mathsf{AF_{PR}}$ sont "complètes" en ce qui concerne les arguments épistémiques. Cela montre aussi que les arguments explicatifs et instrumentaux n'ont pas d'impact sur le statut des croyances.

**Proposition 58** *Si $E$ est une extension stable de $\mathsf{AF_{PR}}$, alors l'ensemble $E \cap \mathcal{A}_b$ est une extension stable de $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.*

Nous montrons aussi que le système $\mathsf{AF_{PR}}$ pour le PR (celui sans tenir compte de la contrainte) possède toujours au moins une extension stable.

**Proposition 59** *Le système $\mathsf{AF_{PR}}$ possède toujours au moins une extension stable.*

Si un argument explicatif appartient à une extension stable de $\mathsf{AF_{PR}}$, alors tous ses sous-arguments appartiennent à la même extension.

**Proposition 60** *Soit $\delta \in \mathcal{A}_d$. Soit $E_i$ une extension stable de $\mathsf{AF_{PR}}$. Si $\delta \in E_i$, alors $\texttt{SUB}(\delta) \subseteq E_i$.*

Cela signifie que les croyances utilisées par cet argument explicatif sont "garanties". De même, on peut montrer que si un argument instrumental appartient à une extension stable alors toutes ses préconditions sont supportées par cette extension.

**Proposition 61** *Soit $\pi \in \mathcal{A}_p$. Soit $E$ une extension stable de $\mathsf{AF_{PR}}$. Si $\pi \in E$, alors $\texttt{Prec}(\pi) \subseteq \bigcup_{\alpha_j \in E \cap \mathcal{A}_b} \texttt{SUPP}(\alpha_j)$.*

Ce résultat est aussi vérifié pour les arguments épistémiques. Sachant qu'un argument explicatif peut être basé sur des croyances contradictoires, on montre qu'un tel argument est alors attaqué par au moins un argument épistémique. En fait, c'est encore pire que cela : un tel argument est attaqué par chaque extension stable du système $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.

**Proposition 62** *Soit $\delta \in \mathcal{A}_d$. Si $\texttt{BELIEFS}(\delta) \vdash \bot$, alors $\forall E_i$ avec $E_i$ est une extension stable de $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$, $\exists \alpha \in E_i$ tel que $\alpha \mathcal{R}_{bd} \delta$.*

Une conséquence directe est qu'un tel argument explicatif n'appartiendra jamais à une extension stable du système $\mathsf{AF_{PR}}$.

**Proposition 63** *Soit $\delta \in \mathcal{A}_d$ avec $\texttt{BELIEFS}(\delta) \vdash \bot$. Sous la sémantique stable, l'argument $\delta$ est rejeté par $\mathsf{AF_{PR}}$.*

Puisqu'un argument explicatif avec des croyances contradictoires est rejeté par $\mathsf{AF_{PR}}$, il l'est aussi par $\mathsf{CoAF_{PR}}$.

**Proposition 64** *Soit $\delta \in \mathcal{A}_d$ avec $\texttt{BELIEFS}(\delta) \vdash \bot$. Sous la sémantique stable, $\delta$ est rejeté par $\mathsf{CoAF_{PR}}$.*

En plus, dans la propriété 53 page 98, nous avons montré que, lorsque deux arguments explicatifs sont basés sur des croyances dont l'union est inconsistante, alors au moins l'un des deux est attaqué par un argument épistémique. Dans ce cas là-aussi, un des deux arguments est attaqué par chaque extension stable du système $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$.

**Proposition 65** *Soit $\delta_1, \delta_2 \in \mathcal{A}_d$ avec $\texttt{BELIEFS}(\delta_1) \nvdash \bot$ et $\texttt{BELIEFS}(\delta_2) \nvdash \bot$.*
*Si $\texttt{BELIEFS}(\delta_1) \cup \texttt{BELIEFS}(\delta_2) \vdash \bot$, alors $\forall E_i$ avec $E_i$ une extension stable de $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$, $\exists \alpha \in E_i$ telle que $\alpha \mathcal{R}_{bd} \delta_1$, ou $\alpha \mathcal{R}_{bd} \delta_2$.*

Nous allons encore plus loin et nous montrons que ces deux arguments ne peuvent pas être acceptés en même temps (être dans la même extension). Cela garantit que le système renvoie des résultats corrects.

**Proposition 66** *Soit $\delta_1, \delta_2 \in \mathcal{A}_d$ avec $\texttt{BELIEFS}(\delta_1) \nvdash \bot$ et $\texttt{BELIEFS}(\delta_2) \nvdash \bot$.*
*Si $\texttt{BELIEFS}(\delta_1) \cup \texttt{BELIEFS}(\delta_2) \vdash \bot$, alors $\nexists E$ avec $E$ une extension $C$-stable de $\mathsf{CoAF_{PR}}$ telle que $\delta_1 \in E$ et $\delta_2 \in E$.*

On peut traiter de la même manière les conflits entre arguments explicatif et instrumental. Nous avons déjà montré que, dans ce cas, au moins l'un des deux est attaqué par un argument épistémique. Nous complétons ce résultat par le fait que les deux ne peuvent appartenir aux mêmes extensions, l'un d'eux étant rejeté par le système.

**Proposition 67** *Soit $\delta \in \mathcal{A}_d$ et $\pi \in \mathcal{A}_p$ avec* $\texttt{BELIEFS}(\delta) \nvdash \bot$. *Si* $\texttt{BELIEFS}(\delta) \cup \texttt{Prec}(\pi) \vdash \bot$ *alors* $\forall E_i$ *avec $E_i$ est une extension stable de $\langle \mathcal{A}_b, \mathcal{R}_b \rangle$,* $\exists \alpha \in E_i$ *telle que* $\alpha \mathcal{R}_{bd} \delta$, *ou* $\alpha \mathcal{R}_{bp} \pi$.

**Proposition 68** *Soit $\delta \in \mathcal{A}_d$ et $\pi \in \mathcal{A}_p$ avec* $\texttt{BELIEFS}(\delta) \nvdash \bot$. *Si* $\texttt{BELIEFS}(\delta) \cup \texttt{Prec}(\pi) \vdash \bot$ *alors* $\nexists E$ *avec $E$ une extension $C$-stable de* $\mathsf{CoAF_{PR}}$ *telle que* $\delta \in E$ *et* $\pi \in E$.

Les résultats suivants sont essentiels. Ils montrent que le système proposé pour le PR satisfait le premier postulat de "rationalité" décrit dans [CA05] (celui sur la "consistance" des extensions). En effet, nous montrons que chaque extension stable de notre système supporte un ensemble consistant de croyances et de désirs.

Pour cela, nous utiliserons les notations suivantes : Soit $E \subseteq \mathcal{A}$.

$$\texttt{Bel}(E) = (\bigcup_{\alpha_i \in E \cap \mathcal{A}_b} \texttt{SUPP}(\alpha_i)) \cup (\bigcup_{\delta_j \in E \cap \mathcal{A}_d} \texttt{BELIEFS}(\delta_j)) \cup (\bigcup_{\pi_k \in E \cap \mathcal{A}_p} \texttt{Prec}(\pi_k))$$

$$\texttt{Des}(E) = (\bigcup_{\delta_j \in E \cap \mathcal{A}_d} \texttt{DESIRS}(\delta_j)) \cup (\bigcup_{\pi_k \in E \cap \mathcal{A}_p} \texttt{CONC}(\pi_k))$$

**Théorème 9** *(Consistance) Soit* $\mathsf{CoAF_{PR}}$ *un système d'argumentation à contrainte pour le PR, et $E_1, \ldots, E_n$ ses extensions $C$-stables. $\forall E_i$, $i = 1, \ldots, n$, on a :*

1. *L'ensemble* $\texttt{Bel}(E_i) = \texttt{Bel}(E_i \cap \mathcal{A}_b)$.
2. *L'ensemble* $\texttt{Bel}(E_i)$ *est un sous-ensemble consistant maximal (pour l'inclusion ensembliste) de $\mathcal{K}$.*
3. *L'ensemble* $\texttt{Des}(E_i)$ *est consistant.*

Une conséquence directe de ce résultat est le fait qu'un ensemble d'intentions est aussi consistant :

**Théorème 10** *Sous la sémantique stable, chaque ensemble d'intentions de* $\mathsf{CoAF_{PR}}$ *est consistant.*

Nous avons aussi montré que notre système vérifie le second postulat de rationalité proposé par [CA05], celui portant sur la "fermeture" des extensions. Ici-aussi, nous utilisons quelques notations particulières :

Soit $E_i$ une extension $C$-stable de $\mathsf{CoAF_{PR}}$. Soit $\mathcal{A}s$ l'ensemble de tous les arguments (épistémiques, explicatifs and instrumentaux) qui peuvent être construits à partir de $\texttt{Bel}(E_i)$, $\texttt{Des}(E_i)$, les plans utilisés dans les arguments de $E_i$, et la base $\mathcal{B}_d$.

**Théorème 11** *(Fermeture) Soit* $\mathsf{CoAF_{PR}}$ *un système d'argumentation à contrainte pour le PR, et $E_1, \ldots, E_n$ ses extensions $C$-stables. $\forall E_i$, $i = 1, \ldots, n$, on a :*

– $\texttt{Arg}(\texttt{Bel}(E_i)) = E_i \cap \mathcal{A}_b$.
– $\mathcal{A}s = E_i$.

#### 3.2.6.8 Discussion

À part notre travail, il existe d'autres tentatives pour modéliser le PR à l'aide de systèmes argumentatifs. Certains de ces modèles (e.g. [Amg03, AK05, HvdT04]) sont des instances du système *abstrait* d'argumentation proposé par Dung [Dun95]. D'autres (e.g. [KM03, SGC04]) sont basés sur un encodage du raisonnement argumentatif en programmes logiques. Et enfin, il existe des travaux basés sur des théories entièrement nouvelles du PR et de la persuasion (e.g. [ABCM04, TP05]). Notre travail s'inscrit donc dans la première catégorie.

Dans [Amg03], on trouve un système d'argumentation pour générer des plans consistants pour des ensembles donnés de désirs et de règles de planification. Dans [AK05], ce système a ensuite été étendu en un système qui génère les désirs

eux-mêmes. Pour cela, la notion de "règles de génération de désir" a été introduite, ces règles permettant de générer les désirs à partir des croyances. Notre proposition généralise cela puisque les règles que nous utilisons permettent de générer des désirs à partir de croyances, mais aussi à partir d'autres désirs. D'autre part, le travail de [AK05] ne permet pas de bien distinguer entre croyances et désirs dans les antécédents et conséquents de ces règles ; cela conduit parfois à faire des inférences erronées sur la base de désirs pas encore satisfaits. Notre approche corrige ce problème en particulier dans la définition de l'attaque entre arguments explicatifs. Le langage logique que nous utilisons a déjà été proposé dans [RA06]. Nous avons étendu en les enrichissant les relations de conflit entre types d'arguments et en calculant en une seule "passe" les arguments à accepter. En effet, dans [RA06], le calcul de l'acceptabilité étant fait indépendamment pour chaque type d'argument, il pouvait arriver qu'un argument explicatif pour un désir donné soit accepté alors que l'argument instrumental qui aurait permis la réalisation de ce même désir ne soit pas accepté.

La notion de "règle de désir" a aussi été utilisée par Hulstijn et van der Torre [HvdT04]. Dans ce cas-là, il ne pouvait y avoir que des désirs dans le conséquent de la règle. Toutefois, cette approche impose que les buts choisis[31] soient supportés par des arbres de but[32] qui contiennent à la fois les règles de désir et les règles portant sur les croyances et qui sont déductivement consistants. Cette contrainte de fermeture déductive consistante ne permet pas de distinguer entre les littéraux pour les désirs et ceux pour les croyances (voir la proposition 2 dans [HvdT04]). Cela veut dire que l'agent ne peut pas croire en $\neg p$ tout en désirant $p$.Dans notre approche, on pourra construire un argument épistémique pour croire en $\neg p$ et aussi un argument explicatif pour justifier $p$.

D'autres travaux, comme ceux d'Atkinson et Bench Capon [ABCM04] portent plutôt sur l'étude des différents schémas d'arguments que l'on peut rencontrer en PR. Leur point de départ est le syllogisme suivant donné par le philosophe Walton dans [Wal96].
– $G$ est un but/désir pour l'agent $X$
– Faire l'action $A$ suffit pour que l'agent $X$ obtienne $G$
– Alors l'agent $X$ doit faire l'action $A$
Ce syllogisme, qui s'applique à l'étape "means-end reasoning", est en fait un argument en faveur de "faire l'action $A$". Par contre, cela ne signifie pas que l'action est garantie, puisque d'autres arguments (les contre-arguments) peuvent être construits contre l'action. Le travail de [ABCM04] consiste à définir plusieurs variantes de ce syllogisme et différentes manières de les attaquer. Par contre, la manière d'intégrer tous ces schémas dans un système unique afin de répondre à la question critique du PR "quelle est la bonne chose à faire dans une situation donnée ?" n'est pas donnée. Dans notre approche, nous répondons à cette question (en calculant les extensions) et en plus nous couvrons la plupart des schémas et des attaques prévues dans [ABCM04]. Par exemple, nous répondons aux les questions suivantes associées au syllogisme de Walton :

1. Existe-t-il d'autres manières de réaliser le but/désir $G$ ?
2. Est-il possible de faire l'action $A$ ?
3. L'agent a-t-il d'autres buts qui peuvent être pris en compte ?
4. Y-a-t-il d'autres conséquences à l'exécution de $A$ qui pourraient être prises en compte ?

La première question consiste à trouver les différent arguments instrumentaux pour le désir $G$ et à tous les prendre en compte dans le raisonnement, *i.e.* quand on calcule les ensembles d'intentions. La seconde question consiste à vérifier si on est dans un état du monde autorisant l'exécution de $A$ (cela correspond aux préconditions des plans). La troisième question est aussi traitée par notre approche puisque on part de l'ensemble des désirs potentiels pour trouver les intentions. Et enfin, la dernière question correspond à la notion de postconditions des plans ainsi que les croyances de la base $\mathcal{K}$.

#### 3.2.6.9 Conclusion

Ce travail consistait à utiliser un système d'argumentation pour résoudre le problème du "practical reasoning" (PR) qui consiste à répondre à la question "quelle est la meilleure chose à faire dans une situation donnée" ? Notre contribution sur ce sujet est en deux parties :

[31] Ce qui correspond à nos désirs justifiés.
[32] Un arbre de but correspondant à notre notion d'argument explicatif.

– premièrement, il s'agit, à notre connaissance, du premier système résolvant le problème du PR en une seule étape, *i.e.* en combinant génération de désirs et réalisation de désirs. Ce qui permet éviter les inconvénients dus à des approches trop décomposées.
– La seconde contribution a consisté à étudier de manière approfondie les propriétés théoriques de ce nouveau système.

Les perspectives de ce travail sont très diverses.
– Tout d'abord, nous pouvons étudier le cas où la relation d'attaque entre arguments instrumentaux ne serait pas binaire.
– Ensuite, nous envisageons d'introduire des préférences dans ce système (les croyances pourraient être incertaines à des degrés divers, il pourrait y avoir des priorités entre les désirs et les plans pourraient avoir des coûts différents). Ainsi, le calcul des intentions pourrait ne produire que ceux qui sont les plus significatives.
– D'autre part, dans [CLS05c, KP01], une autre relation entre arguments a été exhibée, représentant une notion de support entre arguments. Il serait intéressant de voir l'impact de cette relation sur notre système dans le contexte du PR.
– Et enfin, l'étude d"une théorie de la preuve adaptée à ce système serait un champ d'investigation très intéressant. L'idée consiste ici à répondre à la question "est-ce qu'un désir potentiel donné est une intention possible de l'agent ?" sans calculer la totalité des extensions préférées.

## 3.3 Argumentation bipolaire

Dans la plupart des systèmes d'argumentation, il n'existe qu'un seul type d'interaction entre les arguments. C'est ce que l'on appelle la relation d'attaque. Toutefois, depuis quelques années, des études récentes [KP01, Ver03, ACLS04] ont montré l'intérêt de considérer un nouveau type de relation. En effet, si un argument peut attaquer un autre argument, il est aussi envisageable qu'un argument soit capable d'en "aider" certains autres. L'ajout de ce type d'interaction suggère une notion de bipolarité, *i.e.* l'existence de types d'interaction ayant une nature diamétralement opposée et représentant des "forces qui se repoussent".

La bipolarité a déjà été très fortement étudiée et utilisée dans le cadre des bases de connaissances et de la représentation de préférences [Bou94, TP94, LVW02, BDKP02]. Par exemple, dans [BDKP02] deux types de préférences sont décrites : une préférence *positive* représentant ce que l'agent veut réellement, et une préférence *négative* référençant ce que l'agent rejette. Cette distinction entre préférence positive et préférence négative est supportée par des études en psychologie cognitive qui ont montré que ces deux types de préférence étaient complètement indépendantes et prises en compte de manière différente par l'esprit humain. La bipolarité est aussi largement utilisée dans le cadre de la prise de décision. Dans [ABP05, DF05], il a été montré que quand on prend une décision, on s'appuie généralement sur les informations en faveur et en défaveur des décisions parmi lesquelles il faut choisir.

Dans [DP06], une nomenclature de trois types de bipolarité a été proposée en s'appuyant sur des caractéristiques particulières comme *l'exclusivité* (une information peut-elle être en même temps positive et négative), la *dualité* (peut-on calculer l'information négative à partir d'une information positive), *l'exhaustivité* (peut-on avoir des informations ni positives, ni négatives), le calcul des informations positives et négatives se fait-il *sur les mêmes données initiales*, le calcul des informations positives et négatives se fait-il *suivant le même processus*, *existe-t-il une contrainte de consistance* entre les informations positives et négatives.

Le premier type de bipolarité proposé par [DP06] (*bipolarité symétrique univariée*) exprime le fait que la partie négative est un reflet de la partie positive (donc, elles sont mutuellement exclusives et une simple échelle bipolaire univariée suffit pour les représenter toutes les deux à la fois).

Le second type de bipolarité (*bipolarité duale bivariée*) exprime le fait que nous avons besoin de deux échelles séparées pour représenter les deux parties de l'information, bien que ces deux parties soient calculées à partir des mêmes données initiales (donc, une information peut être positive et négative en même temps et il n'y a plus d'exclusivité). Toutefois, une certaine dualité doit exister entre la partie positive et la partie négative.

Et le troisième type de bipolarité (*bipolarité hétérogène*) exprime le fait que les deux parties de l'information ne sont pas calculées à partir des mêmes données initiales, bien qu'il existe une contrainte minimale de consistance entre la partie positive et la partie négative.

Dans cette section, nous allons étudier l'utilisation de la notion de bipolarité dans le domaine de l'argumentation en nous restreignant plus particulièrement au cas où la bipolarité apparaît dans la définition des interactions entre arguments, c'est-à-dire au cas où il existe deux types d'interaction entre arguments, une interaction "positive" et une interaction "négative" (une étude plus complète sur l'usage de la bipolarité en argumentation a été réalisée dans [ACLS04, ACLSL08]).

L'interaction "négative" correspond bien évidemment à la notion d'attaque déjà très largement développée dans le domaine de l'argumentation.

Quant à l'interaction "positive", montrons sur un exemple ce que cela pourrait être : si un argument $a_1$ attaque un argument $a_3$ et que $a_3$ attaque un argument $a_2$, alors on pourrait dire que $a_1$ "appuie" $a_2$ (donc il existerait une interaction "positive" entre $a_1$ et $a_2$). Dans ce cas, la notion d'appui n'a pas besoin d'être formalisée à l'aide d'une autre relation que la relation d'attaque ; elle correspond à la notion de défense introduite par Dung [Dun95]). L'appui et l'attaque sont alors des notions *dépendantes*. C'est une manière très "économique" de définir une interaction positive entre deux arguments, mais cette méthode ne nous paraît pas entièrement correcte car il nous semble qu'il existe une différence essentielle entre "défense" et "appui" que nous allons illustrer sur quelques exemples :

**Exemple 42** *Nous voulons partir en randonnée. Pour cela, nous préférons une météo ensoleillée, et sinon, un temps nuageux mais pas pluvieux. Nous annulerons la randonnée seulement s'il pleut. Toutefois, nous savons que les nuages peuvent être le signe d'une pluie imminente. Nous examinons donc l'état du ciel le matin avant de partir : il est nuageux. Nous échangeons alors les arguments suivants :*

1. *Aujourd'hui nous disposons de suffisamment de temps, faisons la randonnée.*
2. *La météo est nuageuse et les nuages sont signe de pluie, nous devrions annuler la randonnée.*
3. *Ces nuages sont dispersés, il va faire beau, le soleil va prendre le pas sur les nuages et les nuages vont disparaître.*
4. *Les nuages ne vont pas grossir davantage, mais ne disparaîtront pas. Toutefois, il ne pleuvra pas.*

Dans cet échange d'argument, on peut identifier différents points de conflits : l'argument 4 attaque l'argument 3 qui attaque l'argument 2 qui attaque à son tour l'argument 1. Donc, dans le cadre de Dung, l'argument 3 est un défenseur de l'argument 1, et l'argument 4 est un attaquant (indirect) de l'argument 1. Néanmoins, les arguments 3 et 4 sont d'accord pour supporter le projet de randonnée (ils s'appuient donc l'un l'autre). Donc, l'idée qu'il suffit de considérer, dans une chaîne d'arguments et de contre-arguments, les arguments "pairs" comme des attaquants alors que les arguments "impairs" seraient des défenseurs est une simplification trop forte. La notion de défense proposée par [Dun95] ne suffit donc à exprimer un appui entre argument.

L'exemple suivant permet aussi d'illustrer le besoin d'un autre type d'interaction entre arguments. Il s'agit d'une discussion entre journalistes lors de l'élaboration de la Une d'un journal :

**Exemple 43**

**Arg.** $a$ **:** *En supposant qu'on ait l'accord de la personne concernée et qu'il n'y ait pas de censure, l'information $I$ concernant $X$ sera publiée.*

**Arg.** $b_1$ **:** *$X$ est le premier ministre qui peut user du droit de censure.*

**Arg.** $c_0$ **:** *Nous sommes en démocratie et même un premier ministre ne peut user du droit de censure.*

**Arg.** $c_1$ **:** *Je crois que $X$ vient de démissionner. Il n'est donc plus premier ministre.*

**Arg.** $d$ **:** *La démission a été annoncée officiellement hier à la télé.*

**Arg.** $b_2$ **:** *$I$ est une information privée donc $X$ refuse la publication.*

**Arg.** $e$ **:** *$I$ est une information importante concernant le fils de $X$ (elle est donc bien privée).*

**Arg.** $c_2$ **:** *Toute information concernant un premier ministre est publique (et pas privée).*

**répétition de Arg.** $c_1$ **et** $d$ **:** ...

**Arg.** $c_3$ **:** *Mais $I$ est d'intérêt national, donc $I$ ne peut pas être considérée comme une information privée.*

Dans cet exemple, quelques conflits apparaissent : par exemple, $b_1$ (resp. $b_2$) est en conflit avec $a$. Mais on peut aussi considérer que l'argument $d$ donné par l'agent $Ag_1$ appuie l'argument $c_1$ donné par l'agent $Ag_2$. Ce n'est pas seulement un "acte de langage du type dialogue" : une nouvelle information est vraiment donnée et elle est fournie *après* la production de l'argument $c_1$. Donc prendre en compte $d$ conduit soit à modifier $c_1$, soit à trouver une manière plus intuitive de représenter l'interaction entre $d$ et $c_1$. Nous avons choisi ici la seconde méthode, ce qui revient à travailler de manière incrémentale en prenant en compte les arguments donnés par les différents agents sans "réviser" les arguments déjà avancés, mais du coup, il va falloir représenter tous les types d'interaction entre ces arguments.

Le dernier exemple montre comment on pourrait formaliser cette notion d'appui entre deux arguments quand on utilise une structure logique pour représenter les dits-arguments.

**Exemple 44** *Un meurtre a été commis et les trois suspects sont Liz, Mary et Peter. Les informations disponibles sont les suivantes :*

*Le type de meurtre suggère que le tueur est une femme. Le tueur est probablement petit. Liz est grande alors que Mary et Peter sont petits. Le tueur a des cheveux longs et utilise un rouge-à-lèvres. Un témoin dit qu'il a vu le tueur et qu'il était grand. Toutefois, on sait que ce témoin est myope, il n'est donc pas très fiable.*

*On peut formaliser toutes ses informations à l'aide des symboles propositionnels suivants : $sm$ (le tueur est petit), $fem$ (le tueur est une femme), $mary$ (le tueur est Mary), $lglip$ (le tueur a des cheveux longs et utilise un rouge-à-lèvres), $wit$ (le témoin est fiable), $bl$ (le témoin est myope).*

*On choisit de représenter un argument sous la forme d'un ensemble de prémisses qui implique une conclusion. On a donc les arguments suivants :*

- $a_1$ *en faveur de* $mary$*, avec les prémisses* $\{sm, fem, (sm \wedge fem) \rightarrow mary\}$
- $a_2$ *en faveur de* $\neg sm$*, avec les prémisses* $\{wit, wit \rightarrow \neg sm\}$
- $a_3$ *en faveur de* $\neg wit$*, avec les prémisses* $\{bl, bl \rightarrow \neg wit\}$
- $a_4$ *en faveur de* $fem$*, avec les prémisses* $\{lglip, lglip \rightarrow fem\}$

$a_3$ *attaque* $a_2$ *qui attaque* $a_1$*. Donc* $a_3$ *défend* $a_1$ *contre* $a_2$*.*

*De plus,* $a_4$ *confirme la prémisse* $fem$ *de* $a_1$*. Donc,* $a_4$ *appuie* $a_1$ *(dans le sens où* $a_4$ *renforce* $a_1$*). Par opposition,* $a_3$ *défend* $a_1$ *contre* $a_2$ *signifie que* $a_3$ *affaiblit l'attaque de* $a_1$ *par* $a_2$*. Donc, d'un côté, on a* $a_1$ *qui reçoit un support et de l'autre* $a_1$ *qui subit une attaque affaiblie.*

Les exemples présentés précédemment montrent tous que le processus d'argumentation utilise des arguments et des contre-arguments, des relations d'appui et des relations d'attaque, mais pas toujours de la même manière. Les arguments qui sont utilisés dans un processus d'argumentation dynamique utilisent des prémisses qui ne sont pas toujours des connaissances dûment vérifiées et prouvées. Si nous acceptons qu'un nouveau fait puisse remettre en cause une de ces prémisses (ce qui correspond à la relation d'attaque), alors nous devons aussi accepter qu'un nouveau fait puisse aussi renforcer ou confirmer une prémisse (ce qui correspond à la relation d'appui).

Tous ces exemples et toutes les remarques qui en découlent, nous conduisent donc à proposer une version étendue du cadre de Dung permettant d'intégrer cette notion d'appui entre arguments. Nous allons donc modéliser des situations dans lesquelles on aura deux types d'interactions *indépendantes* : une positive et une négative (voir par exemple dans le domaine médical le travail [KP01]). Et, suivant les travaux [KP01, Ver03], nous présentons un nouveau système d'argumentation : un système d'argumentation abstrait bipolaire.

Cette section est organisée de la manière suivante : dans la sous-section 3.3.1 nous introduisons les définitions formelles d'un système d'argumentation abstrait bipolaire (BAF). Puis, nous considérons le problème fondamental consistant à déterminer quels arguments (ou ensemble d'arguments) peuvent être dits acceptables. La méthode formelle pour résoudre ce problème est de définir de nouvelles sémantiques pour l'acceptabilité adaptées au cadre des BAF. La sous-section 3.3.2 page 111 introduit une première version de ces sémantiques, celle basée sur la notion d'extension définie *à la fois* à partir de la relation d'appui *et* de la relation d'attaque. Puis, dans la sous-section 3.3.3 page 114, une autre version est proposée, toujours basée sur la notion d'extension mais cette fois en passant par une étape de traduction qui transforme le BAF en un AF particulier (la relation d'appui sert à construire des coalitions entre lesquelles ne restera plus alors que la relation d'attaque ; dans cet AF, on peut alors appliquer directement les sémantiques de Dung. Et enfin, en sous-section 3.3.4 page 117, un dernier type de sémantiques est donné basé cette fois sur une notion d'étiquetage.

Tout ce travail a été réalisé avec Claudette CAYROL et a donné lieu à de nombreuses publications dans lesquelles le lecteur trouvera les preuves de toutes les propriétés données ici ([ACLS04, ACLSL08, CLS04, CLS05a, CLS05c, CLS07, MCLS05]). L'article [CLS09], à paraître, donnera la synthèse de tous ces travaux.

### 3.3.1 Définition d'un BAF

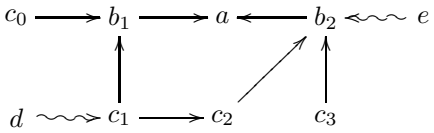
Un système d'argumentation abstrait bipolaire est une extension du système d'argumentation abstrait de base introduit par [Dun95] dans lequel un nouveau type d'interaction entre arguments est représenté par la *relation d'appui* [33]. Cette nouvelle relation est supposée totalement indépendante de la relation d'attaque (*i.e.* elle n'est pas définie à partir d'attaques entre arguments). Ainsi, nous avons une représentation bipolaire des interactions entre arguments.

**Définition 77 (Système d'argumentation abstrait bipolaire – BAF)**
*Un* système d'argumentation abstrait bipolaire $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ *consiste en :*
– *un ensemble $\mathcal{A}$ d'arguments,*
– *une relation binaire $\mathcal{R}_{att}$ sur $\mathcal{A}$ appelée* relation d'attaque
– *et une autre relation binaire $\mathcal{R}_{sup}$ sur $\mathcal{A}$ appelée* relation d'appui.
*Ces relations binaires doivent vérifier la contrainte de consistance suivante : $\mathcal{R}_{att} \cap \mathcal{R}_{sup} = \varnothing$*[34].
*Considérons $a_i$ et $a_j \in \mathcal{A}$, $a_i \mathcal{R}_{att} a_j$ (resp. $a_i \mathcal{R}_{sup} a_j$) signifie que $a_i$ attaque $a_j$ (resp. $a_i$ appuie $a_j$).*
*Soit $a \in \mathcal{A}$, ${\mathcal{R}_{att}}^-(a)$ (resp. ${\mathcal{R}_{sup}}^-(a)$) dénote l'ensemble des attaquants (resp. supporters) de $a$.*

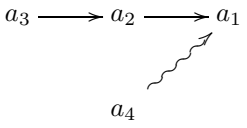
Dans la suite, nous supposerons que $\mathcal{A}$ représente l'ensemble des arguments proposés par un agent rationnel à un moment donné, donc $\mathcal{A}$ sera considéré fini.

Un BAF peut être représenté par un graphe dirigé $\mathcal{G}^b$ appelé le *graphe d'interaction bipolaire*, avec deux types d'arcs, un pour la relation d'attaque ($\rightarrow$) et un pour la relation d'appui ($\rightsquigarrow$).

**Exemple 43 page 107 (suite)** *Pour l'exemple des journalistes, le graphe est :*



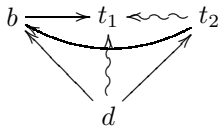

**Exemple 44 page précédente (suite)** *Pour l'exemple du meurtre, le graphe est :*




[33] Si la relation d'appui est supprimée, on retrouve exactement le cadre de Dung.
[34] En argumentation, cette contrainte de consistance est essentielle : il ne nous semble pas rationnel d'autoriser un argument qui pourrait à la fois attaquer et appuyer un autre argument.

**Exemple 45** *Une variante du "projet de randonnée" introduit à l'exemple 42 page 107 de la section 1.2.1 page 13 est la suivante : 3 agents (Tom, Ben et Dan) discutent au sujet d'une randonnée.*
- $t_1$ *: Aujourd'hui, nous avons le temps, faisons une randonnée.*
- $b$ *: Non, la météo est nuageuse et les nuages sont signe de pluie, il est plus prudent d'annuler la randonnée.*
- $t_2$ *: Ces nuages sont clairsemés, la journée va être ensoleillée et sans nuage ; donc nous pouvons faire la randonnée.*
- $d$ *: Non, ces nuages ne sont pas clairsemés. Donc la journée ne va pas être ensoleillée, il y aura des nuages. Toutefois, il ne pleuvra pas donc on peut faire la randonnée.*

*Le BAF associé peut être représenté par :*




Dans la suite, nous ferons abstraction de la structure des arguments et nous considérerons des relations arbitraires indépendantes $\mathcal{R}_{\text{att}}$ et $\mathcal{R}_{\text{sup}}$.

**Définition 78 (Chemin dans le graphe bipolaire)** *Soit* BAF $= \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ *un système d'argumentation abstrait bipolaire et* $\mathcal{G}^b$ *son graphe d'interaction bipolaire. Soit* $a$, $b \in \mathcal{A}$. *Un* chemin de $a$ à $b$ dans $\mathcal{G}^b$ *est une séquence* $(a_1, \ldots, a_n)$ *d'éléments de* $\mathcal{A}$ *telle que* $n \geq 2$, $a = a_1$, $b = a_n$, $a_1 \mathcal{R} a_2$, …, $a_{n-1} \mathcal{R} a_n$, *avec* $\mathcal{R} = \mathcal{R}_{att}$ *ou* $\mathcal{R}_{sup}$. *La longueur de ce chemin sera* $n - 1$.

*Remarquons que si* $n = 2$ *et* $a = b$ *alors le chemin est une* boucle *et si la relation* $\mathcal{R}$ *utilisée dans la boucle est* $\mathcal{R}_{att}$ *alors* $a$ *est dit* auto-attaqué.

L'usage de la bipolarité signifie qu'il existe plusieurs types d'interaction entre arguments et qu'ils vont donc avoir un impact sur la notion d'attaquant : dans l'exemple 43 page 107, le fait que $d$ appuie un attaquant de $b_1$ pourrait être considéré comme une sorte d'interaction négative entre $d$ et $b_1$, qui est toutefois plus faible qu'une attaque directe. Si on adopte un point de vue prudent, de tels arguments ne devraient pas apparaître ensemble dans la même extension. Pour traiter ce problème, dans [CLS05a, CLS05c] a été introduite une notion d'attaque combinant une séquence d'appuis avec une attaque directe.

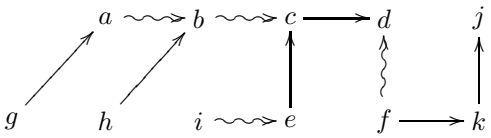
**Définition 79 (Séquence d'appuis)** *Soit* $a$, $b \in \mathcal{A}$. *Il existe une* séquence d'appuis pour $b$ par $a$ *(on dit que* $a$ appuie $b$*) si et seulement s'il existe une séquence* $(a_1, \ldots, a_n)$ *d'éléments de* $\mathcal{A}$ *telle que :* $n \geq 2$, $a = a_1$, $b = a_n$, $a_1 \mathcal{R}_{sup} a_2$, …, $a_{n-1} \mathcal{R}_{sup} a_n$.

**Définition 80 (Attaque appuyée)** *Une* attaque appuyée *pour un argument* $b$ *par un argument* $a$ *est une séquence* $(a, x, b)$ *d'arguments de* $\mathcal{A}$ *telle que* $a$ *appuie* $x$[35] *et* $x \mathcal{R}_{att} b$.

Dans l'exemple 43 page 107, il y a une attaque appuyée de $b_1$ par $d$.

Les définitions précédentes sont illustrées sur l'exemple suivant :

**Exemple 46** *Le graphe suivant représente un BAF.*




*Les chemins dans le graphe* $a - b - c - d$ *et* $i - c$ *correspondent à des attaques appuyées.*

On peut alors donner les définitions suivantes pour des ensembles d'arguments :

[35] Au sens de la définition 79.

**Définition 81 (Attaque et appui par des ensembles)** *Soit $S \subseteq \mathcal{A}$, soit $a \in \mathcal{A}$.*

$S$ attaque $a$ *si et seulement s'il existe une attaque appuyée ou une attaque directe sur $a$ à partir d'un élément de $S$.*

$S$ appuie $a$ *si et seulement s'il existe une séquence d'appuis pour $a$ à partir d'un élément de $S$.*

**Exemple 46 page précédente (suite)** *L'ensemble $\{a, h\}$ attaque $d$ et $b$ et appuie $b$ et $c$.*

### 3.3.2 Acceptabilité

Dans le cadre de Dung, *l'acceptabilité* d'un argument dépend de son appartenance à certains ensembles, appelés ensembles acceptables ou extensions. Ces extensions sont caractérisées par des propriétés particulières. Il s'agit d'une acceptabilité collective.

Dans le cadre bipolaire, nous allons suivre la même démarche que Dung et proposer pour les BAFs des sémantiques pour l'acceptabilité s'appuyant sur des propriétés caractéristiques qu'un ensemble d'arguments devra vérifier pour pouvoir être sélectionné en tant qu'extension du BAF. En général ces propriétés correspondent à une notion de cohérence et une notion de force : une extension doit être cohérente et avoir assez de force pour gagner une discussion. On demande aussi dans certains cas de respecter une notion de maximalité.

Nous considérons un système d'argumentation abstrait bipolaire dénoté par $\langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ et nous utilisons les notions "d'attaque" et "d'appui" donnés par la définition 81. Le premier point étudié va être la notion de cohérence, puis nous proposerons de nouvelles sémantiques pour l'acceptabilité dans un BAF.

#### 3.3.2.1 Gestion des conflits

Dans le cas d'un système d'argumentation basique (avec seulement des attaques), la notion de cohérence d'un ensemble d'arguments repose entièrement sur l'absence de conflits, donc d'attaque directe entre deux arguments de l'ensemble. Dans le cas bipolaire, ce concept de cohérence peut être étendu de deux manières différentes :

- en interdisant non seulement les attaques directes mais aussi les attaques appuyées ; cela permet de renforcer une sorte de cohérence *interne* : nous n'acceptons pas un ensemble $S$ d'arguments qui attaquerait un de ses éléments.
- en étendant la contrainte de consistance entre les relations d'attaque et d'appui ; cela conduit à une cohérence *externe* : nous n'acceptons pas d'ensemble $S$ d'arguments qui attaquerait *et* appuierait un même argument.

La première définition généralise la notion d'absence de conflit proposée par Dung :

**Définition 82 (Ensemble +sans-conflit)** *Soit $S \subseteq \mathcal{A}$. $S$ est* +sans-conflit[36] *si et seulement si $\nexists a, b \in S$ tels que $\{a\}$ attaque $b$.*

**Exemple 46 page précédente (suite)** *L'ensemble $\{h, b\}$ n'est pas +sans-conflit (il existe une attaque directe). L'ensemble $\{b, d\}$ n'est pas +sans-conflit puisque $d$ subit une attaque appuyée de la part de $b$. Par contre, $\{a, h\}$ et $\{b, f\}$ sont +sans-conflit.*

La cohérence externe est formalisée par[37] :

**Définition 83 (Ensemble sain)** *Soit $S \subseteq \mathcal{A}$. $S$ est* sain *si et seulement si $\nexists b \in \mathcal{A}$ tel que $S$ attaque $b$ et soit $S$ appuie $b$, soit $b \in S$.*

**Exemple 46 page ci-contre (suite)** *L'ensemble $\{a, h\}$ n'est pas sain puisque $a$ appuie $b$ et $h$ attaque $b$. L'ensemble $\{b, f\}$ n'est pas sain puisque $d$ subit une attaque appuyée de la part de $b$ et que $f$ appuie $d$. Par contre, $\{g, i, h\}$ est sain.*

Remarquons que la notion d'ensemble sain recouvre celle d'ensemble +sans-conflit :

[36] Cette notation signifie que vérifier si un ensemble est +sans-conflit permet de détecter davantage de conflits que vérifier qu'un ensemble est juste sans-conflit au sens de Dung.

[37] Cette définition est inspirée par [Ver03] et par la définition d'un argument controversé proposé par [Dun95].

**Proposition 69 ([CLS05c])** *Soit $S \subseteq \mathcal{A}$. Si $S$ est sain, alors $S$ est +sans-conflit. Si $S$ est +sans-conflit et fermé pour $\mathcal{R}_{sup}$ alors $S$ est sain.*

**Exemple 46 page 110 (suite)** *L'ensemble $\{g, h, i, e\}$ est +sans-conflit et fermé pour la relation $\mathcal{R}_{sup}$. Donc il est sain.*

#### 3.3.2.2 Nouvelles sémantiques pour l'acceptabilité

En suivant la méthodologie proposée par [Dun95], deux notions sont importantes pour définir des sémantiques basées sur les extensions : la notion de cohérence, et la notion de défense (c'est-à-dire l'attaque d'un attaquant). Dans un BAF, il existe plusieurs notions de cohérence, et deux types d'attaque (directe et appuyée). La notion de défense pourrait donc être étendue de différentes manières. Toutefois, nous avons choisi de garder la définition de la défense classique (celle proposée par Dung) pour les raisons suivantes. Tout d'abord parce qu'il s'agit ici de présenter les principes pris en compte dans le cadre bipolaire, plutôt que de faire une étude exhaustive de tous les cas possibles. Et enfin, parce que la plupart des travaux sur la bipolarité considèrent qu'un appui n'a pas la même force qu'une attaque et donc qu'un argument ne peut être défendu que si ses attaquants directs sont directement attaqués.

Cette remarque est illustrée par l'exemple suivant : $a_1 \rightleftarrows a_2 \rightsquigarrow a_3$

Il y a une attaque appuyée sur $a_1$ par $a_3$ et pas d'attaque sur $a_3$. Toutefois, $a_1$ attaque directement $a_2$ et cela paraît suffisant pour réinstaller $a_1$.

Dans la suite, on va étendre le concept d'admissibilité en renforçant la cohérence. Puis on donnera la définition des extensions pour la sémantique préférée qui respectera exactement les principes proposés par Dung : ce seront des ensembles admissibles maximaux pour l'inclusion ensembliste.

Rappelons juste la définition de la défense donnée dans [Dun95]

**Définition 84 (Défense par un ensemble d'arguments)** *Soit $S \subseteq \mathcal{A}$. Soit $a \in \mathcal{A}$.* $S$ défend $a$ *si et seulement si $\forall b \in \mathcal{A}$, si $b\mathcal{R}_{att}a$ alors $\exists c \in S$ tel que $c\mathcal{R}_{att}b$.*

On peut donner trois définitions différentes pour l'admissibilité, de la plus générale à la plus spécifique. La première est juste une traduction de la définition de Dung dans le cadre des BAF et correspond à la d-admissibilité[38].

**Définition 85 (Ensemble d-admissible)** *Soit $S \subseteq \mathcal{A}$.* $S$ *est* d-admissible *si et seulement si $S$ est +sans-conflit et défend tous ses éléments.*

Si on prend en compte la cohérence externe, on obtient la s-admissibilité[39].

**Définition 86 (Ensemble s-admissible)** *Soit $S \subseteq \mathcal{A}$.* $S$ *est* s-admissible *si et seulement si $S$ est sain et défend tous ses éléments.*

Et enfin, la cohérence externe peut être renforcée en exigeant que l'ensemble soit fermé pour la relation $\mathcal{R}_{\text{sup}}$. Cela donne la définition de la c-admissibilité[40].

**Définition 87 (Ensemble c-admissible)** *Soit $S \subseteq \mathcal{A}$.* $S$ *est* c-admissible *si et seulement si $S$ est +sans-conflit, clos pour $\mathcal{R}_{sup}$ et défend tous ses éléments.*

Il existe bien-sûr des liens entre tous ces ensembles : tout ensemble c-admissible est aussi s-admissible, et tout ensemble s-admissible est aussi d-admissible.

[38] "d" signifie "au sens de Dung".
[39] "s" signifie "sain".
[40] "c" signifie "clos pour $\mathcal{R}_{\text{sup}}$".

**Définition 88 (Extensions préférées)** *Un ensemble $S \subseteq \mathcal{A}$ est une* extension d-préférée *(resp.* s-préférée, c-préférée*) si et seulement si $S$ est maximal (pour l'inclusion ensembliste) parmi les sous-ensembles d-admissibles (resp. s-admissibles, c-admissibles) de $\mathcal{A}$.*

**Exemple 45 page 110 (suite)** *Dans ce cas, les trois sémantiques donnent le même résultat : $\{d, t_1\}$ est l'unique extension d-préférée, s-préférée et c-préférée.*

**Exemple 46 page 110 (suite)** *L'ensemble $\{g, h, i, e, f, d, j\}$ est l'unique extension c-préférée.*

**Exemple 47** *Considérons le système d'argumentation représenté par $a \rightsquigarrow b \leftarrow h$. L'ensemble $\{a, h\}$ est l'unique extension d-préférée. Il y a deux extensions s-préférées ($\{a\}$ et $\{h\}$) et une seule extension c-préférée ($\{h\}$).*

Une des propriétés les plus intéressantes concerne l'existence des extensions : l'existence de toutes ces extensions est garantie puisque l'ensemble vide est d-admissible (resp. s-admissible, c-admissible) et que tout ensemble d-admissible (resp. s-admissible, c-admissible) est inclus dans une extension d-préférée (resp. s-préférée, c-préférée).

Rappelons que des définitions analogues pour l'admissibilité pourraient être proposées en utilisant une notion de défense plus forte (par exemple en remplaçant dans la définition 84 page ci-contre, l'attaque directe par l'attaque au sens de la définition 81 page 111 – attaque directe ou attaque appuyée).

Voyons maintenant une autre sémantique bien connue, la sémantique stable. Ici-aussi, on garde le même type de définition que Dung mais en remplaçant l'attaque directe par l'attaque au sens de la définition 81 page 111.

**Définition 89 (Extension stable)** *$S$ est une* extension stable *si et seulement si $S$ est +sans-conflit et $\forall a \notin S$, $S$ attaque $a$.*

Dans la suite, nous allons nous restreindre à des BAFs acycliques afin d'obtenir certaines propriétés intéressantes. En effet, nous savons que dans le cadre de Dung quand un AF est acyclique il existe toujours une extension stable unique qui est aussi l'unique extension préférée. Dans les BAFs acycliques nous avons la même propriété : il existe une unique extension stable[41].

Par contre, cette unique extension stable n'est pas toujours saine.

**Exemple 47 (suite)** *L'ensemble $\{a, h\}$ est l'unique extension stable, et il n'est pas sain.*

Donc, un BAF acyclique peut ne pas avoir d'extension stable saine. On a la propriété suivante :

**Proposition 70 ([CLS05c])** *Soit $S$ une extension stable. Si $S$ est saine, alors $S$ est close pour $\mathcal{R}_{sup}$.*

Et la conséquence :

**Consequence 3** *Soit $S$ une extension stable. Alors $S$ est saine si et seulement si $S$ est close pour $\mathcal{R}_{sup}$.*

Exactement comme pour les extensions stables, la définition de la d-admissibilité implique l'existence d'une unique extension d-préférée pour un BAF acyclique.

Et enfin, nous avons les résultats suivants concernant la caractérisation des extensions s-préférées et c-préférées quand le BAF est acyclique :

**Proposition 71 ([CLS05c])** *Soit $S$ l'unique extension stable d'un BAF acyclique.*

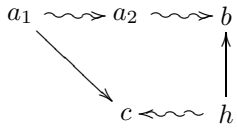
1. *$S$ est aussi l'unique extension d-préférée du BAF.*
2. *Les extensions s-préférées et c-préférées du BAF sont des sous-ensembles de $S$.*
3. *Toute extension s-préférée du BAF qui est close pour $\mathcal{R}_{sup}$ est aussi une extension c-préférée du BAF.*

[41] Il suffit d'instancier le cadre de Dung avec la relation d'attaque correspondant à la définition 81 page 111 et le graphe résultat est encore acyclique.

4. *Si $S$ est saine, alors $S$ est l'unique extension c-préférée du BAF et aussi l'unique extension s-préférée du BAF.*
5. *Si $\mathcal{A}$ est finie, alors chaque extension c-préférée du BAF est incluse dans une extension s-préférée du BAF.*
6. *Si $S$ n'est pas saine, les extensions s-préférées sont les sous-ensembles s-admissibles de $S$ qui sont maximaux (pour l'inclusion ensembliste).*
7. *Si $S$ n'est pas saine, et que $\mathcal{A}$ est finie, il existe une seule extension c-préférée.*

**Exemple 47 page précédente (suite)** *$\{h\}$ est la seule extension s-préférée qui est aussi close pour $\mathcal{R}_{sup}$. Donc, $\{h\}$ est l'unique extension c-préférée.*

**Exemple 48** *Considérons le BAF suivant :*




*$\{a_1, a_2, h\}$ est la seule extension d-préférée. $\{a_1, a_2\}$ et $\{h\}$ sont les deux seules extensions s-préférées. Aucune d'entre elles n'est close pour $\mathcal{R}_{sup}$. $\varnothing$ est l'unique extension c-préférée.*

*Si on ajoute un argument isolé $a_3$ (donc sans aucune interaction avec les autres arguments du BAF), on a alors : $\{a_1, a_2, a_3, h\}$ est la seule extension d-préférée. $\{a_1, a_2, a_3\}$ et $\{h, a_3\}$ sont les deux seules extensions s-préférées. Aucune n'est close pour $\mathcal{R}_{sup}$, et $\{a_3\}$ est l'unique extension c-préférée.*

En conclusion, dans le cas d'un BAF acyclique, deux sémantiques possèdent des propriétés intéressantes : la sémantique stable et la sémantique c-préférée. Si nous sommes davantage préoccupées par la cohérence interne, nous pouvons alors calculer l'unique extension stable, qui est aussi l'unique extension d-préférée. Si nous souhaitons respecter un concept de cohérence plus contraint, nous pouvons alors choisir la sémantique c-préférée, pour laquelle il existera aussi une unique extension.

### 3.3.3 Coalitions en argumentation bipolaire

Les sémantiques pour l'acceptabilité basées sur les extensions et introduites en section 3.3.2 page 111 utilisent explicitement les deux relations d'attaque et d'appui à travers en particulier la notion d'attaque appuyée. Ici, nous allons proposer une autre manière de définir des sémantiques pour l'acceptabilité basées sur les extensions. Cette nouvelle méthode se décompose en deux phases. Tout d'abord, nous allons traduire un BAF en un AF particulier appelé méta-système. Ce méta-système est constitué d'un ensemble de méta-arguments (appelés coalitions), et d'une relation de conflit entre ces méta-arguments. La relation d'attaque du BAF initial n'apparaîtra qu'au niveau du méta-système. Donc, un méta-argument permet de rassembler des arguments qui ne sont pas en conflit. Par opposition, la relation d'appui du BAF initial n'apparaîtra plus au niveau du méta-système, et sera utilisée uniquement pour construire les coalitions. L'idée exploitée ici est qu'un méta-argument n'a de sens que si ses éléments sont liés par la relation d'appui.

Donnons tout d'abord les principes de base pris en compte pour la définition d'une coalition :
– principe de cohérence : il n'existe pas d'attaque directe entre deux arguments d'une même coalition ;
– principe d'appui : si deux arguments appartiennent à une même coalition, ils doivent être reliés (directement ou indirectement) par la relation d'appui.

#### 3.3.3.1 Le concept de coalition

Soit $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ un BAF représenté par le graphe $\mathcal{G}^b$. $\mathcal{G}^b_{\text{sup}}$ dénotera le graphe partiel représentant le système partiel $\langle \mathcal{A}, \mathcal{R}_{\text{sup}} \rangle$[42]. AF dénotera le système d'argumentation partiel $\langle \mathcal{A}, \mathcal{R}_{\text{att}} \rangle$ associé à BAF et représenté par le graphe partiel dénoté $\mathcal{G}^b_{\text{att}}$.

[42] Dans toute cette section, nous considérons que le lecteur connaît les concepts de base de la théorie des graphes (notions de chaîne, de connexité, ...). Voir [Ber73] pour un état de l'art sur cette thématique.

**Définition 90 (Coalitions)** $C \subseteq \mathcal{A}$ *est une* coalition *de* BAF *si et seulement si*

**(i)** *Le sous-graphe de* $\mathcal{G}^b_{sup}$ *induit par* $C$ *est connexe ;*

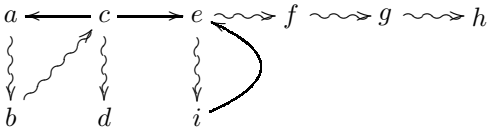
**(ii)** $C$ *est sans-conflit*[43] *pour* AF *;*

**(iii)** $C$ *est maximal (pour* $\subseteq$*) parmi les ensembles vérifiant les conditions* **(i)** *et* **(ii)**.

Remarquons que quand $\mathcal{R}_{\text{att}}$ est vide, les coalitions sont exactement les composantes connexes[44] du graphe partiel $\mathcal{G}^b_{\text{sup}}$.

**Proposition 72 ([CLS07])** *Tout argument qui ne s'auto-attaque pas appartient au moins à une coalition.*

**Exemple 49** *Soit* BAF *représenté par :*




*Les coalitions sont :* $C_1 = \{b, c, d\}$, $C_2 = \{i\}$, $C_3 = \{a, b\}$, $C_4 = \{e, f, g, h\}$.

Les propriétés suivantes montrent que les coalitions peuvent être définies à partir des composantes connexes d'un sous-graphe particulier, ce qui donne une manière constructive de les calculer.

**Proposition 73 ([CLS07])** $C \subseteq \mathcal{A}$ *est une coalition de* BAF *si et seulement si*

**(i)** *Il existe* $S \subseteq \mathcal{A}$ *maximal (pour* $\subseteq$*) sans-conflit pour* AF *tel que* $C$ *est une composante connexe du sous-graphe de* $\mathcal{G}^b_{sup}$ *induit par* $S$

**(ii)** *et* $C$ *est maximal (pour* $\subseteq$*) parmi les sous-ensembles de* $\mathcal{A}$ *vérifiant* **(i)**.

La proposition 73 suggère une procédure pour calculer les coalitions de BAF :

**Étape 1 :** Considérons AF et trouvons les ensembles maximaux sans-conflit de AF.

**Étape 2 :** Pour chaque ensemble d'arguments $S_i$ obtenu à l'étape 1, trouver les composantes connexes du sous-graphe de $\mathcal{G}^b_{\text{sup}}$ induit par $S_i$.

**Étape 3 :** Garder les ensembles maximaux (pour $\subseteq$) parmi ceux obtenus à l'étape 2.

Les ensembles sans-conflit correspondent aux ensembles indépendants :

**Proposition 74 ([CLS07])** *Soit* $S \subseteq \mathcal{A}$. $S$ *est sans-conflit pour* AF *si et seulement si* $S$ *est un sous-ensemble indépendant de* $\mathcal{A}$ *dans le graphe* $\mathcal{G}^b_{att}$.

Donc, $S$ est maximal (pour $\subseteq$) sans-conflit pour AF si et seulement si $S$ est un ensemble indépendant maximal (pour $\subseteq$) de sommets dans le graphe $\mathcal{G}^b_{\text{att}}$ et l'étape 1 de la procédure de calcul consiste à trouver tous les sous-ensembles indépendants maximaux (pour $\subseteq$) de $\mathcal{G}^b_{\text{att}}$. Remarquons que la complexité temporelle des meilleurs algorithmes réalisant ce calcul reste exponentielle dans le pire des cas[45]. Remarquons aussi qu'il existe dans la littérature plusieurs algorithmes pour trouver tous les ensembles indépendants maximaux[46] (voir par exemple le travail de J.M. Nielsen [Nie02]).

Nous savons aussi que :
– Pour l'étape 2, une exploration du graphe en profondeur permet de calculer les composantes connexes en un temps linéaire égal à $\mathcal{O}$(nombre de sommets + nombre d'arcs).
– Et pour l'étape 3, la maximisation pour $\subseteq$ est aussi un processus exponentiel.

[43] Au sens de base donné par Dung.

[44] Soit $G = (V, E)$ un graphe. Soit $S \subseteq V$. $S$ est une *composante connexe de* $G$ si et seulement si le sous-graphe de $G$ induit par $S$ est connexe et qu'il n'existe pas $S' \subseteq V$ tel que $S \subset S'$ et que le sous-graphe de $G$ induit par $S'$ soit connexe.

[45] En effet, le problème de décision suivant est NP-complet : "Soit $k$ un entier. Existe-t-il un ensemble indépendant maximal de taille $k$ ?"

[46] On part d'un sommet $v$, et on essaye de former un ensemble indépendant contenant $v$ ; pour cela, on trouve un sommet non adjacent à $v$ et on l'ajoute à $\{v\}$, puis on trouve un sommet adjacent à aucun des sommets déjà sélectionnés et on l'ajoute, et ainsi de suite, jusqu'à ce qu'on ne puisse plus rien ajouter ; à ce moment-là, l'ensemble est indépendant maximal.

#### 3.3.3.2 Un méta-système pour l'argumentation

Soit $C(\mathcal{A})$ dénotant l'ensemble des coalitions de BAF. Nous définissons la relation de conflit sur $C(\mathcal{A})$ de la manière suivante :

**Définition 91 (Relation C-attaque)** *Soit $C_1$ et $C_2$ deux coalitions de* BAF. $C_1$ C-attaque $C_2$ *si et seulement si il existe un argument $a_1$ dans $C_1$ et un argument $a_2$ dans $C_2$ tel que $a_1 \mathcal{R}_{att} a_2$.*

On peut montrer que :

**Proposition 75 ([CLS07])** *Soit $C_1$ et $C_2$ deux coalitions distinctes de* BAF. *Si $C_1$ et $C_2$ ne sont pas disjointes alors $C_1$ C-attaque $C_2$ ou $C_2$ C-attaque $C_1$.*

Ainsi, un nouveau système d'argumentation $\mathsf{CAF} = \langle C(\mathcal{A}), \text{c-attacks} \rangle$ peut être défini ; c'est le système à coalition associé à BAF.

**Exemple 49 page précédente (suite)** *Dans cet exemple,* CAF *peut être représenté par (pour simplifier les notations, $\rightarrow$ représente à la fois la relation d'attaque dans* BAF *et la relation C-attaque dans* CAF*) :*

$$C_3 \longleftarrow C_1 \longrightarrow C_4 \longleftarrow C_2$$

Les définitions de Dung s'appliquent directement sur CAF, et on a :

**Proposition 76 ([CLS07])** *Soit $\{C_1, \ldots, C_p\}$ un ensemble fini de coalitions distinctes deux à deux. $\{C_1, \ldots, C_p\}$ est sans-conflit pour* CAF *si et seulement si $C_1 \cup \ldots \cup C_p$ est sans-conflit pour* AF.

Donc, CAF est un "méta-système d'argumentation" avec un ensemble de "méta-arguments" (l'ensemble des coalitions $C(\mathcal{A})$) et une relation de "méta-attaque" sur ces coalitions (la relation C-attaque). Une coalition rassemble les arguments qui sont proches au sens de l'appui et qui peuvent être produits ensemble (pas d'attaque). Toutefois, comme les coalitions peuvent être en conflit, on peut appliquer la méthodologie de Dung, et calculer les extensions préférées et stables de CAF. De telles extensions contiendront les coalitions acceptables collectivement. La dernière étape consistera alors à rassembler les éléments des différentes coalitions d'une extension de CAF. Ainsi, les meilleurs groupes d'arguments du BAF (du point de vue des interactions données) auront été sélectionnés.

**Définition 92 (Cp-(Cs-)extension)** *Soit $S \subseteq \mathcal{A}$.*

*$S$ est une* Cp-extension[47] *de* BAF *si et seulement si il existe $\{C_1, \ldots, C_p\}$ une extension préférée de* CAF *telle que $S = C_1 \cup \ldots \cup C_p$.*

*$S$ est une* Cs-extension[48] *de* BAF *si et seulement si il existe $\{C_1, \ldots, C_p\}$ une extension stable de* CAF *telle que $S = C_1 \cup \ldots \cup C_p$.*

Quand l'extension préférée de CAF est l'ensemble vide, par définition nous considérons alors que l'ensemble vide est aussi l'unique Cp-extension de BAF.

**Exemple 49 page précédente (suite)** *Il n'existe qu'une extension préférée de* CAF*, qui est aussi stable : $\{C_1, C_2\}$. Donc, $S = \{b, c, d, i\}$ est la Cp-extension (et aussi la Cs-extension) de* BAF.

Certaines propriétés intéressantes du cadre de Dung sont conservées :

- Un BAF possède toujours (au moins) une Cp-extension. C'est une conséquence de la définition 92.
- Par contre, il n'existe pas toujours une Cs-extension de BAF. Cela vient du fait qu'il n'existe pas toujours une extension stable de CAF.
- Chaque Cs-extension est aussi une Cp-extension. L'inverse est faux.
- Il ne peut pas exister deux Cp-extensions telles que l'une contiendrait strictement l'autre. Cela provient des définitions 90 page précédente et 92.

[47] Cp signifie Coalition-préférée.
[48] Cs signifie Coalition-stable.

Malheureusement, d'autres propriétés importantes sont perdues. L'exemple suivant montre qu'une Cp-extension n'est pas toujours admissible pour AF, et qu'une Cs-extension n'est pas toujours une extension stable de AF.

**Exemple 50** *Soit* BAF *représenté par :*

$$a \longrightarrow b \leftsquigarrow c \rightsquigarrow d \longrightarrow e$$

*Les coalitions sont :* $C_1 = \{a\}$, $C_2 = \{b, c, d\}$, $C_3 = \{e\}$. *Et* CAF *peut être représenté par :*

$$C_1 \longrightarrow C_2 \longrightarrow C_3$$

*Il n'existe qu'une seule extension préférée de* CAF*, qui est aussi stable :* $\{C_1, C_3\}$. *Donc,* $S = \{a, e\}$ *est la Cp-extension (et aussi la Cs-extension) de* BAF. *Nous avons* $d\mathcal{R}_{att}e$*, mais* $a$ *ne défend pas* $e$ *contre* $d$ *(ni par une attaque directe, ni par une attaque appuyée, bien que* $a$ *attaque un élément de la coalition qui attaque* $e$*). Donc,* $S$ *n'est pas admissible pour* AF. $S$ *ne contient pas* $c$*, mais il n'existe aucune attaque (ni directe, ni appuyée) d'un élément de* $S$ *contre* $c$*. Donc,* $S$ *n'est pas une extension stable de* AF.

Remarquons que cet inconvénient s'explique par le fait qu'une coalition est considérée comme un tout et que ses membres ne peuvent pas être utilisés séparément dans le processus d'attaque.

L'exemple 50 suggérerait que l'admissibilité est perdue à cause de la taille de la coalition $\{b, c, d\}$, et qu'il pourrait être plus profitable de considérer deux coalitions indépendantes $\{c, b\}$ et $\{c, d\}$. Une nouvelle formalisation de coalitions en terme de chemins d'appuis maximaux sans-conflit a été proposée dans [CLS07], mais elle ne permet pas de retrouver toutes les propriétés de Dung.

En ce qui concerne la perte de l'admissibilité au sens de Dung, ce n'est pas surprenant puisque :
– l'admissibilité est une notion qui prend en compte les attaques et les défenses "individuelles" ;
– alors que, avec les coalitions et le méta-système d'argumentation, on considère les attaques et les défenses "collectives".

### 3.3.4 Évaluation d'arguments

Cette section propose pour les BAFs quelques sémantiques pour l'acceptabilité basées sur la notion d'étiquetage. Une telle sémantique consiste à définir un ensemble d'étiquettes et une procédure permettant d'étiqueter les arguments. Puis, on choisit les arguments acceptés en fonction des étiquettes attribuées. Un exemple de ce type de sémantique dans le cadre de base de l'argumentation est donné dans [JV99b] avec la sémantique dite robuste. Plus généralement, de nombreuses approches ont été proposées pour évaluer les arguments dans un cadre classique (par exemple [KAEF95, Par97, PS97, BH01, CLS03c, Amg99]). Dans certains d'entre eux, la valeur d'un argument dépend de ses interactions avec les autres arguments ; dans d'autres, elle dépend de la force intrinsèque de l'argument.

Dans un cadre bipolaire, il existe aussi quelques travaux. Par exemple, Karacapilidis & Papadias [KP01] ont proposé un étiquetage pour un graphe bipolaire représentant un débat en vue d'une prise de décision, mais avec seulement deux étiquettes : active et inactive.

Dans cette section, nous nous intéresserons aux évaluations basées sur les interactions dans le cadre des BAFs (l'étape suivante serait d'exploiter ces évaluations pour définir complètement les sémantiques de manière analogue à ce qui avait été proposé pour un système d'argumentation unipolaire – voir [CLS05b] et les sections 3.2.2 page 38 et 3.2.3 page 55).

Nous allons tout d'abord rappeler très brièvement les concepts utilisés dans HERMES [KP01]. Puis nous présenterons une approche générique permettant d'étiqueter un BAF.

#### 3.3.4.1 Étiquetages dans HERMES

Karacapilidis & Papadias [KP01] proposent un outil web basé sur l'argumentation pour faire de la prise de décision dans le domaine médical. Cet outil, appelé HERMES, permet l'expression et la pondération d'arguments, vérifie la

cohérence des préférences entre arguments et évalue les arguments. Les éléments de base de ce système sont : les *issues* (questions dont la réponse est ouverte à la discussion[49]), des *positions* exprimant l'appui ou l'opposition à une solution, à une autre position, ou à une contrainte (une position donne une information pour la discussion) et des *contraintes* exprimant une préférence entre deux positions (donc, permettant de comparer des ensembles de positions). HERMES peut étiqueter les solutions et les positions avec le statut "active" ou "inactive". À la fin de la discussion, les positions "actives" (resp. "inactives") sont acceptées (resp. rejetées). Une solution "active" est un choix recommandé parmi les autres solutions concernant une même issue. Différents étiquetages sont proposés dans HERMES. Ils sont récursifs : l'étiquette d'un élément $e$ dépendant des étiquettes des éléments liés à $e$ dans le graphe de discussion. Dans HERMES, ce graphe de discussion est acyclique, la valeur d'une position $p$ dépendant uniquement des positions *actives* liées à $p$ ; de plus, la valeur d'une position est toujours binaire, même quand des contraintes de préférence sont prises en compte.

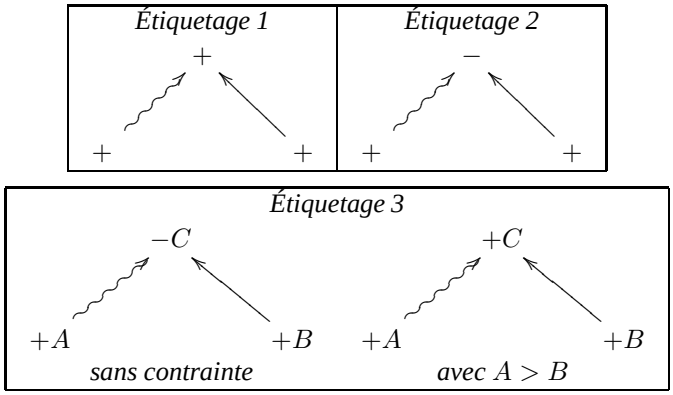
**Étiquetage 1 :** Une position est active si et seulement si il n'existe ni appui, ni attaque sur cette position, ou si elle est appuyée par une position active.

**Étiquetage 2 :** Une position est active si et seulement si elle n'est pas attaquée par une position active.

Remarquons que les étiquetages 1 et 2 ne permettent pas de prendre en compte en même temps les appuis et les attaques sur une position.

**Étiquetage 3 :** Une position est active si et seulement si il n'existe ni appui, ni attaque sur cette position, ou si son score est strictement positif. Le score d'une position $p$ est défini par : $\Sigma_i w(p_i) - \Sigma_j w(p_j)$ avec $p_i$ les positions *actives* qui appuient $p$ et $p_j$ les positions *actives* qui attaquent $p$. Chaque position a le même poids initial et la prise en compte des préférences entre positions permet de modifier les poids relatifs des positions.

**Exemple 51** *Une position active (resp. inactive) sera dénotée par + (resp. –).*




#### 3.3.4.2 Évaluation bipolaire graduelle

L'approche présentée ici (voir [CLS05a]) se place dans le cadre suivant :
- Le processus d'évaluation se fera avant le processus de sélection.
- Le processus d'évaluation utilisera un ensemble de valeurs assez riche (pas seulement deux comme dans HERMES). Donc, il s'agira d'une évaluation graduelle.
- Les valeurs attribuées aux arguments prendront en compte tous les attaquants et supporters directs de ces arguments (ce qui n'est pas le cas dans l'outil HERMES dans lequel la valeur d'un argument dépend seulement des positions *actives*). Ce sera donc une évaluation locale.

[49] Par exemple, si le patient $Y$ a la pathologie $X$, quel est le traitement approprié ? Une issue est un ensemble de *solutions*. Les exemples de solutions peuvent être ici une opération chirurgicale ou l'utilisation de médicaments.

Cette proposition étend à des BAFs, tels que définis dans la section 3.3.1 page 109, le travail déjà réalisé par [JV99b, BH01, CLS05b]. Elle va respecter le même type de principes que ceux déjà décrits dans [CLS03c] et dans la section 3.2.2 page 38 auxquels on rajoute bien-sûr de nouveaux principes pour prendre en compte la relation d'"appui" :
– **P1** : L'évaluation d'un argument dépend des valeurs de ses attaquants directs et de ses supporters directs.
– **P2** : Si la qualité de l'appui (resp. attaque) augmente alors la valeur de l'argument augmente aussi (resp. décroît).
– **P3** : Si la quantité des appuis (resp. attaques) augmente alors la qualité de l'appui (resp. attaque) augmente aussi.
La valeur d'un argument est le résultat de la composition de plusieurs fonctions :

– une pour agréger les valeurs de tous les attaquants directs ; cette fonction calcule la valeur de l'"attaque" ;
– une pour agréger les valeurs de tous les supporters directs ; cette fonction calcule la valeur de l'"appui" ;
– une pour calculer les effets *combinés* de l'attaque et de l'appui sur la valeur de l'argument.

Dans le respect des principes donnés, nous supposerons qu'il existe un ensemble totalement ordonné $\mathcal{V}$ possédant un élément minimum $\mathcal{V}_{\min}$ et un élément maximum $\mathcal{V}_{\max}$. Dans ce cadre, la définition formelle pour une évaluation bipolaire graduelle sera :

**Définition 93** *Soit $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ un système d'argumentation bipolaire. Soit $a \in \mathcal{A}$ avec ${\mathcal{R}_{att}}^-(A) = \{b_1, \ldots, b_n\}$ et ${\mathcal{R}_{sup}}^-(A) = \{c_1, \ldots, c_p\}$.*

*Une évaluation bipolaire graduelle sur $\langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ est une fonction $v : \mathcal{A} \to \mathcal{V}$ telle que :*

$$v(a) = g(h_{sup}(v(c_1), \ldots, v(c_p)), h_{att}(v(b_1), \ldots, v(b_n)))$$ *avec*

– *la fonction $h_{att}$ (resp. $h_{sup}$) : $\mathcal{V}^* \to \mathcal{H}_{att}$ (resp. $\mathcal{V}^* \to \mathcal{H}_{sup}$)*[50] *évaluant la qualité de l'attaque (resp. appui) sur $a$,*
– *et la fonction $g : \mathcal{H}_{sup} \times \mathcal{H}_{att} \to \mathcal{V}$ avec $g(x, y)$ croissante en $x$ et décroissante en $y$.*

*La fonction $h$, $h = h_{att}$ ou $h_{sup}$, doit vérifier les propriétés suivantes :*

1. *si $x_i \geq x'_i$ alors $h(x_1, \ldots, x_i \ldots, x_n) \geq h(x_1, \ldots, x'_i \ldots, x_n)$,*
2. *$h(x_1, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_n)$,*
3. *$h() = \alpha \leq h(x_1, \ldots, x_n) \leq \beta$, pour tout $x_1, \ldots, x_n$*[51].

La définition 93 produit une évaluation locale générique, dont nous pouvons donner quelques instances à titre d'exemple pour illustrer les divers principes choisis.

– Une première instance est définie par $\mathcal{H}_{\text{att}} = \mathcal{H}_{\text{sup}} = \mathcal{V} = [-1, 1]$ intervalle de réels, $h_{\text{att}}(x_1, \ldots, x_n) = h_{\text{sup}}(x_1, \ldots, x_n) = \mathcal{M}\mathtt{ax}(x_1, \ldots, x_n)$, et $g(x, y) = \frac{x-y}{2}$ (donc, on a $\alpha = -1$, $\beta = 1$ et $g(\alpha, \alpha) = 0$).
– Une autre instance se définit par $\mathcal{V} = [-1, 1]$ intervalle de réels, $\mathcal{H}_{\text{att}} = \mathcal{H}_{\text{sup}} = [0, \infty[$ intervalle de réels, $h_{\text{att}}(x_1, \ldots, x_n) = h_{\text{sup}}(x_1, \ldots, x_n) = \Sigma_{i=1}^{n} \frac{x_i+1}{2}$, et $g(x, y) = \frac{1}{1+y} - \frac{1}{1+x}$ (donc, on a $\alpha = 0$, $\beta = \infty$ et $g(\alpha, \alpha) = 0$[52]).

Le tableau suivant donne pour les deux instances les valeurs obtenues sur quelques cas de bases :

| Exemple | avec la 1ère instance | avec la 2ème instance |
|---|---|---|
| Pas d'attaque, pas d'appui : $a$ | $v(a) = 0$ | $v(a) = 0$ |
| Attaque directe : $a \longrightarrow b$ | $v(b) = -0.5$ | $v(b) = -0.33$ |
| Appui direct : $a \rightsquigarrow b$ | $v(b) = 0.5$ | $v(b) = 0.33$ |
| Défense : $a \longrightarrow b \longrightarrow c$ | $v(c) = -0.25$ | $v(c) = -0.25$ |
| Séquence d'appuis : $a \rightsquigarrow b \rightsquigarrow c$ | $v(c) = 0.75$ | $v(c) = 0.4$ |
| Attaque appuyée : $a \rightsquigarrow b \longrightarrow c$ | $v(c) = -0.75$ | $v(c) = -0.4$ |

[50] $\mathcal{V}^*$ dénote l'ensemble des séquences finies d'éléments de $\mathcal{V}$, incluant la séquence vide. $\mathcal{H}_{\text{att}}$ et $\mathcal{H}_{\text{sup}}$ sont des ensembles ordonnés.

[51] Donc, $\alpha$ est la valeur minimale pour l'attaque (resp. appui) – *i.e.* quand il n'existe aucune attaque (resp. aucun appui) –, et $\beta$ est la valeur maximale pour l'attaque (resp. appui).

[52] Remarquons que $h_{\text{att}}(x_1, \ldots, x_n, x_{n+1}) \geq h_{\text{att}}(x_1, \ldots, x_n)$ car $\frac{x_{n+1}+1}{2} \geq 0$ quand $x_{n+1} \in [-1, 1]$ (idem pour $h_{\text{sup}}$). On a aussi $h_{\text{att}}() = h_{\text{sup}}() = \alpha$, $\alpha$ étant la valeur minimale de $[0, \infty[$, et $\beta$ étant la valeur maximale de $[0, \infty[$. On peut aussi vérifier que $g(\alpha, \beta) = g(0, \infty) = -1$ et que $g(\beta, \alpha) = g(\infty, 0) = 1$ (1 et $-1$ étant respectivement $\mathcal{V}_{\min}$ et $\mathcal{V}_{\max}$).

**Exemple 45 page 110 (suite)** *Avec la première (resp. seconde) instance,* $v(t_1) = \frac{1}{4}$ *(resp.* $\frac{37}{154}$*).*

**Exemple 47 page 113 (suite)** *Avec les deux instances,* $v(b) = 0$*. Dans ce cas, il existe un équilibre parfait entre attaque et appui*[53].

Une évaluation locale graduelle telle que nous l'avons définie respectera les propriétés suivantes [CLS05a] :
- Si $\mathcal{R}_{\text{att}}^{-}(a) = \mathcal{R}_{\text{sup}}^{-}(a) = \varnothing$ alors $v(a) = g(\alpha, \alpha)$.
- Si $\mathcal{R}_{\text{att}}^{-}(a) \neq \varnothing$ et $\mathcal{R}_{\text{sup}}^{-}(a) = \varnothing$ alors $v(a) = g(\alpha, y) \leq g(\alpha, \alpha)$ pour $y \geq \alpha$.
- Si $\mathcal{R}_{\text{att}}^{-}(a) = \varnothing$ et $\mathcal{R}_{\text{sup}}^{-}(a) \neq \varnothing$ alors $v(a) = g(x, \alpha) \geq g(\alpha, \alpha)$ pour $x \geq \alpha$.

Et nous pouvons utiliser alors l'échelle comparative suivante :

$$\mathcal{V}_{\min} \leq \underset{(\text{pour } y \geq \alpha)}{g(\alpha, y)} \leq g(\alpha, \alpha) \leq \underset{(\text{pour } x \geq \alpha)}{g(x, \alpha)} \leq \mathcal{V}_{\max}$$

De plus, l'évaluation proposée dans la définition 93 page précédente satisfait les principes **P1** à **P3** (voir [CLS05a] pour une discussion plus détaillée).

## 3.3.5 Conclusion sur la bipolarité

Dans cette section, nous avons présenté une extension du système abstrait proposé par [Dun95] qui soit capable de prendre en compte des interactions bipolaires, c'est-à-dire à la fois des interactions négatives (les attaques) et des interactions positives (les supports). Dans ce cadre, le travail réalisé couvre :
- la définition de nouvelles sémantiques, soit en renforçant la notion de cohérence utilisée dans la définition de l'admissibilité, soit en définissant un méta-système basé sur la notion de coalitions ;
- et la proposition de nouvelles évaluations.

Il faut noter que sur le sujet des systèmes d'argumentation bipolaires, j'ai réalisé d'autres travaux que je n'ai pas détaillés ici (en particulier, le traitement des arguments controversés dans un BAF, travail que j'ai fait en collaboration avec Claudette CAYROL et Caroline DEVRED – voir [CDLS06a]).

En ce qui concerne les 3 principaux points abordés dans cette section (acceptabilité, coalition et évaluation dans un cadre bipolaire), il existe dans la littérature d'autres travaux qui les ont abordés aussi.

### 3.3.5.1 Acceptabilité et bipolarité

*Deflog [Ver03] :* Dans ce système on trouve la possibilité d'exprimer des supports et des attaques sur des phrases du langage. Cela se fait par l'intermédiare de connecteurs spécifiques (1 pour chaque type d'interaction). Des phrases possibles de ce langage (avec les symboles $\rightarrow$ pour l'attaque et $\rightsquigarrow$ pour le support) sont par exemple : $a$, $b$, $(a \rightsquigarrow b)$, $(a \rightarrow b)$, $(c \rightsquigarrow (a \rightsquigarrow b))$, $(d \rightarrow (a \rightsquigarrow b))$. Dans DEFLOG, on retrouve les notions de séquence de supports et d'attaque supportée mais au niveau du langage (donc entre les phrases). Un autre point commun est le fait que la notion d'ensemble sans-conflit proposé dans DEFLOG correspond à notre notion d'ensemble sain.

Il y a d'autre part des liens très forts entre les extensions données par le système DEFLOG et les extensions définies dans [Dun95] (voir [Ver03]).

*Argumentation basée sur l'évidence [ON08] :* Dans ce travail, l'hypothèse de base est qu'un argument ne pourra être accepté que s'il est soutenu par une évidence (dans le sens d'une preuve, d'un témoignage). Cela donne donc une classe particulière d'arguments : les arguments *prima-facie* (qui n'ont besoin d'aucun support).

A partir de là, les arguments peuvent être acceptables s'ils sont supportés (indirectement) par des arguments *prima-facie* (on appelle cela le support par évidence). On peut aussi avoir des arguments supportés qui en attaquent d'autres. Cela donne une notion de défense un peu plus complexe que celle que nous utilisons : un ensemble d'arguments $S$ défend un argument $a$ si $S$ fournit un support par évidence pour $a$ et que $S$ invalide chaque attaque sur $a$ (soit par une attaque directe de l'attaquant de $a$ soit en supprimant le support de cette attaque).

[53] Notons que cela n'est pas nécessairement le cas et que le choix de la fonction $g$ permet d'attribuer plus ou moins d'importance à l'un des facteurs (l'attaque ou l'appui).

Notre proposition est un peu plus souple car nous n'imposons pas l'existence des supports. Toutefois la proposition faite dans [ON08] peut sans problème être traduite sous la forme d'un BAF de la manière suivante : Soit $X$ un ensemble d'arguments (considérés comme *prima-facie*), le support par évidence peut être défini comme une séquence de supports venant d'un des arguments de $X$. Puis, la notion d'attaquant peut se réduire aux éléments de $X$, ou à ceux qui recoivent un support par évidence de $X$. On modifiera alors en conséquence la notion de défense.

#### 3.3.5.2 Coalitions d'arguments

Nous avons utilisé cette notion pour proposer un méta-système d'argumentation en considérant qu'une coalition était un ensemble d'arguments relié par la relation support et n'étant pas en conflit.

Ce concept de coalition a déjà été utilisé en argumentation.

*Argumentation collective [Boc03, NP06] :* Un système d'argumentation collective est un système abstrait dans lequel les données initiales sont un ensemble d'arguments et une relation binaire d'"attaque" entre *ensembles* d'arguments. L'idée principale est qu'un ensemble d'arguments peut générer une attaque contre d'autres arguments, qui ne serait pas réductible à une attaque entre arguments particuliers. Cela est en accord avec notre idée de coalition que nous considérons comme un tout capable d'attaquer une autre coalition. Les propositions de Nielsen et Parsons et de Bochman sont similaires et, partant de cette idée d'attaque entre ensembles d'arguments, proposent des sémantiques sur le même principe que Dung (ensemble de propriétés à vérifier). Notre proposition est un peu différente de ces travaux sur deux points. Premièrement, bien que nous appliquions exactement la même méthode que Dung, nous le faisons sur le système d'argumentation construit à partir des coalitions. Et deuxièmement, nous attribuons une signification particulière à la notion de coalition : elle doit être la plus grande possible et elle sera utilisée dans sa globalité lors du processus de défense.

*Génération de structures de coalition dans les systèmes multi-agents (MAS) [DJ04, Amg05] :* Dans les MAS, la formation de coalitions est un processus réalisé par des agents indépendants et autonomes qui souhaitent agir de concert. Une structure de coalition (CS) est une partition de l'ensemble des agents en coalitions. Chaque coalition se voit attribuer une valeur représentant la différence pour chaque agent entre ses profits et ses dépenses dues à l'appartenance de l'agent à la coalition. Dans ce cadre, la génération d'une CS optimisant la valeur globale de la structure est un problème important. Pour le résoudre, [Amg05] a proposé un système d'argumentation abstrait dans lequel les données initiales sont un ensemble de coalitions et une relation de conflits entre ces coalitions (par exemple, il peut y avoir conflit entre deux coalitions si un même agent appartient aux deux coalitions ou si ces coalitions doivent réaliser la même tâche). On peut alors définir une CS préférée comme un ensemble de coalitions sans-conflit et qui se défend contre toute attaque.

Par contre, le problème de la création des coalitions n'est pas traité dans [Amg05]. De par sa généricité, notre travail pourrait donc servir à la formation de ces coalitions, en amont du travail fait par [Amg05] ; dans ce cas, les arguments seraient les agents et les interactions représenteraient les volontés (relation de support) ou les refus (relation d'attaque) de coopérations entre agents.

#### 3.3.5.3 Evaluation et bipolarité

Il existe quelques autres propositions d'évaluation de système d'argumentation bipolaire, mais la plupart du temps elles ont été développées pour des applications spécifiques.

*Applications médicales :* Le travail le plus important ici correspond au système HERMES décrit dans [KP01]. Toutefois ce système ne propose pas de valeurs graduelles et ne prend en compte qu'une partie des arguments en interaction lors du calcul de la valeur (voir la section 3.3.4.1 page 117).

*Cartes valuées d'argumentations :* Il s'agit de travaux dans le cadre de l'annotation collective de documents utilisant la notion de bipolarité.

Cette annotion collective se fait à l'aide de fils de discussion. Chaque fil démarre par une annotation sur un sujet donné. Puis les utilisateurs du fil répondent en mettant d'autres annotations qui confirment ou réfutent des annotations

déjà existantes. Ces annotations sont associées à une validation sociale qui donne une vue synthétique des discussions. Le but de cette validation est d'identifier les annotations confirmées par la discussion tout en prenant en compte les valeurs intrinsèques des annotations.

Dans [CCCJ05, CCCJ07] un fil de discussion est modélisé par un BAF. Chaque argument correspond à un nœud du fil de discussion. Les supports (resp. les attaques) correspondent aux réponses données dans le fil qui confirment (resp. réfutent). La validation sociale est alors calculée avec une évaluation bipolaire graduelle. Cette évaluation est issue de nos travaux et a été légèrement modifiée pour prendre en compte la valeur intrinséque des annotations.

## 3.4 Conclusion sur le thème 3

Le thème de l'argumentation a été un des thèmes centraux de ma recherche depuis de nombreuses années. Ce thème se compose d'au moins cinq aspects tous très riches en potentialité de recherche et en travaux déjà réalisés.

Tout d'abord, il s'agit de construire les arguments à partir d'un ensemble d'informations en général incohérent. Dans cet aspect, on retrouve le cadre élargi de la restauration de cohérence, puisque les arguments peuvent être vus comme des sous-ensembles cohérents issus de l'ensemble incohérent initial (même s'il existe de très nombreuses façons de représenter ces arguments et donc plusieurs langages de représentation qui ne sont pas toujours des langages logiques).

Ensuite, il faut identifier les interactions existant entre tous ces arguments sachant qu'en général on ne traite que le cas des interactions binaires.

Je n'ai absolument pas travaillé sur ces deux aspects. Je suis partie, comme beaucoup d'autres chercheurs, de l'hypothèse suivante : les arguments et leurs interactions sont considérés comme donnés (le système d'argumentation est dit abstrait).

Par contre, les deux aspects suivants ont été au centre de mon travail sur ce sujet : l'évaluation des arguments et leur sélection. Dans ce contexte, j'ai étudié deux types de systèmes abstraits, ceux dans lesquels il n'existe qu'une seule interaction de type conflictuelle (les systèmes d'argumentation unipolaires), et ceux dans lesquels il existe aussi une interaction de type support (les systèmes d'argumentation bipolaires).

L'évaluation peut se faire à partir de divers éléments. Par exemple, on pourrait prendre en compte des préférences sur les connaissances ayant servi à créer les arguments. Toutefois mon hypothèse initiale d'un système abstrait ne me permet pas d'utiliser ce type d'information. Je n'ai donc à ma disposition qu'un graphe orienté non valué avec des sommets (les arguments) et des arcs (les interactions). J'ai donc proposé des évaluations d'arguments basées sur les interactions pour des systèmes d'argumentation unipolaires et bipolaires. Il est intéressant de noter que cette proposition recouvre et même généralise des propositions déjà existantes.

Pour la partie sélection (appelée acceptabilité), j'ai aussi travaillé sur les deux types de systèmes, unipolaires et bipolaires, et tous mes travaux ont visé soit à exploiter les évaluations proposées précédemment, soit à généraliser les travaux de Dung sur ce sujet. L'idée de Dung était de sélectionner des arguments par paquets "cohérents"[54] et représentant certaines propriétés. Il s'agit donc d'une acceptabilité collective (un argument sera sélectionné parce qu'il appartient à un ensemble d'arguments respectant des contraintes données). J'ai gardé cette idée d'acceptabilité collective ainsi que la liste des propriétés à satisfaire que j'ai enrichie par de nouvelles propriétés afin de pouvoir prendre en compte de nouveaux facteurs dont le plus important a été la bipolarité du système d'argumentation.

Quant au dernier aspect du processus d'argumentation, j'avoue de pas avoir eu le temps de m'y intéresser. Il est pourtant essentiel pour "reboucler" sur le thème de l'inférence non monotone qui fût à l'origine de mes recherches et il sera à prendre aussi en compte afin de caractériser et d'axiomatiser le processus d'argumentation dans sa globalité (voir mon projet de recherche décrit au chapitre 5 page 153).

Mon travail en argumentation a aussi consisté à "porter" dans ce thème des mécanismes du raisonnement non monotone, comme la fusion ou la révision. Dans le premier cas, il s'agissait de définir la fusion de plusieurs systèmes

[54] Ici on se situe au niveau d'un ensemble d'arguments et plus au niveau d'un ensemble d'informations comme on a pu le faire en restauration de cohérence (voir les thèmes 1 et 2 – chapitres 1 page 9 et 2 page 21). Ceci étant, l'idée est assez semblable puisque l'ensemble complet des arguments est "incohérent" au sens où il existe des conflits/attaques entre ces arguments. On pourrait presque dire, par abus de langage, que l'approche proposée par Dung est de la "restauration de cohérence".

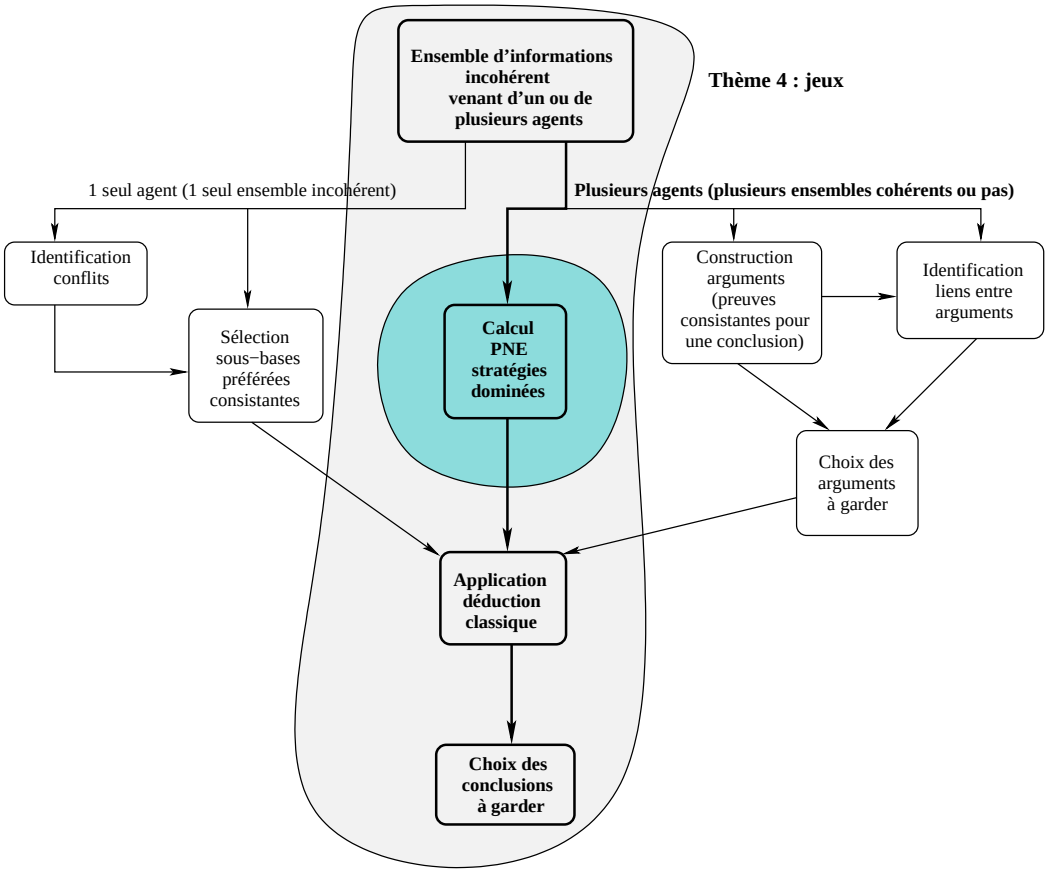
d'argumentation (toujours abstraits mais uniquement unipolaires) ; cela nous a conduit à introduire un nouveau type d'interaction, "l'ignorance", et donc un nouveau type de système d'argumentation. Et dans le second cas, nous avons proposé une étude de la notion de révision d'un système d'argumentation ; cela nous a permis d'identifier et de caractériser divers types de révision.

Et enfin, mes derniers travaux sur ce thème ont été un peu plus "applicatifs", l'un concernant l'utilisation d'un système d'argumentation pour résoudre un problème particulier, celui du *practical reasoning* et l'autre est fortement lié au thème 4, les jeux, puisqu'il s'agit de traduire un système d'argumentation en un jeu booléen. Ce dernier travail n'a pas été présenté dans ce chapitre mais sera décrit en section 4.3.3 page 145. Ce qui m'a amené à m'intéresser alors à la théorie des jeux est le fait que ces deux thèmes sont déjà fortement liés dans la littérature par l'aspect "échange d'informations" du processus d'argumentation qui est très proche de la notion de jeu. Dans le chapitre suivant, je vais donc creuser un peu plus cette question.

# Chapitre 4

# Thème 4 : raisonnement à base de jeux

Tous les travaux présentés dans ce chapitre correspondent à la partie “entourée foncée” donnée sur le schéma suivant :




L'étude sur les jeux que j'ai menée en collaboration avec Jérôme LANG, Élise BONZON, Bruno ZANUTTINI et Denis SIREYJOL a débuté tout d'abord comme une suite de mon travail sur l'argumentation. En effet, l'argumentation peut être vue comme un processus d'interaction entre agents.

Or, la théorie des jeux a pour principal intérêt la modélisation de ces interactions, et cela dans beaucoup de contextes différents (elle est en particulier beaucoup utilisée en économie). Elle a été introduite par les travaux de John Von Neumann et Oskar Morgenstern [vNM44], puis ceux de John Nash [Nas50]. Elle permet d'étudier des situations dans lesquelles le sort de chaque agent (appelé **joueur**) dépend non seulement des décisions qu'il prend, mais également

des décisions prises par d'autres joueurs. Comme chaque joueur n'est pas totalement maître de son sort, on dit que les agents sont en situation d'*interaction stratégique*, ce qui correspond tout-à-fait à ce qui se produit en argumentation. On suppose dans un jeu en interaction stratégique que les joueurs se connaissent : ils savent combien il y a de joueurs, et qui ils sont. Du fait que le gain de chacun dépend en partie des actions des autres, un joueur ne peut pas se contenter de choisir ses propres plans d'actions, en négligeant ce que font les autres. Il doit au contraire se faire une idée aussi précise que possible des stratégies choisies, ou susceptibles d'être choisies, par les autres joueurs. Pour cela, on admet que **les agents sont rationnels**, c'est-à-dire que chaque joueur s'efforce de prendre les meilleures décisions pour lui-même, et sait que les autres joueurs font de même.

Je ne donnerai pas ici un état de l'art exhaustif de la théorie des jeux, mais j'introduirai en section 4.1 quelques concepts utiles pour la suite : une taxonomie partielle des jeux, les types de représentation des jeux, et enfin quelques concepts de solution.

Sachant que mon "point d'entrée" dans la théorie des jeux a été l'argumentation, je citerai ensuite dans la section 4.2 page 129, les différents liens qui existent entre la théorie des jeux et celle de l'argumentation. Ici-aussi, le but ne sera pas l'exhaustivité, mais plutôt d'illustrer par quelques exemples à quel point ces deux théories sont liées.

Et enfin, la section 4.3 page 130 décrira un type de jeux particulier, les jeux booléens dont l'intérêt est multiple :
– utiliser la logique classique pour modéliser des interactions entre agents,
– réutiliser des procédés de représentation de préférence issus de l'inférence non monotone,
– et fournir une autre manière d'exploiter le lien entre argumentation et jeux.

## 4.1 Concepts de base de la théorie des jeux

De manière succincte, un jeu est un ensemble de joueurs, avec un ensemble d'issues du jeu (appelées profils de stratégies, chaque stratégie étant un choix possible pour un joueur) et une fonction d'utilité qui attribue à chaque joueur un gain en fonction de l'issue du jeu.

### 4.1.1 Taxonomie des jeux

Il existe au moins 4 types de jeux, les jeux statiques, dynamiques, coopératifs et non coopératifs, sachant qu'un même jeu peut réunir plusieurs de ces caractéristiques.

Un jeu est dit **statique** lorsque les joueurs choisissent **simultanément** leurs actions, et reçoivent ensuite leurs gains respectifs.

Un jeu **dynamique** est un jeu qui se déroule en **plusieurs étapes**. On se place ici dans le cadre des **jeux dynamiques en information parfaite**[1], c'est-à-dire que l'on admet que toutes les actions passées sont *observables* et *connues* de tous les joueurs. Dans ce cadre, en intervenant à des étapes antérieures du jeu, certains joueurs ont le pouvoir d'affecter directement les gains d'autres joueurs de manière irréversible (notons que cette notion d'information parfaite n'a pas de sens dans le cas des jeux statiques).

Un jeu est **coopératif** lorsque les joueurs peuvent passer entre eux des accords qui les lient de manière contraignante. On dit alors qu'ils forment une **coalition** dont les membres agissent de concert. Dans le cas contraire, c'est-à-dire lorsque les joueurs n'ont pas la possibilité de former des coalitions, le jeu est **non coopératif**.

On dit qu'un jeu coopératif est à **utilité transférable** s'il est possible d'additionner les utilités des joueurs et de les redistribuer aux membres d'une coalition (il existe une "monnaie" commune à tous avec laquelle on peut effectuer des transferts).

Les jeux non coopératifs se divisent en deux grandes familles : les jeux à somme nulle, et ceux à somme non nulle (en économie, cette notion de jeu à somme nulle est importante, car elle correspond à l'absence de production, ou de destruction, de produits).

[1] Il existe aussi la notion d'information complète : chaque joueur connaît lors de la prise de décision ses possibilités d'action, celles des autres joueurs, les gains résultants de toutes ces actions et les motivations des autres joueurs.

Les **jeux à somme nulle** sont tous les jeux où la somme "algébrique" des gains des joueurs est constante : ce que gagne l'un est nécessairement perdu par un autre (donc stricto sensu, il est possible que les jeux ne soient pas à somme nulle, à partir du moment où la somme reste constante : l'enjeu est de répartir entre tous les joueurs un total de gains préalablement fixé).

### 4.1.2 Quelques exemples

**Le dilemme du prisonnier**, que nous allons présenter maintenant, est un exemple célèbre de la théorie des jeux.

**Exemple 52** *Deux suspects sont retenus dans des cellules séparées, et ne peuvent donc pas communiquer. La police ne dispose pas d'éléments de preuve suffisants pour obtenir leur condamnation, l'aveu d'au moins un des deux est donc indispensable. La police propose à chacun d'entre eux le marché suivant :*
- *si vous avouez et que votre complice n'avoue pas, vous aurez une remise de peine, tandis que votre complice aura la peine maximale (10 ans) ;*
- *si vous avouez tous les deux, vous serez condamnés à une peine plus légère (5 ans) ;*
- *si aucun de vous n'avoue, la peine sera minimale (6 mois), faute d'éléments au dossier.*

*On peut formaliser cette situation par un jeu à 2 joueurs, chaque joueur ayant 2 stratégies possibles : avouer (dénotée par $A$) ou se taire (dénotée par $T$).*
- $N = \{1, 2\}$,
- *le prisonnier 1 a deux stratégies possibles :* $s_{1_1} = A$ *et* $s_{1_2} = T$*. Il en est de même pour le joueur 2 :* $s_{2_1} = A$ *et* $s_{2_2} = T$*.*
- *Ce jeu a donc 4 profils de stratégies possibles :* $AA$, $AT$, $TA$ *et* $TT$*.*
- *On peut donc calculer les utilités de chacun des joueurs pour chacune des issues possibles du jeu :*
  - $u_1(AA) = u_2(AA) = -5$,
  - $u_1(TT) = u_2(TT) = -0.5$,
  - $u_1(AT) = u_2(TA) = 0$,
  - $u_1(TA) = u_2(AT) = -10$.

*Ce jeu est un jeu statique (une seule étape), non coopératif et à somme non nulle : les deux prisonniers sont enfermés dans des cellules séparées, ne peuvent pas communiquer, et donc ne peuvent pas passer un accord et former une coalition, mais ce que gagne l'un n'est pas forcément perdu par l'autre.*

**La bataille des sexes** est un autre exemple de jeux qui peut être vue comme un jeu coopératif à utilité non transférable [LR57] :

**Exemple 53** *Lucas et Elsa veulent aller au cinéma. Ils ont le choix entre un film d'horreur et une comédie romantique. Pour les deux, ce qui compte avant tout, c'est d'être ensemble. Néanmoins, Elsa a une préférence pour le film d'horreur et Lucas pour la comédie romantique.*

*On peut formaliser cette situation par un jeu à 2 joueurs, chaque joueur ayant 2 stratégies possibles : aller voir un film d'horreur (dénotée par $H$) ou une comédie romantique (dénotée par $R$).*
- $N = \{1, 2\}$,
- *Elsa* $(1)$ *a 2 stratégies possibles :* $s_{1_1} = H$ *et* $s_{1_2} = R$*. Lucas* $(2)$ *a les même possibilités :* $s_{2_1} = H$ *et* $s_{2_2} = R$*.*
- *On peut donc calculer les utilités de chacun des joueurs pour chacune des issues possibles du jeu :*
  - $u_1(H, H) = u_2(R, R) = 2$,
  - $u_1(R, R) = u_2(H, H) = 1$,
  - $u_1(H, R) = u_1(R, H) = u_2(H, R) = u_2(R, H) = 0$.

*La bataille des sexes peut être formalisée par le jeu coopératif à utilité non transférable à 2 joueurs suivant :*
- $N = \{1, 2\}$,
- $v$ *est une* **fonction caractéristique** *qui associe un vecteur* $v(C)$ *à chaque coalition* $C$ *de* $N$*, chaque élément* $v_i$ *du vecteur* $v(C)$ *correspondant à l'utilité obtenue par le joueur* $i$ *dans la coalition* $C$*. Ici,* $v$ *peut être défini par :*

$$\begin{cases} v(12) = \{(v_1, v_2)\,; 1 \leq v_1 \leq 2, 1 \leq v_2 \leq 2\} \\ v(1) = \{(v_1)\,; 0 \leq v_1 \leq 2\} \\ v(2) = \{(v_2)\,; 0 \leq v_2 \leq 2\} \end{cases}$$

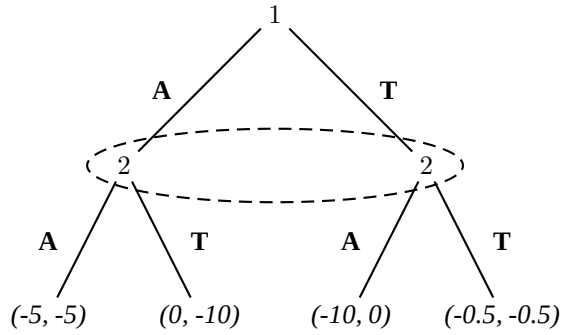
*En effet, si Lucas et Elsa se coordonnent, ils sont sûrs d'aller au même endroit, et donc auront chacun au moins une utilité de* 1*. Par contre, s'ils ne se coordonnent pas, ils ont toutes les chances de ne pas être au même endroit, et donc de n'en retirer aucune satisfaction (même s'ils peuvent avoir de la chance et que l'un des deux se dévoue, en espérant que l'autre n'ait pas eu la même idée).*

### 4.1.3 Représentation des jeux

Un jeu stratégique peut être représenté de deux façons différentes mais équivalentes : sous forme normale (dite aussi stratégique) et sous forme extensive (dite aussi développée).

Un jeu sous **forme extensive** est défini par un **arbre** de décision décrivant les actions possibles des joueurs à chaque étape du jeu, la séquence de tours de jeu des joueurs ainsi que l'information dont ils disposent à chaque étape pour prendre leur décision. Chaque nœud de l'arbre spécifie le joueur qui doit choisir une action (ou stratégie) à ce moment du jeu, ainsi que l'information dont il dispose. Les gains que chaque joueur peut réaliser après avoir suivi un des chemins possibles au sein de l'arbre, correspondant à chaque profil de stratégies, sont associés à chaque feuille de l'arbre. Par exemple, la forme extensive du dilemme du prisonnier est représentée par :

**Exemple 52 page précédente (suite)**




*Chaque nœud de cet arbre correspond à un joueur, et chaque branche partant de ce nœud correspond à une stratégie possible de ce joueur.*
*Dans les feuilles, le premier élément de chaque couple représente l'utilité du prisonnier 1, tandis que le second élément représente celle du prisonnier 2.*
*Le cercle en pointillé entourant les deux occurrences du joueur 2 signifie que ce dernier ne sait pas dans quelle situation il se trouve, il ne sait pas si son complice a choisi d'avouer ou de se taire.*

*Ce jeu a une seconde forme extensive, plaçant le second joueur en haut de l'arbre. Ces deux représentations sont équivalentes car les deux joueurs jouent simultanément.*

Un jeu sous **forme normale** est la donnée de l'ensemble des joueurs, de l'ensemble des stratégies pour chaque joueur et des paiements associés à toute combinaison possible de stratégies. On peut alors représenter ces jeux **sous forme matricielle**, en associant à chaque profil de stratégies $s$ un $n$-uplet donnant l'utilité obtenue par chaque joueur dans l'ordre : $(u_1(s), u_2(s), \ldots, u_n(s))$. Par exemple, la forme normale du dilemme du prisonnier est représentée par :

**Exemple 52 page précédente (suite)**

| 1 \ 2 | $A$ | $T$ |
|---|---|---|
| $A$ | *(-5, -5)* | *(0, -10)* |
| $T$ | *(-10, 0)* | *(-0.5, -0.5)* |

*Comme précédemment, le premier élément de chaque couple représente l'utilité du prisonnier 1, tandis que le second élément représente celle du prisonnier 2.*

### 4.1.4 Concepts de solution

Il existe plusieurs concepts de solution mais je ne présenterai ici que les équilibres de Nash qui s'appliquent à des jeux statiques non coopératifs.

L'**équilibre de Nash**, introduit par John Nash en 1950 [Nas50], est un concept fondamental en théorie des jeux. Il décrit une issue du jeu dans laquelle aucun joueur ne souhaite modifier sa stratégie étant donnée la stratégie de chacun de ses rivaux. En d'autres termes, un équilibre de Nash est un profil de stratégie dont aucun joueur n'a intérêt à dévier s'il suppose que les autres joueurs ne dévieront pas non plus. Il existe plusieurs versions de cet équilibre : en stratégies pures ou en stratégies mixtes (une stratégie mixte étant une stratégie définissant les probabilités avec lesquelles les joueurs choisissent chacune de leurs stratégies pures). Ici, nous n'utiliserons que la première version.

**Exemple 52 page 127 (suite)** *Le profil de stratégies $AA$ est un équilibre de Nash en stratégies pures du dilemme du prisonnier. En effet, on peut vérifier que l'on a : $u_1(AA) \geq u_1(TA)$ et $u_2(AA) \geq u_2(AT)$.*

*$AA$ est le seul équilibre de Nash de ce jeu. En effet, $AT$ ne peut pas être un PNE car $AA$ est une meilleure stratégie pour le joueur $2$ : $u_2(AA) > u_2(AT)$ ; $TA$ non plus car $AA$ est une meilleure stratégie pour le joueur $1$ : $u_1(AA) > u_1(TA)$ ; et de même pour $TT$ car $AT$ est une meilleure stratégie pour le joueur $1$ : $u_1(TT) < u_1(AT)$.*

Ni l'existence, ni l'unicité de cet équilibre de Nash en stratégies pures ne sont garanties.

## 4.2 Argumentation et théorie des jeux

Les premiers liens entre argumentation et jeux ont été identifiés par Dung lui-même dans [Dun95]. Ils ont ensuite été utilisés par diverses personnes pour définir des théories de la preuve en argumentation.

### 4.2.1 Travaux de Dung

[Dun95] utilise des concepts de la théorie des jeux développée par Von Neumann et Morgenstern dans [vNM44] afin de proposer certains parallèles entre l'argumentation et les jeux coopératifs.

Parmi les concepts utilisés, on trouve celui de "domination" : une issue $s'$ domine une autre issue $s$ si et seulement si il existe une coalition (non vide) telle que les membres de la coalition gagnent plus avec $s'$ qu'avec $s$ et avec laquelle l'issue $s$ est réalisable.

[vNM44] définit une solution (NM-solution) d'un jeu coopératif comme un ensemble $S$ d'issues vérifiant deux postulats : (1) aucune issue de $S$ ne domine une autre issue de $S$, et (2) toute issue hors de $S$ est dominée par au moins une issue de $S$.

La version "argumentative" de ce type de jeu est la suivante : dans leur propre intérêt, les joueurs vont tenter d'imposer la coalition qui optimise leur gain ; chaque issue du jeu est potentiellement un argument en faveur de la construction de telle coalition ; par ailleurs, une issue $s$ attaque une autre issue $s'$ quand $s$ domine $s'$ ; nous retrouvons là des notions d'arguments et de relation d'attaque.

À partir d'un jeu coopératif, Dung a alors proposé un système d'argumentation particulier dont l'ensemble d'arguments est l'ensemble des issues du jeu et dont la relation d'attaque est la relation de domination. Il a alors montré que les NM-solutions d'un jeu coopératif sont exactement les extensions stables du système d'argumentation associé à ce jeu.

Il a alors appliqué ces résultats au problème du mariage stable qui est un jeu coopératif bien connu : soit un ensemble de $n$ hommes et un ensemble de $n$ femmes ; chacun a une liste strictement ordonnée de préférence contenant tous les membres du sexe opposé ; le problème du mariage stable consiste à trouver la manière de marier tous les hommes et toutes les femmes qui satisferait au mieux leurs critères de préférence.

Dans [Dun95], Dung a aussi identifié d'autres liens entre argumentation et jeux coopératifs basés sur d'autres concepts de solutions mais nous ne rentrerons pas dans le détail ici.

De manière générale, Dung s'est intéressé à la manière d'utiliser la théorie de l'argumentation (notamment les concepts d'extensions) pour approfondir certains types de jeux. En effet, Dung imagine les issues possibles du jeu comme des arguments et il propose l'extraction des solutions du jeu comme un problème de décision en argumentation grâce aux concepts d'extension et d'acceptabilité.

### 4.2.2 Théorie de la preuve dialectique

La notion de jeu apparaît aussi dès que l'on cherche à définir une théorie de la preuve pour des problèmes d'argumentation.

L'idée est ici de "prouver" si un argument $x$ est acceptable ou non en simulant un "dialogue" entre 2 agents : un partisan $PRO$ qui soutient l'acceptabilité de $x$ et un opposant $OPP$ qui soutient la non-acceptabilité de $x$. Ce dialogue est donc un **jeu à deux joueurs**, qui donne raison à celui des deux qui aura le dernier mot. Ces idées sont à l'origine de [JV99a] et ont été raffinées dans [Dou02].

Ce type de dialogue induit que chaque argument avancé par un des joueurs est pertinent. Et la pertinence d'un argument dépend du type de problème que l'on cherche à résoudre. C'est le rôle de la fonction *coup-légal* que de déterminer l'ensemble des coups possibles à partir de tous les coups joués auparavant.

Il faut aussi pouvoir conclure le dialogue et donc savoir qui a gagné : cela sera déterminé à l'aide du *critère de gain*. Par exemple, dans un problème d'acceptabilité d'un argument $x$, au moins 2 critères de gain sont possibles :

1. Il existe un *dialogue gagné* par $PRO$ portant sur $x$.
2. Il existe une *stratégie gagnante* pour $x$, c'est-à-dire une façon pour $PRO$ de défendre $x$ contre toutes les attaques de $OPP$.

Quand on veut décider telle acceptabilité de tel argument sous une sémantique donnée, il faut construire une théorie de la preuve dont le critère de gain et la fonction coup-légal caractérisent le type d'acceptabilité et la sémantique du problème posé.

Cela a été fait dans [Dou02] pour le problème de l'acceptabilité crédule sous les sémantiques préférée et basique ; et dans [JV99a] pour d'autres sémantiques sur lesquelles je ne m'étendrai ici.

Le travail auquel j'ai participé sur ce sujet a été d'identifier les caractéristiques d'un jeu correspondant à une théorie de la preuve dialectique pour l'argumentation (voir[Sir04]). Cette recherche a été menée avec Jérôme LANG, Sylvie DOUTRE et Denis SIREYJOL et a été appliquée aux théories de la preuve dialectiques proposées par [Dou02]. On a pu ainsi caractériser entre autre les ensembles de stratégies des deux joueurs ($PRO$ et $OPP$) à l'aide de la fonction *coup-légal* et du *critère de gain* choisis pour prouver l'acceptabilité d'un argument donné pour une sémantique particulière.

## 4.3 Jeux booléens

Après avoir commencé à travailler un peu sur des jeux liés à l'argumentation, la question s'est posée d'utiliser une catégorie de jeux proches de la logique avec lesquels on pourrait donc effectuer un raisonnement logique éventuellement non monotone et finir peut-être par "reboucler" sur l'argumentation (qui est aussi un processus de raisonnement non monotone pouvant être basé sur la logique). D'autre part, l'emploi de la logique permet de réduire considérablement la taille du jeu. En effet, dans le cas des jeux statiques, les deux modes de représentation usuels des jeux (forme extensive et forme normale) coïncident, et doivent fournir la description explicite de la fonction d'utilité de chaque agent. Or, cette description est de taille exponentielle en fonction du nombre d'agents[2]. Cette explosion combinatoire est encore plus flagrante lorsqu'à la fois l'ensemble des agents et l'ensemble des stratégies pour chacun des agents sont de taille importante.

[2]Par exemple, si $n$ agents ont chacun un choix entre deux actions possibles, il faudra spécifier $n \times 2^n$ valeurs numériques ; si deux agents contrôlent chacun un ensemble de $p$ variables booléennes (il suffit de penser à de telles variables comme à des boutons que l'agent peut choisir d'enfoncer ou non), chaque agent a $2^p$ stratégies possibles et il faudra donc expliciter $2 \times (2^p)^2 = 2^{2p+1}$ valeurs numériques.

Jérôme LANG, Élise BONZON et moi-même avons donc choisi d'étudier les **jeux booléens** introduits par Harrenstein, van der Hoek, Meyer et Witteveen dans [HvdHMW01, Har04b]. Ce sont des jeux à deux joueurs et à somme nulle, la fonction d'utilité du joueur 1 (et donc celle du joueur 2 qui est son opposé) est représentée par une formule de la logique propositionnelle, appelée *forme booléenne* du jeu. Ce type de jeu est très compact à représenter, malheureusement il est aussi très simple. Notre travail a donc consisté à l'enrichir sur différents points.

Il est important de noter que ce travail a été réalisé essentiellement par Élise BONZON puisqu'il s'agissait de son travail de thèse que j'ai co-encadrée avec Jérôme LANG. C'est en particulier elle qui a proposé les généralisations des jeux booléens à $n$ joueurs et avec préférences non-dichotomiques. Par la suite, l'étude des liens entre jeux booléens et argumentation a été menée en collaboration avec Elise et Caroline DEVRED (LERIA) (voir la section 4.3.3 page 145).

### 4.3.1 Jeux booléens à $n$ joueurs

Ils généralisent donc la version à deux joueurs proposée par [HvdHMW01, Har04b].

#### 4.3.1.1 Définitions de base

Étant donné un ensemble de variables propositionnelles $V$, un jeu booléen à $n$ joueurs sur $V$ est un jeu à $n$ joueurs pour lequel les actions disponibles de chaque joueur consistent à assigner une valeur de vérité à toutes les variables d'un sous-ensemble donné de $V$. Les préférences de chaque joueur $i$ sont représentées par une formule propositionnelle $\varphi_i$ formée sur les variables de $V$.

**Définition 94** *Un* **jeu booléen à $n$ joueurs** *est un 5-uplet $(N, V, \pi, \Gamma, \Phi)$, avec*
– $N = \{1, 2, \ldots, n\}$ *l'ensemble des joueurs (appelés aussi agents) ;*
– $V$ *un ensemble de variables propositionnelles ;*
– $\pi : N \mapsto V$ *une fonction d'affectation de contrôle définissant une partition de $V$ ;*
– $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ *l'ensemble des contraintes, chaque $\gamma_i$ étant une formule propositionnelle de $L_{\pi(i)}$ satisfiable ;*
– $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ *l'ensemble des buts, chaque $\varphi_i$ étant une formule de $L_V$ satisfiable.*
*Un 4-uplet $(N, V, \pi, \Gamma)$, avec $N, V, \pi, \Gamma$ définis comme ci-dessus est appelé un* **pré-jeu booléen**.

La *fonction d'affectation de contrôle* $\pi$ associe à chaque joueur les variables qu'il contrôle. On note $\pi_i$ l'*ensemble des variables contrôlées* par le joueur $i$. Chaque variable est contrôlée par un et un seul joueur. Ainsi, $\{\pi_1, \ldots, \pi_n\}$ forme une partition de $V$.

Chaque $\gamma_i$ représente ici les contraintes d'un agent sur l'ensemble des variables qu'il contrôle. Ce choix de représentation est assez intuitif : en règle générale, les contraintes d'une personne reposent sur les actions qu'il lui est possible d'effectuer. Si une de mes contraintes pèse sur une action d'un autre joueur, je ne peux pas être assurée de la respecter, ce n'est pas de mon ressort. Ainsi, cette représentation des contraintes permet de respecter l'indépendance des agents : chaque agent gère ses variables, et les contraintes qui y pèsent, sans être tributaire d'un autre agent pour cela.

L'utilisation des jeux booléens permet d'avoir une représentation compacte du problème. Pour illustrer ce propos, nous allons utiliser une variante de l'exemple 52 page 127 du dilemme des prisonniers : nous considérons ici $n$ prisonniers qui ne peuvent bénéficier que d'un seul type de remise de peine afin de simplifier le problème.

**Exemple 54** *Dans le jeu simplifié du prisonnier à $n$ joueurs, $n$ détenus (notés $1, \ldots, n$) sont emprisonnés dans des cellules séparées. La police fait à chacun d'eux le même marché :*

> *"Tu as le choix entre couvrir tes complices en te taisant (noté $T_i$, $i = 1, \ldots, n$) ou les trahir en les dénonçant (noté $\neg T_i$, $i = 1, \ldots, n$). Si tu les dénonces, tu auras une remise de peine et tes partenaires purgeront le maximum (sauf si l'un d'eux t'a dénoncé aussi, auquel cas il bénéficiera comme toi d'une remise de peine). Mais si vous vous couvrez tous mutuellement, vous aurez tous une remise de peine*[3]*."*

[3] L'emploi de préférences binaires ne nous permet pas ici d'exprimer le fait qu'un prisonnier préfère la situation dans laquelle il dénonce ses complices tandis que les autres le couvrent à la situation dans laquelle tout le monde couvre tout le monde, elle-même préférée à la situation dans laquelle tout le monde dénonce. Pour faire cela, nous aurons besoin d'un langage plus sophistiqué, *cf.* section 4.3.2 page 136.

*La représentation de ce jeu en forme normale pour $n = 3$ est la suivante :*

| $3 : T_3$ | | |
|---|---|---|
| 1 \ 2 | $T_2$ | $\neg T_2$ |
| $T_1$ | *(1, 1, 1)* | *(0, 1, 0)* |
| $\neg T_1$ | *(1, 0, 0)* | *(1, 1, 0)* |

| $3 : \neg T_3$ | | |
|---|---|---|
| 1 \ 2 | $T_2$ | $\neg T_2$ |
| $T_1$ | *(0, 0, 1)* | *(0, 1, 1)* |
| $\neg T_1$ | *(1, 0, 1)* | *(1, 1, 1)* |

*Dans cette forme normale, les $n$-uplets (ici des triplets) donnent le résultat obtenu par les $n$ joueurs dans l'ordre : (résultat joueur 1, résultat joueur 2, ...). Le 0 (resp. 1) signifie que le joueur concerné perd (resp. gagne).*

*On constate ici que pour $n$ prisonniers, on aura une matrice à $n$ dimensions, chaque dimension étant égale à 2, donc $2^n$ n-uplets à spécifier*[4].

*Or, ce jeu peut être traduit très simplement par le jeu booléen $G = (N, V, \pi, \Gamma, \Phi)$ suivant :*
- $N = \{1, 2, \ldots, n\}$,
- $V = \{T_1, \ldots, T_n\}$,
- $\forall i \in \{1, \ldots, n\}, \pi_i = \{T_i\}$,
- $\forall i \in \{1, \ldots, n\}, \gamma_i = \top$, *et*
- $\forall i \in \{1, \ldots, n\}, \varphi_i = (T_1 \wedge T_2 \wedge \ldots \wedge T_n) \vee \neg T_i$.

*L'utilisation des jeux booléens permet donc de réduire de manière très significative la taille de la représentation du problème.*

**Définition 95** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un jeu booléen. Une* **stratégie** *$s_i$ pour un joueur $i$ de $G$ est une $\pi_i$-interprétation satisfaisant $\gamma_i$. L'***ensemble des stratégies du joueur** *$i$ est représenté par $S_i = \{s_i \in 2^{\pi_i} | s_i \models \gamma_i\}$.*

*Un* **profil de stratégies** *$s$ pour $G$ est un $n$-uplet $s = (s_1, \ldots, s_n)$, avec pour tout $i$, $s_i \in S_i$. $S = S_1 \times \ldots \times S_n$ est l'ensemble de tous les profils de stratégies.*

En d'autres mots, une stratégie pour le joueur $i$ est une affectation à vrai ou faux des variables qu'il contrôle. Pour chaque joueur $i$, $\gamma_i$ est l'ensemble des contraintes restreignant l'ensemble des profils de stratégies possibles pour ce joueur.

Pour tout ensemble non vide de joueurs (appelé **coalition**) $I \subseteq N$, la projection de $s$ sur $I$ est définie par $s_I = (s_i)_{i \in I}$. Si $I = \{i\}$, la projection de $s$ sur $\{i\}$ est dénotée par $s_i$ au lieu de $s_{\{i\}}$.

Les notations suivantes sont usuelles en théorie des jeux. Soit $s = (s_1, \ldots, s_n)$ et $s' = (s'_1, \ldots, s'_n)$ deux profils de stratégies.

On note $s_{-i}$ le *profil de stratégies $s$ privé de la stratégie du joueur $i$* : $s_{-i} = (s_1, s_2, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$. On note $(s_{-i}, s'_i)$ *le profil de stratégies $s$ dans lequel on a remplacé la stratégie du joueur $i$ par celle du profil $s'$* : $(s_{-i}, s'_i) = (s_1, s_2, \ldots, s_{i-1}, s'_i, s_{i+1}, \ldots, s_n)$. Similairement, si $I \subseteq N$, on note $s_{-I} = s_{N \setminus I}$.

$\pi_I$ représente l'ensemble des variables contrôlées par $I$, et $\pi_{-I} = \pi_{N \setminus I}$. Si $I = \{i\}$, on note $\pi_{-i}$ *l'ensemble des variables contrôlées par tous les joueurs sauf le joueur $i$* : $\pi_{-i} = V \setminus \pi_i$. $\{\pi_1, \ldots, \pi_n\}$ formant une partition de $V$, un profil de stratégies $s$ est une interprétation pour $V$, *i.e.* $s \in 2^V$. $\pi^{-1}$ représente la fonction inverse de $\pi$.

L'ensemble des stratégies de $I \subseteq N$ est dénoté par $S_I = \times_{i \in I} S_i$, et l'ensemble des buts de $I \subseteq N$ par $\Phi_I = \bigwedge_{i \in I} \varphi_i$.

Étudions l'exemple suivant pour illustrer ces notions.

**Exemple 55** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un jeu booléen, avec*
- $V = \{a, b, c\}$, $N = \{1, 2\}$,
- $\pi_1 = \{a, b\}$ *et* $\pi_2 = \{c\}$.
- $\gamma_1 = \neg a \vee \neg b$, $\gamma_2 = \top$

[4] Ou un arbre binaire à $2^n$ feuilles si on utilise une représentation sous forme extensive.

– $\varphi_1 = (a \leftrightarrow b) \vee (\neg a \wedge b \wedge \neg c)$,
– $\varphi_2 = (\neg a \wedge b \wedge c) \vee (a \wedge \neg b)$,
*Le joueur* $1$ *a trois stratégies possibles :* $s_{1_1} = a\overline{b}$, $s_{1_2} = \overline{a}b$, $s_{1_3} = \overline{a}\overline{b}$. *La stratégie* $ab$ *ne satisfait pas les contraintes du joueur* $1$, *et donc n'est pas une stratégie acceptable.*

*Le joueur* $2$ *a deux stratégies possibles :* $s_{2_1} = c$ *ou* $s_{2_2} = \overline{c}$.

$G$ *possède donc 6 profils de stratégies :* $S = \{a\overline{b}c, a\overline{b}\overline{c}, \overline{a}bc, \overline{a}b\overline{c}, \overline{a}\overline{b}c, \overline{a}\overline{b}\overline{c}\}$.

*Les profils de stratégies* $\overline{a}\overline{b}c$, $\overline{a}b\overline{c}$ *et* $\overline{a}\overline{b}\overline{c}$ *donnent la victoire à* $1$, *tandis que* $a\overline{b}c$, $a\overline{b}\overline{c}$ *et* $\overline{a}bc$ *permettent à* $2$ *de gagner.*

Le but $\varphi_i$ du joueur $i$ est une relation de préférence dichotomique et compacte, correspondant donc à une fonction d'utilité binaire[5] : un joueur $i$ est satisfait, et a une utilité de 1, si et seulement si son but $\varphi_i$ est satisfait. Il a une utilité de 0 sinon. Les buts $\{\varphi_i, i = 1, \ldots, n\}$ jouent donc le rôle des fonctions d'utilité.

**Définition 96** *Pour chaque joueur* $i$, *la* **fonction d'utilité** *induite par le but de ce joueur est la fonction* $u_i : S \rightarrow \{0, 1\}$ *telle que :*

$$u_i(s) = \begin{cases} 0 \; si \; s \models \neg\varphi_i \\ 1 \; si \; s \models \varphi_i \end{cases}$$

Nous avons :

– $s$ est au moins aussi bon que $s'$ pour $i$, dénoté par $s \succeq_i s'$, si $u_i(s) \geq u_i(s')$, ou de façon équivalente si $s \models \neg\varphi_i$ implique $s' \models \neg\varphi_i$ ;
– $s$ est strictement meilleur que $s'$ pour $i$, dénoté par $s \succ_i s'$, si $u_i(s) > u_i(s')$, ou, de façon équivalente, $s \models \varphi_i$ et $s' \models \neg\varphi_i$.
– $i$ est indifférent entre $s$ et $s'$, dénoté par $s \sim_i s'$, si $s \geq_i s'$ et $s' \geq_i s$, ou de façon équivalente si $s \models \varphi_i$ si et seulement si $s' \models \varphi_i$.

Définissons à présent la notion de stratégie gagnante pour un joueur.

**Définition 97** *Soit* $G = (N, V, \pi, \Gamma, \Phi)$ *un jeu booléen, avec* $\Phi = \{\varphi_1, \ldots, \varphi_n\}$, *et* $N = \{1, \ldots, n\}$. *La stratégie* $s_i$ *est une* **stratégie gagnante** *pour le joueur* $i$ *si, quels que soient les choix effectués par ses adversaires,* $i$ *est sûr de gagner en choisissant cette stratégie.*

$$\forall s_{-i} \in S_{-i}, (s_{-i}, s_i) \models \varphi_i$$

#### 4.3.1.2 Dépendance

De nombreuses structures graphiques sont cachées dans un jeu booléen : la satisfaction de chaque joueur $i$ dépend des joueurs qui contrôlent des variables de $\varphi_i$. Il est donc possible de représenter graphiquement les dépendances entre joueurs en utilisant la nature syntaxique des buts. Intuitivement, si le but d'un joueur $i$ ne dépend d'aucune variable contrôlée par le joueur $j$, alors la satisfaction de $i$ ne dépendra pas directement de $j$. Toutefois, ce n'est qu'une condition suffisante : il se peut que l'expression syntaxique de $\varphi_i$ contienne une variable contrôlée par $j$ , mais que cette variable ne joue aucun rôle dans la satisfaction de $\varphi_i$, comme c'est le cas pour la variable $y$ dans $\varphi_i = x \wedge (y \vee \neg y)$. En utilisant une notion plus forte d'indépendance entre une formule et une variable que la simple notion d'indépendance syntaxique [LM98, LLM03], nous avons ainsi identifié quelques concepts intéressants : les variables et les joueurs utiles et le graphe de dépendance.

**Définition 98** *Soit* $G = (N, V, \Gamma, \pi, \Phi)$ *un jeu booléen.*

*L'ensemble des* **variables utiles** *pour un joueur* $i$, *dénoté par* $RV_(i)$[6], *est l'ensemble de toutes les variables* $v \in V$ *telles que* $\varphi_i$ *n'est pas indépendant de* $v$.

*L'ensemble des* **joueurs utiles** *pour un joueur* $i$, *dénoté par* $RP_i$, *est l'ensemble des joueurs* $j \in N$ *tels que* $j$ *contrôle au moins une variable utile de* $i$ *:* $RP_i = \bigcup_{v \in RV_i} \pi^{-1}(v)$.

[5] Nous verrons comment lever cette contrainte dans la section 4.3.2 page 136.
[6] Normalement, $G$ devrait apparaître dans toutes les notations, mais pour simplifier les notations nous écrirons simplement $RV_i$.

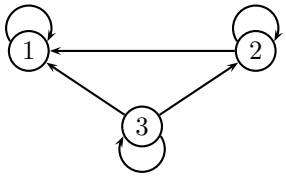
Cette relation entre les joueurs peut être vue comme un graphe orienté contenant un nœud pour chaque joueur, et un arc de $i$ à $j$ si $j \in RP_i$, c'est-à-dire si $j$ est un joueur utile pour $i$.

**Définition 99** *Le* **graphe de dépendance d'un jeu booléen** $G = (N, V, \Gamma, \pi, \Phi)$ *est un graphe orienté* $\mathcal{P} = \langle N, R \rangle$*, avec* $\forall i, j \in N$, $(i, j) \in R$ *(dénoté par* $R(i, j)$*) si* $j \in RP_i$*.*

**Exemple 56** *Axel, Kevin et Virginie sont invités à une fête. Axel veut y aller. Kevin aussi, mais si et seulement si Axel y va. Quant à Virginie, elle voudrait y aller, que Kevin y aille mais pas Axel. Cette situation peut être modélisée par le jeu booléen* $G = (N, V, \Gamma, \pi, \Phi)$*, défini par*
– $N = \{1, 2, 3\}$*, Axel étant le joueur* 1*, Kevin le* 2 *et Virginie* 3 *;*
– $V = \{a, b, c\}$*, où* $a$ *signifie "*1 *va à la fête", et de même pour* $b$ *et* 2*, et pour* $c$ *et* 3 *;*
– *pour chaque* $i$*,* $\gamma_i = \top$ *;*
– $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$*,* $\pi_3 = \{c\}$ *;*
– $\varphi_1 = a$*,*
– $\varphi_2 = a \leftrightarrow b$ *et*
– $\varphi_3 = \neg a \wedge b \wedge c$*.*
*On peut voir que la satisfaction d'Axel ne dépend que de lui-même, celle de Kevin depend de la décision d'Axel et de la sienne, tandis que celle de Virginie dépend des décisions de tous les joueurs.*

*On a :* $RV_1 = \{a\}$*,* $RV_2 = \{a, b\}$*,* $RV_3 = \{a, b, c\}$*,* $RP_1 = \{1\}$*,* $RP_2 = \{1, 2\}$*,* $RP_3 = \{1, 2, 3\}$*.*

*Le graphe de dépendance* $\mathcal{P}$ *induit par* $G$ *est représenté par :.*




$R(i)$ est donc l'ensemble des joueurs nécessaires à $i$ pour satisfaire son but : $j \in R(i)$ si et seulement si $j \in RP_i$. Il faut pourtant remarquer que $j \in R(i)$ n'implique pas que $i$ *a besoin de* $j$ pour que son but soit satisfait. Par exemple, si $\pi_1 = \{a\}$, $\pi_2 = \{b\}$ et $\varphi_1 = a \vee b$, alors $2 \in R(1)$. Pourtant, 1 a une stratégie lui permettant de satisfaire son but (en mettant $a$ à vrai) et n'a donc pas besoin de 2.

La fermeture transitive de $R$ est notée $R^*$. $R^*(i, j)$ représente le chemin de $i$ à $j$ dans $R$. $R^*(i)$ représente donc tous les joueurs qui ont une influence directe ou indirecte sur $i$. $R^{*-1}(i)$ représente tous les joueurs sur lesquels $i$ a une influence directe ou indirecte.

Ces notions sont liées à la notion d'ensemble stable.

**Définition 100** *Soit* $G = (N, V, \pi, \Gamma, \Phi)$ *un jeu booléen.* $B \subseteq N$ *est* **stable**[7] *pour* $R$ *si et seulement si* $R(B) \subseteq B$*, c'est-à-dire* $\forall j \in B$*,* $\forall i$ *tels que* $i \in R(j)$*, alors* $i \in B$*.*

**Définition 101** *Soit* $G = (N, V, \pi, \Gamma, \Phi)$ *un jeu booléen, et* $B \subseteq N$ *un ensemble stable pour* $R$*. La* **projection** *de* $G$ *sur* $B$ *définie par* $G_B = (B, V_B, \pi_B, \Gamma_B, \Phi_B)$ *(où* $V_B = \cup_{i \in B} \pi_i$*,* $\pi_B : B \rightarrow V_B$ *telle que* $\pi_B(i) = \{v | v \in \pi_i\}$*,* $\Gamma_B = \{\gamma_i | i \in B\}$*, et* $\Phi_B = \{\varphi_i | i \in B\}$*) est un jeu booléen.*

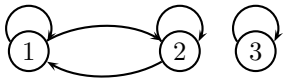
L'application directe de ce résultat est la décomposition possible d'un jeu booléen en sous-jeux booléens.

**Exemple 57** *Soit* $G = (N, V, \pi, \Gamma, \Phi)$ *le jeu booléen défini par* $V = \{a, b, c\}$*,* $N = \{1, 2, 3\}$*,* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$*,* $\pi_3 = \{c\}$*, pour chaque* $i$*,* $\gamma_i = \top$*,* $\varphi_1 = a \leftrightarrow b$*,* $\varphi_2 = a \leftrightarrow \neg b$ *et* $\varphi_3 = \neg c$*.*

*Nous avons :* $RV_1 = \{a, b\}$*,* $RV_2 = \{a, b\}$*,* $RV_3 = \{c\}$*,* $RP_1 = \{1, 2\}$*,* $RP_2 = \{1, 2\}$*,* $RP_3 = \{3\}$*.*

*Le graphe de dépendance* $\mathcal{P}$ *de* $G$ *est représenté par :*

[7] Attention, la notion d'ensemble "stable" que nous définissons ici n'est pas la même que celle utilisée en Théorie des graphes : nous ne parlons pas ici des ensembles de sommets deux-à-deux non adjacents. Et c'est aussi différent de la notion de stabilité utilisée en argumentation qui se rapproche de celle de la Théorie des graphes.




*Les ensembles de joueurs $B = \{1, 2\}$ et $C = \{3\}$ sont stables. On peut décomposer $G$ en 2 jeux booléens :*
– *$G_B = (B, V_B, \pi_B, \Gamma_B, \Phi_B)$, avec $B = \{1, 2\}$, $V_B = \{a, b\}$, $\pi_1 = a$, $\pi_2 = b$, $\gamma_1 = \gamma_2 = \top$, $\varphi_1 = a \leftrightarrow b$, $\varphi_2 = a \leftrightarrow \neg b$.*
– *$G_C = (C, V_C, \pi_C, \Gamma_C, \Phi_C)$, avec $C = \{3\}$, $V_C = \{c\}$, $\pi_3 = c$, $\gamma_3 = \top$, $\varphi_3 = \neg c$.*
*Cette décomposition nous permet ainsi d'étudier deux jeux plus simples, $G_B$ et $G_C$, plutôt que d'étudier le jeu $G$. Nous verrons dans la suite que nous pouvons ainsi calculer les équilibres de Nash sur les sous-jeux plutôt que sur le jeu complet.*

#### 4.3.1.3 Équilibres de Nash

La définition des équilibres de Nash en stratégie pures (PNE) pour les jeux booléens à $n$ joueurs est la même que la définition classique en théorie des jeux, en ayant à l'esprit que les fonctions d'utilité sont induites par les buts des joueurs $\varphi_1, \ldots, \varphi_n$. Formellement, on a donc :

**Définition 102** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un jeu booléen, avec $N = \{1, \ldots, n\}$ l'ensemble des joueurs. $s = (s_1, \ldots, s_n)$ est un* **équilibre de Nash en stratégies pures** *(PNE) si et seulement si :*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, u_i(s) \geq u_i(s_{-i}, s'_i)$$

Dans le cadre des jeux booléens, on peut alors donner une caractérisation simple des équilibres de Nash aux stratégies pures.

**Proposition 77** *Soit $s \in S$. $s$ est un équilibre de Nash en stratégies pures de $G$ si et seulement si pour tout $i \in N$, on a :*
– *soit $s \models \varphi_i$,*
– *soit $s_{-i} \models \neg\varphi_i$.*

Cette caractérisation simple permet de calculer plus facilement les équilibres de Nash en stratégies pures des jeux booléens. Ainsi, un profil de stratégies $s$ sera un PNE si et seulement si, pour chacun des joueurs $i$, soit $s$ permet d'obtenir le but du joueur $i$, soit le joueur $i$ ne peut obtenir son but si les autres joueurs maintiennent leurs stratégies.

Le graphe de dépendance entraîne également quelques propriétés sur les équilibres de Nash.

**Proposition 78** *Soit $G$ un jeu booléen. Si $\forall i \in N$, $i \notin RP_i$ alors tout $s \in S$ est un PNE.*

Si un joueur ne contrôle aucune des variables qui lui sont utiles, il n'a pas d'influence sur son propre but, et donc aucune préférence pour une ou l'autre de ses stratégies. Si c'est le cas pour tous les agents, aucun profil de stratégies n'est préféré, et tous sont donc des équilibres de Nash en stratégies pures.

On a aussi la propriété suivante qui permet d'exploiter la décomposition du jeu en sous-jeux :

**Proposition 79** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un jeu booléen, et soit $B_1 \ldots B_p$ $p$ ensembles stables de joueurs, tels que $B_1 \cup \ldots \cup B_p = N$. Soit $G_{B_1}, \ldots, G_{B_p}$ les $p$ jeux booléens associés.*

*Si $\exists s_{B_1} \ldots s_{B_p}$ des PNEs de $G_{B_1}, \ldots, G_{B_p}$ tels que $\forall i, j \in \{1, \ldots p\}$, $\forall k \in B_i \cap B_j$, $s_{B_i,k} = s_{B_j,k}$, alors $s = (s_{B_1}, \ldots, s_{B_p})$ est un PNE de $G$.*

Dans [BLSL07b, BLSL07d, BLSL07c, BLSL08b, BLSL08a, Bon07], ont aussi été réalisés des travaux sur la notion de coalition efficace et son lien avec la notion de dépendance.

### 4.3.2 Jeux booléens à préférences non-dichotomiques

Ce choix d'utilités binaires (pour lequel les agents peuvent seulement exprimer leur totale satisfaction ou leur total mécontentement, sans niveaux intermédiaires) est une vraie perte de généralité. En effet, on ne peut par exemple pas exprimer dans un jeu booléen classique des préférences du type "Je préférerais boire un café, si je ne peux pas un thé, et si ce n'est pas possible non plus je prendrai un chocolat chaud", qui nécessite 3 niveaux d'utilité ; tout comme on ne peut pas exprimer le dilemme du prisonnier classique (présenté dans l'exemple 52 page 127), qui lui nécessite 4 niveaux d'utilité.

On aimerait donc incorporer des préférences non dichotomiques aux jeux booléens afin que chaque joueur puisse représenter ses préférences de manière plus souple. Nous voudrions pouvoir associer une relation de préférence quelconque à l'ensemble des profils de stratégies $S$.

Ces préférences peuvent être de différents types : elles peuvent être numériques (on parlera alors de préférences *cardinales*), ou encore *ordinales*. Sachant que les notions qui nous intéressent ici (équilibres de Nash en stratégies pures) peuvent être définies avec des préférences ordinales, nous choisissons d'étudier tout particulièrement ces dernières.

Il nous faut donc choisir à présent une représentation permettant d'éviter la description explicite de la fonction d'utilité de chaque agent, qui est, comme nous l'avons déjà vu, de taille exponentielle en fonction du nombre d'agents. Pour les préférences dichotomiques, nous avons utilisé la logique propositionnelle, qui est le langage de représentation compacte le plus naturel pour ces préférences. Pour les préférences non dichotomiques, il nous faut donc trouver un langage permettant d'exprimer les préférences de chacun des joueurs de façon aussi concise, ou compacte, que possible. Dans la littérature, il en existe un certain nombre et nous avons choisi de nous intéresser à deux d'entre eux : les CP-nets et les buts à priorité.

Mais avant tout donnons la définition d'un jeu booléen avec préférences non-dichotomiques et celle des équilibres de Nash en stratégies pures correspondants.

**Définition 103** *Soit un langage propositionnel $L$ pour une représentation compacte de préférences. Un $L$-***jeu booléen** *est défini par un 5-uplet $G = (N, V, \pi, \Gamma, \Phi)$, avec*
- $N = \{1, \ldots, n\}$, $V$, $\pi$ *et* $\Gamma$ *étant définis comme précédemment, et*
- $\Phi = \langle \Phi_1, \ldots, \Phi_n \rangle$. *Pour chaque $i$, $\Phi_i$ est une représentation compacte, dans $L$, de la relation de préférences de $\succeq_i$ de l'agent $i$ sur*[8] $S$. *On note $Pref_G = \langle \succeq_1, \ldots, \succeq_n \rangle$.*

**Définition 104** *Soit un langage propositionnel $L$ pour une représentation compacte de préférences.*

*Soit $G = (N, V, \Gamma, \pi, \Phi)$ un $L$-jeu booléen, et $Pref_G = \langle \succeq_1, \ldots, \succeq_n \rangle$ l'ensemble des préférences de chaque joueur.*

*Deux définitions, une faible et une forte, peuvent ici caractériser les équilibres de Nash.*

$s = (s_1, \ldots, s_n)$ *est un* **équilibre de Nash faible en stratégies pures** *si et seulement si :*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, (s'_i, s_{-i}) \not\succ_i (s_i, s_{-i}) \tag{4.1}$$

$s = (s_1, \ldots, s_n)$ *est un* **équilibre de Nash fort en stratégies pures** *si et seulement si :*

$$\forall i \in \{1, \ldots, n\}, \forall s'_i \in S_i, (s'_i, s_{-i}) \preceq_i (s_i, s_{-i}) \tag{4.2}$$

*L'ensemble des équilibres de Nash forts (resp. faibles) en stratégie pure sera noté $NE_{fort}$ (resp. $NE_{faible}$).*

#### 4.3.2.1 Cas des CP-nets

Les CP-nets appartiennent à une famille de langages de représentation dits "graphiques". Ce langage est fondé sur le critère de comparaison *Ceteris Paribus* : quand un agent exprime en langage naturel une préférence telle que "une table ronde sera mieux dans le salon qu'une table carrée", il ne veut sans doute pas dire que n'importe quelle table

[8] On rappelle que dans un jeu booléen, $S_i = \{s_i \in 2^{\pi_i} | s_i \models \gamma_i\}$, et que $S = S_1 \times \ldots \times S_n$.

ronde sera préférée à n'importe quelle table carrée. Il veut exprimer le fait qu'il préférera une table ronde à une table carrée si ces deux tables ne diffèrent pas significativement dans leurs autres caractéristiques (la taille, la couleur, le bois utilisé, les finitions ou encore le prix). Le principe qui est à l'œuvre dans l'interprétation de telles préférences est que les alternatives doivent être comparées *toutes choses étant égales par ailleurs*, ou encore *Ceteris Paribus*.

Cela conduit à la notion d'indépendance suivante :

**Définition 105** *Soit $X$, $Y$ et $Z$ trois ensembles non vides disjoints formant une partition de $V$. $X$ et $Y$ sont* **conditionnellement préférentiellement indépendants étant donné** *$Z$ si et seulement si $\forall z \in D(Z)$, $\forall x_1, x_2 \in D(X)$ et $\forall y_1, y_2 \in D(Y)$ on a :*

$$x_1 y_1 z \succeq x_2 y_1 z \text{ si et seulement si } x_1 y_2 z \succeq x_2 y_2 z$$

Pour une valeur de $Z$ fixée, la relation de préférence sur les instanciations de $X$ est la même quelles que soient les valeurs des instanciations de $Y$.

Cette indépendance préférentielle conditionnelle est utilisée dans les CP-nets introduits dans [BBHP99] comme un outil pour représenter compactement les relations de préférences qualitatives. Ils ont été principalement étudiés dans [Dom02], [BBD$^+$04a] ou encore [BBD$^+$04b].

Les CP-nets peuvent être utilisés dans le cadre des jeux booléens de façon à représenter les préférences de chaque joueur. On ne considérera plus que les buts des joueurs sont représentés par une formule logique, mais directement par un CP-net propositionnel (pour lequel les domaines des variables sont binaires). Dans ce contexte, $\forall x_i \in V, D(x_i) = \{x_i, \overline{x_i}\} = 2^{x_i}$ et $D(\{x_1 \dots x_p\}) = 2^{\{x_1,\dots,x_p\}}$. Ainsi, un élément de $D(x_i)$ correspond à une $\{x_i\}$-interprétation, et un élément de $D(\{x_1 \dots x_p\})$ est une $\{x_1 \dots x_p\}$-interprétation.

**Définition 106** *Pour chaque variable $X \in V$, un ensemble de* **variables parents**, *noté $Pa(X)$, est spécifié. Ces variables sont celles qui influent sur les préférences de l'agent entre les différentes valeurs de $X$. Formellement, $X$ et $V \setminus (\{X\} \cup Pa(X))$ sont conditionnellement préférentiellement indépendantes étant donné $Pa(X)$.*

*La* **table de préférence conditionnelle** *(notée CPT) décrit les préférences de l'agent sur les valeurs de la variable $X$, étant donné toutes les combinaisons des valeurs des variables parents.*

*Pour chaque instanciation de $Pa(X)$, $CPT(X)$ spécifie* un ordre total *sur $D(X)$ tel que $\forall x_i, x_j \in D(X)$ on ait soit $x_i \succ x_j$, soit $x_j \succ x_i$.*

*Soit $V = \{X_1, \dots, X_n\}$ un ensemble de variables. $\mathcal{N} = \langle \mathcal{G}, \mathcal{T} \rangle$ est un* **CP-net** *sur V, $\mathcal{G}$ étant un graphe orienté sur $V$, et $\mathcal{T}$ étant un ensemble de tables de préférences conditionnelles $CPT(X_i)$ pour chaque $X_i \in V$. Chaque table de préférence conditionnelle $CPT(X_i)$ est associée à un ordre total $\succ_p^i$, selon chaque instanciation $p \in D(Pa(X_i))$.*

Formellement, la relation de préférence induite par le CP-net $\mathcal{N}$, représentée par le graphe de préférences induit, est définie comme suit :

**Définition 107** *La* **relation de préférence** *sur les états induite par un CP-net $\mathcal{N}$ est dénotée par $\succ_{\mathcal{N}}$, et est définie par $\forall o, o' \in D(V)$, $o \succ_{\mathcal{N}} o'$ si et seulement si $\mathcal{N} \models o \succ o'$.*

Illustrons ces définitions avec un exemple simple :

**Exemple 58** *Soit le CP-net présenté sur la figure 4.1 page suivante qui exprime mes préférences sur le menu du dîner. Ce CP-net est composé de deux variables, $S$ et $V$ qui correspondent respectivement à la soupe et au vin. Je préfère strictement manger une soupe de poisson ($S_p$) plutôt qu'une soupe de légumes ($S_l$), tandis que mes préférences sur le vin rouge ($V_r$) ou blanc ($V_b$) dépendent de la soupe que je mangerai. En effet je préfère du vin rouge avec une soupe de légumes, mais du blanc avec une soupe de poisson, ce qui se traduit par les formules $S_p : V_b \succ V_r$ et $S_l : V_r \succ V_b$. On a donc $D(S) = \{S_p, S_l\}$ et $D(V) = \{V_r, V_b\}$.*

*La figure 4.2 page suivante représente la relation de préférence induite par ce CP-net. L'élément du bas ($S_l \wedge V_b$) représente le pire état tandis que l'élément du haut ($S_p \wedge V_b$) est le meilleur état possible.*

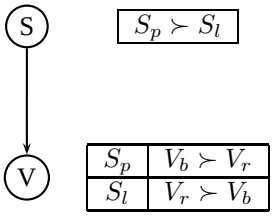



FIG. 4.1 – CP-net “Mon dîner simple”

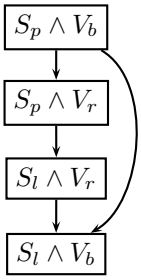



FIG. 4.2 – Graphe de préférences induit par le CP-net “Mon dîner simple”

*Nous remarquons ici que nous avons une flèche entre les nœuds $(S_l \wedge V_b)$ et $(S_p \wedge V_b)$ car nous comparons les états 2 par 2, toutes choses étant égales par ailleurs.*

*Nous pouvons donc ordonner totalement les états possibles (du plus préféré au moins préféré) :*

$$(S_p \wedge V_b) \succ (S_p \wedge V_r) \succ (S_l \wedge V_r) \succ (S_l \wedge V_b)$$

*Cette relation $\succ$ est la seule relation de préférence qui satisfasse ce CP-net.*

L’introduction des CP-nets dans les jeux booléens donne les définitions suivantes

**Définition 108** *La* **table de préférence conditionnelle** *pour un jeu booléen (notée $CPT_i(X)$) décrit les préférences du joueur $i$ sur les valeurs de la variable $X$, étant donné toutes les combinaisons des valeurs des variables parents. Pour chaque instanciation $p$ de $Pa_i(X)$, $CPT_i(X)$ spécifie un ordre total tel que l’on ait soit $x \succ_{i,p} \overline{x}$, soit $\overline{x} \succ_{i,p} x$.*

**Définition 109** *Un* **CP-jeu booléen** *est un 5-tuple $G = (N, V, \Gamma, \pi, \Phi)$, où $N = \{1, \ldots, n\}$ est un ensemble de joueurs, $V = \{x_1, \ldots, x_n\}$ est un ensemble de variables, $\Phi = \langle \mathcal{N}_1, \ldots, \mathcal{N}_n \rangle$, chaque $\mathcal{N}_i$ étant un CP-net sur $V$ ayant pour graphe $\mathcal{G}_i$ et vérifiant, pour tout $i \in N$, $\succeq_i = \succeq_{\mathcal{N}_i}$.*

L’exemple suivant montre la manière dont on peut exploiter ces notions :

**Exemple 59** *Soit le jeu $G = (N, V, \Gamma, \pi, \Phi)$ suivant :*
– $N = \{1, 2\}$
– $V = \{a, b, c\}$, *avec* $D(a) = \{a, \overline{a}\}$, $D(b) = \{b, \overline{b}\}$ *et* $D(c) = \{c, \overline{c}\}$.
– $\gamma_1 = \{a \leftrightarrow b\}$, $\gamma_2 = \top$,
– $\pi_1 = \{a, b\}$, $\pi_2 = \{c\}$,
– *Le but du joueur $1$ est représenté par le CP-net et la relation de préférence induite figure 4.3 page ci-contre,*
– *Le but du joueur $2$ est représenté par le CP-net et la relation de préférence induite figure 4.4 page suivante,*
*On remarque que le joueur $1$ n’exprime pas ses préférences sur $c$ s’il choisit $\overline{a}b$ ou $a\overline{b}$, ces deux états étant inaccessibles. Il est cependant possible de construire la relation de préférence associée en utilisant la relation de transitivité : nous savons ici que les états contenant $ab$ seront toujours préférés aux états contenant $\overline{a}\overline{b}$.*

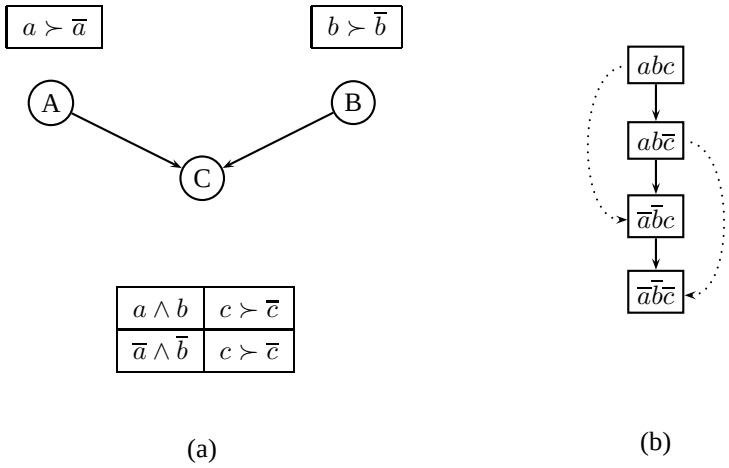



FIG. 4.3 – CP-net des préférences et le pré-ordre associé du joueur 1

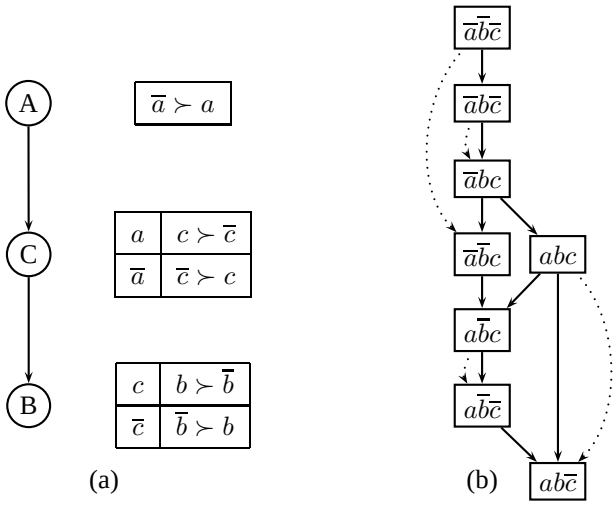



FIG. 4.4 – CP-net des préférences et le pré-ordre associé du joueur 2

*Pour calculer les équilibres de Nash de ce jeu, on ne s'occupe que des profils de stratégies présents dans les relations de préférence des deux joueurs : par exemple, le profil de stratégies $\overline{a}b\overline{c}$ viole les contraintes du joueur $1$ et ne peut donc pas être un PNE. Le calcul des équilibres de Nash faibles et forts nous donne ainsi ici comme résultat :*

$$NE_{faible} = NE_{fort} = \{abc\}$$

Ces CP-jeux booléens présentent quelques propriétés intéressantes en particulier dès que les CP-nets sont acycliques.

**Proposition 80** *Dans un CP-jeu booléen acyclique, les équilibres de Nash faibles et forts sont identiques.*

**Proposition 81** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un CP-jeu booléen tel que les graphes $\mathcal{G}_i$ sont tous identiques ($\forall i, j$, $\mathcal{G}_i = \mathcal{G}_j$) et acycliques. Ce jeu $G$ ainsi défini aura alors un et un seul équilibre de Nash en stratégie pure.*

On va exploiter cette idée en cherchant à rendre les graphes de tous les CP-nets identiques. Pour cela on utilise les notions d'union de graphes et de jeu équivalent.

**Définition 110** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un CP-jeu booléen tel que $\forall i \in N$, les graphes $\mathcal{G}_i$ sont tous acycliques.*

*Pour chaque joueur $i$, $\mathcal{G}_i = \langle V, Arc_i \rangle$, $V$ étant l'ensemble des nœuds du graphe*[9]*, et $Arc_i$ représentant l'ensemble des arcs orientés du graphe représentant le CP-net de $i$.* **L'union des graphes** *de $G$ sera le graphe $\mathcal{G} = \langle V, Arc_1 \cup \ldots \cup Arc_n \rangle$.*

**Définition 111** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un CP-jeu booléen tel que $\forall i \in N$, les graphes $\mathcal{G}_i$ sont tous acycliques.*

*On appelle $G^* = (N, V, \pi, \Gamma, \Phi^*)$ le* **jeu équivalent par réécriture de G** *dans lequel on a remplacé le graphe issu du CP-net représentant le but de chacun des joueurs par l'union de tous les graphes des joueurs.*

*Les tables de préférences conditionnelles sont modifiées de façon à correspondre avec le nouveau graphe, tout en donnant les mêmes préférences qu'avant : si l'arc $(X, Y)$ est ajouté au graphe, la table de préférence conditionnelle de la variable $Y$ sera la même que précédemment pour chaque instanciation $x \in D(X)$. Plus formellement, si on note $\succ_i^y$ la relation associée à la table de préférence conditionnelle $CPT_i(Y)$ pour le CP-net du joueur $i$ dans le jeu $G$, on a pour le jeu $G^*$ : $\forall x \in D(X), \succ_{i,x}^y = \succ_{i,\overline{x}}^y = \succ_i^y$.*

**Proposition 82** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un CP-jeu booléen tel que $\forall i \in N$, les graphes $\mathcal{G}_i$ sont tous acycliques. Soit $G^* = (N, V, \pi, \Gamma, \Phi^*)$ le jeu équivalent par réécriture de $G$. Le jeu $G^*$ permet d'obtenir les mêmes pré-ordres sur les profils de stratégies que le jeu $G$. On dit que $G$* **et** *$G^*$* **sont équivalents**.

Voyons l'utilité de tout cela sur un exemple :

**Exemple 60** *Soit le jeu $G = (N, V, \pi, \Gamma, \Phi)$ suivant :*
- *$N = \{1, 2\}$*
- *$V = \{a, b, c\}$, avec $D(a) = \{a, \overline{a}\}$, $D(b) = \{b, \overline{b}\}$ et $D(c) = \{c, \overline{c}\}$.*
- *$\gamma_1 = \gamma_2 = \top$,*
- *$\pi_1 = \{a, b\}$, $\pi_2 = \{c\}$,*
- *Le but du joueur $1$ est représenté sur la figure 4.5 page ci-contre,*
- *Le but du joueur $2$ est représenté sur la figure 4.6 page suivante.*

*On calcule à présent l'union des graphes issus des CP-nets représentant les buts de ces deux joueurs, et l'on crée ainsi le jeu $G^*$ équivalent par réécriture au jeu $G$. Les buts des deux joueurs sont alors représentés par les CP-nets représentés figures 4.7 page ci-contre et 4.8 page 142.*

*On constate sur la figure 4.7 page ci-contre que pour calculer les équilibres de Nash de ce jeu, il faudrait comparer les profils de stratégies $abc$, $a\overline{b}c$, $\overline{a}bc$ et $\overline{a}\overline{b}c$ pour le joueur $1$, pourtant $a\overline{b}c$ et $\overline{a}bc$ sont incomparables. Cette "incomparabilité" est représentée sur la figure par une ligne avec des pointillés plus larges.*

*L'union des graphes donne ici un graphe acyclique. On peut donc appliquer la propriété 81 : le jeu $G$ a un et un seul équilibre de Nash : $NE = \{abc\}$*

[9] ici identique pour chacun des joueurs puisque c'est l'ensemble des variables du jeu.

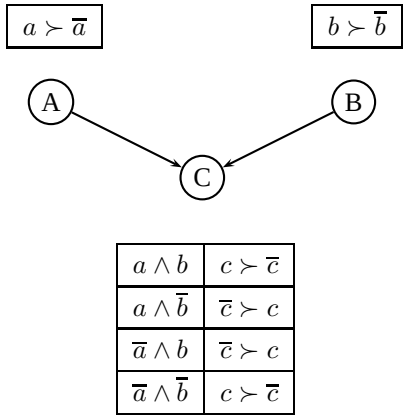



FIG. 4.5 – CP-net des préférences du joueur 1

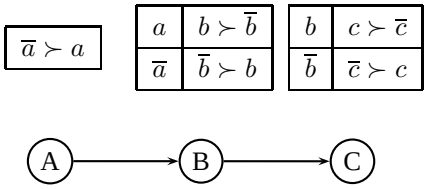



FIG. 4.6 – CP-net des préférences du joueur 2

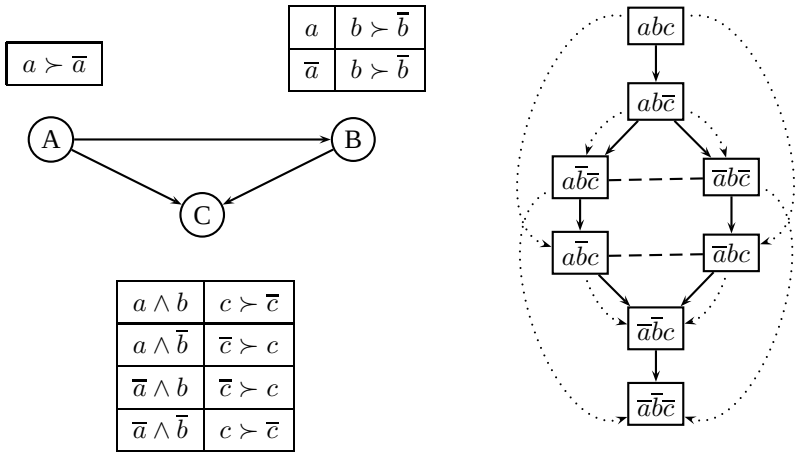
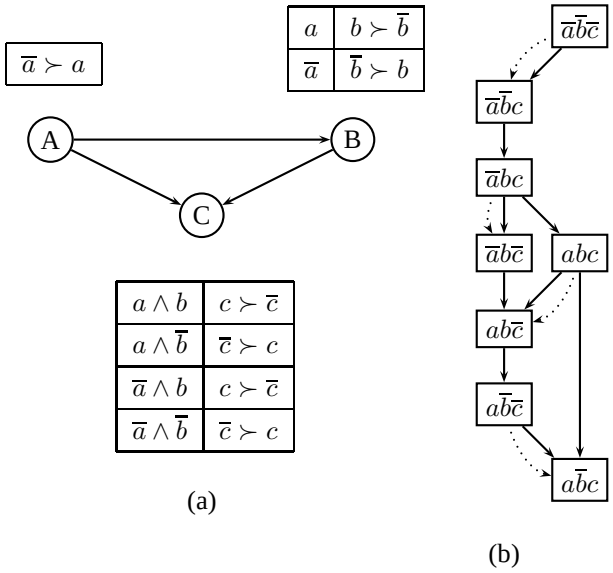



FIG. 4.7 – CP-net des préférences et le pré-ordre associé du joueur 1 pour le jeu $G^*$ équivalent par réécriture de $G$




(a) (b)

FIG. 4.8 – CP-net des préférences et le pré-ordre associé du joueur 2 pour le jeu $G^*$ équivalent par réécriture de G

#### 4.3.2.2 Cas des buts à priorité

Les CP-nets ne permettent pas de représenter toutes les préférences, ce langage n'étant pas totalement expressif. Il est par exemple impossible de dire avec un CP-net "Mes préférences sont, par ordre décroissant, des vacances à la mer en été, puis des vacances à la montagne en hiver, des vacances à la montagne en été, et enfin des vacances à la mer en hiver". Cette relation de préférences est en effet du type $xy \succ \overline{xy} \succ x\overline{y} \succ \overline{x}y$, c'est-à-dire qu'elle compare deux à deux des états n'étant pas identiques "toutes choses étant égales par ailleurs".

Pour cette raison, nous choisissons ici d'étudier un autre langage de représentation de préférences : les buts à priorité. Dans des buts à priorité, les préférences d'un joueur sont exprimées par un ensemble de buts ordonnés selon une relation de priorité. Nous allons donc retrouver les relations de priorité déjà identifiées et utilisées au chapitre 1.1 page 10 pour définir les mécanismes de sélection de sous-bases dans le cadre de la définition des relations d'inférence non monotone. Ces relations sont :

- la relation de préférence appelée *"discrimin"* qui découle des définitions issues de [Bre89, DLP91, Gef92, BCD$^+$93] et qui a fourni le mécanisme de sélection "incl",
- la relation de préférence appelée *"leximin"* qui découle des définitions issues de [DLP91, BCD$^+$93, Leh92] et qui a fourni le mécanisme de sélection "card",
- et la relation de préférence appelée *"best-out"* qui découle des définitions issues de [DLP91, BCD$^+$93] et qui a fourni le mécanisme de sélection "bo".

Nous pouvons alors définir les BP-jeux booléens.

**Définition 112** *Un* **BP-jeu booléen** *est un 5-uplet $G = (N, V, \pi, \Gamma, \Phi)$, où $\Phi = (\Sigma_1, \ldots, \Sigma_n)$ est une collection de bases de buts à priorité.*

*On note $\Sigma_i = \langle \Sigma_i^1 ; \ldots ; \Sigma_i^p \rangle$, $\Sigma_i^j$ représentant la strate $j$ de $\Sigma_i$ (l'ensemble de buts de priorité $j$ pour le joueur $i$).*

L'hypothèse que le nombre de niveaux de priorité est le même ($p$) pour chaque joueur n'entraîne pas de perte de généralité (ajouter des strates vides à une base de buts à priorité ne modifie pas la relation de préférence induite). Nous utiliserons les notations suivantes :

- Si $G$ est un BP-jeu booléen et que $c \in \{disc, lex, bo\}$, alors $Pref_G^c = \langle \succeq_1^c, \ldots, \succeq_n^c \rangle$.
- $NE_{faible}^{disc}(G)$ (resp. $NE_{fort}^{disc}(G)$) représente l'ensemble des équilibres de Nash faibles (resp. forts) pour $Pref_G^{disc}$. Comme $\succeq^{bo}$ et $\succeq^{lex}$ sont des relations de préférence totales, la différence entre équilibres de Nash faibles et forts n'est pas pertinente.

Illustrons ces notions sur l'exemple 54 page 131 du dilemme du prisonnier, mais cette fois-ci dans sa forme classique.

**Exemple 61** *Dans cette version du dilemme du prisonnier,* 2 *détenus sont emprisonnés dans des cellules séparées. La police fait à chacun d'eux le même marché :*

> *"Tu as le choix entre couvrir ton complice en te taisant (noté $T_i$, $i = 1, 2$), ou le trahir en le dénonçant (noté $\neg T_i$, $i = 1, 2$). Si tu le dénonces, et qu'il ne te dénonce pas, tu seras remis en liberté et lui écopera de 10 ans de prison. Si vous vous dénoncez mutuellement, vous écoperez tous les deux de 5 ans de prison. Si aucun de vous ne dénonce l'autre, vous aurez tous les deux 6 mois de prison."*

*Voila la forme normale de ce jeu, avec 2 prisonniers :*

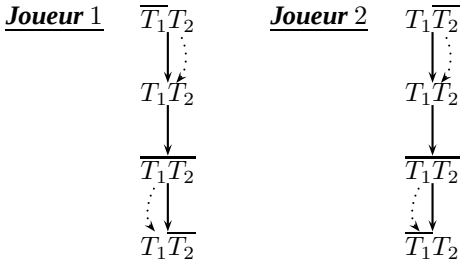
| 1 \ 2 | $T_2$ | $\overline{T}_2$ |
|---|---|---|
| $T_1$ | *(-1/2, -1/2)* | *(-10, 0)* |
| $\overline{T}_1$ | *(0, -10)* | *(-5, -5)* |

*Si l'on calcule l'équilibre de Nash en stratégies pures de ce jeu à partir de cette forme normale, on trouve :*

$$NE = \{\overline{T_1 T_2}\}$$

*Nous pouvons traduire ce jeu par le BP-jeu booléen $G = (N, V, \pi, \Gamma, \Phi)$ suivant :*
- $N = \{1, 2\}$,
- $V = \{T_1, T_2\}$,
- $\pi_1 = \{T_1\}$, $\pi_2 = \{T_2\}$,
- $\gamma_1 = \gamma_2 = \top$,
- $\Sigma_1 = \langle T_2 \,; \neg T_1 \rangle$,
- $\Sigma_2 = \langle T_1 \,; \neg T_2 \rangle$

**Discrimin** *Si l'on applique le critère discrimin à chacun des joueurs, nous trouvons les relations totales suivantes :*

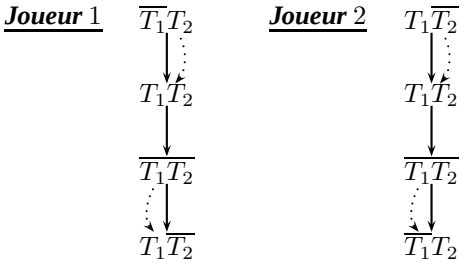



*On constate que ces relations totales, même si elles ne sont qu'ordinales, respectent bien l'ordre donné par la forme normale du jeu.*

*Le calcul des équilibres de Nash faibles et forts nous donne les résultats suivants :*

$$NE^{disc}_{faible} = NE^{disc}_{fort} = \{\overline{T_1 T_2}\}$$

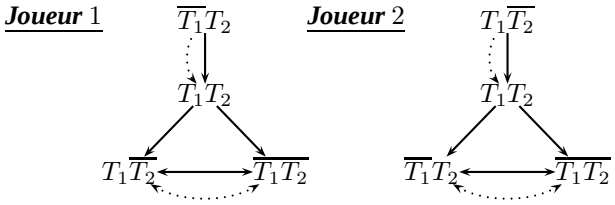
**Leximin** *Si l'on applique le critère leximin à chacun des joueurs, nous trouvons les relations totales suivantes qui sont identiques à celles trouvées avec le discrimin :.*




*Ce jeu a donc le même équilibre de Nash faible et fort :*

$$NE^{lex} = \{\overline{T_1T_2}\}$$

**Best out** *Si l'on applique l'ordre best out à chacun des joueurs, nous trouvons les relations totales suivantes :*




*On trouve :*

$$NE^{bo} = \{T_1\overline{T_2}, \overline{T_1}T_2, \overline{T_1T_2}\}$$

Avec les BP-jeux booléens, on a aussi quelques propriétés intéressantes.

Tout d'abord des liens d'inclusion entre les différents ensembles d'équilibres de Nash :

**Proposition 83** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un BP-jeu booléen et $Pref_G^c = \langle \succeq_1^c, \ldots, \succeq_n^c \rangle$ l'ensemble des relations de préférences sur $G$ à partir du critère $c \in \{disc, lex, bo\}$.*

1. $NE_{fort}^{disc}(G) \subseteq NE^{lex}(G) \subseteq NE^{bo}(G)$
2. $NE^{lex}(G) \subseteq NE_{faible}^{disc}(G) \subseteq NE^{bo}(G)$

Nous pouvons aussi *approximer* un BP-jeu booléen en ne considérant que les $k$-premières strates de chaque joueur. L'objectif est double : obtenir un jeu plus simple (pour le calcul des équilibres de Nash en stratégies pures), et augmenter la possibilité de trouver un équilibre de Nash en stratégies pures significatif qui tient compte des strates les plus prioritaires.

**Définition 113** *Soit $G = (N = \{1, \ldots, n\}, V, \pi, \Gamma, \Phi)$ un BP-jeu booléen. Soit $k \in \{1, p\}$.*
*$G^{[1\to k]} = (N, V, \pi, \Gamma, \Phi^{[1\to k]})$ représente le* **jeu booléen $k$-réduit de** *$G$ dans lequel les buts de chaque joueur sont réduits à leurs $k$ premières strates : $\Phi^{[1\to k]} = \langle \Sigma_1^{[1\to k]}, \ldots, \Sigma_n^{[1\to k]} \rangle$.*

**Proposition 84** *Soit $G$ un BP-jeu booléen, et soit $c \in \{discr, lex, bo\}$. Si $s$ est un équilibre de Nash fort (resp. faible) en stratégies pures pour $Pref_{G^{[1\to k]}}^c$, alors $s$ est aussi un équilibre de Nash fort (resp. faible) pour $Pref_{G^{[1\to k-1]}}^c$.*

Cela donne un résultat intéressant à exploiter pour conclure négativement à l'existence d'un équilibre de Nash :

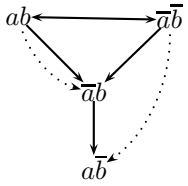
**Proposition 85** *Soit $G$ un BP-jeu booléen, et soit $G^{[1]}$ le jeu booléen 1-réduit associé à $G$. Si, quel que soit le critère utilisé (discrimin, leximin ou best-out), le jeu $G^{[1]}$ n'a pas d'équilibre de Nash en stratégies pures, alors le jeu $G$ n'en a pas non plus.*

**Exemple 62** *Soit $G = (N, V, \pi, \Gamma, \Phi)$ un jeu booléen, avec :*

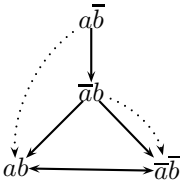
– $N = \{1, 2\}$,
– $V = \{a, b\}$,
– $\pi_1 = \{a\}$, $\pi_2 = \{b\}$,
– $\Sigma_1 = \langle a \rightarrow b\,; b \rightarrow a \rangle$,
– $\gamma_1 = \gamma_2 = \top$,
– $\Sigma_2 = \langle a \leftrightarrow \neg b\,; \neg b \rangle$

*Si l'on applique l'ordre best out à chacun des joueurs, nous trouvons les relations totales suivantes :*

***Joueur*** 1




***Joueur*** 2




*Ce jeu n'a pas d'équilibre de Nash.*

*Si on étudie à présent le jeu booléen $G^{[1]}$ associé dans lequel les buts des joueurs sont réduits à la première strate de leur but dans $G$, on obtient le jeu suivant : $G^{[1]} = \{N, V, \pi, \Gamma, \Phi^{[1]}\}$ un jeu booléen, avec :*
– $N = \{1, 2\}$,
– $V = \{a, b\}$,
– $\pi_1 = \{a\}$, $\pi_2 = \{b\}$,
– $\gamma_1 = \gamma_2 = \top$,
– $\Sigma_1^{[1]} = \{a \rightarrow b\}$,
– $\Sigma_2^{[1]} = \{a \leftrightarrow \neg b\}$

*On peut à présent construire la forme normale de $G^{[1]}$ :*

| 1 \ 2 | $\overline{b}$ | $b$ |
|---|---|---|
| $\overline{a}$ | *(1, 0)* | *(1, 1)* |
| $a$ | *(0, 1)* | *(1, 0)* |

*Ce jeu a un équilibre de Nash : $\overline{a}b$.*

### 4.3.3 Retour sur l'argumentation

L'idée que nous avons ici cherché à appliquer consiste à traduire un système d'argumentation AF en un CP-jeu booléen $G$, afin d'utiliser les outils de la théorie des jeux et les propriétés spécifiques des CP-jeux booléens pour calculer les extensions de AF.

Cette idée est venue des faits suivants :
– l'argumentation et les jeux ont déjà des liens forts, en particulier par le fait qu'ils permettent de représenter les interactions entre agents rationnels ;
– un jeu statique devrait permettre la représentation du AF abstrait (dans les deux cas, les agents donnent leurs arguments/ leurs stratégies sans chercher à analyser ceux/celles des autres agents, l'analyse se faisant ensuite lors du calcul des extensions/des PNEs ;
– l'aspect graphique des CP-nets fait écho à l'aspect graphique du graphe d'argumentation.

Le but de cette idée est beaucoup plus d'établir un nouveau lien entre argumentation et jeux que d'obtenir des algorithmes plus efficaces par l'intermédiaire de cette traduction. Il existe en effet déjà dans la littérature des algorithmes de calcul des extensions préférées qui peuvent être très efficaces (voir par exemple [DM01, CDM03]).

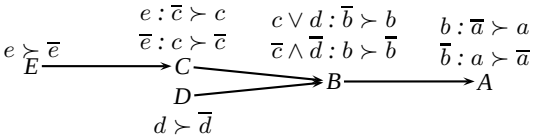
Notons que la partie "contrainte" d'un CP-jeu booléen ne paraît pas utile à la traduction que nous cherchons à réaliser. Nous utiliserons donc dans cette section une version simplifiée des CP-jeux booléens (sans la contrainte $\Gamma$).

#### 4.3.3.1 Traduction d'un AF en un CP-jeu booléen

Cette traduction s'effectue à l'aide de l'algorithme 2 qui présuppose l'existence de deux autres algorithmes :
– EstCyclique qui renvoie *vrai* si le graphe d'argumentation est cyclique[10],
– SupImpCircuits pour supprimer les circuits impairs s'ils existent[11].
L'exécution de ces deux algorithmes peut être vue comme une étape de précompilation de l'algorithme 2. Le fait que l'AF que l'on cherche à traduire ne contienne pas de circuit impair est important car on sait que, dans ce cas-là, on a des propriétés intéressantes du point de vue de l'argumentation et on va voir qu'on en a aussi du point de vue des CP-jeux booléens. Bien-sûr, il existe certains circuits impairs dans des graphes d'argumentation dont on peut considérer qu'ils ont un sens. Mais ce travail étant préliminaire, nous avons choisi la simplicité et opté pour la suppression de ces circuits dans *tous les cas* en utilisant le fait qu'*en général* ces circuits peuvent être considérés comme des paradoxes (voir la section 4.3.3.3 page 149).

Soit AF un système d'argumentation ne contenant pas de circuit impair, les principes de l'algorithme 2 sont les suivants :
– chaque argument de AF est une variable de $G$ ;
– chaque variable est contrôlée par un joueur différent (donc il existe autant de joueurs que de variables) ;
– les CP-nets de tous les joueurs sont définis de la même manière :
  – le graphe du CP-net est exactement le graphe de AF ;
  – les préférences de chaque variable $v$ non attaquée sont $v \succ \overline{v}$ (si un argument n'est pas attaqué, on veut le protéger),
  – les préférences de chaque variable $v$ attaquée par l'ensemble de variables $\mathcal{R}^{-1}(v)$ dépend de ces variables : si au moins l'une d'entre elles $w \in \mathcal{R}^{-1}(v)$ est satisfaite alors $v$ ne pourra pas être satisfaite (on a donc $\bigvee_{w\in\mathcal{R}^{-1}(v)} w : \overline{v} \succ v$) ; et si aucune d'entre elle n'est satisfaite alors $v$ peut être satisfaite (et on a $\bigwedge_{w\in\mathcal{R}^{-1}(v)} \overline{w} : v \succ \overline{v}$).
La construction d'un CP-jeu booléen $G$ pour AF se fait en temps polynomial (même si AF est cyclique et que nous devons d'abord supprimer les circuits impairs). Illustrons cette traduction sur un exemple.

**Exemple 63** *Considérons* $\mathsf{AF} = \langle\{a,b,c,d,e\},\{(b,a),(c,b),(d,b),(e,c)\}\rangle$ *(*AF *est acyclique) et traduisons-le dans le CP-jeu booléen* $G = (N,V,\pi,\Phi)$*. L'application de l'algorithme 2 nous donne* $V = \{a,b,c,d,e\}$ *et* $N = \{1,2,3,4,5\}$*, avec* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$*,* $\pi_3 = \{c\}$*,* $\pi_4 = \{d\}$ *et* $\pi_5 = \{e\}$*. Le CP-net suivant représente les préférences de tous les joueurs :*




Prenons maintenant l'exemple d'un AF avec des circuits mais uniquement pairs.

**Exemple 64** *Considérons* $\mathsf{AF} = \langle\{a,b\},\{(a,b),(b,a)\}\rangle$*. L'application de l'algorithme 2 donne* $V = \{a,b\}$ *et* $N = \{1,2\}$*, avec* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$*. Le CP-net suivant représente les préférences de tous les joueurs :*

Et enfin, donnons l'exemple d'un AF avec des circuits impairs.

[10] Cet algorithme est linéaire : (Étape 1) suppression de tous les sommets n'ayant pas de prédécesseurs ; (Étape 2) réitérer l'étape 1 jusqu'à ce que tous les sommets restant aient un prédécesseur (il existe un circuit), ou que le graphe soit vide (il n'y a pas de circuit).

[11] Cet algorithme est polynomial : (Étape 1) calcul de la matrice d'adjacence booléenne correspondant à tous les chemins minimaux de longueur impaire ; Pour cela il suffit de partir de la matrice d'adjacence booléenne du graphe d'attaque $\mathcal{M}$ ($\mathcal{M}(i,j) = 1$ s'il existe une attaque de $i$ sur $j$ dans AF) et de calculer $\mathcal{M}^{\text{olc}} = \mathcal{M}^1 + \mathcal{M}^3 + \ldots + \mathcal{M}^{2n-1}$ avec $n = |\mathcal{A}|$ (la borne $2n-1$ est obtenue par un résultat général de la théorie des graphes : s'il existe un chemin de $a$ vers $b$ dans un graphe dirigé alors il existe aussi un chemin simple de $a$ vers $b$ – c'est-à-dire un chemin dans lequel chaque sommet n'apparaît qu'une seule fois –) ; (Étape 2) suppression de tous les arguments pour lesquels la valeur diagonale de $\mathcal{M}^{\text{olc}}$ est 1 ; (Étape 3) suppression de tous les arcs dont l'une des extrémités a été supprimée à l'étape 2.

**Algorithme 2:** Traduction d'un AF en un CP-jeu booléen

**begin**

/* INPUT : $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$ un système d'argumentation */
/* OUTPUTS : $G = (N, V, \pi, \Phi)$ un CP-jeu booléen, AF après suppression des circuits impairs */
/* VARIABLES LOCALES : $i$ = agent courant, $a$ = argument courant */

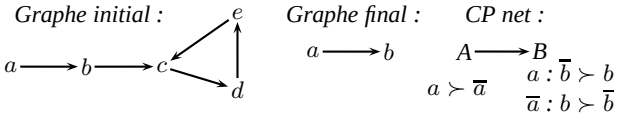
/* si nécessaire, suppression des circuits impairs */
**if** EstCyclique(AF) **then** AF = SupImpCircuits(AF)
/* plus de circuit impair dans AF */
/* calcul des CPTs de chaque argument */
**for** $a \in \mathcal{A}$ **do**
  **if** $\mathcal{R}^{-1}(a) = \varnothing$ **then** $CPT(a) = a \succ \overline{a}$
  /* argument non attaqué */
  **else**
  /* les autres arguments */
    $CPT(a) = \{\bigvee_{v \in \mathcal{R}^{-1}(a)} v \ : \overline{a} \succ a\} \cup \{\bigwedge_{v \in \mathcal{R}^{-1}(a)} \overline{v} \ : a \succ \overline{a}\}$

/* calcul du CP-net $\mathcal{N}$ */
$\mathcal{N} = \langle \mathsf{AF}, \cup_{a \in \mathcal{A}} CPT(a) \rangle$ /* graphe d'attaque */
/* après suppression des circuits impairs, */
/* associé à la CPT pour chaque argument */

/* calcul de $N$, $V$, $\pi$ et $\Phi$ */
$i = 1$, $N = \varnothing$, $V = \mathcal{A}$ /* chaque argument est une variable */
**for** $a \in \mathcal{A}$ **do**
  $N = N \cup \{i\}$ /* un agent pour chaque argument */
  $\pi_i = \{a\}$ /* $i$ contrôle seulement cet argument */
  $\mathcal{N}_i = \mathcal{N}$ /* le même CP-net pour chaque agent */
  $i = i + 1$
**return** ($G = (N, V, \pi, \langle \mathcal{N}_1, \ldots, \mathcal{N}_{|V|} \rangle)$, AF)

**end**

**Exemple 65** *Soit* $\mathsf{AF} = \langle \{a, b, c, d, e\}, \{(a, b), (b, c), (c, d), (d, e), (e, c)\} \rangle$ *qui contient des circuits impairs à supprimer. Après suppression* AF *ne contient plus que* $a$ *et* $b$*. Et l'application de l'algorithme 2 donne* $V = \{a, b\}$*,* $N = \{1, 2\}$*, avec* $\pi_1 = \{a\}$*,* $\pi_2 = \{b\}$ *et le CP-net suivant représente les préférences de tous les joueurs :*




Nous avons bien-sûr montré qu'il existe un lien entre les extensions préférées de AF et les équilibres de Nash de $G$ :

**Proposition 86** *Soit* $\mathsf{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$*. Soit* $G = (N, V, \pi, \Phi)$ *le CP-jeu booléen et* $\mathsf{AF}'$ *le système d'argumentation obtenus à partir de* AF *par l'application de l'algorithme 2.*

*$s$ est une extension préférée de* $\mathsf{AF}'$ *si et seulement si $s$ est un SPNE pour* $G$[12].

**Exemple 63 page précédente (suite)** *$G$ a un seul SPNE* $\{ed\overline{c}\overline{b}a\}$ *et* AF *a une seule extension préférée* $\{e, d, a\}$.

**Exemple 64 page ci-contre (suite)** *$G$ a deux SPNEs* $\{a\overline{b}, \overline{a}b\}$ *et* AF *a aussi deux extensions préférées* $\{a\}$, $\{b\}$.

**Exemple 65 (suite)** *$G$ a un seul SPNE* $\{a\overline{b}\}$ *et* AF *(*après suppression des circuits impairs*) a une seule extension préférée* $\{a\}$.

[12] Rappelons qu'un SPNE est une $V$-interprétation, c'est-à-dire que si $s = a\overline{b}c$ alors $s$ correspond à l'ensemble $\{a, c\}$.

#### 4.3.3.2 Calcul des extensions préférées

Puisque les extensions préférées correspondent exactement aux SPNEs, les principales propriétés concernant le calcul des SPNE dans le cadre des CP-jeux booléens peuvent être utilisées. La première concerne le cas des AFs acycliques :

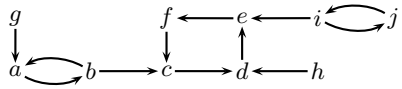
**Proposition 87** *Soit* AF *un AF. Soit* $G$ *le CP-jeu booléen et* AF$'$ *l'AF obtenus à partir de* AF *en appliquant l'algorithme 2.*

*Si* AF$'$ *est acyclique, alors* AF$'$ *possède une seule extension préférée calculable en temps polynomial en utilisant* $G$*.*

Cette proposition est bien-sûr vérifiée dans le cas simple où AF est initialement acyclique.

Le calcul des SPNE(s) pour des AFs cycliques est un peu plus complexe. Nous proposons ici deux algorithmes (les algorithmes 3 et 4) qui permettent d'effectuer ce calcul quand l'AF ne contient que des circuits de longueur paire.

Ces algorithmes présupposent l'existence de l'algorithme CompIntCircuitForProp qui renvoie le circuit (ou l'un des circuits s'il en existe plusieurs) qui permet d'atteindre le plus de variables possible dans un ensemble donné de variables[13]. Par exemple, sur le graphe suivant :

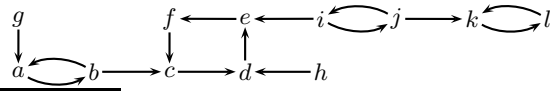



$\{a, b\}$ permet d'atteindre les variables $a$, $b$, $c$, $d$, $e$, $f$, et $\{i, j\}$ permet d'atteindre les variables $i$, $j$, $c$, $d$, $e$, $f$. Ces circuits sont plus intéressants que tous les autres du point de vue de la propagation de valeurs à travers le graphe (cela signifie que s'ils sont le point d'entrée d'un processus de propagation alors cette propagation sera plus efficace).

Soit $\mathcal{N}$ le CP-net représentant les buts des joueurs d'un CP-jeu booléen, les principes des algorithmes 3 et 4 sont les suivants :

- instanciation de toutes les variables non-attaquées (celles qui n'ont pas de parents dans $\mathcal{N}$ et qui seront donc satisfaites dans le SPNE) ;
- propagation de ces instanciations le plus loin possible ;
- une fois que toutes ces instanciations ont été réalisées, faire la boucle suivante :
  - si toutes les variables ont été instanciées, on renvoie le SPNE construit ;
  - sinon, avec l'algorithme CompIntCircuitForProp, le circuit $C$ restant le plus intéressant est calculé (il en existe forcément un, sinon toutes les variables auraient été instanciées) ;
  - à partir de l'état courant du SPNE courant, on crée deux nouveaux SPNEs ; le premier contient une variable de $C$ instanciée à *vrai*, le second contient cette même variable instanciée à *faux* ;
  - propagation de ces instanciations pour chacun de ces SPNEs aussi loin que possible.

**Exemple 66** *Considérons le graphe suivant :*




[13] Cet algorithme utilise la notion de matrice d'adjacence booléenne tout comme l'algorithme SupImpCircuits :
- calcul de la matrice d'adjacence booléenne $\mathcal{M}^{\text{ap}}$ correspondant à tous les chemins minimaux dans le graphe réduit à un ensemble donné de variables : $\mathcal{M}^{\text{ap}} = \mathcal{M} + \mathcal{M}^2 + \mathcal{M}^3 + \ldots + \mathcal{M}^{2n}$ avec $n = |V|$ ; $\mathcal{M}^{\text{ap}}(i)$ dénotera $(\mathcal{M}^{\text{ap}}(i,1), \ldots, \mathcal{M}^{\text{ap}}(i,n))$ ;
- ToSee $= V$ ; $C = \varnothing$ ; end? $=$*faux* ;
- Boucle : tant que NON(end?) faire
  $v =$ top(ToSee) ; ToSee $=$ ToSee $\setminus \{v\}$ ;
  si ($\nexists w \in$ ToSee telle que $\mathcal{M}^{\text{ap}}(v) \subset \mathcal{M}^{\text{ap}}(w)$) alors
  /* aucune autre variable ne permet d'atteindre plus de variables que $v$ */
  $C = C \cup \{v\}$ ; end? $=$*vrai* ;
  $\forall w \in$ ToSee faire si $\mathcal{M}^{\text{ap}}(v) = \mathcal{M}^{\text{ap}}(w)$ alors $C = C \cup \{w\}$ ;
  sinon si ToSee est vide alors end? $=$*vrai* ;
- Renvoyer $C$

*les étapes du processus de calcul sont les suivantes :*
- *$g$ et $h$ sont instanciées à vrai (état courant du SPNE $= gh$) ;*
- *puis $a$ et $d$ sont instanciées à faux (état courant du SPNE $= gh\overline{a}\overline{d}$) ;*
- *puis $b$ est instanciée à vrai (état courant du SPNE $= gh\overline{a}\overline{d}b$) ;*
- *puis $c$ est instanciée à faux (état courant du SPNE $= gh\overline{a}\overline{d}b\overline{c}$) ;*
- *à ce point la propagation simple s'arrête ; nous devons alors calculer les circuits les plus intéressants dans l'ensemble des variables restantes $(e, f, i, j, k, l)$ et on obtient le circuit $(i, j)$ ;*
- *le processus de propagation redémarre avec les états courants des deux SPNEs courants : $gh\overline{a}\overline{d}b\overline{c}i$ et $gh\overline{a}\overline{d}b\overline{c}\overline{i}$ ;*
- *...*
- *à la fin du processus de propagation, on obtient $gh\overline{a}\overline{d}b\overline{c}i\overline{e}f\overline{j}k\overline{l}$, $gh\overline{a}\overline{d}b\overline{c}i\overline{e}f\overline{j}kl$ et $gh\overline{a}\overline{d}b\overline{c}\overline{i}e\overline{f}j\overline{k}l$. Il existe donc trois SPNEs qui correspondent aux trois extensions préférées $\{g, h, b, e, j, l\}$, $\{g, h, b, i, f, l\}$ et $\{g, h, b, f, i, k\}$.*

**Algorithme 3:** Calcul des SPNEs d'un CP-jeu booléen obtenu à partir d'un AF

**begin**

/* INPUT : un CP-jeu booléen $G = (N, V, \pi, \Phi)$, avec $\Phi = \langle \mathcal{N}_1, \ldots, \mathcal{N}_n \rangle$ */
/* OUTPUT : un ensemble de SPNEs $SP$ */
/* VARIABLES LOCALES : $v =$ variable courante, $In =$ (resp. $Out =$) ensemble des variables instanciées à *vrai* (resp. *faux*), $R =$ ensemble des variables restant à instancier */

$In = \varnothing$, $Out = \varnothing$, $R = V$ /* Initialisation */
/* Instanciation de toutes les variables sans parents */
**for** $v \in R$ **do if** $Pa(v) = \varnothing$ **then** $R = R \setminus \{v\}$, $In = In \cup \{v\}$
/* propagation par un processus récursif */
**return** CompSPNERec($G$, $R$, $In$, $Out$)

**end**

La proposition suivante montre que les algorithmes 3 et 4 calculent exactement l'ensemble des SPNEs du CP-jeu booléen.

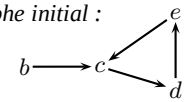
**Proposition 88** *Soit $G$ un CP-jeu booléen obtenu en appliquant l'algorithme 2. Soit $SP$ l'ensemble des profils de stratégie de $G$ obtenu en appliquant les algorithmes 3 et 4. $s \in SP$ si et seulement si $s$ est un SPNE pour $G$.*

#### 4.3.3.3 Gestion des circuits impairs

La suppression des circuits impairs a une influence importante sur le calcul des SPNE(s) et ce point pourrait être considéré comme problématique par ceux qui n'adhèrent pas à notre hypothèse simplificatrice de départ : *en général*, un circuit impair peut être considéré comme un paradoxe (nous sommes toutefois conscientes qu'il existe parfois des circuits impairs, en particulier, quand ils ne sont pas stricts, qui ont un sens et qu'il serait donc judicieux de ne pas supprimer ; cela sera pris en compte dans une future version de ce travail).

**Exemple 67** *Considérons* AF $= \langle \{b, c, d, e\}, \{(b, c), (c, d), (d, e), (e, c)\} \rangle$. *L'*AF *initial est cyclique, et contient un circuit impair qui sera supprimé par l'application de l'algorithme 2 ; l'*AF *final ne contiendra plus que $b$ et le CP-jeu booléen correspondant sera composé de $V = \{b\}$, $N = \{1\}$, avec $\pi_1 = \{b\}$ et du CP-net suivant représentant les préférences de tous les joueurs :*

*Graphe initial :*




*Graphe final :* $b$

*CP net :* $B$ : $b \succ \overline{b}$

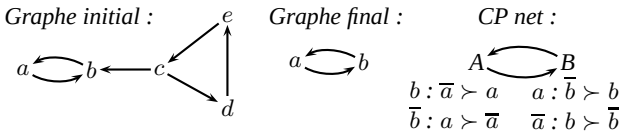
*Donc, $G$ a un seul SPNE $\{b\}$ qui correspond à l'extension préférée de l'*AF *final. Toutefois, il ne correspond pas à l'extension préférée de l'*AF *initial qui était l'ensemble $\{b, d\}$. Si nous considérons qu'un circuit impair est un paradoxe, alors ses arguments ne sont pas significatifs, et nous pouvons considérer que $\{b\}$ est une extension plus réaliste que $\{b, d\}$ (cette approche n'est malheureusement pas validée par la plupart des sémantiques pour l'acceptabilité).*

**Algorithme 4:** COMPSPNEREC : Calcul récursif des SPNEs d'un CP-jeu booléen obtenu à partir d'un AF

```
begin
    /* INPUTS : un CP-jeu booléen G = (N, V, π, Φ),
         R = ensemble des variables restant à instancier,
         In = ensemble des variables déjà instanciées à vrai,
         Out = ensemble des variables déjà instanciées à faux*/
    /* OUTPUT : un ensemble de SPNEs SP */
    /* VARIABLES LOCALES : v = variable courante, n = cardinal de R, C = ensemble de variables formant un
    circuit*/

    if R = ∅ then
        /* toutes les variables sont instanciées : un SPNE a été trouvé */
        return {(In Out̄)}
    else
        n = |R|  /* n = nombre des variables restant à instancier */
        for v ∈ R do
            /* propagation simple */
            if Pa(v) ⊆ Out then
                /* tous les parents sont instanciés à faux */
                In = In ∪ {v}, R = R \ {v}
            else
                if (Pa(v) ∩ In) ≠ ∅ then
                    /* au moins 1 parent instancié à vrai */
                    Out = Out ∪ {v}, R = R \ {v}
        if n = |R| then
            /* aucune variable instanciée dans l'instruction For */
            C = COMPINTCIRCUITFORPROP(G, R)
            v = TOP(C)
            return(COMPSPNEREC(G, R \ {v}, In ∪ {v}, Out) ∪ COMPSPNEREC(G, R \ {v}, In, Out ∪ {v}))
        else
            /* au moins 1 variable instanciée dans l'instruction For */
            return COMPSPNEREC(G, R, In, Out)
end
```

**Exemple 68** *Considérons* AF $= \langle\{a,b,c,d,e\}, \{(a,b),(b,a),(c,b),(c,d),(d,e),(e,c)\}\rangle$. *L'*AF *initial est cyclique, et contient un circuit impair à supprimer ; donc l'*AF *final contiendra seulement* $a$ *et* $b$ *et le CP-jeu booléen sera composé de* $V = \{a,b\}$, $N = \{1,2\}$, *avec* $\pi_1 = \{a\}$, $\pi_2 = \{b\}$ *et du CP-net suivant représentant les préférences de tous les joueurs :*




*Ici,* $G$ *possède deux SPNEs* $\{a\bar{b}, \bar{a}b\}$ *qui correspondent aux deux extensions préférées de l'*AF *final. Toutefois, ils ne correspondent pas à l'extension préférée de l'*AF *initial qui était l'ensemble* $\{a\}$. *Dans ce cas, la prise en compte de l'extension* $\{b\}$ *signifie que l'attaque de* $c$ *sur* $b$ *est considérée comme non significative puisque* $c$ *appartient à un circuit impair qui est un paradoxe, donc* $c$ *est jugé non significatif et donc ne peut fournir une attaque réaliste d'un autre argument.*

## 4.4 Conclusion sur le thème 4

Dans ce chapitre, nous avons présenté la notion de jeux booléens et nous l'avons utilisée pour simuler les interactions entre agents rationnels. Ainsi, nous avons à la fois proposé un autre mode de traitement de l'incohérence qui s'approche de l'idée de la restauration de cohérence (à partir d'informations incohérentes, définition de paquets d'informations cohérents sur lesquels on peut appliquer un raisonnement classique) et établi un lien supplémentaire entre argumentation et théorie des jeux.

En effet, ces jeux ont la particularité de décrire les préférences des agents sous la forme de formules booléennes plus ou moins élaborées (cela va de la simple formule propositionnelle à des bases de formules avec priorité ou à des CP-nets). Ainsi, on retrouve le fait que plusieurs agents se partagent (pas forcément de manière exclusive) un ensemble de connaissances pouvant être inconsistant. Le calcul des concepts de solution d'un tel jeu revient à identifier les ensembles consistants et "préférés" de buts atteints par les joueurs, ce qui est exactement la finalité de la restauration de cohérence.

Ce type de jeu peut alors être utilisé pour modéliser un système d'argumentation et permettre ainsi le calcul des ensembles d'arguments "acceptables" (les extensions) par l'intermédiaire du calcul des concepts de solution du jeu (les équilibres de Nash en stratégie pure – PNE). Il s'agit essentiellement d'établir un nouveau lien entre argumentation et jeux, cette traduction ne permettant pas d'obtenir des algorithmes plus efficaces que ceux déjà existants. Cette modélisation n'est pour l'instant pas complète puisqu'elle ne traite pas le cas des systèmes d'argumentation à circuit impair. Cet inconvénient ne nous paraît pas bloquant pour l'instant puisque la plupart de ces systèmes d'argumentation sont connus comme pouvant être problématiques (un circuit impair est très souvent considéré comme un paradoxe et un système d'argumentation contenant des circuits impairs risque de ne pas avoir certaines propriétés intéressantes).

Le travail réalisé ici n'en est encore qu'à une phase préliminaire et mérite de se poursuivre.

# Chapitre 5

# Projet de recherche

La grande majorité de mes travaux ont concerné l'argumentation et je souhaiterais poursuivre dans cette thématique qui me paraît recéler encore beaucoup de potentialités.

En effet, l'argumentation vue dans sa globalité comporte plusieurs aspects :

1. il s'agit tout d'abord de construire des "arguments" à partir d'ensembles d'informations,
2. puis d'identifier les "interactions" (positives ou négatives) entre ces arguments,
3. ensuite, on peut "évaluer" les arguments ou les interactions en utilisant divers mécanismes (par exemple, les préférences entre connaissances),
4. on peut alors utiliser tout cela pour choisir, sélectionner les "meilleurs" arguments,
5. et enfin, on peut "conclure" en utilisant les arguments sélectionnés (la nature de la conclusion dépendant bien-sûr du problème pour lequel on a argumenté).

Toutes ces étapes ont fait l'objet de très nombreux travaux et personnellement j'ai beaucoup travaillé sur les étapes 3 et 4.

Tous ces travaux ont donné de multiples systèmes d'argumentation que l'on peut schématiquement regrouper en deux grandes familles :

– La première famille regroupe les systèmes d'argumentation *abstraits* qui ignorent les deux premières étapes du processus d'argumentation en considérant que les arguments et les interactions sont donnés ; ces travaux se focalisent sur les notions d'évaluation et d'acceptabilité ; tous mes travaux sur l'argumentation entrent dans ce cadre mais ce ne sont pas les seuls (voir par exemple les travaux décrits dans [Dun95, Cam06]).
– La seconde famille regroupe des systèmes d'argumentation plus *concrets* dans le sens où ils prennent en compte toutes les étapes du processus d'argumentation ; malheureusement, il s'agit en général de systèmes développés pour des applications particulières (par exemple, pour gérer l'inconsistance dans des bases de connaissances ou pour expliquer des décisions) ; de nombreux exemples de ce type de systèmes sont donnés dans [SL92, PS97, AC02b, BH01] ...

Dès lors, un problème essentiel apparaît : comment faire le tri parmi tous ces systèmes ? Il existe déjà quelques études comparatives qui ont permis de mettre en évidence des liens, des similarités, des différences entre certains systèmes appartenant à la même famille.

Par exemple, pour les systèmes abstraits, [BG07] a proposé une comparaison des différentes acceptabilités basée sur une collection de critères.

De même, pour les systèmes concrets, des comparaisons ont été établies entre systèmes destinés au même type d'application. Par exemple, dans le projet ASPIC (voir http ://aspic.acl.icnet.uk), on peut trouver une comparaison des systèmes gérant l'inconsistance, alors que dans [AP08] la comparaison concerne les systèmes destinés à faire de la décision argumentée. Le plus souvent, ces études comparatives s'appuient sur les conclusions (les sorties) des systèmes en laissant de côté les langages utilisés ainsi que la manière dont les arguments et leurs interactions sont définis et dont on effectue le choix entre les arguments ; cela paraît être un peu trop succinct et générateur d'erreurs.

De plus, malgré l'abondante littérature qui existe dans ce domaine, il reste quelques questions en suspens :

– Quels sont les critères permettant de comparer deux systèmes d'argumentation appartenant à une même famille ?
– Quels sont les critères permettant de comparer deux systèmes d'argumentation destinés à une même application ?
– Comment comparer deux systèmes développés pour deux applications distinctes ?
– Comment comparer deux systèmes appartenant à deux familles distinctes ?
– Quelle est la classe de systèmes la plus appropriée pour un type donné d'application ?

Mon projet serait donc le suivant : répondre à ces diverses questions en proposant
– l'unification des systèmes d'argumentation existants dans un cadre général qui prendrait en compte toutes les étapes d'un processus d'argumentation,
– la définition de postulats de rationalité devant être satisfaits par un "bon" système d'argumentation,
– la définition d'un ensemble de critères permettant d'évaluer et de comparer les systèmes d'argumentation (l'un des critères étant évidemment le respect des postulats de rationalité),
– et enfin la validation à la fois théorique et expérimentale de ces critères.

## 5.1 État de l'art sur le sujet

Il existe actuellement quelques travaux sur certains aspects de ce projet qui justifient d'autant plus l'intérêt du dit-projet.

**Existence de critères de comparaison et de postulats de rationalité pour l'argumentation en IA** Quelques études existent déjà sur ce thème. On trouve tout d'abord de très rares travaux sur la partie axiomatisation de systèmes d'argumentation particuliers (voir [CA07] qui propose deux postulats, un pour la consistance et un pour la fermeture). Il existe, par contre, beaucoup plus de travaux concernant la comparaison de systèmes d'argumentation (voir [Cam06, JV99b, BG07]).

**Existence de liens entre argumentation et inférence non monotone** Une étude semblable à celle qui m'intéresse a déjà été menée dans le cadre de l'inférence non monotone et je me suis même appuyée dessus lors de mes travaux de thèse (voir chapitre 1 page 9 ainsi que tous les travaux [Gab85, KLM90, LM92, GM94]). Or, il existe un lien fort entre ces deux domaines, autant du point de vue théorique (voir [Dun95, Cay95, BDKT97, NOC08, TKI08]) que du point de vue pratique (voir [EGW08] ainsi que les résultats du projet ASPIC). La question se pose alors d'exploiter ces liens pour étudier l'application des postulats de rationalité de l'inférence non monotone au domaine de l'argumentation. L'idée à exploiter serait alors de définir des mécanismes de traduction permettant de passer d'un domaine à l'autre.

**Existence de critères de comparaison et de postulats de rationalité pour l'argumentation en psychologie cognitive et en philosophie** L'argumentation n'est pas seulement une composante de l'IA, c'est aussi un domaine de recherche lié fortement à la philosophie et à la psychologie cognitive. Dans ces domaines, il existe déjà quelques travaux proposant des ensembles de règles ou de caractéristiques des processus d'argumentation qui pourraient sûrement avoir un sens en IA (voir par exemple [Rip98, OCH08, HO07, Tha03]). Un autre aspect particulièrement intéressant est lié au point précédent : lors des travaux sur les postulats de rationalité en inférence non monotone, une validation expérimentale de ces postulats avait été menée sous la forme d'expérimentations psychologiques (voir [NBR02, BBN04, BBN05]). Il semble donc intéressant de renouveler l'expérience dans le cadre de l'argumentation.

## 5.2 Originalité du projet

Ce projet est original sur au moins trois aspects :

**Une tentative de vision globale du processus d'argumentation en IA** Aucun des travaux existants sur ce sujet ne traite le processus d'argumentation dans son intégralité (soit on reste au niveau du langage et des arguments, soit on passe à un niveau abstrait en occultant la manière dont les arguments sont construits).

**La nécessité de la pluridisciplinarité** Pour mener à bien ce projet, il va falloir à la fois des informaticiens spécialisés en Intelligence Artificielle mais aussi des pyschologues-cogniticiens. Les premiers seront à même d'identifier des propriétés théoriques propres aux mécanismes d'argumentation en IA alors que les seconds pourront valider d'un point de vue cognitif les dites-propriétés et, pourquoi pas, extraire de nouvelles propriétés des expérimentations cognitives menées sur le terrain afin d'enrichir la partie théorique.

**La réalisation d'un outil de traduction des mécanismes d'argumentation vers les mécanisme d'inférence non monotone et vice-versa** Un tel outil faciliterait évidemment l'étude de la "portabilité" des postulats de rationalité issus de l'inférence non monotone au domaine de l'argumentation et serait un plus pour étudier tous les points liant ces deux thèmes.

## 5.3 Organisation du projet

Ce projet pourrait se décomposer en quatre étapes :

**Étape 1** Étude des travaux existants sur le sujet et concernant aussi bien l'aspect théorique qu'expérimental.

**Étape 2** Définition des principes généraux des systèmes d'argumentation en suivant trois axes de recherche : étude théorique des mécanismes d'argumentation en IA, impact des postulats de rationalité issus de l'inférence non monotone (à l'aide de l'outil de traduction) et découverte d'autres principes par l'expérimentation psychologique.

**Étape 3** Formalisation des critères de comparaison entre systèmes d'argumentation (parmi ces critères on trouvera sûrement le respect des principes identifiés à l'étape 2).

**Étape 4** Validation à la fois théorique et expérimentale des principes et des critères.

Ce projet concernerait bien-sûr des informaticiens spécialisés en IA avec des connaissances solides en argumentation et en raisonnement non monotone mais il nécessiterait aussi la participation de psychologues afin d'étudier dans un premier temps la littérature sur ce sujet dans leur domaine de compétence et de mener ensuite des expérimentations destinées tout d'abord à découvrir quelques principes des systèmes d'argumentation, puis à valider l'ensemble des principes et des critères identifiés dans le projet.

Ce projet a été déposé en tant que projet ANR blanc en janvier 2009 et il concerne 11 personnes réparties dans 4 équipes de recherche en France.

**Mon rôle dans ce projet :** Je ne serais pas que la coordinatrice de projet, car je suis aussi très intéressée par l'aspect théorique concernant la définition des postulats de rationalité et la formalisation de critères de comparaison entre systèmes d'argumentation. Il me semble que cela me permettrait de valoriser à la fois mes connaissances sur ce sujet acquises dans le cadre plus restreint de l'inférence non monotone, et celles concernant les mécanismes de l'argumentation.

# Bibliographie


[ABCM04] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Justifying practical reasoning. In C. Reed F. Grasso and G. Carenini, editors, *Proceedings of the Fourth Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 87–90, 2004.

[ABP05] Leila Amgoud, Jean-François Bonnefon, and Henri Prade. An argumentation-based approach to multiple criteria decision. In *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 269–280, Barcelone, Spain, 2005.

[AC98] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In G. F. Cooper and S. Moral, editors, *Proc. of the $14^{th}$ Uncertainty in Artificial Intelligence*, pages 1–7, Madison, Wisconsin, 1998. Morgan-Kaufmann.

[AC02a] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, 29 :125–169, 2002.

[AC02b] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34 :197–216, 2002.

[ACLS04] Leila Amgoud, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In J. Delgrande and T. Schaub, editors, *Proc. of the $10^{th}$ NMR workshop (Non Monotonic Reasoning), Uncertainty Framework subworkshop*, pages 1–9, Whistler, BC, Canada, 2004.

[ACLSL08] Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23 :1062–1093, 2008.

[ADLS08] Leila Amgoud, Caroline Devred, and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In Lin Padgham and David Parkes, editors, *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), Estoril - Portugal, 12/05/08-16/05/08*, pages 429–436. ACM, 2008.

[ADLS09] Leila Amgoud, Caroline Devred, and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In Iyad Rahwan and Pavlos Moraïtis, editors, *Argumentation in multi-agent systems*, volume "LNAI 5384", pages 35–54. Springer Verlag, 2009.

[AGM85] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change : partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50 :510–530, 1985.

[AK05] Leila Amgoud and Souhila Kaci. On the generation of bipolar goals in argumentation-based negotiation. In Iyad Rahwan, Pavlos Moraitis, and Chris Reed, editors, *Argumentation in Multi-Agent Systems : (Proceedings of the First International Workshop (ArgMAS'04) : Expanded and Invited Contributions)*, volume 3366 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin, Germany, 2005.

[ALS07] Leila Amgoud and Marie-Christine Lagasquie-Schiex. A constrained argumentation system for practical reasoning. Technical Report IRIT/RR- -2007-14- -FR, IRIT-UPS, June 2007.

[AM02] L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proc. of NMR*, pages 409–417, 2002.

[Amg99] Leila Amgoud. *Contribution à l'intégration des préférences dans le raisonnement argumentatif*. PhD thesis, Université Paul Sabatier, Toulouse, July 1999.

[Amg03] Leila Amgoud. A formal framework for handling conflicting desires. In *Proc. of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2003*, pages 552–563, 2003.

[Amg05] L. Amgoud. Towards a formal model for task allocation via coalition formation. In *Proc. of AAMAS*, pages 1185–1186, 2005.

[AMP00] Leila Amgoud, Nicolas Maudet, and Simon Parsons. Arguments, Dialogue and Negotiation. In W. Horn, editor, *Proc. of the $14^{th}$ ECAI (European Conference of Artifical Intelligence)*, pages 338–342, Berlin, Germany, August 2000.

[AP02] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.J. Meyer and M. Tambe, editors, *Intellignet Agents VIII : Agent theories, architectures and languages*, pages 190–205, Berlin, Germany, 2002. Springer.

[AP04] L. Amgoud and H. Prade. Reaching agreement through argumentation : A possibilistic approach. In *Proc. of the $9^{th}$ International Conference on the principles of knowledge representation and reasoning (KR)*, 2004.

[AP07] Leila Amgoud and Henri Prade. Practical reasoning as a generalized decision making problem. In *Proc. of MFI'07*, pages 15–24. Annals of LAMSADE, 2007.

[AP08] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 2008.

[BBD+04a] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. CP-nets : A Tool for Representing and Reasoning with Conditional *Ceteris Paribus* Preference Statements. *Journal of Artificial Intelligence Research*, 21 :135–191, 2004.

[BBD+04b] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. Preference-Based Constrained Optimization with CP-nets. *Computational Intelligence*, 20(2) :137–157, 2004. Special Issue on Preferences.

[BBHP99] Craig Boutilier, Ronen I. Brafman, Holger H. Hoos, and David Poole. Reasoning with Conditional Ceteris Paribus Preference Statements. In *Uncertainty in Artificial Intelligence (UAI'99)*, pages 71–80, 1999.

[BBN04] S. Benferhat, J. F. Bonnefon, and R. M. Da Silva Neves. An experimental analysis of possibilistic default reasoning. In D. Dubois, C. Welty, and M. Williams, editors, *In Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning (KR 2004)*, pages 130–139, 2004.

[BBN05] S. Benferhat, J. F. Bonnefon, and R. M. Da Silva Neves. An overview of possibilistic handling of default reasoning : an experimental study. *Synthese*, pages 53–70, 2005.

[BBPW04] J. Bennaim, S. Benferhat, O. Papini, and E. Würbel. An answer set programming encoding of prioritized removed sets revision : application to GIS. In J. Alferes and J. Leite, editors, *Proceedings of JELIA (LNAI : Logics for AI)*, pages 604–616, Lisbonne, Portugal, September 2004. Springer Verlag.

[BC98] T. Bench-Capon. Specification and implementation of toulmin dialogue game. In *Proceedings of JURIX 98*, pages 5–20, GNI, Nijmegen, 1998.

[BC02] Trevor J.M. Bench-Capon. Value based argumentation frameworks. In Benferhat and Giunchiglia, editors, *Proc. of the $9^{th}$ International Workshop on Nonmonotonic Reasoning (session on Argument, Dialogue and Decision)*, pages 444–453, Toulouse, France, 2002.

[BCD+93] Salem Benferhat, Claudette Cayrol, Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency management and prioritized syntax-based entailment. In Ruzena Bajcsy, editor, *Proc. of the $13^{th}$ IJCAI*, pages 640–645, Chambéry, France, 1993. Morgan-Kaufmann.

[BCDD07] T. Bench-Capon, S. Doutre, and P.E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171 :42–71, 2007.

[BDKP02] Salem Benferhat, Didier Dubois, Souhila Kaci, and Henri Prade. Bipolar representation and fusion of preferences in the possibilistic logic framework. In *Proceedings of the eighth International Conference on Principle of Knowledge Representation and Reasoning (KR'02)*, pages 158–169, 2002.

[BDKT97] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93 :63–101, 1997.

[BDP93] Salem Benferhat, Didier Dubois, and Henri Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In David Heckerman and Abe Mamdani, editors, *Proc. of the $9^{th}$ UAI*, pages 411–419, Washington, DC, 1993. Morgan-Kaufmann.

[BDP94] Salem Benferhat, Didier Dubois, and Henri Prade. Some syntactic approaches to the handling of inconsistent knowledge bases : a comparative study. Rapport de recherche 94-55R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), December 1994.

[Ber73] C. Berge. *Graphs and Hypergraphs*. North-Holland Mathematical Library, 1973.

[BG98] Salem Benferhat and Laurent Garcia. Traitement de bases de connaissances incohérentes localement stratifiées et application au raisonnement tolérant les exceptions. In LASMEA, editor, *Proc. of the $11^{th}$ RFIA*, pages 233–242, Clermont Ferrand, France, 1998. AFCET.

[BG07] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10–15) :675–700, 2007.

[BH01] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128 (1-2) :203–235, 2001.

[BLSL05] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Jeux booléens statiques et représentation compacte de préférences (rapport préliminaire). In Andreas Herzig, Yves Lespérance, and Abdel-Illah Mouaddib, editors, *Proc. of the $3^{rd}$ MFI (Journées francophones des Modèles formels de l'interaction)*, Caen, France, 2005.

[BLSL06a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Compact preference representation for Boolean games. In *Pacific Rim International Conference on Artificial Intelligence (PRICAI), Guilin, China, 07/08/2006-11/08/2006*, volume 4099, pages 41–50, http ://www.springerlink.com/, août 2006. Springer-Verlag.

[BLSL06b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Jeux booléens statiques et représentation compacte de préférences. Rapport de recherche IRIT/RR–2006-13–FR, IRIT, Université Paul Sabatier, Toulouse, juillet 2006.

[BLSL07a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Compact preference representation in Boolean games. KNAW Academy Colloquium, Amsterdam, 05/02/07-07/02/07, 2007.

[BLSL07b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Dependencies between players in Boolean games. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Hammamet, Tunisie, 31/10/2007-02/11/2007*, pages 743–754, http ://www.springerlink.com, novembre 2007. Springer.

[BLSL07c] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. Rapport de recherche IRIT/RR–2007-13–FR, IRIT, Université Paul Sabatier, Toulouse, juin 2007.

[BLSL07d] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. Workshop on Logics and Collective Decision making (LCD'07), Lille, 13/03/2007-14/03/2007, 2007.

[BLSL08a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Coalitions efficaces dans les jeux booléens. In *Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA), Amiens, 23/01/08-25/01/08*, page (support électronique). Laboratoire Modélisation, Information et Systèmes (MIS) - Université de Picardie Jules Vernes, janvier 2008.

[BLSL08b] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Efficient coalitions in Boolean games. In Krzysztof Apt and Robert van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 283–297. Amsterdam University Press, 2008.

[BLSLZ06] Elise Bonzon, Marie-Christine Lagasquie-Schiex, Jérôme Lang, and Bruno Zanuttini. Boolean games revisited. In *European Conference on Artificial Intelligence (ECAI), Riva del Garda, Italy, 28/08/2006-01/09/2006*, pages 265–269, http ://www.springerlink.com/, 2006. Springer-Verlag.

[BLSLZ09] Elise Bonzon, Marie-Christine Lagasquie-Schiex, Jérôme Lang, and Bruno Zanuttini. Compact preference representation and boolean games. *Journal of Autonomous Agents and Multi-Agent Systems*, 18(1) :1–35, feb 2009.

[Boc03] A. Bochman. Collective argumentation and disjunctive programming. *Journal of Logic and Computation*, 13 (3) :405–428, 2003.

[Bon07] Elise Bonzon. *Modélisation des interactions entre agents rationnels : les jeux booléens.* Thèse de doctorat, Université Paul Sabatier, Toulouse, France, novembre 2007.

[Bou94] Craig Boutilier. Towards a logic for qualitative decision theory. In *Proc. of the $4^{th}$ KR*, pages 75–86, Bonn, Germany, 1994.

[Bre89] Gerhard Brewka. Preferred subtheories : An extended logical framework for default reasoning. In N.S. Sridharan, editor, *Proc. of the $11^{th}$ IJCAI*, pages 1043–1048, Detroit, MI, 1989. Morgan-Kaufmann.

[Bre94] Gerhard Brewka. Reasoning about priorities in default logic. In *Proc. of AAAI-94*, pages 940–945, Seattle, WA, 1994.

[CA05] Martin Caminada and Leila Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the 20th National Conference on Artificial Intelligence(AAAI 2005)*, pages 608–613. AAAI Press, 2005.

[CA07] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5–6) :286–310, 2007.

[Cam06] Martin Caminada. Semi-stable semantics. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *Proc. of the $1^{st}$ COMMA (Computational models of arguments)*, pages 121–128. IOS Press, 2006.

[Cas96] Thierry Castell. Computation of prime implicates and prime implicants by a variant of the Davis and Putnam procedure. In *Proceedings of the Eighth International Conference on Tools with Artificial Intelligence*, pages 428–429. IEEE, 1996.

[Cas97] Thierry Castell. *Consistance et déduction en logique propositionnelle*. Thèse, Université Paul Sabatier, Toulouse, France, 1997.

[Cay95] Claudette Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In Chris S. Mellish, editor, *Proc. of the $14^{th}$ IJCAI*, pages 1443–1448, Montréal, Canada, 1995. Morgan Kaufmann.

[CCCJ05] Guillaume Cabanac, Max Chevalier, Claude Chrisment, and Christine Julien. A social validation of collaborative annotations on digital documents. In *Proc. of IWAC*, pages 31–40, 2005.

[CCCJ07] Guillaume Cabanac, Max Chevalier, Claude Chrisment, and Christine Julien. Collective annotation : Perspectives for information retrieval improvement. In *Proc. of RIAO*, 2007.

[CDLS06a] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Handling controversial arguments in bipolar argumentation systems. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *Proc. of Computational Models of Argument (COMMA)*, pages 261–272. IOS Press, 2006.

[CDLS06b] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Prise en compte des arguments controversés dans des systèmes d'argumentation bipolaires. Technical Report IRIT/RR–2006-01–FR, CRIL-2006003, IRIT, CRIL, 2006.

[CDLS07] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Handling Ignorance in Argumentation : Semantics of Partial Argumentation Frameworks. In Khaled Mellouli, editor, *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Hammamet, Tunisie, 31/10/07-02/11/07*, number 4724 in LNAI, pages 259–270, http ://www.springerlink.com, 2007. Springer.

[CDLSM02a] Claudette Cayrol, Sylvie Doutre, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Défense minimale : un raffinement de la sémantique préférée pour les systèmes d'argumentation. Rapport de recherche 2002-03-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, January 2002.

[CDLSM02b] Claudette Cayrol, Sylvie Doutre, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Minimal defence : a refinement of the preferred semantics for argumentation frameworks. In Benferhat and Giunchiglia, editors, *Proc. of the $9^{th}$ International Workshop on Nonmonotonic Reasoning (session on Argument, Dialogue and Decision)*, pages 408–415, Toulouse, France, 2002.

[CDM03] C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Journal of logic and computation*, 13 :377–403, 2003.

[CdSCLS08] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Revision of an argumentation system. In Gerhard Brewka and Jerome Lang, editors, *Proc. of KR 2008*, pages 124–134. AAAI Press, 2008.

[CLS93] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Comparaison de relations d'inférence non-monotone : étude de complexité. Rapport de recherche 93-23R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, September 1993.

[CLS94a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Classification de relations d'inférence non-monotone : la prudence et les propriétés de déduction. Rapport de recherche 94-49R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, November 1994.

[CLS94b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the complexity of non-monotonic entailment in syntax-based approaches. In *Proc. of the $11^{th}$ ECAI workshop on Algorithms, Complexity and Commonsense Reasoning*, Amsterdam, Nederland, 1994.

[CLS95] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Non-monotonic syntax-based entailment : A classification of consequence relations. In C. Froidevaux and J. Kohlas, editors, *Lecture Notes in Artificial Intelligence 946 (Proc. of ECSQARU-95)*, pages 107–114, Fribourg, Switzerland, 1995. Springer Verlag.

[CLS97] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Quelques réflexions sur des relations d'inférence non-monotone graduelles. Rapport de recherche 97-28R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, June 1997.

[CLS00] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Une approche décisionnelle du traitement de l'incohérence inspirée par le raisonnement non monotone. In *Proc. of the $12^{th}$ RFIA*, pages 277–285 (vol. 3), Paris, France, 2000.

[CLS01] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Évaluation graduelle d'arguments. Rapport de recherche 2001-18-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, November 2001.

[CLS02a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In *Proc. of the $9^{th}$ IPMU (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, pages 83–90, Annecy, France, 2002.

[CLS02b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Vers une acceptabilité graduelle des arguments dans les sytèmes d'argumentation. Rapport de recherche 2002-42-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, December 2002.

[CLS03a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Critique et amélioration de l'évaluation graduelle par tuples pour le traitement des circuits. Rapport de recherche 2003-13-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, September 2003.

[CLS03b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual acceptability in argumentation systems. In *Proc. of the $3^{rd}$ CMNA (International workshop on computational models of natural argument)*, pages 55–58, Acapulco, Mexique, 2003.

[CLS03c] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In B. Bouchon-Meunier, L.Foulloy, and R.R. Yager, editors, *Intelligent Systems for Information Processing : From representation to Applications*, chapter Reasoning, pages 179–190. Elsevier, 2003.

[CLS03d] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Vers une acceptabilité graduelle des arguments dans les systèmes d'argumentation. In *Proc. of the $2^{nd}$ MFI (Journées francophones des Modèles formels de l'interaction)*, pages 41–50, Lille, France, 2003.

[CLS04] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolarité en argumentation. Rapport de recherche 2004-07-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, February 2004.

[CLS05a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In Luis Godo, editor, *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 366–377, Barcelone, Spain, 2005. Springer-Verlag.

[CLS05b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23 :245–297, 2005.

[CLS05c] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In Luis Godo, editor, *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 378–389, Barcelone, Spain, 2005. Springer-Verlag.

[CLS07] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments in bipolar argumentation frameworks. In *Proc. of the $7^{th}$ CMNA (International workshop on Computational Models of Natural argument)*, pages 14–20, Hyderabad, India, 2007.

[CLS09] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolar abstract argumentation systems. In Iyad Rahwan and Guillermo Simari, editors, *Argumentation in Artificial Intelligence*. Springer, 2009. À paraître.

[CLSM00] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. Une approche méthodologique de la résolution de conflits. In *Proc. of the $12^{th}$ RFIA*, pages 215–224 (vol. 1), Paris, France, 2000.

[CLSS96a] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Non-monotonic reasoning : from complexity to algorithms. In *Proc. of the $4^{th}$ International Symposium on Artificial Intelligence and Mathematics*, pages 25–28, Fort Lauderdale, Florida, 1996.

[CLSS96b] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Non-monotonic reasoning : from complexity to algorithms. Rapport de recherche 96-07R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, March 1996.

[CLSS98] Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Thomas Schiex. Nonmonotonic reasoning : from complexity to algorithms. *Annals of Mathematics and Artificial Intelligence*, 22(3-4) :207–236, 1998.

[CMDK$^+$05] Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, and Pierre Marquis. Argumentation and fusion. In *Proc. of the $20^{th}$ AAAI*, Pittsburg, Pennsylvania, 2005.

[CMDK$^+$07] Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, and Pierre Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence, Argumentation in Artificial Intelligence*, 171(10-15) :730–753, octobre 2007.

[CMDM05] S. Coste-Marquis, C. Devred, and P. Marquis. Inference from contreversial arguments. In *Proceedings of the 12th International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2005)*, pages 606–620, 2–6 december 2005.

[CMDM06] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR' 06)*, pages 112–122, 2006.

[CRS93] Claudette Cayrol, Véronique Royer, and Claire Saurel. Management of preferences in assumption-based reasoning. In R. Yager and B. Bouchon, editors, *Advanced methods in AI. Lecture notes in computer science 682*, pages 13–22. Springer Verlag, 1993. Extended version in Technical Report IRIT-CERT, 92-13R (University Paul Sabatier Toulouse).

[DBC01] Paul E. Dunne and Trevor J.M. Bench-Capon. Coherence in finite argument systems. Technical report 01-006, University of Liverpool, Department of Computer Science (U.L.C.S.), 2001.

[DBC02] Paul E. Dunne and Trevor J.M. Bench-Capon. Coherence in finite argument system. *Artificial Intelligence*, 141(1-2) :187–203, 2002.

[DF05] Didier Dubois and Hélène Fargier. On the qualitative comparison of sets of positive and negative affects. In *Proc. of the eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU) – LNAI 3571*, pages 305–316, Barcelone, Spain, 2005.

[DJ04] V.D. Dang and N.R. Jennings. Generating coalition structures with finite bound from the optimal guarantees. In *Proc. of AAMAS*, pages 564–571, 2004.

[DLP91] Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency in possibilistic knowledge bases - to live or not to live with it. In L.A. Zadeh and J. Kacprzyk, editors, *Fuzzy logic for the Management of Uncertainty*, pages 335–351. Wiley and sons, 1991.

[DM01] S. Doutre and J. Mengin. Preferred Extensions of Argumentation Frameworks : Computation and Query Answering. In A. Leitsch R. Goré and T. Nipkow, editors, *IJCAR 2001*, volume 2083 of *LNAI*, pages 272–288. Springer-Verlag, 2001.

[Dom02] Carmel Domshlak. *Modeling And Reasoning About Preferences With CP-nets*. PhD thesis, Ben-Gurion University of the Negev, Negev, Israel, 2002.

[Dou02] Sylvie Doutre. *Autour de la sémantique préférée des systèmes d'argumentation*. Thèse, Université Paul Sabatier, IRIT, 2002.

[DP06] Didier Dubois and Henri Prade. A bipolar possibilitic representation of knowledge and preferences and its applications. In *Fuzzy Logic and Applications, 6th International Workshop, WILF 2005, Crema, Italy, September 15-17, 2005, Revised Selected Papers*, volume 3849 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2006.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77 :321–357, 1995.

[EGFK93a] M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as logical uncertainty. In *Proc. of ECSQARU'93 – LNCS 747*, pages 85–90, 1993.

[EGFK93b] Morten Elvang-Goransson, John Fox, and Paul Krause. Dialectic reasoning with inconsistent information. In David Heckerman and Abe Mamdani, editors, *Proc. of the $9^{th}$ UAI*, pages 114–121, Washington, DC, 1993. Morgan-Kaufmann.

[EGH95] M. Elvang-Gøransson and A. Hunter. Argumentative logics : Reasoning with classically inconsistent information. *Data and Knowledge Engineering*, 16 :125–145, 1995.

[EGW08] U. Egly, S. A. Gaggl, and S. Woltran. Aspartix : Implementing argumentation frameworks using answer-set programming. In *In Proceedings of ICLP*, 2008.

[FGS04] M. Falappa, A. García, and G. Simari. Belief dynamics and defeasible argumentation in rational agents. In *Proc. of NMR*, pages 164–170, 2004.

[Gab85] D. M. Gabbay. Theoretical foundations for nonmonotonic reasoning in expert systems. *Logics and models of Concurrent Systems*, 1985.

[Gar98] Laurent Garcia. *Raisonnement local dans les bases de connaissances ordonnées : Applications au traitement des incohérences et au raisonnement plausible*. Thèse, Université Paul Sabatier, IRIT, 1998.

[Gef92] H. Geffner. *Default reasoning : Causal and Conditional Theories*. MIT Press, 1992.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-completeness*. W.H. Freeman and Company, New York, 1979.

[GM94] Peter Gärdenfors and David Makinson. Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65 :197–245, 1994.

[GNT04] M. Ghallab, D. Nau, and P. Traverso. *Automated planning, theory and practice*. Elsevier, Morgan Kaufmann, 2004.

[Gor95] T. F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[Got92] Georg Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3) :397–425, 1992.

[Har04a] G. Harman. Practical aspects of theoretical reasoning. *The Oxford Handbook of Rationality*, pages 45–56, 2004.

[Har04b] Paul Harrenstein. *Logic in Conflict*. PhD thesis, Utrecht University, 2004.

[HO07] U. Hahn and M. Oaksford. The rationality of informal argumentation : A bayesian approach. *Psychological Review*, 114 :704–732, 2007.

[Hun04] A. Hunter. Making argumentation more believable. In *Proc. of AAAI*, pages 269–274, 2004.

[HvdHMW01] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. Boolean Games. In J. van Benthem, editor, *Proceedings of the 8th International Conference on Theoretical Aspects of Rationality and Knowledge (TARK'01)*, volume Theoretical Aspects of Rationality and Knowledge, pages 287–298. San Francisco, Morgan Kaufmann, 2001.

[HvdT03] Joris Hulstijn and Leendert van der Torre. Combining goal generation and planning in an argumentation framework. In Tom Heskes, Peter Lucas, Louis Vuurpijl, and Wim Wiegerinck, editors, *Proceedings of the 15th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2003)*, pages 155–162, Katholieke Universiteit Nijmegen, October 2003.

[HvdT04] Joris Hulstijn and Leendert van der Torre. Combining goal generation and planning in an argumentation framework. In Anthony Hunter and Jerome Lang, editors, *Proceedings of the Workshop on Argument, Dialogue and Decision at the International Workshop on Non-monotonic Reasoning (NMR 2004)*, Whistler, Canada, June 2004.

[JK90] Ulrich Junker and Kurt Konolige. Computing the Extensions of Autoepistemic and Default Logics with a Truth Maintenance System. In *Proc. of the $8^{th}$ AAAI*, pages 278–283, 1990.

[JV99a] Hadassa Jakobovits and Dick Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law (ICAIL'99)*, pages 53–62, 1999.

[JV99b] Hadassa Jakobovits and Dick Vermeir. Robust semantics for argumentation frameworks. *Journal of logic and computation*, 9(2) :215–261, 1999.

[KAEF95] Paul Krause, Simon Ambler, Morten Elvang, and John Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1) :113–131, 1995.

[Kle86] Johan De Kleer. An assumption-based TMS. *Artificial Intelligence*, 28 :127–162, 1986.

[KLM90] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44 :167–207, 1990.

[KM03] A.C. Kakas and P. Moraïtis. Argumentation based decision making for autonomous agents. In *Proc. of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 883–890. ACM Press, 2003.

[KP01] Nikos Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making : the HERMES system. *Information systems*, 26(4) :259–277, 2001.

[KP02] S. Konieczny and R. Pino Pérez. Merging information under constraints : a qualitative framework. *Journal of Logic and Computation*, 12(5) :773–808, 2002.

[Laf03] Pascal Lafourcade. Application de la résolution de conflits logiques, aux conflits des problèmes d'ordonnancement. Rapport de DEA. DEA RCFR, Université Paul Sabatier Toulouse, 2003.

[Lan94] Jérôme Lang. Syntax-based default reasoning as probabilistic model-based diagnosis. In *Proc. of the $10^{th}$ UAI*, pages 391–398, Seattle, Washington, 1994.

[Leh92] Daniel Lehmann. Another perspective on default reasoning. Rapport de recherche 92-12, Leibniz Center for Research in Computer Science. Hebrew University of Jerusalem, Israel, July 1992.

[Leh95] Daniel Lehmann. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence*, 15 :61–82, 1995.

[Lev92] François Levy. Reason maintenance systems and default theories. Pré-publication 92-2, LIPN, Université Paris Nord, 1992.

[LLM03] Jérôme Lang, Paolo Liberatore, and Pierre Marquis. Propositional Independence - Formula-Variable Independence and Forgetting. *Journal of Artificial Intelligence Research*, 18 :391–443, 2003.

[LM92] Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail ? *Artificial Intelligence*, 55 :1–60, 1992.

[LM98] Jérôme Lang and Pierre Marquis. Complexity Results for Independence and Definability in Propositional Logic. In *International Conference on Knowledge Representation and Reasoning,(KR'98)*, pages 356–367, 1998.

[LR57] R. Duncan Luce and Howard Raiffa. *Games and Decisions : Introduction and Critical Survey*. Wiley & Sons, 1957.

[LS92] Marie-Christine Lagasquie-Schiex. Étude et évaluation de MVL (MultiValued Logics) théorie et outil de M.L.Ginsberg. Rapport de recherche 92-39R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, October 1992.

[LS94] Marie-Christine Lagasquie-Schiex. La complexité des relations d'inférence non-monotone dans les approches syntaxiques. In *Proc. of the $2^{nd}$ RJCIA*, pages 65–72, Marseille, France, 1994.

[LS95a] Marie-Christine Lagasquie-Schiex. *Contribution à l'étude des relations d'inférence non-monotone combinant inférence classique et préférences*. Thèse, Université Paul Sabatier, IRIT, 1995.

[LS95b] Marie-Christine Lagasquie-Schiex. Implémentation de quelques algorithmes en raisonnement non-monotone (annexe de la thèse numéro 2217). Rapport de recherche 95-22-R, IRIT, France, December 1995.

[LS95c] Marie-Christine Lagasquie-Schiex. Une classification des relations d'inférence non-monotone dans les approches syntaxiques. In *Proc. des $5^{emes}$ journées nationales du PRC-GDR Intelligence Artificielle*, page 373, Nancy, France, 1995. Présentation effectuée en tant que chercheur invité (meilleur papier jeune chercheur en Intelligence Artificielle).

[LVW02] Jérôme Lang, Leendert Van der Torre, and Emil Weydert. Utilitarian desires. *Journal of Autonomous Agents and Multi-Agents Systems*, 5(3) :329–363, 2002.

[Mac79] J. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical logic*, 8 :117–133, 1979.

[MCLS05] Matthieu Mardi, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. Bipolarité en argumentation : acceptabilité et algorithmes. Rapport de recherche 2005-20-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, October 2005.

[Men95] Jérôme Mengin. A theorem prover for default logic based on an extended resolution principle and prioritized conflict resolution. In *Lecture Notes in Artificial Intelligence 946 (Proc. of ECSQARU-95)*, 1995.

[Nas50] John Nash. Equilibrium points in $n$-person games. *Procedings of the National Academy of Sciences*, 36 :48–49, 1950.

[NBR02] R. Neves, J. F. Bonnefon, and E. Raufaste. An empirical test of patterns for nonnomotonic reasoning. *Annals of Mathematics and Artificial Intelligence*, 2002.

[Neb91] Bernhard Nebel. Belief revision and default reasoning : Syntax-based approaches. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. of the $2^{nd}$ KR*, pages 417–428, Cambridge, MA, 1991. Morgan-Kaufmann.

[Nie95] Ilkka Niemelä. Towards Efficient Default Reasoning. In *Proc. of the $14^{th}$ IJCAI*, pages 312–318, 1995.

[Nie02] J.M. Nielsen. On the number of maximal independent sets in a graph. Technical Report RS 02-15, Center for Basic Research in Computer Science (BRICS), April 2002.

[NOC08] J. C. Nieves, M. Osorio, and U. Cortés. Preferred extensions as stable models. *Theory and Practice of Logic Programming*, 8(4), 2008.

[NP06] S.H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation. In *Proc. of the $3^{rd}$ WS on Argumentation in multi-agent systems*, 2006.

[OCH08] M. Oaksford, N. Chater, and U. Hahn. Human reasoning and argumentation : The probabilistic approach. In J. E. Adler and L. J. Rips, editors, *Reasoning : Studies of Human Inference and its Foundations*. Cambridge University Pres, 2008.

[ON08] Nir Oren and Timothy J. Norman. Semantics for evidence-based argumentation. In *Proc. of COMMA*, pages 276–284, 2008.

[Pap92] O. Papini. A complete revision function in propositional calculus. In B. Neumann, editor, *Proceeding of the European Conference on Artificial Intelligence (ECAI)*, pages 339–343. John Wiley and sons, Ltd., 1992.

[Par97] Simon Parsons. Normative argumentation and qualitative probability. In *Proc. of the first International Joint Conference on Qualitative and quantitative practical reasoning, ECSQARU-FAPR, LNAI 1244*, pages 466–480, Germany, 1997.

[PC05] F. Paglieri and C. Castelfranchi. Revising beliefs through arguments : Bridging the gap between argumentation and belief revision in mas. In *Argumentation in Multi-Agent Systems*, pages 78–94. Springer, 2005.

[PG00] J. Pollock and A. Gillies. Belief revision and epistemology. *Synthese*, 122(1-2) :69–92, 2000.

[PJ96] S. Parsons and N. R. Jennings. Negotiation through argumentation—a preliminary report. In *Proceedings of the 2nd International Conference on Multi Agent Systems*, pages 267–274, 1996.

[PL92] Gadi Pinkas and Ronald P. Loui. Reasoning from inconsistency : A taxonomy of principles for resolving conflict. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. of the $3^{rd}$ KR*, pages 709–719, Cambridge, MA, 1992. Morgan-Kaufmann.

[Poo88] David Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36 :27–47, 1988.

[PS97] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7 :25–75, 1997.

[RA06] Iyad Rahwan and Leila Amgoud. An Argumentation-based Approach for Practical Reasoning . In Gerhard Weiss and Peter Stone, editors, *5th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2006, Hakodate, Japan*, pages 347–354, New York, USA, 2006. ACM Press.

[Raz78] J. Raz. Practical reasoning. *Oxford, Oxford University Press*, 1978.

[Rei87] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32 :57–95, 1987.

[Rip98] L. J. Rips. Reasoning and conversation. *Psychological Review*, 105 :411–441, 1998.

[RMF+08] N. D. Rotstein, M. O. Moguillansky, M. A. Falappa, A. J. García, and G. R. Simari. Argument theory change : revision upon warrant. In P. Besnard, S. Doutre, and A. Hunter, editors, *Proc. of the $2^{nd}$ International Conference of Computational Models of Argument (COMMA)*, pages 336–347. IOS Press, 2008.

[RN95] Stuart Russell and Peter Norvig. *Artificial intelligence. A modern approach*. Prentice Hall, 1995. ISBN : 0-13-103805-2.

[RPRS08] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation : a game-theoretical investigation. In *Proc. of COMMA*, pages 324–335, 2008.

[SGC04] Guillermo R. Simari, Alejandro J. Garcia, and Marcela Capobianco. Actions, planning and defeasible reasoning. In *Proceedings of the 10th International Workshop on Non-Monotonic Reasoning, Whistler BC, Canada*, pages 377–384, 2004.

[Sir04] Denis Sireyjol. Prallèles entre théorie de l'argumentation et théorie des jeux. Rapport de DEA. DEA RCFR, Université Paul Sabatier Toulouse, 2004.

[SL92] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53 :125–157, 1992.

[Sme93] Philippe Smets. Probability of deductibility and belief functions. Rapport de recherche 93-5.2, IRIDIA Université libre de Bruxelles, Bruxelles, 1993.

[Som94] Léa Sombé. A glance at revision and updating in knowledge bases. *International journal of intelligent systems*, 9(1) :1–27, 1994.

[Tha03] P. Thagard. Why wasn't o. j. convicted : Emotional coherence in legal inference. *Cognition and Emotion*, 17 :361–383, 2003.

[Tho00] Richmond H. Thomason. Desires and defaults : A framework for planning with inferred goals. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR 2000 : Principles of Knowledge Representation and Reasoning, Proceedings of the Seventh International Conference, Breckenridge, Colorado, USA*, pages 702–713. Morgan Kaufmann Publishers, 2000.

[TKI08] M. Thimm and G. Kern-Isberner. A distributed argumentation framework using defeasible logic programming. In *Proceedings of COMMA 2008*, pages 381–392, 2008.

[TP94] S. W. Tan and J. Pearl. Specification and evaluation of preferences under uncertainty. In *Proc. of the $4^{th}$ KR*, pages 530–539, Bonn, Germany, 1994.

[TP05] Yuqing Tang and Simon Parsons. Argumentation-based dialogues for deliberation. In Frank Dignum, Virginia Dignum, Sven Koenig, Sarit Kraus, Munindar P. Singh, and Michael Wooldridge, editors, *Proceedings of the 4th international joint conference on autonomous agents and multi agent systems (AAMAS 2005), Universiteit Utrecht, The Netherlands*, pages 552–559, New York NY, USA, 2005. ACM Press.

[Ver03] Bart Verheij. Deflog : on the logical interpretation of prima facie justified assumptions. *Journal of Logic in Computation*, 13 :319–346, 2003.

[Vid99] Laurent Vidal. Résolution de conflits guidée par les préférences. Rapport de stage, DESS I.R.R. (Intelligence artificielle, Reconnaissance des formes et Robotique), France, June 1999.

[vNM44] John von Newmann and Oskar Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.

[Vre97] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90 :225–279, 1997.

[Wal96] D. Walton. *Argument schemes for presumptive reasoning*, volume 29. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.

[Was99] R. Wassermann. Full acceptance through argumentation - a preliminary report. In *Proc. of IJCAI Workshop Practical Reasoning and Rationality*, 1999.

[Win88] M. Winslett. Reasoning about action using a possible models approach. In *Proc. of AAAI*, pages 89–93, 1988.

[WK95] D. Walton and E. Krabbe. *Commitment in dialogue. Basic concepts of interpersonal reasoning*. State University of New York Press, 1995.

[Woo00] Michael J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, 2000.

[WPJ00] E. Würbel, O. Papini, and R. Jeansoulin. Revision : an application in the framework of GIS. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 505–516, Breckenridge, Colorado, USA, April 2000.

[WvdH97] C. Witteveen and W. van der Hoek. A general framework for revising nonmonotonic theories. In J. Dix, U. Furbach, and A. Nerode, editors, *Proc. of the $4^{th}$ International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR) – LNAI*, volume 1265, pages 258–272. Springer, 1997.

[Xuo92] N.H. Xuong. *Mathématiques discrètes et informatique*. Masson, 1992.

# Annexe A

# Annexes

## A.1 Répartition des encadrements

### A.1.1 Encadrements stagiaires Master 2

| Nom stagiaire | % encadrement | période | sur le thème |
|---|---|---|---|
| L. VIDAL | 50 | 1999 | 2 “Résolution de conflits” |
| L. DUSSEUX | 50 | 2002 | 2 “Résolution de conflits” |
| P. LAFOURCADE | 33 | 2003 | 2 “Résolution de conflits” |
| D. GAY | 50 | 2004 | 3 “Argumentation” |
| D. SIREYJOL | 50 | 2004 | 3-4 “Argumentation et Jeux” |
| M. MARDI | 50 | 2005 | 3 “Argumentation” |
| A. MIQUEL | 33 | 2007 | 3 “Argumentation” |

### A.1.2 Encadrements de doctorants

| Nom doctorant | % encadrement | période | sur le thème |
|---|---|---|---|
| S. DOUTRE<br>(maître de conférences à Toulouse I depuis 2005) | 33 | 1999-2002 | 3 “Argumentation” |
| Titre de la thèse : “Autour de la sémantique préférée des systèmes d’argumentation” | | | |
| E. BONZON<br>(maître de conférences à Paris Descartes depuis 2008) | 50 | 2004-2007 | 4 “Jeux” |
| Titre de la thèse : “Modélisation des interactions entre agents rationnels : les jeux booléens” | | | |

## A.2 Répartition des publications

### A.2.1 Par type de publication

| Type publication | Nombre | année de publication |
|---|---|---|
| revue internationale avec comité de relecture | 5 | 1998 à 2009 |
| contribution à des ouvrages de synthèse | 3 (+ 1 à paraître) | 2002 à 2009 |
| conférence internationale avec comité de relecture et actes publiés | 13 | 1996 à 2008 |
| conférence ou workshop internationaux sans actes publiés (mais avec comité de relecture) | 4 | 2002 à 2007 |
| conférence nationale avec comité de relecture et actes publiés | 6 | 1994 à 2008 |
| autres conférences | 4 | 1994 à 2008 |
| rapport de recherche | 16 | 1992 à 2008 |
| thèse de doctorat | 1 | 1995 |
| TOTAL | 52 (+ 1 à paraître) | 1992 à 2009 |

### A.2.2 Par thème de recherche

| Thème de recherche | Nombre | année de publication |
|---|---|---|
| Inférence non monotone (thème 1) | 14 | 1992 à 2000 |
| Résolution de conflits (thème 2) | 1 | 2000 à 2003 |
| Argumentation (thème 3) | 26 (+ 1 à paraître) | 2001 à 2009 |
| Jeux (thème 4) | 11 | 2004 à 2008 |
| TOTAL | 52 (+ 1 à paraître) | 1992 à 2009 |

En cours de soumission :
- un papier à IJCAI (conférence internationale)
- un papier à IJIS (revue internationale)
- un papier à JOLC (revue internationale)

## A.3 Participation à des projets et collaborations

Participation à l'Action Spécifique AS119 "Les représentations bipolaires en raisonnement et décision" dans le cadre du Réseau Thématique Pluridisciplinaire (RTP 11) "Information et Intelligence : Raisonner et Décider" 2002-2004.

Partenaire du projet européen ASPIC "Argumentation Service Platform with Integrated Components" du 6ème Programme Cadre de Recherche et Développement de la commission européenne, de 2003 à 2007.

En cours de soumission :
- un projet ANR blanc en tant que coordinatrice (projet AxSA), qui correspond à mon projet de recherche (voir chapitre 5 page 153).

Collaborations en dehors des projets :
- en local (dans les équipes RPDMP et LILAC à l'IRIT) : Claudette CAYROL, Leila AMGOUD, Sylvie DOUTRE, Florence DUPIN DE SAINT-CYR, Jérôme LANG, Jérôme MENGIN ;
- au niveau national : Sylvie COSTE-MARQUIS (Laboratoire du CRIL à Lens), Sébastien KONIECZNY (Laboratoire du CRIL à Lens), Pierre MARQUIS (Laboratoire du CRIL à Lens), Caroline DEVRED (Laboratoire du LERIA à Angers), Elise BONZON (Laboratoire CRIP5 à Paris Descartes), Bruno ZANUTTINI (Laboratoire du GREYC à Caens).