Raisonnement & Incohérence

Sébastien Konieczny

RAISONNEMENT & INCOHÉRENCE

Sébastien Konieczny

Habilitation à diriger des recherches

Composition du jury

Gerhard Brewka (rapporteur)
Professeur à l'Université de Leipzig
Marie-Christine Rousset (rapporteur)
Professeur à l'Université de Grenoble
Torsten Schaub (rapporteur)
Professeur à l'Université de Potsdam
Salem Benferhat
Professeur à l'Université d'Artois
Andreas Herzig
Directeur de Recherche au CNRS - IRIT
David Makinson
Professeur à la London School of Economics
Pierre Marquis
Professeur à l'Université d'Artois

Ce qui est créé par l’esprit est plus vivant que la matière.

(Charles Baudelaire)

REMERCIEMENTS

Je voudrais tout d'abord remercier très vivement Gerhard Brewka, Marie-Christine Rousset et Torsten Schaub pour l'intérêt qu'ils ont porté à mon travail en acceptant d'en être rapporteurs, et pour leurs remarques et les discussions que nous avons lors des différentes conférences où nous nous croisons.

Je remercie David Makinson de l'honneur qu'il me fait en acceptant de participer à ce jury. Ses articles sur les relations d'inférence non monotones et sur la révision de croyances, que j'ai lus pendant mon DEA, m'ont beaucoup impressionné (ils m'impressionnent toujours autant), et sont pour beaucoup dans le choix de ce sujet de recherche.

Mon passage à l'IRIT (Toulouse) de 2001 à 2004 m'a été extrêmement bénéfique, en particulier les discussions avec Leila Amgoud, Salem Benferhat, Philippe Besnard, Didier Dubois, Hélène Fargier, Andreas Herzig, Jérôme Lang et Henri Prade. Je ne pouvais les mettre tous dans ce jury, j'ai donc demandé à Andreas Herzig de les représenter. Cela me fait très plaisir qu'il ait accepté, non seulement parce que les nombreuses discussions que nous avons eues m'ont beaucoup apporté, mais également parce que le travail que nous avons effectué sur les opérateurs de confluence a été initié par une question qu'Andreas m'a posée lors d'un séminaire à Dagsthul.

Je suis très heureux que Salem Benferhat ait accepté de faire partie de ce jury. J'apprécie beaucoup les conversations, scientifiques ou autres, que nous avons régulièrement.

Un grand merci enfin à Pierre Marquis, d'abord pour avoir accepté de faire partie de ce jury, mais surtout pour notre (longue à présent) collaboration, et pour ces nombreuses réunions où nous ne sommes jamais d'accord !

Je remercie également l'ensemble de mes co-auteurs, sans qui une majorité de ces travaux n'auraient pas vu le jour. Merci en particulier à Patricia Everaere, avec qui travailler est toujours un plaisir.

Je voudrais à présent remercier celui qui a été, et qui reste, mon maître, Ramón Pino Pérez. Ramón m'a guidé durant mon DEA et ma thèse. Il m'a beaucoup appris. Et je continue à apprendre à chacune de nos rencontres.

Je souhaite remercier l'ensemble des collègues du laboratoire, et surtout ceux que je croise quotidiennement à la faculté, pour la bonne ambiance de travail. Merci en particulier à Bertrand, Gilles, Jean-Luc, Pierre, Salem, Stéphane, Sylvain et Sylvie pour les pauses divertissantes. Et bravo à Sylvain pour tenir aussi longtemps dans le bureau !

Je voudrais remercier mes amis, et plus particulièrement Bruno, Sonia et Claude, ainsi que toute ma famille pour leur soutien constant, malgré le fait que dans la vie de tous les jours aussi je suis plus doué pour la logique que pour la psychologie...

Merci à Isabelle pour tout, et pour le reste.

Ce document a bénéficié des remarques et critiques judicieuses de Philippe Besnard, Pierre Marquis, Ramón Pino Pérez, Patricia Everaere, Sylvain Lagrue, Isabelle Dieval et Emmanuel Genot, ainsi que de l'aide technique (et TeXnique) de Bertrand Mazure et Bruno Beaufils.

TABLE DES MATIÈRES

Première partie

Synthèse des travaux de recherche

Chapitre I

INTRODUCTION

Le tout est plus grand que la somme des parties

(Aristote)

Ce document est un rapport d'habilitation à diriger des recherches. Il présente une grande partie de mes travaux de recherche, qui ont comme thème commun la résolution de conflits logiques (le raisonnement en présence d'incohérences). Pour des raisons de cohérence je ne détaillerai pas les travaux que j'ai réalisé sur d'autres thèmes, tels que la théorie de la décision [53], la théorie des jeux [34, 57, 70] et la planification multi-agents [33, 69].

1.1. Co-auteurs

Je ne conçois pas la recherche comme une activité solitaire. Il me semble que les échanges lors de réunions scientifiques permettent de faire naître des idées et résultats que chacun des participants n'aurait pas eus seul. Bref, en recherche comme ailleurs, je suis convaincu que Le tout est plus grand que la somme des parties.

La plupart de mes travaux ont été réalisés en collaboration avec d'autres collègues. Dans la suite j'utiliserai un *nous* qui n'est donc pas une figure de style, mais qui illustre ce travail en commun. Je ne citerai pas explicitement les noms des co-auteurs correspondant à chaque travail, pour ne pas alourdir la présentation, mais ces noms sont facilement identifiables dans les articles correspondants. Ces co-auteurs sont [1] : Pierre Marquis (CRIL, Lens), Ramón Pino Pérez (Université des Andes, Mérida, Venezuela), Patricia Everaere (LIFL, Lille), Jérôme Lang (LAMSADE, Paris), Ramzi Ben Larbi (CRIL, Lens), Anthony Hunter (University College London, Royaume Uni), Didier Dubois (IRIT, Toulouse), Henri Prade (IRIT, Toulouse), Olivier Gauwin (LIFL, Lille),

1. Classés par nombre d'articles en commun.

Salem Benferhat (CRIL, Lens), Odile Papini (LSIS, Marseille), Stéphane Janot (LIFL, Lille), Hassan Bezzazi (LIFL, Lille), Sylvie Coste-Marquis (CRIL, Lens), Caroline Devred (LERIA, Angers), Marie-Christine Lagasquie-Schiex (IRIT, Toulouse), Andreas Herzig (IRIT, Toulouse), Philippe Besnard (IRIT, Toulouse), Laurent Perrussel (IRIT, Toulouse), Eric Grégoire (CRIL, Lens).

Dans la suite, nous indiquons les citations de nos articles avec des références numériques (par exemple [35]). Les citations d'articles d'autres auteurs sont alphanumériques (par exemple [Mak94]).

1.2. Présentation de la problématique

Nous nous intéressons au problème du **raisonnement en présence d'incohérence**. Cette incohérence peut avoir différentes causes, que nous détaillons ci-après, mais elle est toujours due à l'**incertitude** inhérente au fait que les agents raisonnent à partir de **croyances** qui sont par nature **incertaines** et donc potentiellement fausses. Nous nous intéressons aux approches **qualitatives**, qui ne nécessitent pas de disposer d'un ensemble très important d'informations. Nous nous focalisons en particulier sur les **approches logiques**, où les croyances des agents sont exprimées en **logique propositionnelle**.

1.2.1. Logique propositionnelle

Un langage de représentation doit satisfaire à un certain nombre de contraintes, dont certaines sont antinomiques. Parmi les nombreuses propriétés que l'on peut attendre d'un langage de représentation, on peut au moins mentionner :

- **la compréhensibilité** : il faut que le langage soit facilement compréhensible par un humain ;
- **l'expressivité** : il faut que le langage permette de représenter des informations complexes ;
- **la concision** : il faut que le langage permette de représenter de manière économique (en terme d'espace) ces informations complexes ;
- **l'efficacité** : il faut que le langage puisse être facilement manipulable pour les tâches de raisonnement qui l'utiliseront (complexité algorithmique).

Il existe de nombreux langages de représentation. La logique propositionnelle est un des langages présentant un bon compromis entre ces critères. C'est un langage qui est facilement compréhensible pour un humain, qui offre une expressivité correcte et une bonne concision. Il existe enfin des méthodes de calcul efficaces en pratique pour tester la cohérence ou pour calculer les inférences possibles à partir d'un ensemble de formules.

1.2.2. Bases de croyances

Une croyance est une information que l'agent croit actuellement, c'est-à-dire une information qu'il considère comme vraie, mais qui ne l'est pas forcément, et qui peut donc être invalidée si l'agent se rend compte qu'il a fait une erreur.

On appelle base de croyances un ensemble d'informations, qui est représenté, dans la plupart de nos travaux, par un ensemble de formules propositionnelles.

Il est courant en intelligence artificielle, et plus généralement en logiques des croyances, de distinguer entre **croyances** et **connaissances**. Nous venons de définir les croyances comme des informations que l'agent croit vraies, mais sans en être assuré. Les connaissances sont des informations que l'agent sait vraies, c'est-à-dire qu'il est assuré que ces informations sont vraies. Cela implique en particulier qu'elles ne pourront plus être remises en question par la suite. Il nous semble que cette hypothèse de disposer d'informations dont on est assuré qu'elles soient vraies est irréaliste. Il s'agit donc simplement d'une hypothèse simplificatrice utile pour représenter la plupart des problèmes d'intelligence artificielle. Mais dans les faits, l'agent dispose juste d'informations dont certaines sont beaucoup plus certaines (et pas juste « certaines ») que d'autres. Nous avons récemment proposé de représenter cela à l'aide de « niveaux de croyances »[40], qui permettent d'avoir des gradualités entre croyances de même niveau, mais également de représenter des niveaux de croyances différents (et donc de généraliser cette distinction croyances/connaissances).

1.2.3. Représentation de l'incertitude

Les croyances dont dispose l'agent sont donc des informations plus ou moins certaines. Il existe plusieurs manières de représenter cette incertitude. En fait, on dispose d'une palette allant de représentations purement quantitatives de l'incertitude à des représentations purement qualitatives. Les approches quantitatives permettent de disposer d'une information plus fine, mais la contrepartie est qu'elles nécessitent automatiquement plus d'informationsque les approches qualitatives.

Parmi les approches principales, on peut distinguer les :

- **Probabilités** : il est possible d'associer une probabilité à chaque information.
- **Possibilités** : il est possible d'associer un degré de plausibilité à chaque information. C'est une information plus fruste que les distributions de probabilité (une distribution de possibilité peut être considérée comme la borne inférieure d'un ensemble de distributions de probabilité).
- **Fonctions ordinales conditionnelles (ou possibilités qualitatives)** : là encore, on associe un degré de plausibilité à chaque information, mais on n'autorise que des opérations « qualitatives » sur ces degrés, comme le $\min$ et le $\max$ par exemple. Ce cadre permet tout de même d'exprimer des intensités dans les plausibilités.
- **Pré-ordres** : il n'est plus possible d'associer une information numérique (un degré) à chaque information, on ne dispose que de la donnée des plausibilités relatives des informations, modélisé par un pré-ordre qui indique si une information est plus plausible qu'une autre.

On peut également mentionner les fonctions de croyances (appelées aussi modèle des croyances transférables), qui sont une généralisation des probabilités : la distribution de poids ne s'effectue plus uniquement sur les éléments (événements élémentaires) mais sur les sous-ensembles d'éléments. Ces fonctions de croyances peuvent modéliser les différents cadres sus-cités. Néanmoins ceux-ci présentent suffisamment d'intérêt pour être étudiés séparément.

Même si l'approche bayésienne (probabiliste) est l'approche dominante en intelligence artificielle (en particulier aux Etats-Unis), pour des raisons assez naturelles, en particulier sa facilité de mise en oeuvre, elle ne constitue pas l'approche universelle car elle manque d'expressivité. Elle ne permet pas en particulier de représenter l'ignorance. Il est possible de représenter une incertitude (quantifiée), par exemple que la probabilité de A est de 50%, celle de B de 30% et celle de C de 20%. Mais il est impossible de représenter (avec une distribution de probabilité unique) l'ignorance entre A, B et C, c'est-à-dire que l'on sait que soit A, soit B, soit C va se réaliser mais que l'on ne connait pas les probabilités correspondantes. Les Bayésiens utilisent souvent dans ce cas l'hypothèse laplacienne, c'est-à-dire qu'ils supposent une équiprobabilité entre les différents événements, donc qu'ils associent à ce cas la distribution de probabilité : A avec 33%, B avec 33% et C avec 33%. Ce qui ne représente clairement pas la même information. Or, l'ignorance se représente très facilement avec les approches plus qualitatives. Celles-ci, étudiées principalement en Europe, permettent également de raisonner avec des informations plus pauvres, c'est-à-dire dans des cas où l'on dispose de certaines informations sur la plausibilité des informations, mais pas suffisamment pour définir une distribution de probabilité. De plus, elles sont plus proches des mécanismes utilisés par les humains, car il est rare qu'un humain raisonne avec une information numérique aussi précise qu'une distribution de probabilité.

Nous nous plaçons dans les cadres les plus qualitatifs, c'est-à-dire que dans la plupart de nos travaux l'incertitude sera modélisée par des pré-ordres.

1.3. Résolution de conflits logiques

La majeure partie de nos travaux peut se regrouper sous la problématique du raisonnement à partir d'un ensemble de formules logiques incohérent. Suivant les hypothèses que l'on peut faire sur cet ensemble de formules, on obtient des cadres différents, qui nécessitent des méthodes de raisonnement spécifiques :

- **Incohérence** : on ne dispose d'aucune information supplémentaire. Il est donc nécessaire d'obtenir des conclusions cohérentes (raisonnables) à partir d'un ensemble d'information incohérent.
- **Révision** : une des formules est plus importante que les autres. Il faut donc garder cette formule, tout en éliminant les incohérences.
- **Fusion** : les formules proviennent de sources différentes. Il faut alors définir une base cohérente à partir de ces informations, en prenant en compte la localisation de ces informations (avec des arguments majoritaires par exemple).
- **Négociation** : les formules proviennent de sources différentes, comme pour la fusion, mais les sources gardent la maîtrise des modifications de leurs formules. Il faut donc tenir compte de possibles interactions (coalitions, etc.).

Nous nous sommes principalement intéressés aux trois premiers points. La négociation est un sujet que nous avons abordé plus récemment. Les résultats obtenus sur ce sujet sont plus préliminaires, mais ce sujet constitue le point que nous comptons développer en priorité à l'avenir. Nous allons structurer notre présentation autour de ces quatre points, en insistant sur nos projets concernant le quatrième, qui constituera donc une partie des perspectives et de notre projet de recherche.

1.4. Plan

Ce document est constitué de deux parties. La première présente une synthèse de nos travaux de recherche. La seconde partie complète la première et regroupe une sélection de publications qui nous semble significative et représentative de notre démarche scientifique.

La partie principale de ce document est la synthèse de nos travaux de recherche. Nous présenterons de manière structurée une grande partie de nos travaux de recherche, mais nous tenterons également d'insister sur la démarche adoptée et sur les buts poursuivis. Afin de placer nos travaux dans leur contexte, nous présenterons un bref état de l'art pour chaque thème abordé. Il ne faut toutefois pas considérer ce document comme une monographie, car la présentation de l'état de l'art et des travaux concernant chaque thème est biaisée vers nos travaux. Nous tenterons tout de même à chaque fois de mentionner les travaux principaux, afin que ce document puisse servir de point d'entrée à ces différents domaines.

Dans le prochain chapitre, nous exposons notre démarche scientifique générale, c'est-à-dire la démarche que nous adoptons pour l'ensemble de nos travaux. Elle constitue la principale justification de l'approche axiomatique que nous privilégions dans la majeure partie de nos travaux.

Les chapitres suivants présentent nos travaux concernant les quatre thèmes présentés à la section précédente : raisonnement en présence d'incohérence, révision de croyances, fusion de croyances, et négociation.

Le chapitre 3 présente le problème du raisonnement en présence d'incohérence. Le point de départ est un ensemble de formules logiques qui est incohérent. Le problème principal est alors celui de l'inférence à partir de cet ensemble incohérent. Une autre question importante est également celle de la mesure de l'incohérence contenue dans cet ensemble.

Le chapitre 4 concerne la révision de croyances. Dans ce cadre une nouvelle formule doit être ajoutée à la base de croyances et cette formule est typiquement incohérente avec les formules de la base. Le cadre de base de la révision de croyances, le cadre AGM [AGM85, Gär88, KM91b], est unanimement accepté. Mais ce cadre ne modélise qu'une seule étape de révision. On a donc besoin de contraindre le comportement des opérateurs lors des itérations. Cela conduit au problème de la révision itérée de croyances.

Le chapitre 5 s'intéresse à la fusion de croyances. On dispose dans ce cas d'un ensemble de sources (agents) qui fournissent des formules propositionnelles dont l'union est incohérente. Il faut alors définir des opérateurs afin d'associer à ces profils de bases de croyances une information cohérente synthétisant au mieux les croyances du groupe de sources.

Le chapitre 6 aborde la problématique de la négociation. Nous nous intéressons à une définition abstraite de la négociation, que nous avons nommée conciliation. L'idée est de ne pas s' « encombrer » de détails pratiques tels que le protocole de communication utilisé, etc., mais d'aborder le problème sous son aspect fonctionnel : un opérateur de négociation (conciliation) est une fonction qui prend en entrée un profil de bases de croyances et qui fournit en sortie un nouveau profil de croyances où le conflit a disparu, ou tout au moins a diminué.

Le chapitre 7 détaille quelques perspectives de ce travail.

La seconde partie du document présente une sélection de huit articles. Réaliser cette sélection n'a pas été chose aisée, mais la règle finalement choisie a été la suivante : nous avons sélectionné deux articles pour chaque thème abordé dans la première partie du mémoire (raisonnement en présence d'incohérence, révision, fusion, négociation), qui illustrent le mieux la démarche scientifique discutée au chapitre 2 et notre approche axiomatique de ces problèmes.

1.5. Notations

Nous considérons un langage propositionnel $\mathcal{L} = \{\alpha, \beta, \ldots\}$, défini à partir d'un ensemble fini de variables propositionnelles $\mathcal{P} = \{a, b, c, \ldots\}$, et des connecteurs usuels.

Une interprétation ω est une fonction de $\mathcal{P}$ dans $\{0, 1\}$. L'ensemble des interprétations est noté $\mathcal{W}$. Une interprétation ω est un modèle d'une formule $\alpha \in \mathcal{L}$ si et seulement si elle la rend vraie. $mod(\alpha)$ représente l'ensemble des modèles de la formule α, i.e. $mod(\alpha) = \{\omega \in \mathcal{W} \mid \omega \models \alpha\}$.

Une **base** K est un ensemble fini [2] de formules, qui représente les croyances d'un agent. Soit $\mathcal{K}$ l'ensemble des bases. Une base $K = \{\alpha_1, \ldots, \alpha_n\}$ est cohérente si la formule $\alpha = \alpha_1 \wedge \ldots \wedge \alpha_n$ est cohérente. Une base K est une théorie si elle contient l'ensemble de ses conséquences logiques, i.e. $K = Cn(K) = \{\alpha \mid K \vdash \alpha\}$. Une base (ou une formule) est dite complète si elle possède un unique modèle.

Un **profil** $E = \{K_1, \ldots, K_n\}$ est un multi-ensemble [3] de bases, qui représente les croyances d'un groupe d'agents. Soit $\mathcal{E}$ l'ensemble des profils. On note $\bigwedge E$ la conjonction des bases de $E = \{K_1, \ldots, K_n\}$, i.e., $\bigwedge E = K_1 \wedge \ldots \wedge K_n$. Un profil E est cohérent si et seulement si $\bigwedge E$ est cohérent. On note $\sqcup$ l'union multi-ensembliste, et pour des profils singletons on note $K \sqcup E$ au lieu de $\{K\} \sqcup E$. On note E^n le profil où E est répété n fois, i.e. $E^n = \underbrace{E \sqcup \ldots \sqcup E}_{n}$. Deux profils sont équivalents, noté $E_1 \equiv E_2$, s'il existe une bijection f entre E_1 et E_2 tel que pour tout $K \in E_1$ $f(K) \equiv K$.

Si $\leq$ est un pré-ordre sur A (i.e. une relation transitive et réflexive), alors $<$ dénote l'ordre strict associé, défini par $x < y$ si et seulement si $x \leq y$ et $y \not\leq x$, et $\simeq$ dénote la relation d'équivalence associée, définie par $x \simeq y$ si et seulement si $x \leq y$ et $y \leq x$. Un pré-ordre est *total* si $\forall x, y \in A$, $x \leq y$ ou $y \leq x$. Un pré-ordre qui n'est pas total est dit *partiel*. Soit un pré-ordre $\leq$ sur A, et $B \subseteq A$, alors $\min(B, \leq) = \{x \in B \mid \nexists y \in B\ y < x\}$.

Si X est un ensemble, $|X|$ représente le cardinal de l'ensemble X. Nous utiliserons $\subseteq$ pour noter l'inclusion ensembliste et $\subset$ pour noter l'inclusion stricte (i.e. $A \subset B$ ssi $A \subseteq B$ et $B \nsubseteq A$).

2. Sauf à la section 4.1 où l'on travaille avec des théories.
3. Un multi-ensemble est un ensemble où un même élément peut apparaître plusieurs fois.

Chapitre II

DÉMARCHE SCIENTIFIQUE

On ne peut se passer d'une méthode pour se mettre en quête de la vérité des choses.

(Descartes)

Afin de comprendre et justifier notre approche, qui est principalement axée vers la caractérisation logique des processus de raisonnement, il peut être utile d'expliquer la démarche scientifique la motivant.

2.1. Démarche

Il nous semble que l'étude et la compréhension de tout processus de raisonnement nécessite les étapes suivantes :

1. **Spécification du problème** : cette première étape est habituellement assez directe, il s'agit de définir formellement le problème que l'on veut résoudre.
2. **Proposition de solutions** : la deuxième étape consiste à rechercher des solutions *ad hoc* afin de résoudre le problème.
3. **Caractérisation logique** : la troisième étape consiste en l'étude des propriétés logiques des solutions au problème et en la recherche de caractérisations logiques de ces solutions.

Souvent, en intelligence artificielle, on se contente de la deuxième étape, et l'on ne voit pas forcément l'utilité de la troisième. Pourtant, cette troisième étape est indispensable si l'on veut prétendre comprendre et résoudre un problème donné. Elle est également indispensable pour pouvoir comparer les méthodes *ad hoc* proposées lors de la deuxième étape.

En effet, proposer une solution *ad hoc* à un problème, même si celle-ci est difficile à trouver et à concevoir, n'est pas synonyme de comprendre le problème. Considérons par exemple le problème de l'inférence non monotone, c'est-à-dire l'inférence à partir d'un ensemble d'informations simplement plausibles. Une solution *ad hoc* à ce problème est par exemple d'utiliser la logique des défauts. Cette solution permet de définir une relation d'inférence non monotone. Mais ce n'est pas parce que l'on dispose de cette solution que l'on peut prétendre avoir compris ce qu'est le problème de l'inférence non monotone et encore moins l'avoir résolu.

A titre de comparaison, prenons un autre problème qui est celui de faire voler des objets. Une solution *ad hoc* à ce problème est d'utiliser un avion. Mais ce n'est pas parce que l'on a construit un avion que l'on a compris ce qu'est l'aérodynamique. Historiquement d'ailleurs, des avions ont volé bien avant que l'on puisse démontrer pourquoi.

Cela n'enlève rien de l'importance et de la nécessité de la deuxième étape, car concevoir une relation d'inférence non monotone, comme concevoir un avion, sont deux problèmes importants, demandant de résoudre des questions difficiles. Mais on ne peut s'arrêter à cette étape. Pour le vol, il faut une théorie scientifique capable de modéliser le phénomène. Et c'est exactement la même chose pour l'inférence non monotone, et pour l'ensemble des processus de raisonnement.

Une difficulté supplémentaire que l'on a dans le cadre d'un problème d'intelligence artificielle est que la théorie que l'on veut mettre en place n'a pas pour objet d'étude des processus physiques, comme c'est le cas pour les sciences expérimentales. Il n'est donc pas suffisant d'utiliser une approche hypothético-déductive classique, et de valider la théorie proposée en observant qu'elle se vérifie dans l'ensemble des expériences.

Lorsque ce que l'on veut modéliser est un processus cognitif, une possibilité est d'utiliser des expériences de psychologie pour tester que la théorie proposée est conforme au comportement humain (voir par exemple [BBN04, BBN05, BNDP08]). Cela nécessite de considérer l'être humain comme modèle de rationalité. Or, il est assez facile de constater que l'homme est un bien piètre agent rationnel. Cela est évident dans nos interactions quotidiennes, et cela est illustré dans des dizaines de travaux en économie expérimentale.

La question est alors de savoir si l'intelligence artificielle consiste en la modélisation du comportement d'un être humain, ou en celui d'un agent rationnel idéal [1]. Il nous semble que les deux problèmes de la modélisation du comportement humain et celui de la modélisation d'un agent rationnel idéal sont deux questions intéressantes. Mais également qu'il est nécessaire de d'abord étudier le comportement d'un agent rationnel idéal, avant de passer au problème plus complexe de la modélisation du comportement d'un être humain.

On peut d'ailleurs noter qu'en économie, et particulièrement en théorie des jeux, la même démarche a été adoptée, avec des travaux théoriques faisant une hypothèse de rationalité (idéale) des agents afin d'obtenir des résultats intéressants, suivis de travaux exportant ces résultats dans le cadre de la rationalité limitée.

1. Cette question a donné lieu à des discussions très intéressantes lors des journées d'Intelligences Artificielles Fondamentales (IAF) de 2008.

Si l'on ne considère pas l'être humain comme modèle de rationalité, cela ne permet donc pas de concevoir des expériences afin de valider les théories proposées. En revanche, il est possible d'étudier en quoi les méthodes *ad hoc* proposées lors de la deuxième étape sont en accord avec la théorie proposée lors de la troisième étape. On peut donc alors utiliser cette théorie pour justifier ces méthodes.

Cela ne signifie pas que la troisième étape est plus importante que la deuxième, car il est de toute façon nécessaire de disposer de méthodes pratiques de résolution du problème en question. Et cela permet également d'obtenir les premières intuitions afin de trouver des propriétés caractérisant ces méthodes. Il faut donc voir ces étapes comme complémentaires l'une de l'autre. Mais la troisième étape est au moins aussi importante pour ce qui est de la compréhension du problème en question. Et, afin de résoudre de manière convaincante un problème donné, il est nécessaire de réaliser les trois étapes.

Nous allons à présent illustrer l'intérêt de cette démarche sur quelques exemples significatifs. Et comme nous avons utilisé l'inférence non monotone comme illustration, nous commencerons par celle-ci.

2.2. Inférence non monotone

1. **Spécification du problème** : comment permettre l'inférence (de « sens commun ») à partir d'informations simplement plausibles ou de règles générales (du genre « normalement les oiseaux volent »), dont les conclusions peuvent être remises en cause si l'on ajoute de nouvelles informations (d'où la non monotonie) ?

2. **Proposition de solutions** : de nombreuses méthodes ont été proposées pour résoudre ce problème. On peut citer en particulier les méthodes à base d'hypothèses de monde clos, la logique des défauts, la circonscription, etc.

3. **Caractérisation logique** : Gabbay a été le premier à suggérer que de ne caractériser ces logiques que par le fait qu'elles ne satisfaisaient pas la propriété de monotonie de la logique classique n'était pas suffisant. Il a donc suggéré de tenter d'étudier les propriétés communes et souhaitables de ces différentes méthodes [Gab85]. Cela a été développé en particulier par Makinson [Mak94], puis par Kraus, Lehmann et Magidor [KLM90, LM92]. On a pu montrer que l'ensemble des propriétés minimales que l'on peut attendre d'une relation d'inférence non monotone (le système P de [KLM90]) était équivalent à une sémantique très naturelle : la sémantique préférentielle [Sho87].

Cette caractérisation est indispensable non seulement pour comprendre ce qu'est une logique non monotone, mais également pour comparer les méthodes *ad hoc* proposées. Cela permet d'étudier quelles sont les propriétés satisfaites par ces méthodes, au lieu de les comparer en échangeant des exemples qu'une méthode résout alors que l'autre non. Ceci ne permet pas d'établir de conclusions d'une portée suffisante, mais a été longtemps l'usage dans les articles traitant de non monotonie.

2.3. Révision

1. **Spécification du problème** : comment incorporer aux croyances d'un agent une nouvelle information qui remet en question une partie de ses croyances actuelles ?
2. **Proposition de solutions** : c'est un problème qui se pose dès que l'on doit gérer les croyances d'un agent, ou une base de données. Plusieurs méthodes dédiées ont donc été proposées [FUV83, Dal88, Bor85, KW85, Win88].
3. **Caractérisation logique** : Alchourrón, Gärdenfors and Makinson ont proposé des propriétés logiques pour les opérateurs de révision, ainsi qu'une caractérisation logique montrant que les opérateurs satisfaisant ces propriétés correspondent à des méthodes de révision très naturelles [AGM85, Gär88]. Depuis, de nombreux théorèmes de représentation ont été énoncés, illustrant le rôle fondamental de cette caractérisation logique.

2.4. Vote

Nous voudrions à présent souligner que cette approche n'est pas utile juste dans le cadre du raisonnement et de l'intelligence artificielle, mais qu'elle s'est déjà montrée féconde dans d'autres champs disciplinaires, en particulier en économie.

1. **Spécification du problème** : comment définir une bonne méthode de vote ?
2. **Proposition de solutions** : de très nombreux systèmes de vote ont été proposés, et ont généré beaucoup de débats afin de déterminer quels étaient les meilleurs d'entre eux. On peut faire remonter ces discussions au moins jusqu'au XVIII^{e} siècle avec les échanges entre Condorcet [Con85] et Borda [Bor81].
3. **Caractérisation logique** : un cap a été franchi lorsque Arrow a décidé, plutôt que d'étudier un système de vote particulier, d'étudier les propriétés que l'on peut attendre de ces systèmes. Il a ainsi montré un théorème d'impossibilité prouvant qu'il n'existe aucune « bonne » méthode de vote [Arr63]. Cela a permis de franchir un palier dans la compréhension des méthodes de vote et a marqué un tournant dans la discipline (et a valu un prix Nobel à Arrow en 1972).

Nous nous limitons à cet exemple de choix social, mais les approches axiomatiques ont également donné des résultats intéressants en théorie des jeux.

2.5. Conclusion

Il nous semble donc qu'il est nécessaire, afin de prétendre comprendre un processus de raisonnement, de réaliser ces trois étapes : spécification, résolution (proposition de solutions), caractérisation.

L'ordre que nous avons utilisé pour présenter la deuxième étape (résolution) et la troisième (caractérisation) est l'ordre usuel, car il y a souvent une solution *ad hoc* de proposée avant que l'on arrive à caractériser ces solutions. Mais fréquemment ces deux étapes sont menées en parallèle, et des échanges entres ces deux étapes sont très utiles, afin de s'assurer que la modélisation réalisée lors de la troisième étape capture bien l'ensemble des solutions définies lors de la deuxième étape, et de l'améliorer si ce n'est pas le cas ; et afin d'utiliser la troisième étape pour trouver de meilleures solutions *ad hoc* ou pour améliorer les propositions existantes. Cela illustre la complémentarité de ces deux étapes.

Notre approche des différents problèmes concerne principalement la troisième étape, c'est-à-dire l'étude des propriétés logiques des processus en question et leur caractérisation logique.

Ainsi, lorsque nous proposons de nouvelles méthodes *ad hoc* pour un problème particulier, en particulier pour des problèmes où ces méthodes ne sont pas nombreuses, nous considérons ces travaux comme un pas vers la troisième étape et la caractérisation logique.

Chapitre III

RAISONNEMENT EN PRÉSENCE D'INCOHÉRENCE

Autant que savoir, douter me plaît.

(Dante)

Lorsque l'on doit raisonner à partir d'un ensemble de formules incohérent, on ne peut pas utiliser la logique classique, qui trivialise. Il faut donc trouver des mécanismes dédiés. Le problème principal est celui de l'inférence, afin de pouvoir exploiter la base malgré ses incohérences. Un autre problème important est celui de la mesure de l'incohérence. Il peut en effet être utile d'évaluer à quel point une base est incohérente. De même, savoir quelles sont les formules qui apportent beaucoup ou peu d'incohérence peut être très utile pour de nombreuses tâches.

3.1. Inférence en présence d'incohérence

Il s'agit ici de définir des relations qui permettent d'obtenir des conclusions non triviales à partir d'un ensemble incohérent de formules, ce qui n'est pas possible avec la logique (propositionnelle) classique, qui trivialise en inférant toutes les formules du langage.

Pour éviter la trivialisation, il faut affaiblir la puissance inférentielle de la logique classique. Il n'y a que deux possibilités :

- affaiblir l'ensemble des hypothèses (ensemble de formules).
- affaiblir la logique (la relation d'inférence).

On peut situer les relations d'inférence en présence d'incohérence par rapport aux autres logiques (classiques, non monotones) en utilisant la définition de Tarski [Tar36, TCW83] d'une inférence en terme d'inclusion de modèles.

Ainsi l'inférence en logique classique $A \models B$ peut être définie comme l'inclusion des modèles de A dans les modèles de B. Cette inférence est donc vérifiée si à chaque fois que A est vrai (i.e. pour chaque interprétation qui satisfait A) B l'est également :

$$A \models B \text{ ssi } mod(A) \subseteq mod(B)$$

On peut définir l'inférence non monotone de manière similaire. La principale différence est que, au lieu de demander que tous les modèles de A soient des modèles de B, on ne s'intéresse qu'aux modèles les plus typiques (c'est le sens de la sémantique préférentielle de [Sho87]) ; soit $\gamma(X)$ une fonction de sélection qui retourne un sous-ensemble de X :

$$A \mid\!\sim B \text{ ssi } \gamma(mod(A)) \subseteq mod(B)$$

Le problème lorsque l'on veut définir une relation d'inférence en présence d'incohérence est que A n'a pas de modèle classique. Il faut donc trouver un moyen d'associer un ensemble non vide de modèles à A. Une première solution est d'affaiblir les hypothèses, c'est-à-dire d'utiliser une procédure afin de ne sélectionner qu'un sous ensemble [1] des formules de A :

$$A \models_H B \text{ ssi } mod(\gamma(A)) \subseteq mod(B)$$

Une deuxième solution est d'affaiblir la relation d'inférence. En effet, si A n'a pas de modèle classique, une possibilité est d'utiliser des interprétations (multi-valuées par exemple) qui garantissent que A aura toujours des modèles. Cela aura pour conséquence d'affaiblir la logique, car cela ne permettra pas d'avoir autant de conséquences. Mais cela permettra dans tous les cas d'avoir une relation d'inférence non triviale. On utilise donc exactement la définition générale de Tarski, mais avec une notion de modèle différente de celle de la logique classique :

$$A \models_M B \text{ ssi } mod_M(A) \subseteq mod_M(B)$$

Cette description des logiques classiques, non monotones, et sous incohérence est intéressante pour bien comprendre les différentes problématiques en présence. Cela montre en particulier que l'inférence non monotone est une inférence supra-classique : le problème est que l'on a trop de modèles pour A et que l'on ne veut raisonner qu'à partir d'un sous-ensemble de ces modèles. Et, qu'au contraire, l'inférence sous incohérence est une inférence infra-classique : le problème est que l'on n'a pas assez de modèles pour A.

Mais cette description est un peu simpliste pour rendre compte de l'ensemble des relations d'inférence sous incohérence. En particulier, beaucoup des solutions basées sur l'affaiblissement de la logique sont des logiques paraconsistantes définies purement syntaxiquement, en changeant la signification des connecteurs logiques, ou en contraignant les méthodes de preuves, ce qui ne permet pas de les définir sémantiquement en terme de modèles.

Détaillons à présent les deux solutions.

1. Plus exactement il s'agit habituellement d'un ensemble de sous-ensembles.

3.1.1. Affaiblir les hypothèses

L'affaiblissement des hypothèses conduit aux approches basées sur les ensembles maximaux cohérents. Puisqu'il n'est pas possible de garder l'ensemble des formules, l'idée est de définir des inférences à partir de sous-ensembles de ces formules.

Lorsque toutes les formules ont la même importance, on peut définir les relations d'inférence suivantes [RM70, BDP97] (voir [Bre89, BDP98] lorsque certaines formules sont plus importantes/prioritaires que d'autres). Soit une base de croyances $K = \{\varphi_1, \ldots, \varphi_n\}$:

Définition 1 *K' est un* **sous-ensemble maximal cohérent** *(ou* **maxcons***) de K s'il satisfait :*
- $K' \subseteq K$
- $K' \nvdash \bot$
- *Si* $K' \subset K'' \subseteq K$, *alors* $K'' \vdash \bot$

Notons $\textsc{Maxcons}(K)$ *l'ensemble des maxcons de* K.

Habituellement une base incohérente admet plusieurs sous-ensembles maximaux cohérents. Il faut alors choisir une stratégie afin de définir les conséquences de K à partir des conséquences classiques des maxcons. De nombreuses relations d'inférences sont alors définissables dans ce cadre. Les principales sont :

Définition 2
- *L'***inférence universelle** *[RM70] (ou* **MC-inférence***[BDP97]) où α est une conséquence de K si elle est conséquence de tous les maxcons :*

$$K \vdash_{MC} \alpha \text{ ssi } \forall K' \in \textsc{Maxcons}(K)\ K' \vdash \alpha$$

- *L'***inférence existentielle** *[RM70] où α est une conséquence de K si elle est conséquence d'au moins un maxcons :*

$$K \vdash_{\exists} \alpha \text{ ssi } \exists K' \in \textsc{Maxcons}(K)\ K' \vdash \alpha$$

- *L'***inférence argumentative** *[BDP93] où α est une conséquence de K si elle est conséquence d'au moins un maxcons et si sa négation n'est conséquence d'aucun maxcons :*

$$K \vdash_{A} \alpha \text{ ssi } \exists K' \in \textsc{Maxcons}(K)\ K' \vdash \alpha \text{ et } \nexists K' \in \textsc{Maxcons}(K)\ K' \vdash \neg\alpha$$

- *L'***inférence (universelle) à base de cardinalité** *[BDP97, BKMS92] (ou* **L-inférence** *[BDP97]) où α est une conséquence de K si elle conséquence de tous les plus grands (pour la cardinalité) maxcons, soit* $\textsc{Maxcons}_{card}(K) = \{K_c \mid K_c \in \textsc{Maxcons}(K) \text{ et } \forall K' \in \textsc{Maxcons}(K)\ |K_c| \geq |K'|\}$ *:*

$$K \vdash_{L} \alpha \text{ ssi } \forall K' \in \textsc{Maxcons}_{card}(K)\ K' \vdash \alpha$$

- *L'***inférence libre** [2] *[BDP92] où α est une conséquence de K si elle est conséquence de l'ensemble des formules qui appartiennent à tous les maxcons* [3] *:*

$$K \vdash_{Free} \alpha \text{ ssi } \{\varphi \mid \forall K' \in \textsc{Maxcons}(K)\ \varphi \in K'\} \vdash \alpha$$

2. Qui correspond à l'approche WIDTIO (« *When In Doubt Throw It Out* ») [Win90].
3. Dans [BDP92] cette relation est définie à partir des formules n'appartenant à aucun ensemble minimal incohérent, ce qui est équivalent.

$$
\begin{tikzcd}
 & & \vdash_L \arrow[dr] & & \\
\vdash_{Free} \arrow[r] & \vdash_{MC} \arrow[ur] \arrow[dr] & & \vdash_{\exists} \arrow[r] & \vdash \\
 & & \vdash_A \arrow[ur] & &
\end{tikzcd}
$$

FIGURE 3.1 – Relations d'inférence à base de maxcons

Comme on peut s'y attendre d'après leur définition, les relations présentées sont liées logiquement, comme représenté à la figure 3.1 [BDP97]. En particulier, elles étendent toutes la logique classique, dans le sens où dans le cas où la base K est cohérente, elles donnent toutes exactement les mêmes inférences que la logique classique, et dans le cas où la base K est incohérente elles permettent tout de même d'inférer des conséquences non triviales. C'est une propriété d'autant plus intéressante que dans l'approche consistant à affaiblir la logique (i.e. les logiques paraconsistantes) que nous aborderons dans la section suivante, la plupart de ces logiques ne donnent pas les mêmes conséquences que la logique classique lorsque la base K est cohérente.

La relation la plus délicate à manipuler est l'inférence existentielle qui permet d'inférer à la fois une formule α et sa négation $\neg\alpha$. Cela ne conduit pas à une trivialisation (par explosion) comme dans le cadre de la logique classique, mais il n'est alors par exemple pas possible de prendre la fermeture classique de l'ensemble des conséquences sans risquer l'incohérence.

Beaucoup d'autres relations d'inférences peuvent être imaginées, comme une inférence argumentative à base de cardinalité par exemple.

Des travaux plus récents affaiblissent les hypothèses d'une autre manière qu'en sélectionnant des maxcons. On peut par exemple mentionner l'affaiblissement par oubli de variables [LM02], où l'utilisation de nos opérateurs de fusion (voir section 5.2.1) afin de définir une relation d'inférence qui affaiblit les formules en sélectionnant les modèles les plus proches [Ari08].

Dans ces approches, une hypothèse sous-jacente est que des formules différentes sont liées par un lien moins fort qu'une conjonction logique. Dans le cas où elle est cohérente, une base de croyances $B = \{\varphi_1, \ldots, \varphi_n\}$, où les φ_i sont des formules propositionnelles, est équivalente à la base $B' = \{\varphi_1 \wedge \ldots \wedge \varphi_n\}$. En revanche, dans le cas où cette conjonction est incohérente, ces deux bases ont typiquement des ensembles de conséquences distincts. Cela illustre le fait que la virgule utilisée dans la base B est un connecteur particulier, différent de la conjonction. Nous avons étudié une logique comportant un tel connecteur *« virgule »*. Ce cadre permet de fournir une sémantique à ce connecteur et de généraliser un certain nombre d'approches permettant de raisonner en présence d'informations contradictoires [28, 66].

3.1.2. Affaiblir la logique

L'autre possibilité lorsque l'on veut raisonner non trivialement à partir d'une base classiquement incohérente est de garder l'ensemble des formules mais d'affaiblir la relation d'inférence. Cela conduit aux logiques paraconsistantes. Il y a deux possibilités :

- Soit on fait en sorte de bloquer la trivialisation en affaiblissant les connecteurs logiques.
- Soit on fait en sorte de localiser les incohérences en utilisant plus que les deux valeurs de vérité (vrai/faux) usuelles.

Affaiblissement des connecteurs logiques

Cette voie regroupe l'essentiel des travaux sur les logiques paraconsistantes. L'idée est ici de faire en sorte que l'on ne puisse dériver l'ensemble des formules à partir de la base incohérente (principe appelé *ex falso quodlibet sequitur* ou **explosion**). C'est-à-dire qu'on peut définir les **logiques paraconsistantes** comme les logiques $\vdash_P$ qui satisfont la propriété suivante :

$$\exists K\ \exists\alpha\ \exists\beta\ K \vdash_P \alpha \text{ et } \ K \vdash_P \neg\alpha \text{ et } K \nvdash_P \beta$$

Nous ne ferons pas de rappel de ces nombreux travaux, voir [Hun96, Béz00, CM01] pour une introduction sur ce sujet. La caractéristique commune de ces logiques est la nécessité d'affaiblir les connecteurs logiques. En particulier l'affaiblissement de la négation conduit par exemple aux **C-systèmes** de da Costa [CM01]. Le problème avec ce genre de solutions est que l'on s'éloigne de la logique classique : en général, ces logiques ne permettent pas les mêmes inférences qu'avec la logique classique lorsque la base K est cohérente (on peut néanmoins citer l'approche originale de Besnard et Schaub [BS96] consistant à découpler chaque variable de sa négation puis d'ajouter des règles par défaut pour tenter ensuite de recoupler un maximum de ces variables avec leur négation, qui n'a pas ce défaut).

Une autre possibilité est de contraindre les preuves, afin de bloquer l'explosion. Par exemple la **logique quasi-classique** [BH95] interdit l'utilisation des règles comme la résolution après l'utilisation de règles comme l'introduction de disjonction. L'intérêt de cette logique est qu'elle permet (quasiment) les mêmes inférences que la logique classique si la base est cohérente.

Dans ce cadre, nous avons défini [3, 45, 67] une nouvelle logique, nommée **logique quasi-possibiliste**, qui généralise la logique possibiliste et la logique quasi-classique [BH95], permettant de tirer parti des avantages des deux logiques, qui sont toutes les deux proches de la logique propositionnelle classique. Cette logique peut résoudre les conflits apparaissant au même niveau de certitude (comme la logique quasi-classique), mais elle peut également prendre en compte la stratification de la base pour introduire de la gradualité dans l'analyse des conflits (comme en logique possibiliste). Nous avons proposé également des mesures de conflits associées à la logique définie.

Utilisation de logiques multi-valuées

L'utilisation de logiques multi-valuées permet de circonscrire les incohérences aux variables concernées et d'éviter le *ex falso quodlibet sequitur*. Cette approche est plus fine que celles basées sur la sélection de sous-ensembles maximaux cohérents de formules puisqu'avec ces approches ce sont les formules qui sont prises comme unité de base, alors qu'en utilisant une logique multi-valuée ce sont les variables. Cela permet donc également de prendre en compte des informations provenant de formules incohérentes.

La logique la plus typique est la **logique LP** de Priest [Pri91]. C'est une **logique trivaluée**, avec en plus des valeurs de vérité Vrai (T) et Faux (F), une valeur B signifiant incohérent (« *à la fois vrai et faux* »).

Définition 3 • *Une interprétation ω de LP est une fonction qui associe à chaque variable propositionnelle une des valeurs de vérité $\mathsf{F}, \mathsf{B}, \mathsf{T}$. Soit $3^{\mathcal{P}}$ l'ensemble des interprétations de LP.*
On ordonne les valeurs de vérité comme suit : $\mathsf{F} <_t \mathsf{B} <_t \mathsf{T}$.
Les interprétations sont étendues aux formules comme suit :
– $\omega(\top) = \mathsf{T}$, $\omega(\bot) = \mathsf{F}$
– $\omega(\neg\alpha) = \mathsf{B}$ *ssi* $\omega(\alpha) = \mathsf{B}$
  $\omega(\neg\alpha) = \mathsf{T}$ *ssi* $\omega(\alpha) = \mathsf{F}$
– $\omega(\alpha \wedge \beta) = \min_{\leq_t}(\omega(\alpha), \omega(\beta))$
– $\omega(\alpha \vee \beta) = \max_{\leq_t}(\omega(\alpha), \omega(\beta))$

• *L'ensemble des modèles* [4] *d'une formule φ est :*

$$Mod_{LP}(\varphi) = \{\omega \in 3^{\mathcal{P}} \mid \omega(\varphi) \in \{\mathsf{T}, \mathsf{B}\}\}$$

• *La relation d'inférence LP est alors définie par :*

$$K \models_{LP} \varphi \text{ ssi } Mod_{LP}(K) \subseteq Mod_{LP}(\varphi)$$

Le problème est que la relation d'inférence LP ne permet pas d'inférer beaucoup de conséquences. Il est plus intéressant d'utiliser une variante de cette logique avec minimisation, la **logique LP**$_m$ [Pri91] :

Définition 4 • *Soit $\omega!$ l'ensemble des variables « incohérentes » d'une interprétation ω, i.e.*

$$\omega! = \{x \in \mathcal{P} \mid \omega(x) = \mathsf{B}\}$$

Alors les modèles minimaux d'une formules sont les « plus classiques » (i.e. les modèles avec le plus grand nombre de variables propositionnelles (pour l'inclusion ensembliste) affectées à T ou F) :

$$min(Mod_{LP}(\varphi)) = \{\omega \in Mod_{LP}(\varphi) \mid \nexists \omega' \in Mod_{LP}(\varphi) \text{ t.q. } \omega'! \subset \omega!\}$$

La relation d'inférence LP_m est alors définie par :

$$K \models_{LP_m} \varphi \text{ ssi } min(Mod_{LP}(K)) \subseteq Mod_{LP}(\varphi)$$

La logique LP$_m$ permet également d'obtenir les mêmes conséquences que la logique classique lorsque la base est cohérente.

4. Lorsque l'on travaille avec une logique multi-valuée, il faut définir un ensemble de **valeurs de vérité désignées**, c'est-à-dire les valeurs de vérité qui définissent les modèles d'une formule. Dans le cas de la logique LP il s'agit de $\{\mathsf{T}, \mathsf{B}\}$.

Dans [22] nous avons exploré l'utilisation de logiques tri-valuées pour le raisonnement paraconsistant. Nous avons défini et étudié plusieurs formes d'inférences tri-valuées, et nous les avons comparées du point de vue des propriétés logiques, de la prudence et de la complexité algorithmique. L'idée est d'obtenir plus de conséquences qu'avec la logique trivaluée de base LP. Nous avons proposé 3 mécanismes différents : le premier est une inférence argumentative où une formule est inférée si elle est conséquence de la logique sous-jacente et que sa négation ne l'est pas. Le second est basé sur un principe de minimisation de l'incertitude : seules les formules dont la valeur de vérité est T sont inférées (pas celles évaluées à B). Le troisième définit les formules inférées comme les formules « au moins aussi vraies » que la base. En combinant ces mécanismes entre eux, et avec les inférences basiques et celles basées sur une inférence préférentielle, cela a donné 9 nouvelles relations d'inférences différentes.

Une fois que l'on a accepté le principe de travailler avec plus de deux valeurs de vérité, une question importante est de savoir combien de valeurs de vérité sont nécessaires/utiles. Ginsberg a étudié la famille des **logiques multi-valuées à base de bi-treillis** [Gin88]. Les valeurs de vérité sont ordonnées selon deux ordres : l'ordre sur la vérité (plus ou moins vrai ou faux), et l'ordre sur la connaissance (plus ou moins d'information ou d'ignorance). Deux cas particuliers sont les logiques $\mathcal{FOUR}$, qui est la logique multi-valuée de base pour le raisonnement paraconsitant proposée par Belnap [Bel77b, Bel77a], (B représente intuitivement la valeur incohérente *« vrai et faux »* et U la valeur inconnue *« ni vrai ni faux »*) et la **logique** $\mathcal{SEVEN}$ qui correspond à la **logique des défauts supernormaux** (dF représente la valeur *« faux par défaut »*, dT représente la valeur *« vrai par défaut »*, dB représente la valeur *« vrai et faux par défaut »*).

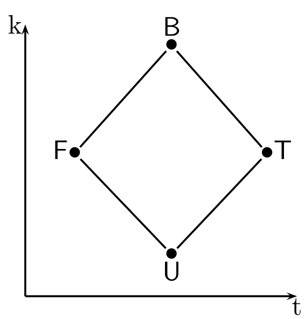


FIGURE 3.2 – (FOUR, $\leq_t, \leq_k$)

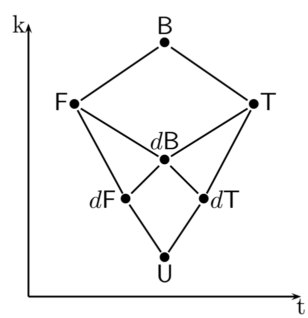


FIGURE 3.3 – (SEVEN, $\leq_t, \leq_k$)

On peut utiliser des logiques avec autant de valeurs de vérité que l'on veut. Mais Arieli et Avron ont montré dans un très bon article [AA98] que 4 valeurs (donc les logiques issues de la **logique de Belnap** [Bel77b, Bel77a]) étaient suffisantes, dans le sens où toute relation d'inférence basée sur une logique multi-valuée à base de bi-treillis peut être définie dans une logique à 4 valeurs.

Nous avons également étudié [10] la définition de relations d'inférence basées sur une interprétation bipolaire des logiques multi-valuées à base de treillis. La motivation de ce travail était d'étudier les conséquences d'une interprétation bipolaire des valeurs de vérité, c'est-à-dire qu'au lieu de considérer l'ordre de vérité et l'ordre de connaissance usuels, on utilise deux ordres où l'on décorelle le vrai et le faux, on a donc un ordre du vrai (verum) et un ordre du faux (falsum). On peut représenter cela graphiquement par une rotation des valeurs de vérité comme illustré dans les figures 3.4 et 3.5 sur la logique $\mathcal{FOUR}$. Utiliser alors ces nouveaux pré-ordres pour définir des relations d'inférences préférentielles (à base de minimisation) permet de définir de nouvelles relations d'inférence.

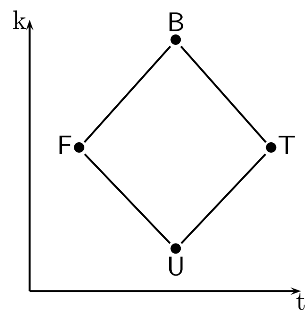


FIGURE 3.4 – $(\mathsf{FOUR}, \leq_t, \leq_k)$

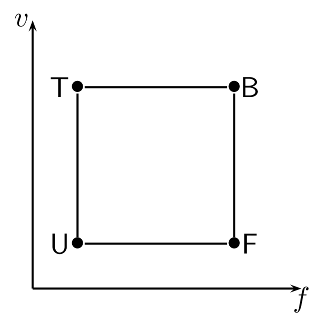


FIGURE 3.5 – $(\mathsf{FOUR}, \leq_v, \leq_f)$

3.2. Mesure de l'incohérence

Il est usuel en logique classique d'utiliser une mesure binaire de la contradiction : une base est soit cohérente, soit contradictoire. Cette dichotomie est évidente lorsque le seul outil déductif est l'inférence classique, puisque une base contradictoire n'est alors d'aucune utilité. Mais nous disposons d'un nombre important de logiques développées pour permettre des inférences non triviales à partir de bases contradictoires. Sous cet angle, cette dichotomie n'est pas suffisante pour donner une mesure de la contradiction d'une base de croyance. Une mesure plus fine est nécessaire.

Nous avons écrit un chapitre d'état de l'art [15] sur les **mesures d'information** et les **mesures de contradiction** des bases logiques, comme par exemples les mesures proposées par Hunter [Hun02], Knight [Kni02, Kni03], Lozinskii [Loz94] ou par nous [26]. Il y a différentes définitions pour ces mesures, ce qui est naturel dans la mesure où il n'existe pas de logique paraconsistante unique et incontestée mais plutôt un ensemble de logiques ayant chacune des avantages et des inconvénients. Il est donc normal que plusieurs définitions soient possibles pour ces mesures de contradiction. C'est à notre connaissance le premier article récapitulatif sur ces mesures de contradiction et d'information (pour des bases incohérentes).

Nous allons nous focaliser dans la suite sur les mesures de contradiction, mais les mesures d'information pour des bases incohérentes (que nous avons traitées également dans [15]) sont également intéressantes. L'idée est que si l'on veut importer les idées de Shannon sur la mesure de l'information dans des cadres logiques, la mesure de l'information revient plus ou moins à compter le nombre de modèles. Or une base incohérente, qui n'a aucun modèle, ne contient, suivant cette définition, aucune information.

Cela n'est pas raisonnable dès que l'on arrive à distinguer des bases incohérentes (ce qui est le cas avec toute logique pour le raisonnement en présence d'incohérence). Il faut donc généraliser cette notion dans ce cadre. Voir [15] pour un aperçu des travaux de [Loz94, Kni03] et [26].

On peut classer les mesures de contradiction dans deux classes, suivant que l'on prend comme « unité de mesure » les formules ou les variables. Nous allons présenter brièvement ces deux classes.

3.2.1. Mesures basées sur les formules

L'idée de ces mesures est que plus l'incohérence demande de formules de la base, moins cette incohérence est importante. Cette intuition peut-être justifiée par des exemples tels que le **paradoxe de la loterie**.

Exemple 1 *Il y a un certain nombre de billets de loterie dont l'un est le ticket gagnant. Représentons par w_i le fait que le ticket i est le ticket gagnant. On représente alors la phrase précédente par $w_1 \vee \ldots \vee w_n$. De plus pour chaque ticket i on suppose de manière pessimiste (ou simplement probabilistiquement réaliste si le nombre de tickets est important) que ce ticket n'est pas le ticket gagnant, donc que $\neg w_i$. On a donc la base K_{L_n} suivante :*

$$K_{L_n} = \{\neg w_1, \ldots, \neg w_n, w_1 \vee \ldots \vee w_n\}$$

Clairement, s'il n'y a que deux ou trois tickets pour la loterie, cette base est hautement incohérente. Mais s'il y a des millions de tickets, il n'y a intuitivement pratiquement pas de conflit dans la base.

Cette idée est le point de départ de la mesure proposée par Knight [Kni01, Kni03, Kni02].

Définition 5 *Une fonction de probabilité sur $\mathcal{L}$ est une fonction $P : \mathcal{P} \rightarrow [0, 1]$ t.q. :*
- *si $\vdash \alpha$, alors $P(\alpha) = 1$*
- *si $\vdash \neg(\alpha \wedge \beta)$, alors $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$*

Voir [Par94] pour plus de détails sur cette définition. Dans le cas fini, cette définition donne une distribution de probabilité sur les interprétations, et la probabilité d'une formule est alors la somme des probabilités de ses modèles.

La mesure d'incohérence proposée par Knight [Kni01] est définie par :

Définition 6 *Soit une base de croyances K.*
- *K est $\eta-$**cohérent** ($0 \leq \eta \leq 1$) si il existe une fonction de probabilité P telle que $P(\alpha) \geq \eta$ pour tout $\alpha \in K$.*
- *K est **maximalement** $\eta-$**cohérent** si η est maximal (i.e. si $\gamma > \eta$ alors K n'est pas $\gamma-$consistant).*

La notion de *η-cohérence maximale* est alors une mesure d'incohérence de la base. On peut vérifier rapidement que cette définition formalise l'idée que plus on a besoin de formules pour générer l'incohérence, moins celle-ci est problématique. On peut vérifier qu'une base est maximalement 0-cohérente si et seulement si elle contient une formule contradictoire, et qu'elle est maximalement 1-cohérente si et seulement si elle est cohérente. Voyons également ce que cela donne sur quelques exemples.

Exemple 2 *Soit* $K_1 = \{a, b, \neg a \vee \neg b\}$.
K_1 *est maximalement* $\frac{2}{3}-$*cohérent.*
Soit $K_2 = \{a \wedge b, \neg a \wedge \neg b, a \wedge \neg b\}$.
K_2 *est maximalement* $\frac{1}{3}-$*cohérent, et chacune de ses sous-bases de cardinalité* 2 *est maximalement* $\frac{1}{2}-$*cohérent.*

Pour un ensemble minimal incohérent de formules, calculer cette mesure est facile :

Définition 7 *Un* **ensemble minimal incohérent**[5] M *de* K *est défini par :*
- $M \subseteq K$
- $M \vdash \perp$
- *Si* $M' \subset M$ *alors* $M' \nvdash \perp$

Soit $\mathsf{MI}(K)$ *l'ensemble des ensembles minimalement incohérents de* K.

Théorème 1 ([Kni01]) *Si* $K' \in \mathsf{MI}(K)$, *alors* K' *est maximalement* $\frac{|K'|-1}{|K'|}-$*consistant.*

En général, cette mesure est plus compliquée à calculer, mais il est possible d'utiliser la méthode du simplexe pour le calcul [Kni01].

3.2.2. Mesures basées sur les variables

Une autre méthode pour évaluer l'incohérence d'un ensemble de formules est de regarder quelle est la proportion du langage concernée par l'incohérence. Il n'est donc pas possible d'utiliser la logique classique à cette fin puisque l'incohérence contamine l'ensemble de la base (et du langage). Mais si l'on compare les deux bases $K_1 = \{a, \neg a, b \wedge c, d\}$ et $K_2 = \{a, \neg a, b \wedge \neg c, c \wedge \neg b, d, \neg d\}$, on remarque que dans K_1 l'incohérence concerne principalement la variable a, alors que dans K_2 toutes les variables sont incluses dans un conflit. C'est ce genre de distinctions que ces approches permettent.

Une méthode afin de circonscrire l'incohérence aux variables directement concernées est d'utiliser des logiques multi-valuées, avec au moins une troisième « valeur de vérité » indiquant qu'il y a un conflit sur la valeur de vérité (vrai ou faux) de la variable.

Nous n'avons pas ici la place de détailler l'ensemble des mesures qui ont été proposées [Gra78, Hun02, GH06] [26], voir [15] pour plus de détails sur ces approches. Nous ne donnerons donc qu'un exemple illustratif : le **degré d'incohérence** proposé par Grant [Gra78].

Soit X une interprétation quadri-valuée (T, F, B, U), où B signifie intuitivement *« à la fois vrai et faux »* , et U signifie *« ni vrai ni faux »* , on définit deux fonctions :

5. Ces ensembles sont également appelés MUS (pour Minimal Unsatisfiable Subformula) dans la littérature sur SAT [GMP06, GMP09].

- $\mathsf{CCount}(X) = \{a \in \mathcal{P} \mid X(a) = \mathsf{T}$ ou $X(a) = \mathsf{F}\}$
- $\mathsf{ICount}(X) = \{a \in \mathcal{P} \mid X(a) = \mathsf{B}\}$

Le degré d'incohérence [6] est un rapport entre les conflits contenus dans la base (ICount) et l'information contenue dans la base (CCount) :

Définition 8 *Le degré d'incohérence d'une interprétation quadri-valuée X est :*

$$Inc_G(X) = \frac{\mathsf{CCount}(X)}{\mathsf{CCount}(X) + 2 * \mathsf{ICount}(X)}$$

On peut alors définir le degré d'incohérence comme le degré d'incohérence maximum de ses modèles [7] :

$$Inc_G(K) = \max_{X \models_4 K} Inc_G(X)$$

Le problème avec cette mesure est qu'elle ne donne pas des résultats très satisfaisants, parce qu'elle considère trop de modèles. Comme nous l'avons dit dans la section 3.1.2, les logiques multi-valuées sans minimisation admettent de trop nombreux modèles. Pour améliorer le comportement de cette mesure il est nécessaire d'utiliser des logiques multi-valuées avec minimisation. En utilisant la logique quasi-classique [BH95] par exemple, cela conduit à la définition de la mesure de **coherence** proposée par Hunter [Hun02].

Le problème que nous voyons dans ce genre de mesures est qu'elles mélangent quantité de contradiction et quantité d'information, puisque ICount mesure les conflits (nombre de variables « incohérentes ») et CCount mesure l'information (nombre de variables sur lesquelles on a une information). Ces mesures sont donc des mesures composites, définies à partir d'une mesure de contradiction et d'une mesure d'information. Elles peuvent présenter de l'intérêt, mais nous pensons qu'elles ne forment pas de bonnes mesures en tant que mesures de contradiction puisqu'elles ne se contentent pas de mesurer le conflit.

Nous avons proposé [26, 46] une approche à base de tests pour définir dans un cadre uniforme des mesures d'information et de contradiction de bases de croyances. Intuitivement, dans cette approche, le degré de contradiction d'une base est le coût d'un plan de test préféré pour purifier la base alors que le degré d'information d'une base est le coût d'un plan de test préféré pour déterminer quel est le monde réel. Plusieurs critères de préférence peuvent être utilisés. Les propriétés des **mesures d'incohérence à base de plans de tests** introduites ont été analysées et elles ont été comparées avec diverses mesures proposées jusque-là dans la littérature (par Hunter [Hun02], Knight [Kni02, Kni03], Lozinskii [Loz94] notamment). Un point fort de l'approche proposée est que le cadre développé est suffisamment général pour que diverses approches pour l'inférence en présence de contradictions puissent être prises en compte. Voyons un exemple de ces plans de tests dans le cadre de la logique LP_m.

6. On devrait plutôt parler de degré de cohérence puisque la valeur maximale est obtenue pour une base cohérente.

7. Soit $\models_4$ la relation de satisfaction de la logique de Belnap [Bel77b, Bel77a]

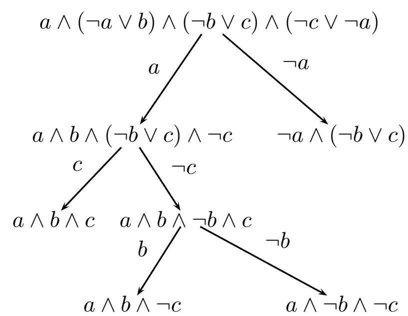


FIGURE 3.6 – Degré de contradiction dans LP_m

Exemple 3 *Soit la base $K = \{a \wedge (\neg a \vee b) \wedge (\neg b \vee c) \wedge (\neg c \vee \neg a)\}$. La figure 3.6 présente un plan de coût minimal (dans le contexte atomique standard[8]) qui purifie (i.e. permet de supprimer tous les conflits de la base) la base K. Tous les autres plans de test ont un coût plus élevé. Le degré de contradiction de K est de 3 (c'est la profondeur[9] du plan de tests de la figure 3.6).*

3.2.3. Valeurs d'incohérence

Jusqu'ici nous avons discuté de mesures d'incohérence, c'est-à-dire de fonctions qui permettent de mesurer à quel point une base est incohérente, et donc de comparer la quantité d'incohérence de plusieurs bases.

Il est intéressant de disposer de ce type de mesures par exemple pour s'en servir afin de définir des méthodes de raisonnement diverses, où il peut être utile de disposer d'un critère afin de choisir, parmi un ensemble de bases, la base la moins contradictoire.

Une mesure bien plus utile serait, au lieu de quantifier l'incohérence au niveau, global, de la base, de pouvoir évaluer à quel point chaque formule de la base est conflictuelle. C'est-à-dire d'évaluer la responsabilité de chaque formule dans les incohérences de la base. Cela permettrait d'utiliser ces mesures par exemple pour identifier les formules les plus conflictuelles, afin de résoudre les conflits : on peut par exemple définir une inférence à base de maximaux cohérents où la maximalité est calculée à partir de cette mesure, ou des opérateurs de révision gardant en priorité les formules les moins conflictuelles, etc.

Pour distinguer les deux types de mesures, nous nommons **mesures d'incohérences** les mesures au niveau des bases, et **valeurs d'incohérences** les mesures au niveau des formules.

De manière assez naturelle, si l'on dispose d'une valeur d'incohérence, permettant de définir le conflit de chaque formule, on peut utiliser cette valeur pour définir une mesure d'incohérence, en agrégeant les valeurs d'incohérence des formules de la base (en prenant leur maximum par exemple).

8. C'est-à-dire que l'on considère que les seuls tests disponibles consistent à tester la valeur d'une variable et que tous ces tests ont le même coût.

9. Comme tous les tests ont le même coup, on compare les plans de tests par la profondeur maximale de leurs branches.

Définir des mesures d'incohérence à partir de valeurs d'incohérences présente un avantage supplémentaire :

L'idée vient du constat que les mesures existantes se partagent en deux classes, ayant chacune ses inconvénients. La première classe regroupe les mesures qui tiennent compte du nombre de formules intervenant dans la production de l'incohérence (section 3.2.1). Cela permet de prendre en compte la répartition des conflits entre les différentes formules de la base, mais ne permet pas d'inspecter plus finement les conflits dans les formules. La deuxième classe permet d'inspecter plus finement les conflits au niveau des variables propositionnelles, plutôt qu'au niveau des formules, mais cela se paye par une incapacité à prendre en compte la répartition du conflit entre les différentes formules de la base (section 3.2.2). Jusqu'ici il semblait difficile de prendre en compte ces deux facteurs, et aucune des mesures proposées n'y parvenait.

Nous avons défini [32] de nouvelles mesures d'incohérence basées sur la valeur de Shapley (une notion issue de la théorie des jeux). Nous avons utilisé une idée simple : utiliser une mesure de la deuxième classe (section 3.2.2), permettant un examen fin de l'incohérence au niveau des variables propositionnelles, et l'utiliser pour définir un jeu sous forme coalitionnelle. On utilise ensuite la valeur de Shapley sur le jeu ainsi défini afin de redistribuer cette incohérence sur les différentes formules responsables du conflit. Les deux dimensions sont donc bien prises en compte dans la détermination de la valeur d'incohérence de la formule.

Nous donnons à présent quelques exemples afin d'illustrer le fait que les valeurs que l'on obtient sur des cas simples sont très intuitives. L'exemple ci-dessous est celui de la **valeur d'incohérence de Shapley** la plus simple S_{I_d}. La mesure d'incohérence associée $\hat{S}_{I_d}$ est simplement calculée à partir du maximum des valeurs d'incohérence des formules de la base.

Exemple 4
- $K_1 = \{a, \neg a, b\}$.
La valeur d'incohérence est $S_{I_d}(K_1) = (\frac{1}{2}, \frac{1}{2}, 0)$. Et la mesure d'incohérence de la base est $\hat{S}_{I_d}(K_1) = \frac{1}{2}$.
b est une formule libre, elle n'est engagée dans aucun conflit, elle a donc la valeur 0 (c'est une des propriétés de ces valeurs d'incohérence). Les deux autres formules (a et $\neg a$) sont aussi responsables l'une que l'autre du conflit, elles se partagent donc la responsabilité du conflit, leur valeur d'incohérence est donc $\frac{1}{2}$.
- $K_2 = \{a, b, b \wedge c, \neg b \wedge d\}$.
La valeur est $S_{I_d}(K_2) = (0, \frac{1}{6}, \frac{1}{6}, \frac{4}{6})$. Et la mesure d'incohérence de la base est $\hat{S}_{I_d}(K_2) = \frac{2}{3}$.
La dernière formule est clairement celle qui apporte le plus d'incohérence à la base puisque enlever simplement cette formule restaure la cohérence de la base.
- $K_4 = \{a \wedge \neg a, b, \neg b, c\}$.
La valeur est $S_{I_d}(K_4) = (\frac{4}{6}, \frac{1}{6}, \frac{1}{6}, 0)$. Et la mesure d'incohérence de la base est $\hat{S}_{I_d}(K_4) = \frac{2}{3}$.
La première formule est une contradiction, il faut donc absolument la retirer de la base si on veut supprimer l'incohérence. C'est donc cette formule qui est la plus conflictuelle. Ensuite il faut enlever soit b soit $\neg b$, ce deux formules ont

donc une responsabilité dans les conflits de la base. Ce qui n'est pas le cas de la dernière formule.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 9).

Nous avons également travaillé à la définition de **valeurs d'incohérences basées sur le calcul des ensembles minimaux incohérents** [36]. Cela se justifie par le fait que ces ensembles représentent l'essence des conflits de la base. Or, nous avons montré qu'une de ces valeurs d'incohérence était également une valeur d'incohérence de Shapley (celle de l'exemple ci-dessus). Ce résultat nous a permis de donner une caractérisation logique complète de cette mesure. C'est à notre connaissance la première caractérisation logique d'une mesure d'incohérence.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 10).

Chapitre IV

Révision

Pour atteindre la connaissance,
ajoute des choses chaque jour.
Pour atteindre la sagesse,
retire des choses chaque jour.

(Lao Tzu, Tao-te Ching, ch. 48)

Le problème posé par la révision de croyances peut être résumé par la question suivante : étant donnée une base de croyances représentant les croyances d'un agent à propos du monde et une nouvelle information à propos de ce monde, pouvant remettre en cause une partie de ses croyances, quels changements cette nouvelle information va-t-elle produire dans la base de croyances ?

Alchourrón, Gärdenfors et Makinson (AGM) ont proposé un ensemble de postulats qu'un opérateur de révision « raisonnable » doit satisfaire [AGM85].

Une critique envers le cadre AGM est qu'il ne permet pas d'assurer un bon comportement lors de l'itération du processus de révision. De nombreuses solutions ont été proposées [DP97, Bou93, Leh95, Wil94] mais aucune n'est totalement satisfaisante.

4.1. Le cadre AGM

Dans cette section, nous considérons des théories, c'est-à-dire des bases K closes pour la déduction logique (i.e. $Cn(K) = K$).

Si l'on considère une formule α, pour un agent avec les croyances K, cette formule α ne peut avoir que 3 statuts épistémiques différents :

- Soit $\alpha \in K$, c'est-à-dire que l'agent croit que α est vraie. On dit que α est acceptée par l'agent.
- Soit $\neg\alpha \in K$, c'est-à-dire que l'agent croit que α est fausse. On dit que α est

rejetée par l'agent.

- Soit $\alpha \notin K$ et $\neg\alpha \notin K$, on dit que α est indéterminée (contingente) pour l'agent.

Les opérateurs de changement de croyances peuvent alors être définis comme des transitions entre ces différents statuts, comme illustré à la figure 4.1.

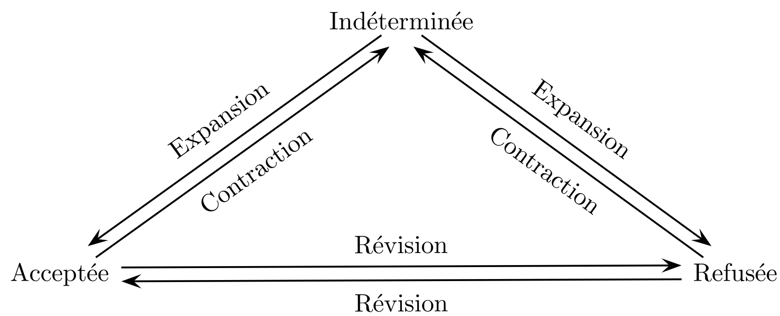


FIGURE 4.1 – Transitions entre statuts épistémiques

Lorsque la formule passe d'un statut indéterminée à acceptée (ou symétriquement refusée), cette transition est nommée **expansion**, car on ajoute simplement de l'information. La transition inverse (de acceptée/refusée à indéterminée) est appelée **contraction**, car on souhaite enlever une information des croyances de l'agent. Et lorsque l'on passe directement du statut acceptée à refusée (ou symétriquement de refusée à acceptée), la transition est nommée **révision**. Dans ce cas, on change d'avis sur la véracité d'une information.

Pour ces trois types d'opérateurs de changement de croyances, on souhaite évidemment avoir des propriétés de rationalité, c'est-à-dire que les opérateurs doivent respecter un certain nombre de propriétés garantissant qu'ils réalisent le changement attendu.

Cette approche axiomatique est réellement fondatrice de la théorie de la révision de croyances AGM. L'idée est, au lieu de définir des opérateurs particuliers, d'énumérer un ensemble de propriétés que tout opérateur de changement *raisonnable* devrait satisfaire. Puis on regarde s'il existe des opérateurs satisfaisant l'ensemble de ces propriétés et on essaye éventuellement de les caractériser, c'est-à-dire de montrer que des opérateurs satisfaisant un ensemble de propriétés donné s'expriment tous sous une certaine forme. C'est ce qu'on appellera des **théorèmes de représentation** (on peut également parler de caractérisation logique).

En ce qui concerne les opérateurs de changement de croyances, les propriétés attendues sont assez faciles à exprimer intuitivement, par 3 principes :

- **Principe de succès** (appelé aussi **primauté de la nouvelle information**) : le changement doit réussir, c'est-à-dire qu'après l'opération, l'information doit avoir le statut voulu.
- **Principe de cohérence** : on veut que la base résultante de l'opération soit une base cohérente (on veut donc éviter la trivialisation).
- **Principe de changement minimal** : on veut modifier les croyances de l'agent le moins possible, pour assurer qu'on n'élimine aucune information de manière inconsidérée, et qu'on n'ajoute aucune information non souhaitée.

Alchourrón, Gärdenfors et Makinson ont proposé une formalisation logique de ces propriétés [AGM85, Gär88]. Voyons donc à présent comment formaliser ces principes pour chacun des opérateurs de changement.

4.1.1. Postulats AGM

Expansion

Un opérateur d'**expansion** $+$ est une fonction de $\mathcal{K} \times \mathcal{L}$ vers $\mathcal{K}$ qui vérifie les propriétés suivantes :

(K+1) $K + \alpha$ est une théorie **(clôture)**

(K+2) $\alpha \in K + \alpha$ **(succès)**

(K+3) $K \subseteq K + \alpha$ **(inclusion)**

(K+4) Si $\alpha \in K$, alors $K + \alpha = K$ **(vacuité)**

(K+5) Si $K' \subseteq K$, alors $K' + \alpha \subseteq K + \alpha$ **(monotonie)**

(K+6) $K + \alpha$ est la plus petite base satisfaisant (K+1)-(K+5) **(minimalité)**

L'explication intuitive de ces axiomes est la suivante : (K+1) assure que le résultat de l'expansion est bien une théorie. (K+2) dit que la nouvelle information doit être vraie dans la nouvelle base de croyance. La motivation du nom expansion peut être expliquée par (K+3) qui certifie que l'on garde toutes les informations de l'ancienne base. (K+4) dit que si la nouvelle information appartient déjà à la base de croyance alors il n'y a rien à faire pour l'accepter. Le postulat (K+5) exprime la monotonie de l'expansion. Et le dernier postulat (K+6) exprime la minimalité du changement, c'est-à-dire qu'il s'assure que la nouvelle base de croyance ne contient pas de croyance non justifiée par l'ajout de la nouvelle information.

Il n'y a qu'un seul opérateur d'expansion :

Théorème 2 ([Gär88]) *L'opérateur d'expansion $+$ satisfait les postulats (K+1)-(K+6) ssi $K + A = Cn(K \cup A)$.*

Contraction

Un opérateur de **contraction** $\div$ est une fonction de $\mathcal{K} \times \mathcal{L}$ vers $\mathcal{K}$ qui vérifie les propriétés suivantes :

(K$\div$1) $K \div \alpha$ est une théorie **(clôture)**

(K$\div$2) $K \div \alpha \subseteq K$ **(inclusion)**

(K$\div$3) Si $\alpha \notin K$, alors $K \div \alpha = K$ **(vacuité)**

(K$\div$4) Si $\nvdash \alpha$, alors $\alpha \notin K \div \alpha$ **(succès)**

(K$\div$5) Si $\alpha \in K$, alors $K \subseteq (K \div \alpha) + \alpha$ **(restauration)**

(K$\div$6) Si $\vdash \alpha \leftrightarrow \beta$, alors $K \div \alpha = K \div \beta$ **(préservation)**

(K$\div$7) $(K \div \alpha) \cap (K \div \beta) \subseteq K \div (\alpha \wedge \beta)$ **(intersection)**

(K$\div$8) Si $\alpha \notin K \div (\alpha \wedge \beta)$, alors $K \div (\alpha \wedge \beta) \subseteq K \div \alpha$ **(conjonction)**

(K÷1) assure que le résultat de la contraction est bien une théorie. (K÷2) garantit que lors de la contraction aucune nouvelle information n'est ajoutée à la base de croyance. (K÷3) dit que si l'information α n'est pas acceptée par K, il n'y a rien à faire pour retirer α de K. Le postulat (K÷4) assure le succès de la contraction, c'est-à-dire que si α n'est pas une tautologie, alors la contraction réussit. Le postulat (K÷5) assure que la contraction de K par α suivie de l'expansion par α redonne la théorie K comme résultat (l'inclusion inverse de (K÷5) étant une conséquence de (K÷1)-(K÷4)). (K÷6) dit que le résultat de la contraction ne dépend pas de la syntaxe de l'information. Ces six postulats sont les postulats de base pour les opérateurs de contraction. (K÷7) et (K÷8) sont appelés postulats supplémentaires [1]. (K÷7) dit que si une information est à la fois dans la contraction par α et dans la contraction par β alors elle doit être dans la contraction par la conjonction $\alpha \wedge \beta$. (K÷8) exprime la minimalité du changement pour la conjonction.

Révision

Un opérateur de **révision** $*$ est une fonction de $\mathcal{K} \times \mathcal{L}$ vers $\mathcal{K}$ qui vérifie les propriétés suivantes :

(K*1) $K * \alpha$ est une théorie **(clôture)**

(K*2) $\alpha \in K * \alpha$ **(succès)**

(K*3) $K * \alpha \subseteq K + \alpha$ **(inclusion)**

(K*4) Si $\neg\alpha \notin K$, alors $K + \alpha \subseteq K * \alpha$ **(vacuité)**

(K*5) $K * \alpha = K_\perp$ ssi $\vdash \neg\alpha$ **(cohérence)**

(K*6) Si $\vdash \alpha \leftrightarrow \beta$, alors $K * \alpha = K * \beta$ **(extensionalité)**

(K*7) $K * (\alpha \wedge \beta) \subseteq (K * \alpha) + \beta$ **(inclusion conjonctive)**

(K*8) Si $\neg\beta \notin K * \alpha$, alors $(K * \alpha) + \beta \subseteq K * (\alpha \wedge \beta)$ **(vacuité conjonctive)**

L'interprétation de ces postulats est la suivante : (K*1) s'assure que le résultat de la révision est bien une théorie. (K*2) dit que la nouvelle information est vraie dans la nouvelle base de croyance. Le postulat (K*3) implique que la révision par la nouvelle information ne peut pas ajouter de croyance qui ne soit une conséquence de la nouvelle information et de la base de croyance. Et les postulats (K*3) et (K*4) ensemble signifient que, lorsque la nouvelle information n'est pas contradictoire avec l'ancienne base de croyance, alors la révision de la base de croyance se résume à l'expansion de cette base. (K*5) exprime le fait que la seule façon d'arriver à une base incohérente par une révision est de réviser par une information contradictoire. (K*6) dit que le résultat de la révision ne dépend pas de la syntaxe de la nouvelle information. Ces six postulats sont les postulats de base pour les opérateurs de révision, les deux postulats (K*7) et (K*8) ont été appelés postulats supplémentaires et expriment le bon comportement des opérateurs de révision en terme de minimalité de changement. Ils assurent que la révision par une conjonction de deux informations revient à une révision par la première

1. De fait, certains travaux étudient des opérateurs ne satisfaisant que les postulats de base. Notre opinion est que ces postulats supplémentaires sont au moins aussi importants pour un bon opérateur que les postulats de base.

information et une expansion par la seconde dès que cela est possible (i.e. dès que la seconde information ne contredit aucune croyance issue de la première révision).

Révision en logique propositionnelle

Les propriétés que nous venons de présenter sont valables pour n'importe quelle logique [2]. On peut les exprimer plus simplement dans le cadre de la logique propositionnelle. C'est ce que nous allons voir ici.

Soient deux formules propositionnelles φ et μ. L'opérateur $\circ$ est un opérateur de révision s'il vérifie les postulats suivants :

(R1) $\varphi \circ \mu \vdash \mu$

(R2) Si $\varphi \wedge \mu$ est cohérent alors $\varphi \circ \mu \equiv \varphi \wedge \mu$

(R3) Si μ est cohérent alors $\varphi \circ \mu$ est cohérent

(R4) Si $\varphi_1 \equiv \varphi_2$ et $\mu_1 \equiv \mu_2$ alors $\varphi_1 \circ \mu_1 \equiv \varphi_2 \circ \mu_2$

(R5) $(\varphi \circ \mu) \wedge \psi \vdash \varphi \circ (\mu \wedge \psi)$

(R6) Si $(\varphi \circ \mu) \wedge \psi$ est cohérent alors $\varphi \circ (\mu \wedge \psi) \vdash (\varphi \circ \mu) \wedge \psi$

Soit un opérateur de révision $*$ sur des théories et $\circ$ un opérateur de révision sur des bases de croyances propositionnelles. On dit que l'opérateur $*$ correspond à l'opérateur $\circ$ si quand $K = Cn(\varphi)$, alors $K * \alpha = Cn(\varphi \circ \alpha)$.

Théorème 3 ([KM91b]) *Soit un opérateur de révision $*$ et son opérateur $\circ$ correspondant. Alors $*$ satisfait les postulats $(K * 1) - (K * 8)$ si et seulement si $\circ$ vérifie les postulats $(R1) - (R6)$.*

Identités

Comme le suggère la figure 4.1, il est possible de décomposer la révision, qui est une transition du statut accepté au statut refusé, en une contraction, qui effectuera la transition du statut accepté vers le statut indéterminé, suivie d'une expansion, qui effectuera la transition du statut indéterminé au statut refusé. C'est ce que nous dit l'**identité de Levi** :

- $K * \alpha = (K \div \neg\alpha) + \alpha$ **(Identité de Levi)**

Théorème 4 ([Gär88]) *Si l'opérateur de contraction $\div$ satisfait (K÷1)-(K÷4) et (K÷6) et l'opérateur d'expansion $+$ satisfait (K+1)-(K+6), alors l'opérateur de révision $*$ défini par l'identité de Levi satisfait (K*1)-(K*6). De plus, si (K÷7) est satisfait, alors (K*7) est satisfait pour la révision ainsi définie, et si (K÷8) est satisfait, alors (K*8) est satisfait pour la révision ainsi définie.*

Remarque 1 *Le postulat* **recovery** *(K÷5) n'est pas nécessaire pour ce résultat.*

On peut également définir la contraction à partir de l'opérateur de révision, grâce à l'**identité de Harper** :

2. Plus exactement, n'importe quelle relation de conséquence au sens de Tarski.

- $K \div \alpha = K \cap (K * \neg\alpha)$ **(Identité de Harper)**

Pour comprendre l'idée de cette identité, supposons que α soit pour le moment acceptée par K. Le résultat de la contraction de K par α est l'ensemble des formules qui sont vraies indépendamment de l'acceptation de α, c'est-à-dire les formules qui sont vraies à la fois lorsque α est acceptée (K) et lorsque α est refusée ($K * \neg\alpha$).

Théorème 5 ([Gär88]) *Si l'opérateur de révision $*$ satisfait (K*1)-(K*6), alors l'opérateur de contraction $\div$ défini par l'identité de Harper satisfait (K÷1)-(K÷6). De plus, si (K*7) est satisfait, alors (K÷7) est satisfait pour la contraction ainsi définie, et si (K*8) est satisfait, alors (K÷8) est satisfait pour la contraction ainsi définie.*

Ces deux identités montrent un lien très étroit entre opérateurs de révision et opérateurs de contraction, qui peuvent tous deux être définis à partir de l'autre. On n'a donc pas réellement à étudier les deux types d'opérateurs, mais il suffit d'en étudier un, puis d'utiliser les identités pour définir l'autre. Le choix de l'opérateur de base est principalement une question de goût. Habituellement les logiciens et les philosophes prennent les opérateurs de contraction comme opérateurs de base, puisque l'identité de Levi présente la révision comme une composition d'une contraction suivie d'une expansion. Et la plupart des chercheurs en intelligence artificielle choisissent plutôt les opérateurs de révision comme opérateurs de base, puisque c'est l'opération dont on a le plus besoin dans les systèmes à bases de connaissances.

4.1.2. Théorèmes de représentation

Maintenant que nous avons défini ce que nous attendions des opérateurs de changement de croyances, il est temps de donner des moyens pratiques de définir ces opérateurs. C'est ce que vont nous fournir les théorèmes de représentation.

Nous ne présentons dans la suite que trois théorèmes de représentation : celui utilisant les intersections partielles, celui utilisant les enracinements épistémiques, et celui utilisant les assignements fidèles. Il existe d'autres théorèmes de représentation (voir [AM85, FH96a] par exemple).

Contraction par intersection partielle

L'idée des opérateurs de contraction par intersection est de garder le maximum de formules de l'ancienne base. On va garder l'ensemble de tous les sous-ensembles de la base qui n'impliquent pas l'information que l'on veut retirer. Et l'inférence à partir de cet ensemble sera l'inférence sceptique, c'est-à-dire que la nouvelle base de l'agent sera constituée de l'ensemble des formules que l'on pourra inférer de tous ces sous-ensembles.

Définition 9 *Soient une théorie K et une proposition α. L'ensemble des* **sous-théories maximales de K n'impliquant pas** *α, noté $K \perp \alpha$, est l'ensemble de tous les K' qui vérifient :*
- $K' \subseteq K$
- $K' \nvdash \alpha$

- *Si $K' \subset K'' \subseteq K$ alors $K'' \vdash \alpha$*

Définition 10 *La fonction de* **contraction par intersection totale** *(full meet contraction) $\div_f$ est définie comme*

$$K \div_f \alpha = \begin{cases} \bigcap(K\bot\alpha) & \text{si } K\bot\alpha \text{ n'est pas vide, et} \\ K & \text{sinon} \end{cases}$$

Le problème avec cette définition est que le résultat est beaucoup trop fort, puisque si l'on effectue une contraction (par intersection totale) par α, le résultat est l'ensemble des formules de K qui sont conséquences logiques de $\neg\alpha$. Cela a en particulier comme conséquence que :

Théorème 6 ([AM82]) *Si une fonction de révision $*$ est définie à partir d'une fonction de contraction par intersection totale au moyen de l'identité de Levi, alors pour chaque proposition α telle que $\neg\alpha \in K$, on a $K * \alpha = Cn(\alpha)$.*

C'est-à-dire que l'on oublie toute information à propos des anciennes croyances de l'agent, ce qui n'est pas très souhaitable. Le problème est qu'en gardant l'ensemble de toutes les sous-théories maximales de K n'impliquant pas α, on a retiré trop d'information.

L'idée est alors de n'en garder que certaines (les « meilleures », les « plus typiques », etc.).

Définition 11 *Soit une théorie K, une* **fonction de sélection** *γ est une fonction qui associe à chaque proposition α l'ensemble $\gamma(K\bot\alpha)$, qui est un sous-ensemble non vide de $K\bot\alpha$ si celui-ci n'est pas vide et $\gamma(K\bot\alpha) = \{K\}$ sinon.*

Définition 12 *Une fonction de* **contraction par intersection partielle** *(partial meet contraction) $\div$ est définie comme*

$$K \div \alpha = \bigcap \gamma(K\bot\alpha)$$

Enonçons à présent le théorème de représentation, qui indique que tout opérateur de contraction par intersection partielle satisfait les propriétés logiques attendues pour la contraction, et inversement que tout opérateur satisfaisant ces propriétés logiques peut être défini par un opérateur de contraction par intersection partielle.

Théorème 7 ([AGM85]) *Soit un opérateur $\div$, $\div$ est une fonction de contraction par intersection partielle si et seulement si $\div$ satisfait les postulats (K$\div$1)-(K$\div$6).*

Il manque pour le moment deux propriétés, que l'on va obtenir en contraignant un peu la fonction de sélection.

Définition 13 *Une fonction de sélection γ est* **relationnelle** *si et seulement si pour tout K il existe une relation $\leq$ sur $K \times K$ telle que*

$$\gamma(K\bot\alpha) = \{K' \in K\bot\alpha \mid K' \leq K'',\ \forall\, K'' \in K\bot\alpha\}$$

Si $\leq$ est une relation transitive alors γ est dite **relationnelle transitive***.*

Théorème 8 ([AGM85]) *Soit un opérateur $\div$, $\div$ est une fonction de contraction par intersection partielle relationnelle transitive (transitively relational partial meet contraction function) si et seulement si $\div$ satisfait les postulats (K$\div$1)-(K$\div$8).*

Avec ce résultat on voit en quoi les postulats supplémentaires (K$\div$7)-(K$\div$8) parlent de minimalité du changement. En effet, l'ajout de ces postulats implique l'existence d'une relation de plausibilité $\leq$ guidant le processus de sélection des sous-théories.

Contraction par Enracinements Epistémiques

L'idée est ici d'ordonner les formules de la base de croyances des plus importantes (fiables/crédibles) au moins importantes. Et lorsque l'on doit réaliser une contraction, prendre cet ordre en compte pour n'éliminer que les formules les moins importantes.

Soient deux formules α et β, la notation $\alpha \leq \beta$ signifie "β est au moins aussi enracinée (plausible / importante) que α". $\leq$ est un **enracinement épistémique** (*epistemic entrenchment*) s'il satisfait les propriétés suivantes [Gär88] :

(EE1) Si $\alpha \leq \beta$ et $\beta \leq \gamma$, alors $\alpha \leq \gamma$ **(transitivité)**

(EE2) Si $\alpha \vdash \beta$, alors $\alpha \leq \beta$ **(domination)**

(EE3) $\alpha \leq \alpha \wedge \beta$ ou $\beta \leq \alpha \wedge \beta$ **(conjonction)**

(EE4) Si $K \neq K_\perp$, $\alpha \notin K$ ssi $\forall \beta\ \alpha \leq \beta$ **(minimalité)**

(EE5) Si $\beta \leq \alpha\ \forall \beta$, alors $\vdash \alpha$ **(maximalité)**

Théorème 9 ([Gär88]) *Une fonction de contraction $\div$ satisfait $(K \div 1) - (K \div 8)$ si et seulement si il existe $\leq$ satisfaisant (EE1)-(EE5), où $\beta \leq \alpha$ ssi $\beta \notin K \div \alpha \wedge \beta$ ou $\vdash \alpha \wedge \beta$.*

Intuitivement ce théorème dit que si β n'est pas conséquence du résultat de la contraction par la conjonction $\alpha \wedge \beta$ c'est qu'il n'était pas plus enraciné que α.

Révision par assignements fidèles - Systèmes de sphères

L'idée est d'ordonner les interprétations de la plus plausible à la moins plausible. Donnons d'abord la définition générale en terme de systèmes de sphères donnée par Grove [Gro88].

Définition 14
- *On appelle* **monde possible** *un sous-ensemble maximal cohérent du langage et on note $\mathbf{M}_\mathcal{L}$ l'ensemble des mondes possibles du langage $\mathcal{L}$.*
- *Soit une base de croyances K. Si $K = K_\perp$ alors $[K] = \emptyset$, sinon $[K] = \{M \in \mathbf{M}_\mathcal{L} \mid K \subseteq M\}$.*
- *Soit un ensemble $S \in \mathbf{M}_\mathcal{L}$, l'ensemble $K_S = \bigcap\{M \mid M \in S\}$.*

Définition 15 *Un* **système de sphères centré sur** $[K]$ *[Gro88] est une collection de sous-ensembles $\mathbf{S}$ de $\mathbf{M}_\mathcal{L}$ qui vérifient les conditions suivantes :*

(S1) *Si $S, S' \in \mathbf{S}$, alors $S \subseteq S'$ ou $S' \subseteq S$*

(S2) $[K] \in \mathbf{S}$

(S3) *Si $S \in \mathbf{S}$, alors $[K] \subseteq S$*

(S4) $\mathbf{M}_{\mathcal{L}} \in \mathbf{S}$

(S5) *Si α est une formule et si $[\alpha]$ intersecte une sphère de $\mathbf{S}$, alors il existe une sphère minimale qui intersecte $[\alpha]$ (on note $C(\alpha) = [\alpha] \cap S_\alpha$).*

Théorème 10 ([Gro88]) *Soit une base de croyances K. Il existe un système de sphères $\mathbf{S}$ centré sur $[K]$ tel que pour toute formule α, $K * \alpha = K_{C(\alpha)}$ si et seulement si $*$ est un opérateur de révision satisfaisant $(K * 1) - (K * 8)$.*

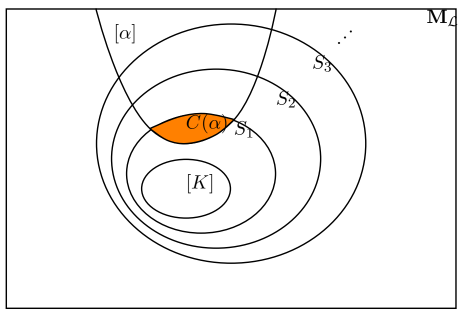


FIGURE 4.2 – Révision par α d'un système de sphères centré sur $[K]$

Graphiquement on peut représenter cela comme sur la figure 4.2 : les mondes possibles qui satisfont la base ($[K]$), sont les mondes les plus plausibles. Puis on ordonne tous les autres mondes suivant leur plausibilité (cela donne les sphères $S_1, S_2, \ldots$). Et lorsque l'on révise par une nouvelle information α, le résultat est l'ensemble des mondes possibles de cette nouvelle information qui sont les plus plausibles vis à vis de K (i.e. du système de sphères centré sur $[K]$).

Lorsque l'on travaille en logique propositionnelle finie ces systèmes de sphères peuvent être représentés par des pré-ordres totaux, ce qui conduit aux assignements fidèles de Katsuno et Mendelzon :

Définition 16 *Un* **assignement fidèle** *(faithful assignment) [KM91b] est une fonction qui associe à chaque base de croyances K un pré-ordre $\leq_K$ sur les interprétations tel que :*

- *Si $\omega \models K$ et $\omega' \models K$, alors $\omega \simeq_K \omega'$*
- *Si $\omega \models K$ et $\omega' \not\models K$, alors $\omega <_K \omega'$*
- *Si $K_1 \equiv K_2$, alors $\leq_{K_1} = \leq_{K_2}$*

Théorème 11 ([KM91b]) *Un opérateur de révision $\circ$ satisfait les postulats (R1)-(R6) si et seulement si il existe un assignement fidèle qui associe à chaque base de croyances K un pré-ordre total $\leq_K$ tel que*

$$mod(K \circ \mu) = \min(mod(\mu), \leq_K)$$

On a donc le même mécanisme, mais le système de sphères est plus simplement représenté par un pré-ordre total. Même si la représentation sous forme de systèmes de sphères est souvent utilisée pour décrire le comportement des opérateurs de révision [Rot09], nous trouvons plus pratique d'utiliser la représentation sous forme de pré-ordres totaux décrite ci-dessous. Non seulement cela évite des erreurs d'interprétation car la représentation sous forme de système de sphères, qui est planaire, peut faire croire à une spatialité qui n'existe pas. Mais surtout cela est beaucoup plus intuitif pour décrire le comportement des opérateurs de révision itérée. Nous utiliserons donc la représentation de la figure 4.3.

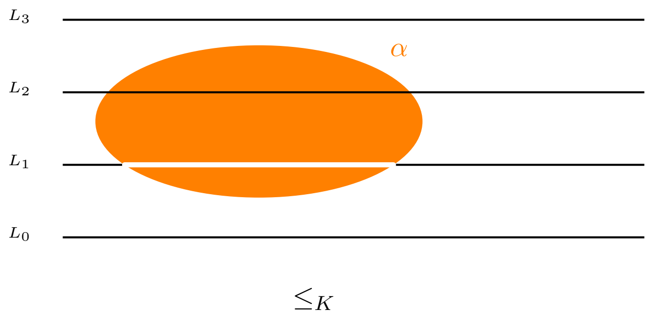


FIGURE 4.3 – Révision de K par α

Les interprétations sont réparties sur différents niveaux, deux interprétations à un même niveau sont aussi plausibles l'une que l'autre (i.e. $\omega \simeq_K \omega'$) et une interprétation ω apparaissant à un niveau inférieur à une autre ω' est strictement plus plausible [3] (i.e. $\omega <_K \omega'$). Bien entendu les interprétations apparaissant au niveau le plus bas sont les modèles de la base de croyances K.

La traduction entre systèmes de sphère et le pré-ordre associé à K par l'assignement fidèle est directe. Chaque niveau correspond à une sphère (L_0 à $[K]$, et chaque L_i à $S_i - S_{i-1}$).

Lorsque l'on révise par une nouvelle information α, le résultat est constitué de l'ensemble des modèles de α les plus plausibles suivant le pré-ordre de plausibilité $\leq_K$ associé à K par l'assignement fidèle, soient les interprétations situées sur le segment blanc sur la figure 4.3.

4.1.3. Liens avec les logiques non monotones

La théorie de la révision AGM a des liens forts avec les **logiques non monotones**, telles que définies par Makinson [Mak94], et Kraus, Lehmann, Magidor [KLM90, LM92], ce qui a fait dire à Gärdenfors [Gär90] : « *Belief revision and nonmonotonic logic are two sides of the same coin.* »

3. Contrairement à la plupart des domaines où les meilleures options sont les plus grandes, en révision de croyances les meilleures options sont souvent les plus petites (et donc on minimise au lieu de maximiser). Cela est peut-être une conséquence de l'obsession d'obtenir un changement *minimal*.

Théorème 12 ([MG89, FL94]) • *Soient un opérateur de révision $*$ et une base de croyances K. Si on définit la relation $\mid\!\sim_K$ comme suit :*

$$\alpha \mid\!\sim^*_K \beta \text{ ssi } \beta \in K * \alpha$$

Alors $\mid\!\sim^*_K$ est une relation rationnelle qui satisfait la règle suivante, nommée préservation de la cohérence :

$$\text{Si } \alpha \mid\!\sim^*_K \bot, \text{ alors } \alpha \vdash \bot$$

- *Soit $\mid\!\sim$ une relation rationnelle qui préserve la cohérence, il existe une base K et un opérateur $*$ tels que $\mid\!\sim \; = \; \mid\!\sim^*_K$.*

On peut également montrer un lien étroit entre la **logique possibiliste** et les enracinements épistémiques, et donc la contraction et la révision : soit une relation $\geq_c$ sur les formules, $\alpha \geq_c \beta$ signifie « α est au moins aussi certain que β ». $\geq_c$ est une **relation de nécessité qualitative** si elle vérifie les propriétés suivantes [Dub86, DP91] :

(D1) $\alpha \geq_c \alpha$ **(réflexivité)**

(D2) $\alpha \geq_c \beta$ ou $\beta \geq_c \alpha$ **(totalité)**

(D3) Si $\alpha \geq_c \beta$ et $\beta \geq_c \gamma$, alors $\alpha \geq_c \gamma$ **(transitivité)**

(D4) $\top >_c \bot$ **(non trivialité)**

(D5) $\top \geq_c \alpha$ **(certitude de la tautologie)**

(D6) Si $\alpha \geq_c \beta$, alors $\alpha \wedge \gamma \geq_c \beta \wedge \gamma$ **(stabilité conjonctive)**

Cette relation de nécessité qualitative entre formules est à la base de la logique possibiliste.

Théorème 13 ([DP91]) *L'ensemble d'axiomes (D1), (D2), (D3), (D5) et (D6) est équivalent à (EE1)-(EE4).*

Comme on peut le voir avec ce théorème, la différence majeure entre les enracinements épistémiques et les relations de nécessité qualitative (à la base de la logique possibiliste), est que les relations de nécessité autorisent l'existence de formules aussi prioritaires que les tautologies, et qui ne pourront donc pas être remises en question par des révisions, ce qui n'est pas possible dans le cas des enracinements épistémiques (comparer $(D5)$ et $(EE5)$). On peut donc voir ces formules comme des contraintes d'intégrité, ou des connaissances (par oppositions aux croyances « usuelles » qui, elles, peuvent être remises en doute).

4.2. Révision itérée

Il est assez facile de montrer que la caractérisation AGM n'est pas suffisante pour modéliser l'itération du processus de révision. Des propriétés supplémentaires sont nécessaires afin de contraindre le comportement des opérateurs de révision itérée. Il est également nécessaire de passer à une représentation des croyances plus riche qu'une simple base de croyances. On utilise donc ce que l'on appelle des **états épistémiques** Ψ. A chaque état épistémique Ψ correspond une base, notée $Bel(\Psi)$, représentant les croyances actuelles de l'agent. Ψ encode également d'autres informations sur la plausibilité des formules que l'agent ne croit pas actuellement. Pour simplifier les notations dans la suite, lorsqu'un état épistémique sera utilisé dans une formule logique, il s'agira de la base de croyances associée, ainsi on écrira par exemple $\Psi \vdash \mu$, $\Psi \wedge \mu$, $\omega \models \Psi$ au lieu respectivement de $Bel(\Psi) \vdash \mu$, $Bel(\Psi) \wedge \mu$, $\omega \models Bel(\Psi)$

On peut alors facilement réécrire les postulats de Katsuno et Mendelzon dans le cadre des états épistémiques [DP97] :

(R*1) $\Psi \circ \mu \vdash \mu$

(R*2) Si $\Psi \wedge \mu$ est cohérent, alors $\Psi \circ \mu \equiv \Psi \wedge \mu$

(R*3) Si μ est cohérent, alors $\Psi \circ \mu$ est cohérent

(R*4) Si $\Psi_1 = \Psi_2$ et $\mu_1 \equiv \mu_2$, alors $\Psi_1 \circ \mu_1 \equiv \Psi_2 \circ \mu_2$

(R*5) $(\Psi \circ \mu) \wedge \varphi \vdash \Psi \circ (\mu \wedge \varphi)$

(R*6) Si $(\Psi \circ \mu) \wedge \varphi$ est cohérent, alors $\Psi \circ (\mu \wedge \varphi) \vdash (\Psi \circ \mu) \wedge \varphi$

On peut alors ajouter d'autres postulats afin de contraindre le comportement des opérateurs lors d'itérations (remarquons que ces postulats, contrairement aux précédents, font apparaître deux opérations de révision successives).

(C1) Si $\alpha \vdash \mu$, alors $(\Psi \circ \mu) \circ \alpha \equiv \Psi \circ \alpha$

(C2) Si $\alpha \vdash \neg\mu$, alors $(\Psi \circ \mu) \circ \alpha \equiv \Psi \circ \alpha$

(C3) Si $\Psi \circ \alpha \vdash \mu$, alors $(\Psi \circ \mu) \circ \alpha \vdash \mu$

(C4) Si $\Psi \circ \alpha \nvdash \neg\mu$, alors $(\Psi \circ \mu) \circ \alpha \nvdash \neg\mu$

L'explication des postulats est la suivante : (C1) dit que si deux informations sont incorporées successivement et si la deuxième implique la première, alors incorporer seulement la seconde donne la même base. (C2) dit que lorsque deux informations contradictoires arrivent, la seconde seule donnerait le même base. (C3) dit qu'une information doit être gardée si l'on effectue une révision par une information qui, étant donnée la base, implique la première. (C4) dit qu'aucune information ne peut contribuer à son propre rejet.

Dans [DP94] les postulats (C1)-(C4) ont d'abord été donnés comme complément aux postulats usuels (R1)-(R6) [KM91b]. Freund et Lehmann [FL94] ont montré que (C2) est incompatible avec les postulats AGM. De plus Lehmann [Leh95] a montré que les postulats (C1) et (R1)-(R6) impliquent (C3) et (C4). Dans [DP97], Darwiche et Pearl ont reformulé leurs postulats (et les postulats AGM) en termes d'états épistémiques ((R*1)-(R*6)) et ont de ce fait enlevé cette contradiction et ces redondances.

On peut alors montrer les théorèmes de représentation suivants [4], dont le premier est une simple généralisation du théorème de Katsuno et Mendelzon :

4. Les assignements fidèles sont généralisés pour les états épistémiques, ils deviennent donc des fonctions de l'ensemble des états épistémiques vers les pré-ordres, avec : 1. Si $\omega \models \Psi$ et $\omega' \models \Psi$, alors $\omega \simeq_\Psi \omega'$ 2. Si $\omega \models \Psi$ et $\omega' \not\models \Psi$, alors $\omega <_\Psi \omega'$ 3. Si $\Psi_1 = \Psi_2$, alors $\leq_{\Psi_1} = \leq_{\Psi_2}$

Théorème 14 ([DP97]) *Un opérateur de révision* $\circ$ *satisfait les postulats (R*1)-(R*6) si et seulement si il existe un assignement fidèle qui associe à tout état épistémique* Ψ *un pré-ordre total sur les interprétations* $\leq_\Psi$ *tel que :*

$$mod(\Psi \circ \mu) = \min(mod(\mu), \leq_\Psi)$$

Le second est plus intéressant car il ajoute les contraintes concernant l'itération :

Théorème 15 ([DP97]) *Soit un opérateur de révision qui vérifie (R*1)-(R*6). L'opérateur vérifie (C1)-(C4) si et seulement si l'opérateur et l'assignement fidèle correspondant vérifient :*

(CR1) *Si* $\omega \models \mu$ *et* $\omega' \models \mu$*, alors* $\omega \leq_\Psi \omega'$ *ssi* $\omega \leq_{\Psi\circ\mu} \omega'$

(CR2) *Si* $\omega \models \neg\mu$ *et* $\omega' \models \neg\mu$*, alors* $\omega \leq_\Psi \omega'$ *ssi* $\omega \leq_{\Psi\circ\mu} \omega'$

(CR3) *Si* $\omega \models \mu$ *et* $\omega' \models \neg\mu$*, alors* $\omega <_\Psi \omega'$ *seulement si* $\omega <_{\Psi\circ\mu} \omega'$

(CR4) *Si* $\omega \models \mu$ *et* $\omega' \models \neg\mu$*, alors* $\omega \leq_\Psi \omega'$ *seulement si* $\omega \leq_{\Psi\circ\mu} \omega'$

Ce théorème de représentation est important car il signifie que l'on peut considérer les opérateurs de révision itérée comme des fonctions de transition entre pré-ordres totaux (avec les contraintes données par (CR1-CR4)), et donc que l'on peut considérer les pré-ordres totaux comme la représentation canonique des états épistémiques, puisque le théorème de représentation exprime le fait que quelle que soit la représentation exacte des états épistémiques, il est possible de modéliser leur comportement au travers d'un assignement fidèle [5].

Boutilier propose [Bou93, Bou96] un opérateur de **révision naturelle**. Cet opérateur est un cas particulier d'**opérateur de Darwiche et Pearl**. Cet opérateur est le seul à satisfaire les propriétés (R*1)-(R*6) et la propriété de **minimisation absolue** (CB) suivante :

(CB) Si $\Psi \circ \psi \vdash \neg\mu$, alors $(\Psi \circ \psi) \circ \mu \equiv \Psi \circ \mu$

Il peut être considéré comme accomplissant un changement minimal dans le pré-ordre associé aux états épistémiques. Malheureusement cette minimalisation du changement se paye par un mauvais comportement de l'opérateur, puisqu'il induit une conditionnalisation des informations successives, comme illustré sur l'exemple suivant [DP97] :

Exemple 5 *Je rencontre un étrange animal qui semble être un oiseau, je crois donc que cet animal est un oiseau. Cet animal se rapproche et je vois que cet animal est rouge, je pense donc que cet animal est rouge. Un expert en animaux étranges passe par là et m'apprend qu'en fait cet animal n'est pas un oiseau mais un mammifère. Je révise donc mes croyances et pense à présent que cet animal est un mammifère. Que dois-je croire à propos de la couleur de l'animal ?*

D'après l'opérateur de révision naturelle, on ne peut plus croire que l'animal est rouge (il suffit de prendre $Bel(\Psi) = oiseau$, $\mu = \neg oiseau$ et $\psi = rouge$). Bien que la couleur et l'espèce de l'animal ne soient a priori absolument pas liées, l'ordre dans

5. Voir [BLP05] pour une généralisation de ce cadre aux pré-ordres partiels.

lequel on apprend ces informations les « conditionne » en quelque sorte pour l'opérateur de révision naturelle, car la croyance « l'animal est rouge » dépend de la croyance « l'animal est un oiseau », puisque remettre en cause cette dernière invalide la première. Alors que si l'on avait appris la couleur de l'animal avant d'avoir les informations sur son espèce, on aurait maintenu cette information sur sa couleur.

Il est un peu problématique que la caractérisation de Darwiche et Pearl autorise un tel opérateur. Il a donc été proposé de définir des opérateurs de révisions admissibles [BM06], afin de l'éliminer. Ces opérateurs sont définis par un postulat d'itération supplémentaire [BM06, JT07] :

(P) Si $\Psi \circ \alpha \not\vdash \neg\mu$ alors $\Psi \circ \mu \circ \alpha \vdash \mu$

Définition 17 *Un opérateur de révision est admissible [BM06] si il satisfait (R*1-R*6)* [6]*, (C1), (C2) et (P).* [7]

Et le théorème de représentation correspondant est :

Théorème 16 ([BM06, JT07]) *Soit un opérateur de révision $\circ$ qui vérifie (R*1)-(R*6). L'opérateur $\circ$ vérifie (P) si et seulement si l'opérateur et l'assignement fidèle correspondant vérifient (CP) :*

(CP) *Si $\omega \models \mu$ et $\omega' \models \neg\mu$, alors $\omega \leq_\Psi \omega'$ seulement si $\omega <_{\Psi\circ\mu} \omega'$*

Beaucoup d'opérateurs particuliers ont été proposés. On peut se référer à [Rot09] [13] pour une présentation des principales approches.

4.3. Itération dans le cadre AGM

Nous avons étudié quelles propriétés d'itération peuvent être exprimées dans le cadre AGM usuel [25]. En effet, le saut représentationnel des bases de croyances vers les états épistémiques a été provoqué par les preuves d'incohérences des postulats de Darwiche et Pearl avec le cadre AGM usuel. Mais il n'existait pas de travail examinant quelles sont les propriétés d'itération possibles dans le cadre AGM. Ces travaux constituent donc une justification *a posteriori* du saut représentationnel effectué par la communauté pour étudier la révision itérée.

Nous avons en particulier mis en évidence les implications des axiomes de Darwiche et Pearl dans le cadre AGM usuel. Nous avons également proposé des propriétés très faibles pour l'itération et exploré leur conséquences. Il se trouve que les résultats obtenus sont négatifs, donnant des résultats d'impossibilité de l'itération dans le cadre AGM usuel.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 12).

6. Dans [BM06] les auteurs renforcent un peu le postulat (R*4).
7. On peut noter que (C3) et (C4) sont conséquences de ces postulats.

4.4. Révision itérée et états épistémiques

Toutes les propositions pour la révision itérée utilisent donc plus ou moins formellement une notion d'état épistémique, c'est-à-dire qu'elles travaillent sur des objets plus complexes qu'une simple base de croyances. Nous avons montré [1, 13] que ces propositions utilisent le même type d'états épistémiques sous différentes sémantiques. Mais, même dans la proposition de Darwiche et Pearl [DP97], qui semble la plus aboutie, la distinction entre la syntaxe et la sémantique de ces états épistémiques n'est pas nette, ce qui conduit à des confusions. Nous avons proposé une syntaxe pour les états épistémiques, définissant ceux-ci comme les états librement engendrés à partir d'un état initial vierge et d'un constructeur. C'est-à-dire que, syntaxiquement, on peut considérer qu'un état épistémique est l'ensemble des informations reçues par l'agent (cette idée était déjà présente dans [Leh95], puisque ses postulats sont exprimés sur des suites de formules). Nous avons ensuite défini une notion de modèle sémantique associée à cette syntaxe et nous avons montré que cela permet de capturer le cadre AGM classique mais également les propositions pour la révision itérée [1, 13]. Nous avons également proposé de nouveaux opérateurs de **révision itérée à mémoire** qui se définissent par la suite d'informations reçues [21], puis des **opérateurs dynamiques** où la notion d'état épistémique est généralisée [44]. Nous avons enfin étudié comment définir des opérateurs de révision itérée où l'on ne révise plus par une simple formule, mais par un état épistémique [19].

4.5. Opérateurs d'amélioration

Les opérateurs de révision AGM [AGM85, Gär88] usuels satisfont le principe de primauté de la nouvelle information [8]. C'est-à-dire que la nouvelle information est toujours acceptée (crue par l'agent) après la révision. Cela est une des hypothèses de base de la révision, parfaitement justifiée dans beaucoup de cas. Mais on pourrait souhaiter que l'impact de cette nouvelle information soit moins important sur les croyances de l'agent. Il y a très peu de travaux dans cette optique, souvent appelée **révision non prioritaire** (voir le numéro spécial de *Theoria* pour un bon aperçu [Han98]), et la plupart du temps cela revient à mettre la nouvelle information au même niveau que les croyances actuelles de l'agent, menant à des problèmes de fusion d'informations. Nous avons défini [35, 58, 73] une nouvelle famille d'opérateurs de changement, que nous avons appelés **opérateurs d'amélioration** (*improvement*), qui ont un comportement plus mesuré que les opérateurs de révision. Lorsque l'on incorpore une nouvelle information, cela améliore la crédibilité de cette information, mais l'agent ne l'accepte pas forcément. Il faudra itérer l'incorporation un certain nombre de fois pour assurer l'acceptation. Ce type d'opérateur de changement est parfaitement raisonnable. Les opérateurs de révision itérée usuels [DP97, BM06, JT07] sont un cas particulier de notre cadre. Notre caractérisation est donc la plus générale pour modéliser le changement itéré. Techniquement la difficulté de la définition de ces opérateurs réside dans le fait que leur caractérisation nécessite d'utiliser la définition de l'opérateur de révision

8. Cela est formalisé par le postulat (R*1).

associé (obtenu avec un nombre suffisant d'itérations identiques). Il nous semble que c'est à cause de cette difficulté que ces opérateurs très naturels n'ont pas été caractérisés jusque là.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 11).

Chapitre V

FUSION

Ce qui est contraire est utile et c'est de ce qui est en lutte que naît la plus belle harmonie ; tout se fait par discorde.

(Héraclite)

La majeure partie de nos travaux concerne la fusion de bases de croyances. Cette section sera donc un peu plus longue que les autres. Nous aborderons tout d'abord la caractérisation logique de la fusion, puis les différentes familles d'opérateurs qui ont été proposées. Nous discutons ensuite le problème de la manipulation de la fusion. Enfin nous présentons des travaux de fusion d'informations plus structurées qu'en logique classique.

5.1. Caractérisation de la fusion

Nous avons proposé [17, 18, 2] une caractérisation logique des opérateurs de fusion, qui est une généralisation du cadre AGM pour la révision [AGM85, Gär88] et qui prend en compte l'aspect agrégation des croyances.

Nous avons défini deux sous-classes d'opérateurs, les opérateurs majoritaires et les opérateurs d'arbitrage. Nous avons donné un théorème de représentation en termes de pré-ordres sur les interprétations. Outre le fait que ce théorème fournit une sémantique à ces opérateurs et en donne une définition plus constructive, il permet également d'étudier plus finement [6] les rapports entre la fusion, la théorie de la décision [Sav54] et la théorie du choix social [Arr63].

Définition 18 *$\triangle : \mathcal{E} \times \mathcal{L} \mapsto \mathcal{K}$ est un opérateur de* **fusion contrainte** *si et seulement si il satisfait les propriétés suivantes :*

(IC0) $\triangle_\mu(E) \vdash \mu$

(IC1) *Si μ est cohérent, alors $\triangle_\mu(E)$ est cohérent*

(IC2) *Si E est cohérent avec μ, alors $\triangle_\mu(E) = \bigwedge E \wedge \mu$*

(IC3) *Si $E_1 \equiv E_2$ et $\mu_1 \equiv \mu_2$, alors $\triangle_{\mu_1}(E_1) \equiv \triangle_{\mu_2}(E_2)$*

(IC4) *Si $K \vdash \mu$ et $K' \vdash \mu$, alors $\triangle_\mu(K \sqcup K') \wedge K \nvdash \bot \Rightarrow \triangle_\mu(K \sqcup K') \wedge K' \nvdash \bot$*

(IC5) $\triangle_\mu(E_1) \wedge \triangle_\mu(E_2) \vdash \triangle_\mu(E_1 \sqcup E_2)$

(IC6) *Si $\triangle_\mu(E_1) \wedge \triangle_\mu(E_2)$ est cohérent, alors $\triangle_\mu(E_1 \sqcup E_2) \vdash \triangle_\mu(E_1) \wedge \triangle_\mu(E_2)$*

(IC7) $\triangle_{\mu_1}(E) \wedge \mu_2 \vdash \triangle_{\mu_1 \wedge \mu_2}(E)$

(IC8) *Si $\triangle_{\mu_1}(E) \wedge \mu_2$ est cohérent, alors $\triangle_{\mu_1 \wedge \mu_2}(E) \vdash \triangle_{\mu_1}(E) \wedge \mu_2$*

Une partie de ces propriétés avait déjà été proposée par Revesz [Rev97] pour caractériser ce qu'il appelle les opérateurs d'**adéquation sémantique** (*model fitting*).

La signification intuitive de ces propriétés est la suivante : (IC0) assure que le résultat de la fusion satisfait les contraintes d'intégrité. (IC1) dit que si les contraintes d'intégrité sont cohérentes alors le résultat de la fusion est cohérent, c'est-à-dire que l'on peut toujours extraire des croyances cohérentes du groupe d'agents. (IC2) demande que, lorsque c'est possible, le résultat de la fusion soit simplement la conjonction des bases de croyances et des contraintes d'intégrité. Donc, lorsqu'il n'y a pas de conflit entre les agents et les contraintes, la fusion est simplement l'union des différentes croyances. (IC3) est le principe d'indépendance de syntaxe, c'est-à-dire que le résultat de la fusion ne dépend pas de la forme syntaxique des croyances mais simplement des opinions exprimées. (IC4) est la propriété d'équité. Elle assure que lorsque l'on fusionne l'opinion de deux agents, l'opérateur ne peut pas donner de préférence à l'un d'eux. (IC5) exprime l'idée suivante : si un groupe E_1 se met d'accord sur un ensemble d'alternatives qui contient l'alternative A, et si un autre groupe E_2 se met d'accord sur un autre ensemble d'alternatives qui contient également A, alors si l'on joint les deux groupes A fera encore partie des alternatives acceptables. Et (IC5) et (IC6) ensemble, expriment le fait que, dès que l'on peut trouver deux sous-groupes qui s'accordent sur au moins une alternative, alors le résultat de la fusion sera exactement l'ensemble des alternatives sur lesquelles ces deux groupes s'accordent. (IC7) et (IC8) sont une généralisation directe des postulats (R5) et (R6) de la révision de croyances (voir section 4.1.1). Ils expriment des conditions sur les conjonctions de contraintes d'intégrité et s'assurent de ce fait que la notion de *proximité* est bien fondée. C'est-à-dire, par exemple, que si une alternative A est préférée parmi un ensemble d'alternatives possibles et si on restreint le nombre d'alternatives possibles tout en gardant l'alternative A, celle-ci sera toujours préférée parmi les alternatives restantes. Cette propriété est assez usuelle dans les différentes théories du choix (décision, choix social, etc.).

Nous allons à présent définir deux sous classes d'opérateurs de fusion, les opérateurs de fusion majoritaires et les opérateurs d'arbitrage.

Un opérateur de **fusion majoritaire** est un opérateur de fusion contrainte qui satisfait la propriété suivante :

(Maj) $\exists n \ \triangle_\mu (E_1 \sqcup E_2^n) \vdash \triangle_\mu(E_2)$

Ce postulat exprime le fait que si une opinion a une large audience, ce sera alors l'opinion du groupe. On peut remarquer que cette propriété est trés générale. Elle ne dit pas exactement le nombre de répétitions nécessaires d'un profil pour s'imposer (cela dépend de l'opérateur), mais elle impose l'existence d'un tel seuil. Les opérateurs de fusion majoritaire tentent donc de satisfaire au mieux le groupe dans son ensemble. D'un autre côté, les opérateurs d'arbitrage tentent de satisfaire chacun des éléments du groupe pris individuellement du mieux possible. Un **opérateur d'arbitrage** est un opérateur de fusion contrainte qui satisfait la propriété suivante :

$$\textbf{(Arb)} \quad \left.\begin{array}{l} \triangle_{\mu_1}(K_1) \equiv \triangle_{\mu_2}(K_2) \\ \triangle_{\mu_1 \leftrightarrow \neg\mu_2}(\{K_1, K_2\}) \equiv (\mu_1 \leftrightarrow \neg\mu_2) \\ \mu_1 \nvdash \mu_2 \\ \mu_2 \nvdash \mu_1 \end{array}\right\} \Rightarrow \triangle_{\mu_1 \vee \mu_2}(\{K_1, K_2\}) \equiv \triangle_{\mu_1}(K_1)$$

Ce postulat dit que si un ensemble d'alternatives préférées sous un ensemble de contraintes d'intégrité μ_1 pour une base de croyances K_1 correspond à l'ensemble des alternatives préférées par la base K_2 sous les contraintes μ_2, et si les alternatives qui n'appartiennent qu'à un des deux ensembles de contraintes d'intégrité sont toutes aussi crédibles pour le groupe ($\{K_1, K_2\}$), alors les alternatives préférées pour le groupe parmi la disjonction des deux ensembles de contraintes seront celles préférées par chacune des bases sous leurs contraintes respectives. Ce postulat est bien plus intuitif lorsqu'il est exprimé sous la forme d'assignement syncrétique (voir condition 8). Il exprime le fait que ce sont les alternatives médianes qui sont favorisées.

A présent que nous disposons d'une définition logique des opérateurs de fusion contrainte, nous allons donner un théorème de représentation qui permet de définir ces opérateurs de manière bien plus intuitive. Ce théorème montre qu'un opérateur de fusion contrainte correspond à une famille de pré-ordres sur les interprétations.

Définition 19 *Un* **assignement syncrétique** *est une fonction qui associe à chaque profil E un pré-ordre $\leq_E$ sur les interprétations telle que pour tous profils E, E_1, E_2 et pour toutes bases K, K' les conditions suivantes sont satisfaites :*

1. *Si $\omega \models \bigwedge E$ et $\omega' \models \bigwedge E$, alors $\omega \simeq_E \omega'$*
2. *Si $\omega \models \bigwedge E$ et $\omega' \not\models \bigwedge E$, alors $\omega <_E \omega'$*
3. *Si $E_1 \equiv E_2$, alors $\leq_{E_1} = \leq_{E_2}$*
4. *$\forall \omega \models K\ \exists \omega' \models K'\ \omega' \leq_{K \sqcup K'} \omega$*
5. *Si $\omega \leq_{E_1} \omega'$ et $\omega \leq_{E_2} \omega'$, alors $\omega \leq_{E_1 \sqcup E_2} \omega'$*
6. *Si $\omega <_{E_1} \omega'$ et $\omega \leq_{E_2} \omega'$, alors $\omega <_{E_1 \sqcup E_2} \omega'$*

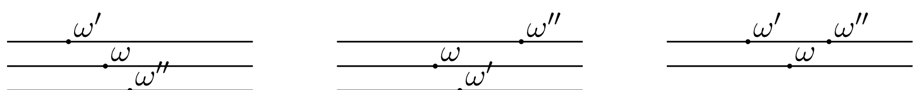
Un **assignement syncrétique majoritaire** *est un assignement syncrétique qui satisfait la condition suivante :*

7. *Si $\omega <_{E_2} \omega'$, alors $\exists n\ \omega <_{E_1 \sqcup E_2^n} \omega'$*

Un **assignement syncrétique juste** *est un assignement syncrétique qui satisfait la condition suivante :*

8. $\left.\begin{array}{l} \omega <_{K_1} \omega' \\ \omega <_{K_2} \omega'' \\ \omega' \simeq_{K_1 \sqcup K_2} \omega'' \end{array}\right\} \Rightarrow \omega <_{K_1 \sqcup K_2} \omega'$

La condition 1 dit que deux modèles du profil sont équivalents pour le pré-ordre associé et la condition 2 dit qu'un modèle du profil est toujours préféré à un contre-modèle. La condition 3 dit que si deux profils sont équivalents alors les deux pré-ordres associés sont équivalents. Ces trois conditions sont une généralisation des conditions de l'assignement fidèle pour les opérateurs de révision [KM91b]. La condition 4 dit que pour le pré-ordre associé à un profil composé de deux bases de croyances, pour chaque modèle de l'une, il existe un modèle de l'autre qui est au moins aussi bon. La condition 5 dit que si une interprétation est au moins aussi bonne qu'une autre pour un profil E_1, et que cette interprétation est également au moins aussi bonne pour un profil E_2, alors elle sera au moins aussi bonne que l'autre pour la réunion des deux profils. La condition 6 renforce un peu ce résultat en exigeant que si une interprétation est strictement meilleure qu'une autre pour un profil E_1, et que cette interprétation est au moins aussi bonne pour un profil E_2, alors cette interprétation doit être strictement meilleure que l'autre pour la réunion des deux profils. La condition 7 dit que si l'on répète un groupe E_2 suffisamment de fois alors les préférences strictes de ce groupe seront respectées. La condition 8 dit que ce sont les choix médians qui sont préférés pour le groupe. Ce comportement est illustré figure 5.1 (les interprétations les plus basses sont les interprétations préférées, par exemple pour K_1 : $\omega'' <_{K_1} \omega <_{K_1} \omega'$). L'interprétation ω, qui n'est jamais aussi mauvaise que ω' et ω'' est préférée à celles-ci pour le résultat de la fusion.



(a) K_1 (b) K_2 (c) $K_1 \sqcup K_2$

FIGURE 5.1 – Arbitrage

Nous pouvons à présent énoncer le théorème de représentation pour les opérateurs de fusion contrainte :

Théorème 17 ([18, 77]) *Un opérateur $\triangle$ est un opérateur de fusion contrainte (respectivement un opérateur majoritaire ou un opérateur d'arbitrage) si et seulement si il existe un assignement syncrétique (respectivement un assignement syncrétique majoritaire ou un assignement syncrétique juste) qui associe à chaque profil E un pré-ordre total $\leq_E$ tel que*

$$mod(\triangle_\mu(E)) = \min(mod(\mu), \leq_E)$$

Voir [CP06] pour une généralisation de ce théorème dans le cas infini.

5.2. Familles d'opérateurs de fusion

Nous allons présenter dans cette section un panorama des principales familles d'opérateurs de fusion de la litterature.

5.2.1. Opérateurs à base de modèles

Nous avons donné une définition générale des opérateurs de fusion à base de modèles, paramétrés par une distance et une fonction d'agrégation. Les opérateurs étudiés par [Rev97, LM99] sont des cas particuliers utilisant la distance de Hamming et les fonctions d'agrégation Σ ou $\max$. Nous avons montré que les propriétés de ces opérateurs étaient les mêmes quelle que soit la distance utilisée. Nous avons également proposé l'utilisation de la fonction d'agrégation GMAX (leximax) [17, 2] et les fonctions utilisant la somme des puissances, qui permettent de choisir le degré de consensualité de l'opérateur [6, 23].

Nous allons rapidement rappeler la définition de ces opérateurs.

Définition 20 *Une* **(pseudo-)distance** [1] *entre interprétations est une fonction $d : \mathcal{W} \times \mathcal{W} \mapsto \mathbb{R}^+$ telle que pour tout $\omega, \omega' \in \mathcal{W}$:*

- *$d(\omega, \omega') = d(\omega', \omega)$, et*
- *$d(\omega, \omega') = 0$ si et seulement si $\omega = \omega'$.*

Définition 21 *Une* **fonction d'agrégation** *f est une fonction qui associe un réel positif à tout n-uplet ple fini de réels positifs tel que pour tout $x_1, \ldots, x_n, x, y \in \mathbb{R}^+$:*

- *si $x \leq y$, alors $f(x_1, \ldots, x, \ldots, x_n) \leq f(x_1, \ldots, y, \ldots, x_n)$ (monotonie)*
- *$f(x_1, \ldots, x_n) = 0$ si et seulement si $x_1 = \ldots = x_n = 0$ (minimalité)*
- *$f(x) = x$ (identité)*

Définition 22 *Soit une distance entre interprétations d et une fonction d'agrégation f. L'opérateur de* **fusion à base de modèles** *$\triangle^{d,f}$ est défini par :*

$$mod(\triangle_{\mu}^{d,f}(E)) = \min(mod(\mu), \leq_E)$$

où le pré-ordre $\leq_E$ sur $\mathcal{W}$ est défini par :

- *$\omega \leq_E \omega'$ si et seulement si $d(\omega, E) \leq d(\omega', E)$, où*
- *$d(\omega, E) = f(d(\omega, K_1) \ldots, d(\omega, K_n))$, où $E = \{K_1, \ldots, K_n\}$.*

Si la fonction d'agrégation f a de bonnes propriétés, comme les fonctions usuelles (comme le maximum, la somme, le leximax, la somme des puissances n^{iemes}, le leximin), les opérateurs à base de modèles générés (quelle que soit la distance) sont des opérateurs de fusion contrainte (ils satisfont toutes les propriétés (IC0-IC8)).

Plus exactement on a les résultats suivants [5] :

Théorème 18 *Soit une distance entre interprétations d et une fonction d'agrégation f, l'opérateur $\triangle^{d,f}$ satisfait les propriétés (IC0), (IC1), (IC2), (IC7) et (IC8).*

Théorème 19 *Soit une distance entre interprétations d et une fonction d'agrégation f, l'opérateur $\triangle^{d,f}$ satisfait les propriétés (IC0-IC8) si et seulement si la fonction d'agrégation f satisfait les propriétés suivantes :*

- *Pour toute permutation σ, $f(x_1, \ldots, x_n) = f(\sigma(x_1, \ldots, x_n))$ (symétrie)*
- *Si $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$, alors $f(x_1, \ldots, x_n, z) \leq f(y_1, \ldots, y_n, z)$. (composition)*
- *Si $f(x_1, \ldots, x_n, z) \leq f(y_1, \ldots, y_n, z)$, alors $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$. (décomposition)*

1. L'inégalité triangulaire n'est pas nécessaire.

5.2.2. Opérateurs à base de formules

Les opérateurs à base de formules sont des opérateurs syntaxiques, c'est-à-dire que le résultat dépend de la présentation syntaxique des bases de croyances intervenant dans la fusion. Lorsque les bases de croyances sont des ensembles de formules, les opérateurs de fusion à base de formules usuels sélectionnent, dans l'union des bases, des sous-ensembles maximaux cohérents de formules [BKMS92, BDL+98]. Ces méthodes ont l'inconvénient de perdre l'origine des informations et ne permettent donc pas de tenir compte de la distribution de l'information pour la fusion. Ceci est très gênant pour tenir compte de la majorité par exemple.

Nous avons donc proposé [20], après avoir étudié les propriétés logiques de ces opérateurs, l'utilisation de fonctions de sélection, inspirées des *transitively relational partial meet contractions functions* (voir section 4.1.2) dans le cadre de la révision, pour tenir compte de cette distribution. Nous avons montré que cela permet d'obtenir des opérateurs avec de meilleures propriétés logiques et donc un meilleur comportement. Ces résultats illustrent l'utilité de la caractérisation logique des opérateurs. Cette caractérisation permet de classer les différents opérateurs existants selon les propriétés satisfaites ou non, et de proposer des améliorations pour certains de ces opérateurs.

La définition formelle de ces opérateurs est la suivante :

Définition 23 *Soit* $\textsc{Maxcons}(K, \mu)$ *l'ensemble des* **maxcons** *de* $K \cup \{\mu\}$ *qui contiennent* μ*, i.e., les sous-ensembles maximaux (pour l'inclusion ensembliste) de* $K \cup \{\mu\}$ *qui contiennent* μ*. Formellement,* $\textsc{Maxcons}(K, \mu)$ *est l'ensemble de tous les* M *tels que :*

- $M \subseteq K \cup \{\mu\}$,
- $\mu \in M$,
- *si* $M \subset M' \subseteq K \cup \{\mu\}$, *alors* $M' \vdash \bot$.

Soit $\textsc{Maxcons}(E, \mu) = \textsc{Maxcons}(\bigcup_{K_i \in E} K_i, \mu)$*. Lorsque la maximalité des ensembles est défini en termes de cardinalité on utilisera l'indice « card », i.e. on notera l'ensemble* $\textsc{Maxcons}_{card}(E, \mu)$*.*

On peut alors définir les opérateurs de **fusion à base de formules** suivants :

Définition 24 *Soit un profil* E *et une formule* μ *:*
$\triangle_\mu^{C1}(E) = \bigvee \textsc{Maxcons}(E, \mu)$.
$\triangle_\mu^{C3}(E) = \bigvee\{M : M \in \textsc{Maxcons}(E, \top)$ *et* $M \cup \{\mu\}$ *cohérent*$\}$.
$\triangle_\mu^{C4}(E) = \bigvee \textsc{Maxcons}_{card}(E, \mu)$.
$\triangle_\mu^{C5}(E) = \bigvee\{M \cup \{\mu\} : M \in \textsc{Maxcons}(E, \top)$ *et* $M \cup \{\mu\}$ *cohérent*$\}$
si cet ensemble est non vide et μ *sinon.*

Les opérateurs $\triangle_\mu^{C1}(E)$, $\triangle_\mu^{C3}(E)$ et $\triangle_\mu^{C4}(E)$ correspondent respectivement aux opérateurs $Comb1(E, \mu)$, $Comb3(E, \mu)$ et $Comb4(E, \mu)$ définis dans [BKMS92]. L'opérateur $\triangle^{C5}$ est une légère modification de $\triangle^{C3}$ afin d'obtenir de meilleures propriétés logiques [20].

Ces opérateurs effectuent une union des bases, puis tentent d'extraire une information cohérente à partir de cette union incohérente. C'est-à-dire qu'ils utilisent les approches d'inférence à partir de base incohérente décrites à la section 3.1.1. Du point de vue de la fusion c'est assez insatisfaisant car on n'utilise pas les informations sur la localisation des informations parmi les différentes bases, c'est-à-dire justement ce qui différencie la fusion de l'inférence en présence d'incohérence.

Nous avons appelé ces opérateurs **opérateurs de combinaison** [2], pour les différencier des **opérateurs de fusion**.

Nous avons également proposé [20] de ne sélectionner que les maxcons qui satisfont au mieux un critère de fusion. Ces fonctions de sélections sont inspirées de celles utilisées dans le cadre de la révision pour la définition des *partial meet contraction functions* (voir section 4.1.2). Dans les deux cas, ces fonctions de sélection ont pour but de ne sélectionner que les « meilleurs » maxcons. L'idée dans ce cadre de fusion est que ces fonctions apportent une évaluation sociale (i.e. tenant compte de la distribution de l'information parmi les bases).

Dans [20] nous avons étudié trois critères particuliers. Le premier ($\triangle^d$) sélectionne les maxcons qui sont cohérents avec le plus de bases possibles. Le deuxième ($\triangle^{S,\Sigma}$) sélectionne les maxcons qui ont la plus petite différence (symétrique) pour la cardinalité avec les bases. Le troisième ($\triangle^{\cap,\Sigma}$) sélectionne les maxcons qui ont la plus grande intersection (pour la cardinalité) avec les bases.

En terme de propriétés logiques on obtient les résultats suivants pour les opérateurs de combinaison ($\surd$ signifie que le postulat est vérifié) :

	IC0	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	MI	Maj
$\triangle^{C1}$	$\surd$	$\surd$	$\surd$	—	$\surd$	$\surd$	—	$\surd$	—	$\surd$	—
$\triangle^{C3}$	—	—	—	—	$\surd$	$\surd$	—	$\surd$	$\surd$	$\surd$	—
$\triangle^{C4}$	$\surd$	$\surd$	$\surd$	—	—	—	—	$\surd$	$\surd$	$\surd$	—
$\triangle^{C5}$	$\surd$	$\surd$	$\surd$	—	$\surd$	$\surd$	—	$\surd$	$\surd$	$\surd$	—

TABLE 5.1 – Propriétés des opérateurs de combinaison

Si l'on utilise les fonctions de sélection, on obtient de meilleures propriétés qu'avec les opérateurs de combinaison :

	IC0	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	MI	Maj
$\triangle^{C1}$	$\surd$	$\surd$	$\surd$	—	$\surd$	$\surd$	—	$\surd$	—	$\surd$	—
$\triangle^{d}$	$\surd$	$\surd$	$\surd$	—	$\surd$	$\surd$	—	$\surd$	—	—	$\surd$
$\triangle^{S,\Sigma}$	$\surd$	$\surd$	$\surd$	—	$\surd$	—	—	$\surd$	$\surd$	—	$\surd$
$\triangle^{\cap,\Sigma}$	$\surd$	$\surd$	$\surd$	—	—	$\surd$	$\surd$	$\surd$	$\surd$	—	$\surd$

TABLE 5.2 – Propriétés des opérateurs à base de fonction de sélection basés sur $\triangle^{C1}$

On peut constater qu'aucun opérateur à base de formules ne parvient à satisfaire l'ensemble des propriétés logiques de la fusion contrainte. C'est pour cette raison que nous avons proposé une alternative décrite à la section 5.2.4.

2. Terme employé dans les articles [BKM91, BKMS92].

Nous pensons qu'il pourrait être intéressant de continuer cette piste pour étudier plus systématiquement les autres opérateurs et autres fonctions de sélection. Une perspective plus générale serait d'obtenir un théorème de représentation avec ces fonctions de sélection, qui serait une généralisation de celui des *partial meet contraction functions* pour la révision.

5.2.3. Opérateurs DA^2

Nous avons défini [5, 24, 63] une nouvelle famille d'opérateurs de fusion, paramétrée par une distance et deux fonctions d'agrégation, appelés opérateurs de fusion DA^2 (Pour 1 **D**istance et **2** fonctions d'**A**grégation). Ces opérateurs sont une généralisation des opérateurs de fusion à base de modèles usuels, mais permettent également de capturer certains opérateurs de fusion à base de formules. L'inconvénient principal des opérateurs de fusion à base de modèles usuels est qu'ils ne permettent pas de prendre en compte les bases incohérentes. Or, dans certains cas, il peut être nécessaire ou simplement utile d'utiliser ces informations. D'un autre côté, les opérateurs à base de formules permettent de prendre en compte les bases incohérentes, mais ne tiennent pas compte de la distribution des informations. Les opérateurs de fusion DA^2 permettent d'éviter ces deux écueils. Ils sont définis similairement aux opérateurs à base de modèles :

Définition 25 *Soit une distance entre interprétations d et deux fonctions d'agrégation f et g. L'opérateur de* **fusion DA^2** $\triangle^{d,f,g}$ *est défini par :*

$$mod(\triangle_{\mu}^{d,f,g}(E)) = \min(mod(\mu), \leq_E)$$

où le pré-ordre $\leq_E$ sur $\mathcal{W}$ est défini par :

- *$\omega \leq_E \omega'$ si et seulement si $d(\omega, E) \leq d(\omega', E)$, où*
- *$d(\omega, E) = f(d(\omega, K_1) \dots, d(\omega, K_n))$, où $E = \{K_1, \dots, K_n\}$.*
- *$d(\omega, K_i) = g(d(\omega, \alpha_1) \dots, d(\omega, \alpha_{m_i}))$, où $K_i = \{\alpha_1, \dots, \alpha_{m_i}\}$.*

La première fonction d'agrégation g permet d'extraire une information cohérente de la base K_i même si celle-ci est incohérente. Puis la seconde fonction f réalise l'agrégation entre sources.

Nous avons étudié ces opérateurs d'un point de vue logique et du point de vue de la complexité algorithmique, à la fois pour des familles générales d'opérateurs mais également pour des opérateurs particuliers. Ces résultats de complexité permettent d'intéressantes conclusions, ils montrent en particulier que d'un point de vue calculatoire, ces opérateurs de fusion à partir de distances ne sont pas plus durs que les opérateurs à base de modèles usuels puisque l'on reste au même niveau de la hiérarchie polynomiale (la généralisation ne coûte donc rien au niveau de la complexité), et ne sont pas plus durs que les opérateurs de révision.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 13).

5.2.4. Opérateurs disjonctifs

Nous avons défini une nouvelle [30, 49, 65] famille d'opérateurs de fusion, les opérateurs de **fusion à quota**. Ces opérateurs sont proches de méthodes de vote, appelées votes par comités [BSZ91]. L'idée est qu'une interprétation est modèle du résultat si elle est modèle d'un nombre suffisant de bases. Nous avons montré que ces opérateurs proposent un bon compromis entre un ensemble de critères importants pour des opérateurs de fusion : les propriétés logiques (section 5.1), la complexité algorithmique, la (non-)manipulabilité (voir section 5.4) et la puissance inférentielle. Nous avons également proposé une nouvelle famille d'opérateurs, appelés opérateurs GMIN [3], et nous avons montré qu'ils sont plus intéressants que les opérateurs de fusion à quota, car ils sont moins prudents et satisfont plus de propriétés logiques (d'un autre côté cela se paye par une complexité algorithmique plus élevée).

Le point commun de tous ces opérateurs est qu'ils sont disjonctifs, c'est-à-dire que le résultat de la fusion est choisi parmi la disjonction des bases.

Cette propriété n'est pas (et ne doit pas être) satisfaite par tous les opérateurs de fusion, car cela empêche de trouver des solutions de compromis, qui n'ont été proposées par aucune des bases.

Mais dans certains cas, il est justifié d'imposer cette propriété de « disjonction », en particulier en ce qui concerne la fusion de croyances [4]. Supposons par exemple que plusieurs médecins proposent des traitements possibles pour une pathologie donnée, il semble clair qu'il vaut mieux choisir parmi les traitements possibles, plutôt que de mélanger les traitements...

Une autre justification de cette propriété de disjonction est qu'elle peut-être expliquée comme l'expression d'une propriété d'unanimité. L'unanimité est une propriété classique lorsque l'on agrège des informations, et elle est considérée comme une propriété indispensable pour les méthodes de vote par exemple. Cette propriété signifie intuitivement que si tous les agents considèrent qu'un candidat est le meilleur, alors il doit être considéré comme le meilleur pour le groupe. Si l'on considère l'expression de cette condition d'unanimité en fusion, il y a deux interprétations possibles. La plus directe est l'expression de l'**unanimité sur les interprétations**. Ce qui s'exprime de la façon suivante :

(UnaM) Si $\omega \models \mu$ et $\forall K \in E, \omega \models K$, alors $\omega \models \triangle_\mu(E)$

Cette condition est une conséquence de la propriété (IC2), elle est donc satisfaite par tout opérateur de fusion contrainte.

Mais, si l'on considère qu'une base est l'ensemble de ses conséquences, on peut exprimer l'**unanimité sur les formules** :

(UnaF) Si $\exists K \in E$ t.q. $\mu \wedge K$ est cohérent, alors si $\forall K \in E, K \models \alpha$, alors $\triangle_\mu(E) \models \alpha$

La condition de (UnaF) s'assure juste qu'il est possible de sélectionner un résultat dans la disjonction des bases qui soit compatible avec les contraintes d'intégrité.

Et cette condition est équivalente au fait d'être un opérateur disjonctif :

3. Ces opérateurs sont des opérateurs à base de modèle (section 5.2.1), avec une nouvelle fonction d'agrégation GMIN (leximin).

4. Voir la section fusion de croyances versus fusion de but (section 5.3).

(Disj) Si $\bigvee E$ est cohérent avec μ, alors $\triangle_\mu(E) \models \bigvee E$

Cette propriété (Disj) est la principale raison pour justifier l'utilisation des opérateurs de fusion à base de formules (voir section 5.2.2). En effet, comparés aux opérateurs à base de modèles, ces opérateurs satisfont bien moins de propriétés logiques et ont une complexité algorithmique souvent plus importante.

Ce travail suggère donc que les opérateurs GMIN sont un bon substitut aux opérateurs à base de formules : comme eux ils satisfont la propriété de disjonction, mais ils présentent de bien meilleures propriétés logiques (ce sont des opérateurs de fusion contrainte), et une meilleure complexité algorithmique.

5.2.5. Opérateurs à base de conflits

Les opérateurs à base de modèles (section 5.2.1) sont basés sur une notion de proximité entre modèles. Cette notion de proximité est capturée par une distance (numérique), telle que la distance de Hamming par exemple. Une autre possibilité est de considérer comme « distance » l'ensemble des variables propositionnelles qui diffèrent entre deux interprétations (cette distinction existe dans le cadre de la révision, si on examine le lien entre l'opérateur de révision de Dalal [Dal88] et l'opérateur de Borgida [Bor85]). On est donc alors beaucoup plus précis qu'avec la distance de Hamming. Cela nous a conduit à définir un premier opérateur de fusion [38, 52], puis une famille entière d'opérateurs de **fusion basée sur les vecteurs de conflits** [39]. Cette famille est beaucoup plus générale que les opérateurs basés sur les modèles usuels (qui sont donc un cas particulier de cette classe), et permettent même de raffiner ces opérateurs.

Voyons un petit exemple illustrant en quoi définir un **vecteur de conflit** permet une meilleure discrimination des interprétations que lorsque l'on résume le conflit par une distance numérique.

Exemple 6 *Soit un langage contenant les variables a, b, c, d, un profil $E = \{K_1, \ldots, K_4\}$ et une contrainte d'intégrité μ telle que $mod(\mu) = \{\omega_1, \omega_2\}$.*
$\mathsf{diff}(\omega, \omega') = \{a \in \mathcal{P} \mid \omega(a) \neq \omega'(a)\}$, *et* $\mathsf{diff}(\omega, K) = \min_{\omega' \models K}(\mathsf{diff}(\omega, \omega'), \subseteq)$.

	$\mathsf{diff}(\omega, K_1)$	$\mathsf{diff}(\omega, K_2)$	$\mathsf{diff}(\omega, K_3)$	$\mathsf{diff}(\omega, K_4)$
ω_1	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$
ω_2	$\{\{a\}\}$	$\{\{b\}\}$	$\{\{c\}\}$	$\{\{d\}\}$

TABLE 5.3 – Différence entre ω_1 et ω_2 ?

La table 5.3 indique le conflit minimal entre chaque modèle des contraintes et chaque base du profil. Par exemple le conflit entre ω_1 et K_1 porte sur la variable a, c'est-à-dire que le modèle de K_1 qui présente le moins de différence par rapport à ω_1 est simplement en désaccord avec la valeur de vérité de la variable a.

Clairement la distance de Hamming ne permet pas de discriminer entre ces deux interprétations, car elles sont toutes deux à une distance 1 de chaque base. On obtient deux vecteurs $\langle 1, 1, 1, 1 \rangle$ qui sont indiscernables quelle que soit la fonction d'agrégation utilisée.

Si l'on utilise les vecteurs de conflits on obtient dans un cas le vecteur $\langle a, a, a, a\rangle$ et dans l'autre $\langle a, b, c, d\rangle$, ce qui montre une claire différence entre les deux situations.

Dans le cas de ω_1 tous les agents s'accordent sur le fait que la variable « problématique » est a, alors que ce n'est pas le cas pour ω_2. Ces deux interprétations peuvent donc être traitées différemment par les opérateurs de fusion à base de vecteurs de conflits.

5.2.6. Opérateurs à base de défauts

Delgrande et Schaub ont proposé deux opérateurs de **fusion à base de défauts** [DS07]. L'idée est d'utiliser un langage spécifique pour chaque base, afin d'assurer que l'union de ces bases soit cohérente, et ensuite d'ajouter autant de règles de défaut que possible afin d'identifier les variables correspondantes dans les différents langages (ce qui rappelle l'approche de Besnard et Schaub pour l'inférence en présence d'incohérence décrite section 3.1.2).

Ces opérateurs sont proches dans l'idée des opérateurs à base de conflits, puisqu'ils examinent les conflits existants variable par variable, mais ils les traitent de manière différente. Ces opérateurs ne sont donc pas définissables à partir d'une des familles précédentes.

La définition formelle de ces opérateurs est la suivante :

Définition 26 *Un i-**renommage** d'un langage $\mathcal{L}$ est le langage $\mathcal{L}^i$, construit à partir de l'ensemble de variables propositionnelles $\mathcal{P}^i = \{p^i \mid p \in \mathcal{P}\}$, où pour chaque $\alpha \in \mathcal{L}$, α^i est le résultat du remplacement dans α de chaque variable propositionnelle $p \in \mathcal{P}$ par la variable correspondante $p^i \in \mathcal{P}^i$. Soit une base K, le i-renommage de (des formules de) K, est noté K^i.*

Définition 27 *Soit un profil $E = \{K_1, \ldots, K_n\}$.*

- *Soit EQ un sous-ensemble de $\{p^k \Leftrightarrow p^l \mid p \in \mathcal{L}$ et $k, l \in \{1 \ldots n\}\}$ maximal (pour l'inclusion ensembliste) tel que $(\bigwedge_{K_i \in E} K_i^i) \wedge EQ$ est cohérent.*
  *Alors $\{\alpha \mid \forall j \in \{1 \ldots n\}\ (\bigwedge_{K_i \in E} K_i^i) \wedge EQ \models \alpha^j\}$ est une **extension cohérente symétrique** de E.*
  La fusion sceptique $\triangle_s(E)$ de E est l'intersection de toutes les extensions cohérentes symétriques de E.
- *Soit EQ un sous-ensemble de $\{p^j \Leftrightarrow p \mid p \in \mathcal{L}$ et $j \in \{1 \ldots n\}\}$ maximal (pour l'inclusion ensembliste) tel que $(\bigwedge_{K_i \in E} K_i^i) \wedge EQ$ est cohérent.*
  *Alors $(\bigwedge_{K_i \in E} K_i^i) \wedge EQ$ est une **extension cohérente projetée** de E.*
  La fusion sceptique $\nabla_s(E)$ de E est l'intersection de toutes les extensions cohérentes projetées de$E$.

Exemple 7 *Soit le profil $E = \{K_1, K_2, K_3\}$, avec $K_1 = (p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge r)$, $K_2 = (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r)$ et $K_3 = \neg q \wedge \neg r$.*

Il y a quatre ensembles maximaux d'équivalences pour $\Delta_s(E)$:

$EQ_1 = \{p^1 \Leftrightarrow p^2, p^1 \Leftrightarrow p^3, p^2 \Leftrightarrow p^3, r^1 \Leftrightarrow r^2, r^1 \Leftrightarrow r^3, r^2 \Leftrightarrow r^3, q^2 \Leftrightarrow q^3\}$

$EQ_2 = \{p^1 \Leftrightarrow p^3, r^1 \Leftrightarrow r^2, r^1 \Leftrightarrow r^3, r^2 \Leftrightarrow r^3, q^1 \Leftrightarrow q^2\}$

$EQ_3 = \{p^2 \Leftrightarrow p^3, r^1 \Leftrightarrow r^2, r^1 \Leftrightarrow r^3, r^2 \Leftrightarrow r^3, q^1 \Leftrightarrow q^2\}$

EQ_4 = $\{p^1 \Leftrightarrow p^2, p^1 \Leftrightarrow p^3, p^2 \Leftrightarrow p^3, r^2 \Leftrightarrow r^3, q^1 \Leftrightarrow q^2\}$

Donc, $\Delta_s(E) \equiv \neg r \vee (\neg p \wedge q)$.

Pour ∇_s*, les ensembles maximaux d'équivalences sont les suivants :*
($p \Leftrightarrow p^1 \Leftrightarrow p^2 \Leftrightarrow p^3$ *sera utilisé comme abbréviation pour* $p \Leftrightarrow p^1, p \Leftrightarrow p^2, p \Leftrightarrow p^3$*)*

$EQ_1 = \{p \Leftrightarrow p^1 \Leftrightarrow p^2 \Leftrightarrow p^3, r \Leftrightarrow r^1 \Leftrightarrow r^2 \Leftrightarrow r^3, q \Leftrightarrow q^2 \Leftrightarrow q^3\}$

$EQ'_1 = \{p \Leftrightarrow p^1 \Leftrightarrow p^2 \Leftrightarrow p^3, r \Leftrightarrow r^1 \Leftrightarrow r^2 \Leftrightarrow r^3, q \Leftrightarrow q^1\}$,

$EQ_2 = \{p \Leftrightarrow p^1 \Leftrightarrow p^3, r \Leftrightarrow r^1 \Leftrightarrow r^2 \Leftrightarrow r^3, q \Leftrightarrow q^1 \Leftrightarrow q^2\}$

$EQ_3 = \{p \Leftrightarrow p^2 \Leftrightarrow p^3, r \Leftrightarrow r^1 \Leftrightarrow r^2 \Leftrightarrow r^3, q \Leftrightarrow q^1 \Leftrightarrow q^2\}$

$EQ_4 = \{p \Leftrightarrow p^1 \Leftrightarrow p^2 \Leftrightarrow p^3, r \Leftrightarrow r^2 \Leftrightarrow r^3, q \Leftrightarrow q^1 \Leftrightarrow q^2\}$

$EQ'_4 = \{p \Leftrightarrow p^1 \Leftrightarrow p^2 \Leftrightarrow p^3, r \Leftrightarrow r^1, q \Leftrightarrow q^1 \Leftrightarrow q^2\}$

Donc, $\nabla_s(E) \equiv (p \wedge \neg r) \vee (\neg p \wedge q)$.

Ces opérateurs de fusion ont été implémentés dans le cadre de la plate-forme COBA [DLST07].

Une critique que l'on peut adresser à cette approche est que, tout comme les opérateurs à base de formules, ces opérateurs ne tiennent pas compte de la distribution des informations parmi les sources. En particulier ils ne sont pas majoritaires, et une information qui serait crue par toutes les bases sauf une ne sera par exemple pas retenue dans le résultat. Mais, comme pour les opérateurs à base de formules (voir section 5.2.2), il nous semble possible de définir des politiques additionnelles afin de prendre ce type d'arguments en compte en utilisant des fonctions de sélection sur les sous-ensembles maximaux EQ. Nous pensons donc qu'il reste des pistes intéressantes à explorer dans ce cadre.

5.2.7. Opérateurs à base de similarité

Récemment Schockaert et Prade [SP09] ont proposé des opérateurs de fusion basés sur une relation de similarité qualitative : pour chaque variable propositionnelle on associe un pré-ordre partiel sur les interprétations, dont cette variable est l'unique minimum, et qui représente la similarité des autres variables. Deux variables peuvent évidemment ne pas être en relation, d'où le pré-orde *partiel*.

Cette relation peut-être extraite d'un graphe entre variables propositionnelles, comme sur la figure 5.2 extraite de [SP09], où la similarité entre deux variables se compte en nombre d'arcs parcourus.

Shockaert et Prade se servent alors de cette relation de similarité pour tenter de trouver les meilleurs compromis lors de la fusion. La justification est alors de supposer que le conflit n'est pas issu d'avis divergents (donc d'un conflit réel), mais en quelque sorte de problèmes d'ontologie, ou d'approximations (par exemple un agent qui ne fait pas de distinction entre Single(x) et Divorced(x).

L'utilisation d'une telle relation de similarité permet d'utiliser des techniques plus fines pour la résolution de conflit, qui ont des points communs avec les techniques utilisées pour la fusion de systèmes de contrainte (voir section 5.5.5).

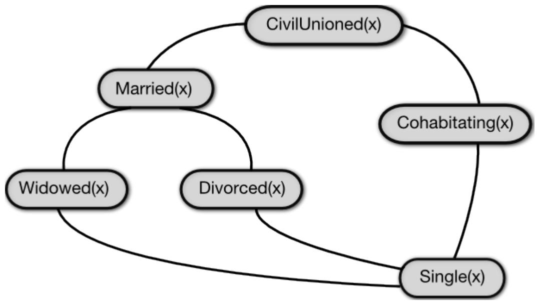


FIGURE 5.2 – Relation de similarité

L'inconvénient est que cette approche nécessite la donnée additionnelle d'une relation de similarité, qui n'est pas toujours disponible.

Cet article est à notre connaissance le premier à aborder l'utilisation de cette notation de similarité pour la fusion de croyances. Nous pensons qu'il reste beaucoup à faire dans cette voie. Il semble intéressant en particulier de tenter de combiner les approches de fusion à base de modèles et cette approche à base de similarité.

5.3. Fusion de croyances versus fusion de buts

Depuis le début du chapitre nous utilisons le terme de « fusion de croyances » de manière générique. Mais les travaux réalisés sur la caractérisation des opérateurs de fusion sont valables autant pour la fusion de croyances proprement dite que pour la fusion de buts. En effet, tous les postulats présentés sont aussi valables dans un cas que dans l'autre. Même s'il peut sembler étrange à première vue que des concepts si différents (croyances ou buts) puissent être traités de façon identique du point de vue de l'agrégation, la pertinence des opérateurs de fusion contrainte pour la fusion propositionnelle (que ce soit de buts ou de croyances) n'a pas encore été remise en cause par l'adjonction de nouveaux postulats qui permettraient de séparer les deux types de fusion.

Les seules nuances que l'on peut trouver est qu'il est plus facile de justifier les opérateurs de fusion disjonctive pour la fusion de croyances que pour la fusion de buts.

Dans l'idée d'étudier plus avant les liens entre fusion et théorie du choix social, nous nous sommes intéressés [71] au problème de l'identification du monde réel (*truth tracking*). Cela permet d'identifier deux types de fusion : le point de vue synthétique et le point de vue épistémique :

Vue synthétique : Sous le point de vue synthétique la fusion a pour objectif de caractériser une base qui représente au mieux les croyances du profil initial. C'est la vue considérée dans les travaux précédents concernant la fusion de croyances.

Vue épistémique : Sous le point de vue épistémique, le but d'un processus de fusion est d'estimer au mieux le monde réel, donc de supprimer autant que possible l'incertitude du groupe d'agents à son sujet.

Il est intéressant de remarquer que la vue épistémique, c'est-à-dire la recherche du monde réel, est un moyen de différencier la fusion de croyances de la fusion de buts. En effet, alors que la recherche du monde réel peut être demandée lors de la fusion de croyances, le concept de recherche du monde réel n'a pas de sens pour des buts. Il est en effet assez clair que la notion de « vrai but », équivalent dans le cadre des buts du monde réel, ne signifie rien.

Nous avons étudié quels opérateurs de fusion étaient adéquats pour la vue épistémique (nous avons proposé un nouveau postulat pour cette recherche du monde réel).

En théorie du choix social, le résultat qui justifie les décisions prises à la majorité est le **théorème du jury** de Condorcet, qui énonce que si, pour répondre à une question binaire (de type Oui/Non), on dispose d'agents fiables (qui ont moins d'une chance sur deux de se tromper) et indépendants, alors se fier à la majorité est la bonne chose à faire (en particulier quand le nombre d'agents tend vers l'infini, la probabilité d'erreur de cette procédure tend vers 0).

C'est ce résultat théorique qui justifie l'utilisation de comités pour prendre des décisions. Or, il est facile de constater que ses hypothèses sont très restrictives. En particulier, on ne peut choisir qu'entre 2 alternatives, et les agents ne peuvent pas avoir d'incertitude (ils ne peuvent pas être indifférents/hésitants entre Oui et Non).

Nous avons démontré, afin de pouvoir l'importer dans le cadre de la fusion, une généralisation du théorème du jury de Condorcet, que nous avons appelé **théorème du jury sous incertitude**, lorsqu'il y a un nombre quelconque d'alternatives et lorsque les agents peuvent être incertains (ils fournissent un ensemble d'alternatives et pas une alternative unique). Nous avons montré que le **vote par approbation** (*approval voting*) [BF83], qui permet aux individus de voter pour un nombre quelconque de candidats et qui élit le candidat ayant reçu le plus de voix, est la méthode à utiliser dans ce cas.

5.4. Manipulation

Nous avons étudié [9, 27, 64] la résistance à la manipulation des opérateurs de fusion. Les opérateurs existants de fusion permettent de définir les croyances/buts du groupe d'agents. Mais, si un agent est capable de conduire des raisonnements sur le résultat de cette opération et sur l'impact qu'il peut avoir sur celui-ci, il peut être tenté de mentir sur ses véritables croyances (ou buts), afin de modifier le résultat conformément à ses intérêts. Nous avons défini des notions de manipulabilité pour la fusion, et nous avons étudié la manipulabilité des principales méthodes de fusion de la littérature.

De manière non surprenante, il est assez difficile d'obtenir des garanties de non manipulabilité. En effet, la fusion de croyances est très proche techniquement de l'agrégation de préférences. Or, un résultat important en choix social, le théorème de Gibbard-Sattherwaite [Gib73, Sat75], montre qu'il est impossible de définir une méthode d'agrégation de préférences non manipulable. Il n'est donc pas étonnant d'avoir très peu de cas de non manipulabilité dans le cas de la fusion.

Nous avons défini la manipulabilité comme suit [9, 27, 64] :

Définition 28 *Soit un indice de satisfaction i, i.e., une fonction de $\mathcal{L} \times \mathcal{L}$ dans $\mathbb{R}$.*

- *Un profil E est manipulable par une base K pour l'indice i étant donné l'opérateur $\triangle$ et la contrainte d'intégrité μ si et seulement si il existe une base K' telle que*
$$i(K, \Delta_\mu(E \sqcup \{K'\})) > i(K, \Delta_\mu(E \sqcup \{K\}))$$
- *Un opérateur de fusion Δ est* **non manipulable** *pour i si et seulement si il n'y a pas de contrainte d'intégrité μ et de profil $E = \{K_1, \ldots, K_n\}$ tel que E est manipulable pour i.*

Cette définition de la manipulabilité est assez standard. La différence avec le cas de l'agrégation de préférence où la comparaison des situations est directe, est que dans le cas de la fusion on doit utiliser un **indice de satisfaction**.

Nous nous sommes concentrés sur trois indices différents (qui sont les plus naturels si l'on ne dispose pas d'informations complémentaires) :

Définition 29

- **Indice drastique faible** *:*
$$i_{dw}(K, K_\triangle) = \begin{cases} 1 & \textit{si } K \wedge K_\triangle \textit{ est cohérent} \\ 0 & \textit{sinon} \end{cases}$$
- **Indice drastique fort** *:*
$$i_{ds}(K, K_\triangle) = \begin{cases} 1 & \textit{si } K_\triangle \models K, \\ 0 & \textit{sinon} \end{cases}$$
- **Indice probabiliste** [5] *:*
$$i_p(K, K_\triangle) = \frac{|[K] \cap [K_\triangle]|}{|[K_\triangle]|}$$

On considère qu'un agent est satisfait par le résultat de la fusion ($K_\triangle$) avec l'indice drastique faible si ce résultat est cohérent avec sa base. Avec l'indice drastique fort, le résultat doit impliquer sa base. L'indice probabiliste permet une mesure plus progressive de la satisfaction, qui dépend de la proportion des modèles en commun entre le résultat de la fusion et la base de l'agent.

On obtient alors les résultats de manipulabilité pour les opérateurs usuels [6] décrits à la table 5.4. La première colonne indique avec quel indice on travaille, la deuxième si l'on restreint la cardinalité du profil à 2 bases ou pas, la troisième si l'on contraint les bases à être complètes (C) ou pas, et la quatrième indique si on autorise (μ) ou pas ($\top$) l'utilisation de contraintes d'intégrité. Les cases où un ■ apparaît indiquent un cas de non manipulabilité. Les lignes en orange indiquent le cas général (sans aucune restriction). On peut remarquer que dans ce cas général très peu d'opérateurs sont non manipulables (en particulier 2 seulement avec l'indice probabiliste).

D'autres résultats ont été obtenus avec d'autres indices et d'autres restrictions (voir [9, 27, 64]).

5. Lorsque $|[K_\triangle]| = 0$, alors $i_p(K, K_\triangle) = 0$.

6. Les opérateurs $\triangle^{\widehat{Ci}}$ sont une petite modification des opérateurs à base de formules $\triangle^{Ci}$ où on effectue d'abord la conjonction de toutes les formules de chaque base K_j du profil avant la fusion.

i	#	K	μ	$\Delta^{d_D,f}$	$\Delta^{d_H,\Sigma}$	$\Delta^{d_H,\text{Gmax}}$	$\triangle^{C1}$	$\triangle^{C3}$	$\triangle^{C4}$	$\triangle^{C5}$	$\triangle^{\widehat{C1}}$	$\triangle^{\widehat{C3}}$	$\triangle^{\widehat{C4}}$	$\triangle^{\widehat{C5}}$
i_{dw}	2	C	$\top$	■	■		■	■		■	■	■	■	■
			μ	■	■		■			■	■		■	■
		NC	$\top$	■	■		■	■		■	■	■	■	■
			μ	■			■				■		■	■
	n	C	$\top$	■	■		■	■		■	■	■	■	■
			μ	■	■		■			■	■		■	■
		NC	$\top$	■			■	■		■	■	■	■	■
			μ	■			■				■		■	
i_{ds}	2	C	$\top$	■	■	■	■	■		■	■	■	■	■
			μ	■	■		■				■		■	■
		NC	$\top$	■	■		■	■		■	■	■	■	■
			μ	■			■				■		■	■
	n	C	$\top$	■	■		■	■		■	■	■	■	■
			μ	■	■		■				■		■	
		NC	$\top$	■			■	■		■	■	■	■	■
			μ	■			■				■		■	
i_p	2	C	$\top$	■	■						■	■	■	■
			μ	■	■						■		■	■
		NC	$\top$	■							■	■	■	■
			μ	■							■		■	■
	n	C	$\top$	■	■								■	
			μ	■	■								■	
		NC	$\top$	■									■	
			μ	■									■	

TABLE 5.4 – Résultats de manipulabilité

5.5. Fusion dans d'autres cadres

La fusion a également été étudiée dans d'autre cadres de représentation. Les opérateurs de fusion présentés jusqu'ici étaient tous définis dans le cadre de la logique propositionnelle et dans le cas où toutes les bases ont la même importance/priorité/fiabilité. On peut avoir besoin de fusionner des informations plus structurées que celles que l'on exprime en logique classique, ce qui génère des problèmes, et des possibilités, supplémentaires.

Nous allons donner un aperçu des travaux les plus proches dans cette section.

5.5.1. Fusion prioritaire, fusion et révision itérée

Dans [DDL06] Delgrande, Dubois et Lang proposent une discussion intéressante sur les opérateurs de **fusion prioritaire**. L'idée est de fusionner un ensemble de formules [7] pondérées. La pondération sert à stratifier ces formules (une formule avec un poids plus grand est plus importante, et le nombre de formules moins prioritaires la contredisant n'importe pas).

Delgrande, Dubois et Lang motivent alors la généralité de leur approche en montrant que les opérateurs de fusion propositionnelle « classiques » (i.e. sur des bases non pondérées) et les opérateurs de révision itérée (à la Darwiche et Pearl) peuvent être considérés comme les deux cas extrêmes de ces opérateurs de fusion prioritaire.

Leur discussion sur les opérateurs de révision itérée est particulièrement intéressante, et rappelle les mise en garde de Friedman et Halpern sur les dangers de définir des opérateurs de changement sans spécifier leur ontologie [FH96b]. L'argument principal est que si on fait l'hypothèse que les nouvelles informations qui arrivent successivement lors d'une suite de révisions concernent un monde statique (hypothèse usuelle), alors il n'y a a priori aucune raison de préférer la dernière. Si ces informations ont des fiabilités différentes, il est possible de représenter ces fiabilités explicitement, afin de les prendre correctement en compte dans le processus de « révision » si elles n'arrivent pas dans l'ordre de leur fiabilité. Et la façon correcte de faire est de réaliser leur fusion prioritaire.

Cette discussion est intéressante car dans un certain nombre de papiers portant sur la révision itérée, il semble que les auteurs confondent l'hypothèse d'informations de plus en plus fiables, avec celle d'informations de plus en plus récentes.

Le cadre de Delgrande, Dubois et Lang suppose que l'on identifie un état épistémique avec la suite de formules que l'agent a reçues jusqu'ici, hypothèse qui avait été proposée dans la définition de révision itérée de Lehmann [Leh95] et dans notre proposition d'opérateurs à mémoire [1, 21].

Delgrande, Dubois et Lang montrent ensuite que les postulats des opérateurs de révision itérée peuvent être retrouvés à partir des postulats de base qu'ils proposent pour la fusion prioritaire. Ils montrent également que l'on obtient une partie des postulats de la fusion contrainte.

Cette proposition est intéressante car elle offre une piste pour la caractérisation logique des opérateurs de fusion prioritaire. Nous pensons qu'il y a encore beaucoup à

7. Chaque formule pouvant représenter une base si on veut faire le parallèle avec la fusion dans le cadre propositionnel.

montrer en suivant cette voie. Les auteurs n'ont en particulier pas donné de théorème de représentation pour leurs opérateurs, et n'ont pas donné l'équivalent dans leur cadre de l'ensemble des postulats de la fusion contrainte. De plus, une de leur hypothèse de base est que la hiérarchisation des formules est stricte, c'est-à-dire qu'une seule formule de poids supérieur l'emporte sur un nombre quelconque de formules de poids inférieur. Cela peut se justifier dans certaines applications, mais dans le cas général, il pourrait être intéressant d'avoir une possible compensation entre formules de différents poids. Finalement, rien dans leur axiomatique n'indique comment doivent être traités les conflits entre formules de même poids, ce qui est l'objet de la fusion propositionnelle. La fusion classique étant présentée comme cas particulier de fusion prioritaire, cela peut donc sembler paradoxal.

Bref, cet article a ouvert une voie intéressante, qui n'a malheureusement pas été exploitée depuis.

5.5.2. Fusion de bases pondérées

Lorsque les informations contenues dans les bases de croyances ne sont pas toutes de même importance, on utilise des approches pondérées. L'approche la plus qualitative est de considérer, pour chaque source/agent, un ensemble de bases (totalement) ordonnées en différentes strates, de la plus importante à la moins importante. Cette situation est habituellement codée en utilisant la **logique possibiliste** [DLP94] ou les **fonctions ordinales conditionnelles** [Spo87]. Dans ce cas on associe un ordinal (habituellement fini, i.e, un entier naturel [8]) à chaque formule.

Dans le cadre de la logique possibiliste Benferhat, Dubois, Kaci et Prade ont étudié plusieurs opérateurs de fusion [KBDP00, BDKP02]. Ils ont en particulier étudié l'extension de nos propriétés logiques dans ce cadre [BK03]. (voir également [QLB06] pour une généralisation de nos opérateurs dans un cadre pondéré).

Dans le cadre des fonctions ordinales conditionnelles, Meyer a également défini différents opérateurs de **combinaison** [Mey01]. Certains d'entre eux sont la traduction des opérateurs de fusion à base de modèles usuels dans ce cadre pondéré,mais d'autres semblent très loin de ce que l'on attend d'un opérateur de fusion.

Tous ces travaux utilisant des bases pondérées reviennent sémantiquement à utiliser pour chaque agent une distribution de possibilités sur les mondes (ou de manière équivalente une fonction ordinale conditionnelle), c'est-à-dire que chaque interprétation est associée à un nombre par l'agent, ce nombre exprimant à quel point l'agent estime l'interprétation plausible.

Les fonctions d'agrégation utilisées pour définir les opérateurs de fusion réalisent donc un calcul à partir de ces nombres. Il se pose alors un problème de **comparaison interpersonnelle des utilités**, c'est-à-dire que cette opération suppose que le nombre 4 chez un agent a la même valeur (i.e. a le même sens) que chez n'importe quel autre agent. C'est ce que l'on appelle l'hypothèse de **commensurabilité**.

Pour être fondée, cette approche nécessite cette hypothèse de commensurabilité. Il y a des cas où cela est parfaitement naturel, par exemple si les sources sont des capteurs identiques. Mais dans des applications où les sources sont des agents autonomes, cette

8. Voir [40] pour une discussion.

hypothèse semble clairement irréaliste. En particulier, lorsque l'on travaille avec des bases pondérées, on est très proche des hypothèses faites en **théorie du choix social** pour les méthodes de vote. Et il est communément accepté que cette hypothèse de commensurabilité n'est pas acceptable dans ce cas [Arr63]. Dans le cadre du vote seule la préférence ordinale de chaque agent est prise en compte, c'est-à-dire l'ordre associé à ces nombres.

Il nous semble que l'étude générale, sans cette hypothèse de commensurabilité, de la fusion de bases pondérées, surtout si on considère des opérateurs majoritaires, revient à étudier des méthodes de vote, et qu'il faut alors se référer à la littérature de choix social [Arr63, ASS02].

Benferhat, Lagrue et Rossit ont étudié des opérateurs de fusion de bases pondérées non majoritaires sans l'hypothèse de commensurabilité [BLR07, BLR09]. Evidemment, cela conduit a des opérateurs beaucoup plus prudents que les opérateurs que l'on définit dans le cas commensurable. Une question intéressante serait alors d'étudier si les opérateurs proposés dans le cadre non commensurable correspondent à des méthodes de vote connues.

5.5.3. Fusion en logique du premier ordre

Lang et Bloch ont proposé de définir les opérateurs de fusion à base de modèle $\triangle^{d,\max}$, utilisant le maximum comme fonction d'agrégation, grâce à un processus de **dilatation** [BL00]. On peut d'ailleurs noter que dans l'article de Dalal [Dal88], son opérateur de révision n'est pas défini à base de distance, mais à partir d'une telle fonction de dilatation.

Gorogiannis et Hunter [GH08] ont étendu cette approche afin de définir les opérateurs de fusion à base de modèles usuels en terme de dilatations. En plus de $\triangle^{d,\max}$, ils ont exprimé la définition de $\triangle^{d,\Sigma}$, $\triangle^{d,Gmax}$, et $\triangle^{d,Gmin}$.

L'intérêt de cette caractérisation de ces opérateurs est que celle-ci peut-être exportée à la logique du premier ordre. En effet, la définition usuelle des opérateurs à base de modèles demande le calcul de distances entre l'ensemble des interprétations. Or dès que l'on passe dans des logiques plus expressives que la logique propositionnelle, comme la logique du premier ordre, le nombre de modèles devient très souvent infini. L'intérêt de la définition en termes de dilatation est que celle-ci peut-être calculée même dans ces cadres. Cela nécessite simplement de choisir la bonne fonction de dilatation. Voir [GH08] pour une discussion et quelques exemples sur ces fonctions de dilatation dans le cadre de la logique du premier ordre.

5.5.4. Fusion de programmes logiques

Certains travaux ont étudié des opérateurs de fusion pour des bases exprimées en programmation logique avec sémantique des modèles stables (*Answer Set Programming*). C'est une question assez naturelle lorsque l'on considère qu'il y a eu beaucoup de travaux sur la révision / mise à jour de programmes logiques (voir par exemple [ZF98, ALS98, ALP$^+$00, DSTW08]), mais jusqu'à récemment aucun sur la fusion.

L'approche de Hué, Papini et Würbel [HPW09] repose sur la suppression d'un certain nombre de formules dans l'union des bases, sélectionnées grâce à une fonction de sélection (un peu comme pour sélectionner les maximaux cohérents dans [20]). Ces opérateurs ne satisfont que très peu des propiétés logiques de la fusion contrainte.

Delgrande, Schaub, Tompits et Woltran [DSTW09] ont également étudié la fusion dans ce cadre. Leurs opérateurs sont basés sur la définition d'une distance entre les modèles stables. Leurs opérateurs satisfont beaucoup plus de propriétés logiques.

Pour comparer rapidement ces deux approches, on peut dire que dans le cadre de la programmation logique, les opérateurs de Hué, Papini et Würbel correspondent aux approches à base de formules, alors que les opérateurs de Delgrande, Schaub, Tompits et Woltran correspondent aux approches à base de modèles. Il n'est donc pas étonnant que ces derniers satisfassent, comme dans le cas de la logique propositionnelle, plus de propriétés logiques. Ces opérateurs étant facilement implémentables, il serait intéressant de comparer leur temps de réponse sur des problèmes pratiques.

5.5.5. Fusion de réseaux de contraintes

Condotta, Kaci, Marquis et Schwind ont proposé des méthodes pour fusionner des réseaux de contraintes qualitatives [CKMS09b, CKMS09a]. Ces méthodes peuvent être très utiles pour fusionner des réseaux de contraintes représentant des régions spatiales ; par exemple dans le cadre de systèmes d'information géographiques (SIG), il peut être nécessaire de tenter de fusionner des bases de données spatiales issues de sources différentes.

Les conflits qui apparaissent sont plus subtils que ceux issus de problèmes exprimés en logique propositionnelle. Dans ce dernier cas, les conflits sont de type vrai/faux, alors que dans le cas des réseaux de contraintes on peut avoir différents types de conflits plus ou moins graves. Par exemple, si l'on s'intéresse à l'algèbre d'Allen, qui permet de représenter les informations spatiales à propos de segments situés sur une droite, on dispose des relations décrites figure 5.3.

Un conflit entre les assertions du type *« Le segment A est avant le segment B »* (A BEFORE B) et *« Le segment A touche le segment B »* (A MEET B) semble beaucoup moins fort que le conflit entre *« Le segment A est avant le segment B »* (A BEFORE B) et *« Le segment B est avant le segment A »* (A iBEFORE B).

Cette « intensité » que l'on sent naître entre les différents conflits permet d'imaginer des politiques de fusion plus variées que dans le cadre propositionnel.

5.5.6. Fusion de systèmes d'argumentation

Nous nous sommes intéressés au problème de la **fusion de systèmes d'argumentation** [8, 29]. Beaucoup de travaux ont étudié l'argumentation comme moyen de raisonner à partir d'informations contradictoires. Fondamentalement on utilise un ensemble d'arguments et une relation de contrariété (attaque) entre les arguments. Un cadre général pour l'argumentation a été proposé par Dung [Dun95]. Mais ces travaux sur l'argumentation ne se préoccupent que d'un seul agent. Nous avons étudié comment généraliser ces cadres pour prendre en compte le fait que les arguments sont distribués parmi un ensemble d'agents. Nous avons examiné le problème posé par le fait que différents agents puissent avoir des systèmes d'argumentation construits à partir

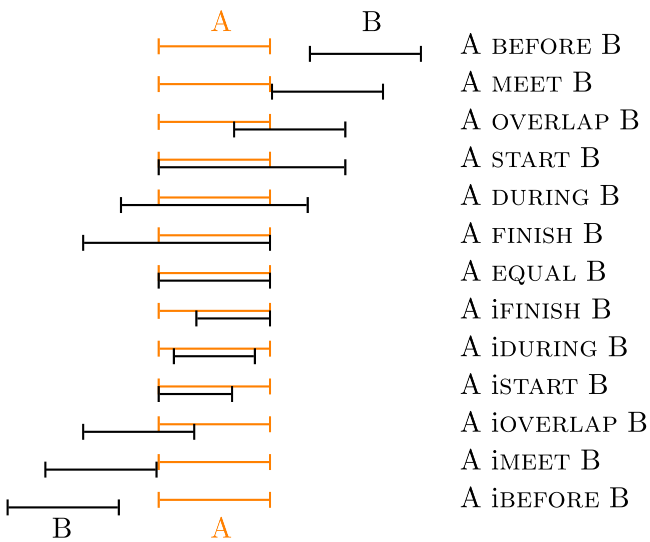


FIGURE 5.3 – Algèbre des intervalles d'Allen

d'arguments différents. Nous avons étudié comment représenter ces systèmes d'argumentation pour pouvoir les comparer, et comment les fusionner, afin de déterminer les arguments acceptables pour le groupe.

Cela nécessite en particulier la définition d'un nouveau cadre d'argumentation, les PAF (pour *Partial Argumentation Framework* ou **cadre d'argumentation partiel**). En effet la définition d'un cadre d'argumentation de Dung est simplement la donnée d'un ensemble d'objets (les arguments) et d'une relation entre ces objets (la relation d'attaque). Ce que signifie l'absence de relation entre deux arguments n'est pas clair. Est-ce que cela signifie que l'agent sait (croit) qu'il n'y a pas d'attaque entre ces deux arguments ou qu'il ne sait pas s'il y a une attaque entre ces deux arguments ? Lorsque l'on regarde la définition des solutions (extensions) dans ce cadre, on se rend compte que l'interprétation par défaut est que l'agent sait qu'il n'y a pas d'attaque entre les arguments. Nous avons donc défini un cadre où au lieu de la dichotomie habituelle attaque / pas d'attaque, on a une trichotomie attaque / pas d'attaque / indéterminé. Cela est nécessaire pour pouvoir représenter l'absence d'information d'un des agents à propos d'un argument.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 14).

Chapitre VI

Négociation

Il faut jouer pour devenir sérieux.

(Aristote)

La problématique de la négociation est une question complexe. Elle a été beaucoup étudiée en économie, où des modèles de théorie des jeux (en particulier dans le cadre des jeux coopératifs), comme le modèle du marchandage (bargaining) étudié par Nash [NJ50], ont apporté quelques réponses. Elle est étudiée également depuis longtemps en intelligence artificielle et plus particulièrement dans le cadre des systèmes multi-agents. Mais ce qui est étudié dans la plupart de ces travaux est la mise en oeuvre de procédures de négociation, c'est-à-dire la proposition de méthodes *ad hoc*. Ces méthodes utilisent une palette d'outils très différents puisque se posent des problèmes de communication, d'argumentation, de modélisation des préférences et des croyances, de recherche de compromis, etc.

Nous pensons qu'avant d'étudier un problème mélangeant des notions aussi différentes, il serait utile de l'abstraire, afin de tenter dans un premier temps de résoudre le problème central, avant d'y ajouter des modalités pratiques qui complexifient beaucoup l'étude.

Nous souhaitons étudier et caractériser des procédures de négociations abstraites, c'est-à-dire des fonctions qui prennent comme donnée un profil de bases (une par agent) et qui produit un nouveau profil, ou les conflits ont disparu, ou tout au moins ont été réduits. Nous avons appelé ces procédures de négociations abstraites, des opérateurs de **conciliation**.

Nous avons déjà exploré des opérateurs particuliers de conciliation que nous détaillons dans les sections ci-dessous. Ainsi, nous n'avons étudié jusqu'ici que la deuxième étape pour ce problème : la proposition de méthodes *ad hoc*. D'autres auteurs ont également proposé des méthodes pour résoudre ce type de problème de négociation [Boo02, Boo06, MFZK04, MPP05, Zha05, ZZ08, Zha07].

Ce que nous comptons faire dans le futur est de passer à la troisième étape, c'est-à-dire la caractérisation logique de ces opérateurs de conciliation. Nous reviendrons sur cette perspective à la dernière section.

6.1. Fusion itérée

Une partie de nos travaux a glissé de la problématique de la fusion de croyances à celle de la négociation. L'exemple symptomatique de ce glissement est celui de la **fusion itérée** que nous détaillons dans cette section, avant d'évoquer dans les deux sections suivantes d'autres travaux utilisant les outils issus de la fusion de croyances pour définir des opérateurs de conciliation.

Nous avons voulu définir des opérateurs de fusion améliorés. L'idée est que si l'on considère le résultat de la fusion comme la croyance « moyenne » du groupe, il doit alors être possible de trouver un résultat plus précis à la fusion de croyances en réalisant une fusion itérée : le principe est, à partir d'un profil de bases de croyances, de calculer le résultat de la fusion de ces bases, puis de répercuter cette fusion dans les croyances de chaque agent (donc chaque agent tient compte des positions des autres) en utilisant un opérateur de révision. On itère ce processus jusqu'à ce que l'on atteigne un point fixe [7, 31, 50]. Malheureusement les opérateurs de fusion itérée ainsi définis n'ont pas de très bonnes propriétés logiques, en tant qu'opérateurs de fusion. En revanche, il peut être intéressant de considérer ce processus comme une opération de conciliation.

6.2. Fusion de croyances comme un jeu entre sources

Nous avons défini une nouvelle famille d'opérateurs de fusion de croyances (ou de buts), basée sur un jeu entre les sources prenant part à la fusion [4, 47, 55]. Le principe de ce jeu est le suivant : à chaque tour, un ensemble de perdants est défini (intuitivement, les perdants sont les sources dont les croyances (ou les buts) sont le plus loin de celles du groupe), ces perdants doivent affaiblir leurs croyances (buts). Le jeu s'arrête lorsque le profil des croyances/buts des agents est cohérent. Un opérateur particulier est donc défini par la méthode de désignation des perdants et par la méthode d'affaiblissement choisie. Cette famille d'opérateurs est intéressante, puisqu'elle peut être vue comme un compromis entre fusion de croyances et négociation. Nous avons comparé ces opérateurs aux opérateurs existants et étudié les propriétés logiques de fusion contrainte satisfaites. Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 15).

Il reste à trouver une caractérisation logique de ces opérateurs et à étudier plus profondément cette piste d'utilisation de la théorie des jeux (qui pour le moment est réduite à sa plus simple expression), afin de fournir un cadre formel à des opérateurs de négociation, ou pour étudier plus finement des phénomènes de coalitions ou de manipulation dans des cadres de fusion.

En particulier nous avons proposé d'utiliser les mesures d'incohérence que nous avons développées (section 3.2.3) pour sélectionner comme perdants les sources les plus conflictuelles [51]. Nous pensons que cette voie est prometteuse.

6.3. Opérateurs de confluence

Beaucoup d'opérateurs de changement ont été étudiés ces dernières années. Un certain nombre d'entre eux sont à présent assez bien cernés, tels que la révision, qui permet d'incorporer de nouvelles informations sur le monde, la mise-à-jour [KM91a, HR99], qui permet d'actualiser les croyances de l'agent pour prendre en compte une évolution du monde, et la fusion qui permet de produire une information cohérente à partir d'un ensemble d'informations contradictoires. Il existe de forts liens entre révision et mise-à-jour et entre révision et fusion. En effet, techniquement, la mise-à-jour peut être vue comme une révision « ponctuelle », et la révision peut être vue comme un cas particulier de fusion. Ces liens suggéraient la possibilité de définir une nouvelle famille d'opérateurs qui sont à la mise-à-jour ce que la fusion est à la révision, et qui pourraient donc être vus comme des opérateurs de fusion « ponctuelle ». Cela est résumé à la figure 6.1.

$$\begin{tikzcd}
\text{Révision} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{Mise-à-jour} \arrow[d, leftrightarrow] \\
\text{Fusion} \arrow[r, leftrightarrow] & \text{Confluence}
\end{tikzcd}$$

FIGURE 6.1 – Révision - Mise-à-jour - Fusion - Confluence

Nous avons défini et caractérisé ces opérateurs, que nous avons nommés **opérateurs de confluence** [37, 54, 72]. Il apparaît que ces opérateurs sont beaucoup moins sélectifs que les opérateurs de fusion, et qu'ils peuvent être utilisés pour définir des opérateurs de négociation en fournissant l'ensemble des points d'accord possibles entre les opinions des différents agents.

Nous ne détaillons pas plus cette approche ici, puisque l'article correspondant est inclus dans la deuxième partie du document (chapitre 16).

6.4. Vers une caractérisation de la conciliation

Nous souhaitons caractériser les opérateurs de conciliation, c'est-à-dire les procédures de négociation abstraite qui prennent comme donnée un profil de bases de croyances (une par agent) et qui produisent un nouveau profil, où les conflits ont disparu, ou tout au moins ont été réduits.

Il peut sembler ambitieux, voire utopique, de vouloir caractériser aussi simplement un processus aussi compliqué que la négociation, nécessitant des échanges entre agents, de l'argumentation, des compromis, etc. Mais on peut remarquer qu'un problème similaire de négociation a été résolu de manière abstraite dans le domaine de la théorie des jeux. Le problème du **marchandage** (*bargaining*) s'énonce comme suit : étant donné un ensemble (convexe et compact) d'issues possibles à la négociation, et un point de désaccord (qui sera le résultat par défaut si les joueurs ne se mettent pas d'accord sur une issue), est-il possible de trouver une issue unique résultat de la négociation ? Nash a montré que cela était possible et a caractérisé axiomatiquement cette solution [NJ50].

Dans l'étude des **jeux de coalitions**, un autre problème abstrait a été résolu. Le but est, étant donné un ensemble de joueurs pouvant former les coalitions qu'ils souhaitent, chacune de ces coalitions gagnant un certain montant, de trouver le juste montant devant revenir à chaque joueur (une allocation). Ce problème semble également compliqué, parce que, pour se mettre d'accord sur une allocation, les joueurs doivent discuter, argumenter, etc. Mais Shapley a montré qu'il était possible de déterminer une telle allocation et il l'a caractérisée axiomatiquement [Sha53].

Ces deux exemples illustrent le fait que l'on peut résoudre ce type de problèmes compliqués de manière convaincante, en capturant l'essence du problème, et en ne se laissant pas encombrer par des modalités pratiques non fondamentales.

Nous souhaitons tenter de trouver une solution similaire pour le problème de la négociation abstraite en logique (conciliation).

Outre le défi scientifique que cela représente, nous pensons que trouver une telle caractérisation est importante pour toutes les personnes travaillant sur la négociation. En effet, la plupart des travaux sur la négociation en intelligence artificielle, notamment dans le domaine des systèmes multi-agents, portent sur la proposition de méthodes *ad hoc*, utilisant différentes techniques (argumentation, etc.). Il est difficile de comparer ces travaux, utilisant souvent des hypothèses ou des techniques assez différentes. Disposer d'une caractérisation logique, et donc d'un ensemble de propriétés logiques caractérisant le processus de conciliation (négociation), permettrait de comparer ces méthodes, en examinant les propriétés satisfaites par chacune d'elles.

Chapitre VII

Conclusion et Perspectives

La science consiste à passer d'un étonnement à un autre.

(Aristote)

Nous avons insisté sur la nécessité d'obtenir des caractérisations logiques pour les différents problèmes étudiés en intelligence artificielle. En ce qui concerne les différents cadres que nous avons étudié, nous avons en particulier obtenus les caractérisations suivantes :

- **Incohérence** : nous avons proposé la première caractérisation d'une valeur d'incohérence. Nous pensons qu'il reste beaucoup à faire dans ce domaine. Ces perspectives seront détaillées ci-dessous.
- **Révision** : nous avons défini et caractérisé les opérateurs d'amélioration, qui sont une généralisation des opérateurs de révision itérée usuels, et qui ouvrent la voie à la définition de nouveaux opérateurs de changement.
- **Fusion** : nous avons caractérisé les opérateurs de fusion contrainte, et étudié d'autres propriétés comme la non manipulabilité ou le « *truth tracking* ».
- **Négociation** : la caractérisation des opérateurs de conciliation est notre principale perspective. Nous avons pour le moment étudié des méthodes particulières. Nous avons caractérisé les opérateurs de confluence. Cela peut peut-être servir de point de départ pour la recherche d'une caractérisation des opérateurs de conciliation.

Nous avons énuméré dans les chapitres correspondants les pistes principales qui nous semblaient prometteuses dans les différents travaux présentés. Nous ne répéterons donc pas ici les perspectives de chacun de ces chapitres.

Nous allons plutôt nous focaliser dans ce chapitre sur l'orientation générale que nous comptons donner à nos travaux, et donc les perspectives que nous comptons étudier. On peut résumer notre projet scientifique sous la formule :

Du conflit entre formules logiques vers le conflit entre agents

En effet, l'orientation générale de notre projet de recherche consiste à continuer à suivre la même démarche scientifique, et principalement les aspects de caractérisation logique de différents problèmes, mais en glissant de problèmes de raisonnement (et de « conflits entre formules logiques ») vers des problématiques de conflits entre agents, et particulièrement vers la négociation et plus généralement la théorie des jeux.

Plus précisément, on peut décrire ces perspectives selon quatre axes principaux :

- Le développement de méthodes de mesure de conflit.
- L'étude des liens qui existent entre la fusion et la théorie du choix social.
- La caractérisation logique de la négociation.
- La théorie des jeux dans des cadres d'incertitude qualitative.

Ces points sont ordonnés selon leur proximité avec nos thèmes de recherche actuels.

Les deux premiers points, concernant les mesures de conflit et les liens entre fusion et théorie du choix social, représentent la continuation de nos travaux actuels, mais en se tournant un peu plus vers la théorie du choix social.

Le troisième point, qui nous semble être le plus prometteur et qui est celui sur lequel nous allons nous concentrer en priorité, concerne la négociation, qui est à l'intersection de problématiques d'intelligence artificielle et d'économie. Nous ne le détaillerons pas ci-dessous puisqu'il a déjà été l'objet du chapitre précédent (voir la section 6.4).

Le dernier point est réellement un problème d'économie (théorie des jeux) pour lequel nous souhaitons apporter un éclairage venant de l'intelligence artificielle (problèmes de représentation de l'incertitude, d'incertitude qualitative, et d'ordinalité).

7.1. Mesures de conflit

La mesure du degré de conflit n'est pas très développée pour le moment, malgré le réel besoin de telles mesures. En effet, pouvoir estimer à quel point un ensemble d'informations (ou d'agents) est conflictuel est aussi important que d'estimer à quel point il est informatif. Or, alors que les mesures d'informations sont étudiées depuis longtemps (citons Shannon (1950)), les premiers travaux sur des mesures de conflits sont très récents. Les perspectives principales de l'utilisation de telles mesures de degré de conflit concernent les coalitions d'agents. Si les coalitions sont déjà fixées, ces mesures peuvent permettre d'indiquer à quel point les coalitions sont proches, ou opposées ; et de quelles coalitions une coalition donnée est la plus proche. Elle peut également renseigner sur la robustesse des différentes coalitions. Si les coalitions ne sont pas encore déterminées, c'est-à-dire que l'on a comme donnée un ensemble d'agents qui peuvent former des coalitions quelconques, les mesures de degré de conflit peuvent renseigner sur la coalition la plus intéressante pour un agent, sur les coalitions les plus probables ou sur les points posant le plus problème entre deux coalitions dans une optique de résolution de conflits par exemple. Ces mesures de conflits peuvent également nous indiquer les agents (ou les coalitions d'agents) qui engendrent le plus de conflits pour le groupe dans son ensemble. Elles peuvent donc être utilisées pour guider un processus de négociation, afin de se concentrer sur les points/agents les plus conflictuels.

Nous avons déjà travaillé à la définition de telles mesures (section 3.2), mais il reste encore beaucoup à faire. Nous comptons continuer à travailler sur ce sujet. Nous sommes convaincus qu'il reste beaucoup de mesures à définir et à caractériser (pour le moment la seule caractérisation d'une mesure d'incohérence est celle de [36]).

Une autre piste concerne le conflit potentiel. Les mesures définies pour le moment s'intéressent au conflit logique déjà contenu dans les bases. Il peut être intéressant dans de nombreux cas de ne pas mesurer le conflit actuel, mais le conflit potentiel, qui permettra d'évaluer à quel point les exigences de différents agents peuvent être (potentiellement) problématiques. En effet, détecter et résoudre les confits potentiels avant qu'ils ne surviennent et qu'il ne soit trop tard est une problématique intéressante. Cela peut être très utile dans des cas de conception de logiciels, lorsque les spécifications proviennent de plusieurs utilisateurs. Il peut être inutile et coûteux de développer le logiciel pour se rendre compte trop tard que les spécifications demandées n'étaient pas compatibles. Il est donc intéressant d'identifier au plus tôt les points susceptibles de générer un conflit potentiel, afin de clarifier les spécifications et d'éviter des problèmes ultérieurs. Plus généralement, la détection de conflits potentiels peut être utile pour les processus de négociation entre agents.

Une dernière piste est de travailler sur des mesures de conflits pour des informations plus « riches », plus structurées, que la logique propositionelle, comme les pré-ordres par exemple. Les préférences des agents sont souvent représentées avec des pré-ordres. Nous souhaitons étudier les mesures de conflits entre pré-ordres, afin de pouvoir également mesurer le conflit entre les préférences de différents agents. Cela nécessitera de se tourner vers la théorie du choix social, en particulier les méthodes de vote qui considèrent de tels pré-ordres.

7.2. Fusion et choix social

Nous désirons continuer à étudier les liens entre fusion et choix social. Nous avons déjà obtenu des résultats intéressants en étudiant ce que la manipulabilité, une notion bien étudiée en choix social, a comme conséquences pour la fusion de croyances [9, 27]. D'autres notions issues du choix social peuvent se montrer intéressantes pour la fusion de croyances.

Il s'est développé ces dernières années en théorie du choix social une théorie de l'agrégation de jugement, qui est un problème proche de la fusion. Pour la fusion de croyances, on dispose comme données de l'ensemble des bases des agents, alors que pour l'agrégation de jugements, on ne dispose que de la réponse des agents à un certain nombre de questions (jugements). Cela change assez sensiblement les données du problème. En particulier la plupart des résultats en agrégation de jugement sont des théorèmes d'impossibilité (voir par exemple [LP04, Lis10]), alors qu'en fusion de croyances des caractérisations logiques montrent la faisabilité de l'approche. Il a d'ailleurs déjà été proposé d'utiliser nos opérateurs de fusion contrainte pour définir des méthodes d'agrégation de jugement [Pig06]. Nous souhaitons examiner les liens entre les deux problèmes, et étudier si les méthodes issues de la fusion de croyances peuvent apporter un nouvel éclairage aux questions posées par l'agrégation de jugement.

Finalement, il nous semble intéressant d'examiner si, parmi l'ensemble important de méthodes de vote existantes, certaines peuvent nous permettre de définir des opérateurs de fusion intéressants. En particulier, en dehors des méthodes de votes majoritaires classiques, il existe un grand nombre de méthodes définissables à partir des graphes de majorité. Ces méthodes définissent des ensembles de « bons » candidats à partir de ces graphes. Nous souhaitons étudier si l'application de ces méthodes dans le cadre de la fusion de croyances conduit à des opérateurs possédant des propriétés intéressantes.

7.3. Economie et représentation des connaissances

Les points précédents étaient consacrés aux apports possibles de l'économie (théorie du choix social et théorie des jeux) à l'intelligence artificielle. Ce dernier point concerne le chemin inverse, c'est-à-dire d'étudier les apports possibles de l'intelligence artificielle en économie.

Nous avons déjà quelques résultats à ce sujet, mais nous ne les avons pas discutés dans cette synthèse car ils ne concernent pas le raisonnement mais la théorie de la décision ou la théorie des jeux [53, 34, 57, 70, 33, 69].

Il s'agit en particulier d'étudier comment les concepts de théorie des jeux supportent des cadres qualitatifs.

En ce qui concerne l'incertitude, la quasi-totalité des travaux en théorie des jeux se basent sur un modèle d'incertitude bayésien (probabiliste). Nous avons déjà étudié deux cas particuliers : la décision sous ignorance totale [53] et la théorie des jeux (non coopérative) sous incertitude stricte [34, 57, 33]. Nous souhaitons continuer dans cette voie et étudier ce que ce changement de cadre d'incertitude a comme conséquences dans d'autres cadres de théorie des jeux, notamment dans le cadre de la théorie des jeux coopérative.

Il est également intéressant d'étudier ce qui survient lorsque l'on utilise un modèle qualitatif (ordinal) pour représenter les préférences des agents. Il est d'usage en théorie des jeux de représenter les préférences des agents à l'aide de fonctions d'utilité numériques. Ce choix permet d'utiliser facilement des fonctions et résultats mathématiques. En particulier un certain nombre de résultats importants dépendent des hypothèses de continuité, de convexité et de compacité obtenues grâce à ce choix de fonctions d'utilité numériques. Nous souhaitons étudier ce que deviennent ces résultats lorsque l'on abandonne ces hypothèses et que l'on se place dans un cadre discret (et qualitatif).

Deuxième partie

Articles

Liste des articles

Raisonnement sous incohérence

- Anthony Hunter, Sébastien Konieczny. Shapley Inconsistency Values. *Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR' 06)*. 249-259. 2006.
- Anthony Hunter, Sébastien Konieczny. Measuring inconsistency through minimal inconsistent sets. *Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*. 358-366. 2008.

Révision

- Sébastien Konieczny, Ramón Pino Pérez. Improvement operators. *Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR' 08)*. 177-186. 2008.
- Andreas Herzig, Sébastien Konieczny, Laurent Perussel. On iterated revision in the AGM framework. *Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03)*. 477-488. 2003.

Fusion

- Sébastien Konieczny, Jérôme Lang, Pierre Marquis. DA^2 merging operators. *Artificial Intelligence*. 157(1-2). 49-79. 2004.
- Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, Pierre Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence*. 171. 740-753. 2007.

Négociation

- Sébastien Konieczny. Belief base merging as a game. *Journal of Applied Non-Classical Logics*. 14(3). 275-294. 2004.
- Sébastien Konieczny, Ramón Pino Pérez. Confluence operators. *Eleventh European Conference on Logics in Artificial Intelligence (JELIA'08)*. 272-284. 2008.

SHAPLEY INCONSISTENCY VALUES

Anthony Hunter, Sébastien Konieczny.
Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR'06).
pages 249-259.
2006.

Shapley Inconsistency Values

Anthony Hunter
Department of Computer Science
University College London, UK
a.hunter@cs.ucl.ac.uk

Sébastien Konieczny
CRIL - CNRS
Université d'Artois, Lens, France
konieczny@cril.univ-artois.fr

Abstract

There are relatively few proposals for inconsistency measures for propositional belief bases. However inconsistency measures are potentially as important as information measures for artificial intelligence, and more generally for computer science. In particular, they can be useful to define various operators for belief revision, belief merging, and negotiation. The measures that have been proposed so far can be split into two classes. The first class of measures takes into account the number of formulae required to produce an inconsistency: the more formulae required to produce an inconsistency, the less inconsistent the base. The second class takes into account the proportion of the language that is affected by the inconsistency: the more propositional variables affected, the more inconsistent the base. Both approaches are sensible, but there is no proposal for combining them. We address this need in this paper: our proposal takes into account both the number of variables affected by the inconsistency and the distribution of the inconsistency among the formulae of the base. Our idea is to use existing inconsistency measures (ones that takes into account the proportion of the language affected by the inconsistency, and so allow us to look inside the formulae) in order to define a game in coalitional form, and then to use the Shapley value to obtain an inconsistency measure that indicates the responsibility/contribution of each formula to the overall inconsistency in the base. This allows us to provide a more reliable image of the belief base and of the inconsistency in it.

Introduction

There are numerous works on reasoning under inconsistency. One can quote for example paraconsistent logics, argumentation frameworks, belief revision and fusion, etc. All these approaches illustrate the fact that the dichotomy between consistent and inconsistent sets of formulae that comes from classical logics is not sufficient for describing these sets. As shown by these works two inconsistent sets of formulae are not trivially equivalent. They do not contain the same information and they do not contain the same contradictions.

Measures of information *à la* Shannon have been studied in logical frameworks (see for example (Kemeny 1953)). Roughly they involve counting the number of models of the set of formulae (the less models, the more informative the set). The problem is that these measures give a null information content to an inconsistent set of formulae, which is counter-intuitive (especially given all the proposals for paraconsistent reasoning). So generalizations of measures of information have been proposed to solve this problem (Lozinskii 1994; Wong & Besnard 2001; Knight 2003; Konieczny, Lang, & Marquis 2003; Hunter & Konieczny 2005).

In comparison, there are relatively few proposals for inconsistency measures (Grant 1978; Hunter 2002; Knight 2001; Konieczny, Lang, & Marquis 2003; Hunter 2004; Grant & Hunter 2006). However, these measures are potentially important in diverse applications in artificial intelligence, such as belief revision, belief merging, and negotiation, and more generally in computer science. Already measuring inconsistency has being seen to be a useful tool in analysing a diverse range of information types including news reports (Hunter 2006), integrity constraints (Grant & Hunter 2006), software specifications (Barragáns-Martínez, Pazos-Arias, & Fernández-Vilas 2004; 2005; Mu *et al.* 2005), and ecommerce protocols (Chen, Zhang, & Zhang 2004).

The current proposals for measuring inconsistecy can be classified in two approaches. The first approach involves "counting" the minimal number of formulae needed to produce the inconsistency. The more formulae needed to produce the inconsistency, the less inconsistent the set (Knight 2001). This idea is an interesting one, but it rejects the possibility of a more fine-grained inspection of the (content of the) formulae. In particular, if one looks to singleton sets only, one is back to the initial problem, with only two values: consistent or inconsistent.

The second approach involves looking at the proportion of the language that is touched by the inconsistency. This allows us to look *inside* the formulae (Hunter 2002; Konieczny, Lang, & Marquis 2003; Grant & Hunter 2006). This means that two formulae (singleton sets) can have different inconsistency measures. But, in these approaches one can identify the set of formulae with its conjunction (i.e. the set $\{\varphi, \varphi'\}$ has the same inconsistency measure as the set $\{\varphi \wedge \varphi'\}$). This can be sensible in several applications, but this means that the distribution of the contradiction among the formulae is not taken into account.

What we propose in this paper is a definition for inconsistency measures that allow us to take the best of the two approaches. This will allow us to build inconsistency measures that are able to look inside the formulae, but also to take into account the distribution of the contradiction among the different formulae of the set. The advantage of such a method is twofold. First, this allows us to know the degree of blame/responsability of each formula of the base in the inconsistency, and so it provides a very detailed view of the inconsistency. Second, this allows us to define measures of consistency for the whole base that are more accurate, since they take into account those two dimensions.

To this end we will use a notion that comes from coalitional game theory: the Shapley value. This value assigns to each player the payoff that this player can expect from her utility for each possible coalition. The idea is to use existing inconsistency measures (that allow us to look inside the formulae) in order to define a game in coalitional form, and then to use the Shapley value to obtain an inconsistency measure with the wanted properties. We will study these measures and show that they are more interesting than the other existing measures.

After stating some notations and definitions in the next section, we introduce inconsistency measures that count the number of formulae needed to produce an inconsistency. Then we present the approaches where the inconsistency measure is related to the number of variables touched by the inconsistency. The next section gives the definition of coalitional games and of the Shapley value. Then we introduce the inconsistency measures based on the Shapley value. The penultimate section sketches the possible applications of those measures for belief change operators. In the last section we conclude and give perspectives of this work.

Preliminaries

We will consider a propositional language $\mathcal{L}$ built from a finite set of propositional symbols $\mathcal{P}$. We will use $a, b, c, \ldots$ to denote the propositional variables, and Greek letters $\alpha, \beta, \varphi, \ldots$ to denote the formulae. An interpretation is a total function from $\mathcal{P}$ to $\{0, 1\}$. The set of all interpretations is denoted $\mathcal{W}$. An interpretation ω is a model of a formula φ, denoted $\omega \models \varphi$, if and only if it makes φ true in the usual truth-functional way. $Mod(\varphi)$ denotes the set of models of the formula φ, i.e. $Mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$. We will use $\subseteq$ to denote the set inclusion, and we will use $\subset$ to denote the strict set inclusion, i.e. $A \subset B$ iff $A \subseteq B$ and $B \not\subseteq A$. We will denote the set of real numbers by $\mathbb{R}$.

A *belief base* K is a finite set of propositional formulae. More exactly, as we will need to identify the different formulae of a belief base in order to associate them with their inconsistency value, we will consider belief bases K as vectors of formulae. For logical properties we will need to use the set corresponding to each vector, so we suppose that we have a function such that for each vector $K = (\alpha_1, \ldots, \alpha_n)$, $\overline{K}$ is the set $\{\alpha_1, \ldots, \alpha_n\}$. As it will never be ambigous, in the following we will omit the $\overline{}$ and write K as both the vector and the set.

Let us note $\mathcal{K}_{\mathcal{L}}$ the set of belief bases definable from formulae of the language $\mathcal{L}$. A belief base is consistent if there is at least one interpretation that satisfies all its formulae.

If a belief base K is not consistent, then one can define the minimal inconsistent subsets of K as:

$$\mathsf{MI}(K) = \{K' \subseteq K \mid K' \vdash \bot \text{ and } \forall K'' \subset K', K'' \nvdash \bot\}$$

If one wants to recover consistency from an inconsistent base K by removing some formulae, then the minimal inconsistent subsets can be considered as the purest form of inconsistency. To recover consistency, one has to remove at least one formula from each minimal inconsistent subset (Reiter 1987).

A *free formula* of a belief base K is a formula of K that does not belong to any minimal inconsistent subset of the belief base K, or equivalently any formula that belongs to any maximal consistent subset of the belief base.

Inconsistency Measures based on Formulae

When a base is not consistent the classical inference relation trivializes, since one can deduce every formula of the language from the base (*ex falso quodlibet*). Otherwise in this case the use of paraconsistent reasoning techniques allows us to draw non-trivial consequences from the base. One possibility is to take maximal consistent subsets of formulae of the base (cf (Manor & Rescher 1970; Benferhat, Dubois, & Prade 1997; Nebel 1991)). This idea can also be used to define an inconsistency measure. This is the way followed in (Knight 2001; 2003).

Definition 1 *A probability function on $\mathcal{L}$ is a function P : $\mathcal{P} \rightarrow [0, 1]$ s.t.:*

- *if $\models \alpha$, then $P(\alpha) = 1$*
- *if $\models \neg(\alpha \wedge \beta)$, then $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$*

See (Paris 1994) for more details on this definition. In the finite case, this definition gives a probability distribution on the interpretations, and the probability of a formula is the sum of the probability of its models.

Then the inconsistency measure defined by Knight (2001) is given by:

Definition 2 *Let K be a belief base.*

- *K is $\eta-$consistent ($0 \leq \eta \leq 1$) if there is a probability function P such that $P(\alpha) \geq \eta$ for all $\alpha \in K$.*
- *K is maximally $\eta-$consistent if η is maximal (i.e. if $\gamma > \eta$ then K is not $\gamma-$consistent).*

The notion of *maximal η-consistency* can be used as an inconsistency measure. This is the direct formulation of the idea that the more formulae are needed to produce the inconsistency, the less this inconsistency is problematic. As it is easily seen, in the finite case, a belief base is maximally 0-consistent if and only if it contains a contradictory formula. And a belief base is maximally 1-consistent if and only if it is consistent.

Example 1 *Let $K_1 = \{a, b, \neg a \vee \neg b\}$.*
K_1 is maximally $\frac{2}{3}-$consistent.
Let $K_2 = \{a \wedge b, \neg a \wedge \neg b, a \wedge \neg b\}$.
K_2 is maximally $\frac{1}{3}-$consistent, whereas each subbase of cardinality 2 is maximally $\frac{1}{2}-$consistent.

For minimal inconsistent sets of formulae, computing this inconsistency measure is easy:

Proposition 1 *If $K' \in \mathsf{MI}(K)$, then K' is maximally $\frac{|K'|-1}{|K'|}$−consistent.*

But in general this measure is harder to compute. However it is possible to compute it using the simplex method (Knight 2001).

Inconsistency Measures based on Variables

Another method to evaluate the inconsistency of a belief base is to look at the proportion of the language concerned with the inconsistency. To this end, it is clearly not possible to use classical logics, since the inconsistency contaminates the whole language. But if we look at the two bases $K_1 = \{a \wedge \neg a \wedge b \wedge c \wedge d\}$ and $K_2 = \{a \wedge \neg a \wedge b \wedge \neg b \wedge c \wedge \neg c \wedge d \wedge \neg d\}$, we can observe that in K_1 the inconsistency is mainly about the variable a, whereas in K_2 all the variables are touched by a contradiction. This is this kind of distinction that these approaches allow.

One way to circumscribe the inconsistency only to the variables directly concerned is to use multi-valued logics, and especially three-valued logics, with the third "truth value" denoting the fact that there is a conflict on the truth value (true-false) of the variable.

We do not have space here to detail the range of different measures that have been proposed. See (Grant 1978; Hunter 2002; Konieczny, Lang, & Marquis 2003; Hunter & Konieczny 2005; Grant & Hunter 2006) for more details on these approaches. We only give one such measure, that is a special case of the degrees of contradiction defined in (Konieczny, Lang, & Marquis 2003). The idea of the definition of these degrees in (Konieczny, Lang, & Marquis 2003) is, given a set of tests on the truth value of some formulae of the language (typically on the variables), the degree of contradiction is the cost of a minimum test plan that ensures recovery of consistency.

The inconsistency measure we define here is the (normalized) minimum number of inconsistent truth values in the LP_m models (Priest 1991) of the belief base. Let us first introduce the LP_m consequence relation.

- An interpretation ω for LP_m maps each propositional atom to one of the three "truth values" $\mathsf{F}, \mathsf{B}, \mathsf{T}$, the third truth value B meaning intuitively "both true and false". $3^{\mathcal{P}}$ is the set of all interpretations for LP_m. "Truth values" are ordered as follows: $\mathsf{F} <_t \mathsf{B} <_t \mathsf{T}$.
  - $\omega(\top) = \mathsf{T}, \omega(\bot) = \mathsf{F}$
  - $\omega(\neg\alpha) = \mathsf{B}$ iff $\omega(\alpha) = \mathsf{B}$
    $\omega(\neg\alpha) = \mathsf{T}$ iff $\omega(\alpha) = \mathsf{F}$
  - $\omega(\alpha \wedge \beta) = \min_{\leq_t}(\omega(\alpha), \omega(\beta))$
  - $\omega(\alpha \vee \beta) = \max_{\leq_t}(\omega(\alpha), \omega(\beta))$
- The set of models of a formula φ is:
  $$Mod_{LP}(\varphi) = \{\omega \in 3^{\mathcal{P}} \mid \omega(\varphi) \in \{\mathsf{T}, \mathsf{B}\}\}$$
  Define $\omega!$ as the set of "inconsistent" variables in an interpretation w, i.e.
  $$\omega! = \{x \in \mathcal{P} \mid \omega(x) = \mathsf{B}\}$$
  Then the minimum models of a formula are the "most classical" ones:
  $$min(Mod_{LP}(\varphi)) = \{\omega \in Mod_{LP}(\varphi) \mid \nexists \omega' \in Mod_{LP}(\varphi) \text{ s.t. } \omega'! \subset \omega!\}$$
  The LP_m consequence relation is then defined by:
  $$K \models_{LP_m} \varphi \text{ iff } min(Mod_{LP}(K)) \subseteq Mod_{LP}(\varphi)$$
  So φ is a consequence of K if all the "most classical" models of K are models of φ.

Then let us define the LP_m measure of inconsistency, noted I_{LP_m}, as:

Definition 3 *Let K be a belief base.*

$$I_{LP_m} = \frac{\min_{\omega \in Mod_{LP}(K)}(\mid \omega! \mid)}{\mid \mathcal{P} \mid}$$

Example 2 $K_4 = \{a \wedge \neg a, b, \neg b, c\}$. $I_{LP_m}(K_4) = \frac{2}{3}$

In this example one can see the point in these kinds of measures compared to measures based on formulae since this base is maximally 0-consistent because of the contradictory formula $a \wedge \neg a$. But there are also non-trivial formulae in the base, and this base is not very inconsistent according to I_{LP_m}.

Conversely, measures based on variables like this one are unable to take into account the distribution of the contradiction among formulae. In fact the result would be exactly the same with $K_4' = \{a \wedge \neg a \wedge b \wedge \neg b \wedge c\}$. This can be sensible in several applications, but in some cases this can also be seen as a drawback.

Games in Coalitional Form - Shapley Value

In this section we give the definitions of games in coalitional form and of the Shapley value.

Definition 4 *Let $N = \{1, \ldots, n\}$ be a set of n players. A game in coalitional form is given by a function $v : 2^N \rightarrow \mathbb{R}$, with $v(\emptyset) = 0$.*

This framework defines games in a very abstract way, focusing on the possible coalitions formations. A coalition is just a subset of N. This function gives what payoff can be achieved by each coalition in the game v when all its members act together as a unit.

There are numerous questions that are worthwhile to investigate in this framework. One of these questions is to know how much each player can expect in a given game v. This depend on her position in the game, i.e. what she brings to different coalitions.

Often the games are super-additive.

Definition 5 *A game is* super-additive *if for each $T, U \subseteq N$ with $T \cap U = \emptyset$, $v(T \cup U) \geq v(T) + v(U)$.*

In super-additive games when two coalitions join, then the joined coalition wins at least as much as (the sum of) the initials coalitions. In particular, in super-additive games, the grand coalition N is the one that brings the higher utility for the society N. The problem is how this utility can be shared among the players[1].

Example 3 *Let $N = \{1, 2, 3\}$, and let v be the following coalitional game:*

$$\begin{array}{ccc} v(\{1\}) = 1 & v(\{2\}) = 0 & v(\{3\}) = 1 \\ v(\{1,2\}) = 10 & v(\{1,3\}) = 4 & v(\{2,3\}) = 11 \\ & v(\{1,2,3\}) = 12 & \end{array}$$

This game is clearly super-additive. The grand coalition can bring 12 to the three players. This is the highest utility achievable by the group. But this is not the main aim for all the players. In particular one can note that two coalitions can bring nearly as much, namely $\{1, 2\}$ and $\{2, 3\}$ that gives respectively 10 and 11, that will have to be shared only between 2 players. So it is far from certain that the grand coalition will form in this case. Another remark on this game is that all the players do not share the same situation. In particular player 2 is always of a great value for any coalition she joins. So she seems to be able to expect more from this game than the other players. For example she can make an offer to player 3 for making the coalition $\{2, 3\}$, that brings 11, that will be split in 8 for player 2 and 3 for player 3. As it will be hard for player 3 to win more than that, 3 will certainly accept.

A solution concept has to take into account these kinds of arguments. It means that one wants to *solve* this game by stating what is the payoff that is "due" to each agent. That requires to be able to quantify the payoff that an agent can claim with respect to the power that her position in the game offers (for example if she always significantly improves the payoff of the coalitions she joins, if she can threat to form another coalition, etc.).

Definition 6 *A* value *is a function that assigns to each game v a vector of payoff $S(v) = (S_1, \dots, S_n)$ in $\mathbb{R}^n$.*

This function gives the payoff that can be expected by each player i for the game v, i.e. it measures i's power in the game v.

Shapley proposes a beautiful solution to this problem. Basically the idea can be explained as follows: considering that the coalitions form according to some order (a first player enters the coalition, then another one, then a third one, etc), and that the payoff attached to a player is its marginal utility (i.e. the utility that it brings to the existing coalition), so if C is a coalition (subset of N) not containing i, player's i marginal utility is $v(C \cup \{i\}) - v(C)$. As one can not make any hypothesis on which order is the correct one, suppose that each order is equally probable. This leads to the following formula:

Let σ be a permutation on N, with σ_n denoting all the possible permutations on N. Let us note

$$p^i_\sigma = \{j \in N \mid \sigma(j) < \sigma(i)\}$$

That means that p^i_σ represents all the players that precede player i for a given order σ.

[1]One supposes the transferable utility (TU) assumption, i.e. the utility is a common unit between the players and sharable as needed (roughly, one can see this utility as some kind of money).

Definition 7 *The Shapley value of a game v is defined as:*

$$S_i(v) = \frac{1}{n!} \sum_{\sigma \in \sigma_n} v(p^i_\sigma \cup \{i\}) - v(p^i_\sigma)$$

The Shapley value can be directly computed from the possible coalitions (without looking at the permutations), with the following expression:

$$S_i(v) = \sum_{C \subseteq N} \frac{(c-1)!(n-c)!}{n!} (v(C) - v(C \setminus \{i\}))$$

where c is the cardinality of C.

Example 4 *The Shapley value of the game defined in Example 3 is $(\frac{17}{6}, \frac{35}{6}, \frac{20}{6})$.*

These values show that it is player 2 that is the best placed in this game, accordingly to what we explained when we presented Example 3.

Besides this value, Shapley proposes axiomatic properties a value should have.

- $\sum_{i \in N} S_i(v) = v(N)$ **(Efficiency)**
- If i and j are such that for all C s.t. $i, j \notin C$, $v(C \cup \{i\}) = v(C \cup \{j\})$, then $S_i(v) = S_j(v)$ **(Symmetry)**
- If i is such that $\forall C\ v(C \cup \{i\}) = v(C)$, then $S_i(v) = 0$ **(Dummy)**
- $S_i(v + w) = S_i(v) + S_i(w)$ **(Additivity)**

These four axioms seem quite sensible. Efficiency states that the payoff available to the grand coalition N must be efficiently redistributed to the players (otherwise some players could expect more that what they have). Symmetry ensures that it is the role of the player in the game in coalitional form that determines her payoff, so it is not possible to distinguish players by their name (as far as payoffs are concerned), but only by their respective merits/possibilities. So if two players always are identical for the game, i.e. if they bring the same utility to every coalitions, then they have the same value. The dummy player axiom says simply that if a player is of no use for every coalition, this player does not deserve any payoff. And additivity states that when we join two different games v and w in a whole super-game $v + w$ ($v + w$ is straightforwardly defined as the function that is the sum of the two functions v and w, that means that each coalition receive as payoff in the game $v + w$ the payoff it has in v plus the payoff it has in w), then the value of each player in the supergame is simply the sum of the values in the compound games.

These properties look quite natural, and the nice result shown by Shapley is that they characterize exactly the value he defined (Shapley 1953):

Proposition 2 *The Shapley value is the only value that satisfies all of Efficiency, Symmetry, Dummy and Additivity.*

This result supports several variations : there are other equivalent axiomatizations of the Shapley value, and there are some different values that can be defined by relaxing some of the above axioms. See (Aumann & Hart 2002).

Inconsistency Values using Shapley Value

Given an inconsistency measure, the idea is to take it as the payoff function defining a game in coalitional form, and then using the Shapley value to compute the part of the inconsistency that can be imputed to each formula of the belief base.

This allows us to combine the power of inconsistency measures based on variables and hence discriminating between singleton inconsistent belief base (like Coherence measure in (Hunter 2002), or like the test action values of (Konieczny, Lang, & Marquis 2003)), and the use of the Shapley value for knowing what is the responsibility of a given formula in the inconsistency of the belief base.

We just require some basic properties on the underlying inconsistency measure.

Definition 8 *An inconsistency measure I is called a basic inconsistency measure if it satisfies the following properties, $\forall K, K' \in \mathcal{K}_{\mathcal{L}}$, $\forall \alpha, \beta \in \mathcal{L}$:*

- $I(K) = 0$ *iff K is consistent* **(Consistency)**
- $0 \leq I(K) \leq 1$ **(Normalization)**
- $I(K \cup K') \geq I(K)$ **(Monotony)**
- *If α is a free formula of $K \cup \{\alpha\}$, then $I(K \cup \{\alpha\}) = I(K)$* **(Free Formula Independence)**
- *If $\alpha \vdash \beta$ and $\alpha \nvdash \bot$, then $I(K \cup \{\alpha\}) \geq I(K \cup \{\beta\})$* **(Dominance)**

We ask for few properties on the initial inconsistency measure. The consistency property states that a consistent base has a null inconsistency measure. The monotony property says that the amount of inconsistency of a belief base can only grow if one adds new formulae (defined on the same language). The free formula independence property states that adding a formula that does not cause any inconsistency cannot change the inconsistency measure of the base. The Dominance property states that logically stronger formulae bring (potentially) more conflicts. The normalization property of the inconsistency measure is not mandatory, it is asked only for simplification purposes.

Now we are able to define the Shapley inconsistency values :

Definition 9 *Let I be a basic inconsistency measure. We define the corresponding Shapley inconsistency value (SIV), noted S_I, as the Shapley value of the coalitional game defined by the function I, i.e. let $\alpha \in K$:*

$$S_I^K(\alpha) = \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$$

where n is the cardinality of K and c is the cardinality of C.

Note that this SIV gives a value for each formula of the base K, so if one considers the base K as the vector $K = (\alpha_1, \ldots, \alpha_n)$, then we will use $S_I(K)$ to denote the vector of corresponding SIVs, i.e.

$$S_I(K) = (S_I^K(\alpha_1), \ldots, S_I^K(\alpha_n))$$

This definition allows us to define to what extent a formula inside a belief base is concerned with the inconsistencies of the base. It allows us to draw a precise picture of the contradiction of the base.

From this value, one can define an inconsistency value for the whole belief base:

Definition 10 *Let K be a belief base, $\hat{S}_I(K) = \max_{\alpha \in K} S_I(\alpha)$*

One can figure out other aggregation functions to define the inconsistency measure of the belief base from the inconsistency measure of its formulae, such as the leximax for instance. Taking the maximum will be sufficient for us to have valuable results and to compare this with the existing measures from the literature. Note that taking the sum as aggregation function is not a good choice here, since as shown by the distribution property of Theorem 3 this equals $I(K)$, "erasing" the use of the Shapley value.

We think that the most interesting measure is S_I, since it describes more accurately the inconsistency of the base. But we define $\hat{S}_I$ since it is a more concise measure, that is of the same type as existing ones (it associates a real to each base), that is convenient to compare our framework with existing measures.

Let us see see now two instantiations of SIVs.

Drastic Shapley Inconsistency Value

We will start this section with the simplest inconsistency measure one can define:

Definition 11 *The drastic inconsistency value is defined as:*

$$I_d(K) = \begin{cases} 0 & \text{if } K \text{ is consistent} \\ 1 & \text{otherwise} \end{cases}$$

This measure is not of great interest by itself, since it corresponds to the usual dichotomy of classical logic. But it will be useful to illustrate the use of the Shapley inconsistency values, since, even with this over-simple measure, one will produce interesting results. Let us illustrate this on some examples.

Example 5 $K_1 = \{a, \neg a, b\}$.
Then $I_d(\{a, \neg a\}) = I_d(\{a, \neg a, b\}) = 1$, and the value is $S_{I_d}(K_1) = (\frac{1}{2}, \frac{1}{2}, 0)$. So $\hat{S}_{I_d}(K_1) = \frac{1}{2}$.

As b is a free formula, it has a value of 0, the two other formulae are equally responsible for the inconsistency.

Example 6 $K_2 = \{a, b, b \wedge c, \neg b \wedge d\}$.
Then the value is $S_{I_d}(K_2) = (0, \frac{1}{6}, \frac{1}{6}, \frac{4}{6})$.
And $\hat{S}_{I_d}(K_2) = \frac{2}{3}$.

The last three formulae are the ones that belong to some inconsistency, and the last one is the one that causes the most problems (removing only this formula restores the consistency of the base).

Example 7 $K_4 = \{a \wedge \neg a, b, \neg b, c\}$.
The value is $S_{I_d}(K_4) = (\frac{4}{6}, \frac{1}{6}, \frac{1}{6}, 0)$. So $\hat{S}_{I_d}(K_4) = \frac{2}{3}$.

$\mathbf{LP_m}$ Shapley Inconsistency Value

Let us turn now to a more elaborate value. For this we use the LP_m inconsistency measure (defined earlier) to define a SIV.

Example 8 *Let* $K_4 = \{a \wedge \neg a, b, \neg b, c\}$
and $K'_4 = \{a \wedge \neg a \wedge b \wedge \neg b \wedge c\}$.
Then $S_{I_{LP_m}}(K_4) = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, 0)$*, and* $\hat{S}_{I_{LP_m}}(K_4) = \frac{1}{3}$.
Whereas $S_{I_{LP_m}}(K'_4) = (\frac{2}{3})$ *and* $\hat{S}_{I_{LP_m}}(K'_4) = \frac{2}{3}$.

As we can see on this example, the SIV value allows us to make a distinction between K_4 and K'_4, since $\hat{S}_{I_{LP_m}}(K'_4) = \frac{2}{3}$ whereas $\hat{S}_{I_{LP_m}}(K_4) = \frac{1}{3}$. This illustrates the fact that the inconsistency is more distributed in K_4 than in K'_4. This distinction is not possible with the original I_{LP_m} value. Note that with Knight's coherence value, the two bases have the worst inconsistency value (maximally 0-consistent).

So this example illustrates the improvement brought by this work, compared to inconsistency measures on formulae and to inconsistency measures on variables, since none of them was able to make a distinction between K_4 and K'_4, whereas for $\hat{S}_{I_{LP_m}}$ K_4 is more consistent than K'_4.

Let us see a more striking example.

Example 9 *Let* $K_5 = \{a, b, b \wedge c, \neg b \wedge \neg c\}$.
Then $S_{I_{LP_m}}(K_5) = (0, \frac{1}{18}, \frac{4}{18}, \frac{7}{18})$*,*
and $\hat{S}_{I_{LP_m}}(K_5) = \frac{7}{18}$.

In this example one can easily see that it is the last formula that is the more problematic, and that $b \wedge c$ brings more conflict than b alone, which is perfectly expressed in the obtained values.

Logical properties

Let us see now some properties of the defined values.

Proposition 3 *Every Shapley Inconsistency Value satisfies:*

- $\sum_{\alpha \in K} S_I(\alpha) = I(K)$ **(Distribution)**
- *If* $\exists \alpha, \beta \in K$ *s.t. for all* $K' \subseteq K$ *s.t.* $\alpha, \beta \notin K'$, $I(K' \cup \{\alpha\}) = I(K' \cup \{\beta\})$*, then* $S_I(\alpha) = S_I(\beta)$ **(Symmetry)**
- *If* α *is a free formula of* K*, then* $S_I(\alpha) = 0$ **(Free Formula)**
- *If* $\alpha \vdash \beta$ *and* $\alpha \nvdash \perp$*, then* $S_I(\alpha) \geq S_I(\beta)$ **(Dominance)**

The distribution property states that the inconsistency values of the formulae sum to the total amount of inconsistency in the base ($I(K)$). The Symmetry property ensures that only the amount of inconsistency brought by a formula matters for computing the SIV. As one could expect, a formula that is not embedded in any contradiction (i.e. does not belong to any minimal inconsistent subset) will not be blamed by the Shapley inconsistency values. This is what is expressed in the Free formula property. The Dominance property states that logically stronger formulae bring (potentially) more conflicts.

The first three properties are a restatement in this logical framework of the properties of the Shapley value. One can note that the Additivity axiom of the Shapley value is not translated here, since it makes little sense to add different inconsistency values.

Let us turn now to the properties of the measure on belief bases.

Proposition 4

- $\hat{S}_I(K) = 0$ *if and only if* K *is consistent* **(Consistency)**
- $0 \leq \hat{S}_I(K) \leq 1$ **(Normalization)**
- *If* α *is a free formula of* $K \cup \{\alpha\}$*, then* $\hat{S}_I(K \cup \{\alpha\}) = \hat{S}_I(K)$ **(Free Formula Independence)**
- $\hat{S}_I(K) \leq I(K)$ **(Upper Bound)**
- $\hat{S}_I(K) = I(K) > 0$ *if only if* $\exists \alpha \in K$ *s.t.* α *is inconsistent and* $\forall \beta \in K$, $\beta \neq \alpha$, β *is a free formula of* K **(Isolation)**

The first three properties are the ones given in Definition 8 for the basic inconsistency measures. As one can easily note an important difference is that the monotony property and the dominance property do not hold for the SIVs on belief bases. It is sensible since distribution of the inconsistencies matters for SIVs. The upper bound property shows that the use of the SIV aims at looking at the distribution of the inconsistencies of the base, so the SIV on belief bases is always less or equal to the inconsistency measure given by the underlying basic inconsistency measure. The isolation property details the case where the two measures are equals. In this case, there is only one inconsistent formula in the whole base.

Let us see, on Example 10, counter-examples to monotony and dominance for SIV on belief bases:

Example 10 *Let* $K_6 = \{a, \neg a, \neg a \wedge b\}$,
$K_7 = \{a, \neg a, \neg a \wedge b, a \wedge b\}$,
and $K_8 = \{a, \neg a, \neg a \wedge b, b\}$.
$\hat{S}_{I_d}(K_6) = \frac{2}{3}$, $\hat{S}_{I_d}(K_7) = \frac{1}{4}$, $\hat{S}_{I_d}(K_8) = \frac{2}{3}$.

On this example one can see why monotony can not be satisfied by SIV on belief bases. Clearly $K_6 \subset K_7$, but $\hat{S}_{I_d}(K_6) > \hat{S}_{I_d}(K_7)$. This is explained by the fact that the inconsistency is more *diluted* in K_7, than in K_8. In K_7 the formula a is the one that is the most blamed for the inconsistency ($S^{K_6}_{I_d}(a) = \hat{S}_{I_d}(K_6) = \frac{2}{3}$), since it appears in all inconsistent sets. Whereas in K_7 inconsistencies are equally caused by a and by $a \wedge b$, that decreases the responsability of a, and the whole inconsistency value of the base.

For a similar reason dominance is not satisfied, we clearly have $a \wedge b \vdash b$ (and $a \wedge b \nvdash \perp$), but $\hat{S}_{I_d}(K_7) < \hat{S}_{I_d}(K_8)$.

Applications for Belief Change Operators

As the measures we define allow us to associate with each formula its degree of responsibility for the inconsistency of the base, they can be used to guide any paraconsistent reasoning, or any repair of the base. Let us quote two such possible uses for belief change operators, first for belief revision and then for negotiation.

Iterated Revision and Transmutation Policies

The problem of belief revision is to incorporate a new piece of information which is more reliable than (and conflicting with) the old beliefs of the agent. This problem has received a nice answer in the work of Alchourron, Gardenfors, Makinson (Alchourrón, Gärdenfors, & Makinson 1985) in the one-step case. But when one wants to iterate revision (i.e. to generalize it to the n-steps case), there are numerous problems and no definitive answer has been reached in the purely qualitative case (Darwiche & Pearl 1997; Friedman & Halpern 1996). Using a partially quantitative framework, some proposals have given interesting results (see e.g. (Williams 1995; Spohn 1987)). Here "partially quantitative" means that the incoming piece of information needs to be labeled by a degree of confidence denoting how strongly we believe it. The problem in this framework is to justify the use of such a degree, what does it mean exactly and where does it come from. One possibility is to use an inconsistency measure (or a composite measure computed from an information measure (Lozinskii 1994; Knight 2003; Konieczny, Lang, & Marquis 2003) and an inconsistency measure) to determine this degree of confidence. Then one can define several policies for the agent (we can suppose that an agent accepts a new piece of information only if it brings more information than contradiction, etc). We can then use the partially quantitative framework to derive revision operators with a nice behaviour. In this setting, since the degree attached to the incoming information is not a given data, but computed directly from the information itself and the agent policy (behaviour with respect to information and contradiction, encoded by a composite measure) then the problem of the justification of the meaning of the degrees is avoided.

Negotiation

The problem of negotiation has been investigated recently under the scope of belief change tools (Booth 2001; 2002; 2006; Zhang *et al.* 2004; Meyer *et al.* 2004; Konieczny 2004; Gauwin, Konieczny, & Marquis 2005). The problem is to define operators that take as input belief profiles (multi-set of formulae[2]) and that produce a new belief profile that aims to be less conflicting. We call these kind of operators conciliation operators. The idea followed in (Booth 2002; 2006; Konieczny 2004) to define conciliation operators is to use an iterative process where at each step a set of formulae is selected. These selected formulae are logically weakened. The process stops when one reaches a consensus, i.e. a consistent belief profile[3]. Many interesting operators can be defined when one fixes the selection function (the function that selects the formulae that must be weaken at each round) and the weakening method. In (Konieczny 2004) the selection function is based on a notion of distance. It can be sensible if such a distance is meaningful in a particular application. If not, it is only an arbitrary choice. It would then be sensible to choose instead one of the inconsistency measures we defined in this paper. So the selection function would choose the formulae with the highest inconsistency value. These formulae are clearly the more problematic ones. More generally SIVs can be used to define new belief merging methods.

[2]More exactly belief profiles are sets of belief bases. We use this simplifying assumption just for avoiding technical details here.

[3]A belief profile is consistent if the conjunction of its formulae is consistent.

Conclusion

We have proposed in this paper a new framework for defining inconsistency values. The SIV values we introduce allow us to take into account the distribution of the inconsistency among the formulae of the belief base and the variables of the language. This is, as far as we know, the only definition that allows us to take both types of information into account, thus allowing to have a more precise picture of the inconsistency of a belief base. The perspectives of this work are numerous. First, as sketched in the previous section, the use of inconsistency measures, and especially the use of Shapley inconsistency values, can be valuable for several belief change operators, for instance for modelizations of negotiation. The Shapley value is not the only solution concept for coalitional games, so an interesting question is to know if other solutions concept can be sensible as a basis for defining other inconsistency measures. But the main way of research opened by this work is to study more closely the connections between other notions of (cooperative) game theory and the logical modelization of belief change operators.

Acknowledgments

The authors would like to thank CNRS and the Royal Society for travel funding while collaborating on this research. The second author is supported by the Région Nord/Pas-de-Calais and the European Community FEDER program.

Proofs

Proof of Proposition 3 : To show distribution, let us recall that

$$\begin{array}{rcl} S_I^K(\alpha) & = & \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\})) \\ & = & \frac{1}{n!}\sum_{\sigma\in\sigma_n} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha) \end{array}$$

where σ_n is the set of possible permutations on $\overline{K}$, and $p_\sigma^\alpha = \{\beta \in K \mid \sigma(\beta) < \sigma(\alpha)\}$. Now

$$\begin{aligned} \sum_{\alpha\in K} S_I(\alpha) &= \sum_{\alpha\in K} \frac{1}{n!}\sum_{\sigma\in\sigma_n} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha) \\ &= \frac{1}{n!}\sum_{\sigma\in\sigma_n}\sum_{\alpha\in K} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha) \end{aligned}$$

Now note that we can order the elements of K accordingly to σ when computing the inside sum, that gives:

$$\begin{aligned} &= \frac{1}{n!}\sum_{\sigma\in\sigma_n}[I(\{\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n)}\}) \\ &\qquad -I(\{\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n-1)}\})] \\ &\quad +[I(\{\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n-1)}\}) \\ &\qquad -I(\{\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n-2)}\})] \\ &\quad +\ldots+[I(\{\alpha_{\sigma(1)}\}) - I(\emptyset)] \\ &= \frac{1}{n!}\sum_{\sigma\in\sigma_n} I(\{\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(n)}\}) - I(\emptyset) \\ &= \frac{1}{n!}\, n!\, I(K) \\ &= I(K) \end{aligned}$$

To show symmetry, assume that there are $\alpha, \beta \in K$ s.t. for all $K' \subseteq K$ s.t. $\alpha, \beta \notin K'$, $I(K' \cup \{\alpha\}) = I(K' \cup \{\beta\})$.

Now by definition

$$S_I^K(\alpha) = \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$$

Let us show that $S_I^K(\alpha) = S_I^K(\beta)$ by showing (by cases) that the elements of the sum are the same:

If $\alpha \notin C$ and $\beta \notin C$, then $I(C) = I(C\setminus\{\alpha\}) = I(C\setminus\{\beta\})$, so $I(C) - I(C \setminus \{\alpha\}) = I(C) - I(C \setminus \{\beta\})$.

If $\alpha \in C$ and $\beta \in C$, then note that by hypothesis, as $\alpha, \beta \notin C \setminus \{\alpha, \beta\}$, we deduce that $I(C \setminus \{\alpha\}) = I(C \setminus \{\beta\})$. So $I(C) - I(C \setminus \{\alpha\}) = I(C) - I(C \setminus \{\beta\})$.

If $\alpha \in C$ and $\beta \notin C$. Then $I(C) - I(C\setminus\{\beta\}) = 0$, and let us note $I(C) - I(C\setminus\{\alpha\}) = a$. Let us note $C = C' \cup \{\alpha\}$, and $C'' = C' \cup \{\beta\}$. Now notice that $I(C'') - I(C''\setminus\{\alpha\}) = 0$, and as we can deduce $I(C \setminus \{\alpha\}) = I(C'' \setminus \{\beta\})$ by the hypothesis, we also have $I(C'') - I(C'' \setminus \{\beta\}) = a$.

To show the free formula property, just note that if α is a free formula of K, then for every subset C of K, by the free formula independence property of the basic inconsistency measure we have that for every C, such that $\alpha \in C$, $I(C) = I(C \setminus \alpha)$, so $I(C) - I(C \setminus \alpha) = 0$. Straightforwardly if $\alpha \notin C$, $I(C) = I(C \setminus \alpha)$. So the whole expression $S_I^K(\alpha) = \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$ sums to 0.

Finally, to show dominance we will proceed in a similar way than to show symmetry. Assume that $\alpha, \beta \in K$ are such that $\alpha \vdash \beta$ and $\alpha \nvdash \bot$. Then, by the dominance property of the underlying basic inconsistency measure, we know that for all $C \subseteq K$, $I(C \cup \{\alpha\}) \geq I(C \cup \{\beta\})$. Now by definition of the SIV $S_I^K(\alpha) = \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$. Let us show that $S_I^K(\alpha) \geq S_I^K(\beta)$ by showing (by cases) that the elements of the first sum are greater or equal to the corresponding elements of the second one:

If $\alpha \notin C$ and $\beta \notin C$, then $I(C) = I(C\setminus\{\alpha\}) = I(C\setminus\{\beta\})$, so $I(C) - I(C \setminus \{\alpha\}) \geq I(C) - I(C \setminus \{\beta\})$.

If $\alpha \in C$ and $\beta \in C$, then let us note $C \setminus \{\alpha\} = C' \cup \{\beta\}$. So we also have $C \setminus \{\beta\} = C' \cup \{\alpha\}$. Now note that by hypothesis $I(C' \cup \{\beta\}) \leq I(C' \cup \{\alpha\})$, so $I(C \setminus \{\alpha\}) \leq I(C\setminus\{\beta\})$. Hence $I(C) - I(C\setminus\{\alpha\}) \geq I(C) - I(C\setminus\{\beta\})$.

If $\alpha \in C$ and $\beta \notin C$. Then $I(C) - I(C\setminus\{\beta\}) = 0$, and let us note $I(C) - I(C\setminus\{\alpha\}) = a$. Let us note $C = C' \cup \{\alpha\}$, and $C'' = C' \cup \{\beta\}$. Now notice that $I(C'') - I(C''\setminus\{\alpha\}) = 0$. So $I(C) - I(C \setminus \{\beta\}) \geq I(C'') - I(C'' \setminus \{\alpha\})$ Note that $I(C)\setminus\{\beta\} = I(C'')\setminus\{\alpha\} = C'$. As we can deduce $I(C) \geq I(C'')$ by the hypothesis, we also have $I(C) - I(C\setminus\{\alpha\}) \geq I(C'') - I(C'' \setminus \{\beta\})$.

□

Proof of Proposition 4 : To prove consistency note that if K is consistent, then for every $C \subseteq K$, $I(C) = 0$ (this is a direct consequence of the consistency property of the underlying basic inconsistency measure). Then for every $\alpha \in K$, $S_I^K(\alpha) = 0$. Hence $\hat{S}_I(K) = \max_{\alpha \in K} S_I(\alpha) = 0$. For the only if direction, by contradiction, suppose that $\hat{S}_I(K) = 0$ and that K is not consistent. As K is not consistent, then by the consistency property of the underlying basic inconsistency measure $I(K) = a \neq 0$. By the distribution property of the SIV we know that $\sum_{\alpha \in K} S_I(\alpha) = a \neq 0$, then $\exists \alpha \in K$ such that $S_I(\alpha) > 0$, so $\hat{S}_I(K) = \max_{\alpha \in K} S_I(\alpha) > 0$. Contradiction.

The normalization property is a consequence of the definition of $\hat{S}_I(K)$ as a maximum of values that are all greater than zero, that ensures $0 \leq \hat{S}_I(K)$, and that are all smaller than 1. An easy way to show $\hat{S}_I(K) \leq 1$ is as a consequence of the upper bound property (shown below) $\hat{S}_I(K) \leq I(K)$ and of $I(K) \leq 1$ obtained by the normalization property of the underlying basic inconsistency measure I.

To show the free formula independence property, just notice that for any formula β that is a free formula of $K \cup \{\beta\}$, it is also a free formula of every of its subsets. It is easy to see from the definition that for any $\alpha \in K$, $S_I^K(\alpha) = S_I^{K \cup \{\beta\}}(\alpha)$. This is easier if we consider the second form of the definition: $S_I^K(\alpha) = \frac{1}{n!}\sum_{\sigma \in \sigma_n} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha)$ where σ_n is the set of possible permutations on $\overline{K}$. Now note that for $S_I^{K \cup \{\beta\}}(\alpha)$, the free formula does not bring any contradiction, so it does not change the marginal contribution of every other formulae. Let us call the extensions of a permutation σ on K by β, all the permutations of $K \cup \{\beta\}$ whose restriction on elements of K is identical to σ, i.e. an extension of $\sigma = (\alpha_1, \ldots, \alpha_n)$ by β is a permutation $\sigma' = (\alpha_1, \ldots, \alpha_i, \beta, \alpha_{i+1}, \ldots, \alpha_n)$. Now note that there are $n+1$ such extensions, and that if σ' is an extension of sigma, $I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha) = I(p_{\sigma'}^\alpha \cup \{\alpha\}) - I(p_{\sigma'}^\alpha)$. So $S_I^{K \cup \{\beta\}}(\alpha) = \frac{1}{(n+1)!}(n+1)\sum_{\sigma \in \sigma_n} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha)$ $= \frac{1}{n!}\sum_{\sigma \in \sigma_n} I(p_\sigma^\alpha \cup \{\alpha\}) - I(p_\sigma^\alpha) = S_I^K(\alpha)$. Now as we have for any $\alpha \in K$, $S_I^K(\alpha) = S_I^{K \cup \{\beta\}}(\alpha)$, we have $\hat{S}_I(K \cup \{\alpha\}) = \hat{S}_I(K)$.

The upper bound property is stated by rewriting $I(K)$ as $\sum_{\alpha \in K} S_I(\alpha)$ with the distribution property of the SIV, and by recalling the definition of $\hat{S}_I(K)$ as $\max_{\alpha \in K} S_I(\alpha)$. Now by noticing that for every vector $a = (a_1, \ldots, a_n)$, $\max_{a_i \in a} a_i \leq \sum_{a_i \in a} a_i$, we conclude $\max_{\alpha \in K} S_I(\alpha) \leq \sum_{\alpha \in K} S_I(\alpha)$, i.e. $\hat{S}_I(K) \leq I(K)$.

Let us show isolation. The if direction is straighforward: As α is inconsistent, K is inconsistent, and by the consistency and normalization properties of the underlying basic inconsistency measure we know that $I(K) > 0$. By the free formula property of SIV, for every free formula β of K we have $S_I(\beta) = 0$. As by the distribution property we have $\sum_{\alpha \in K} S_I(\alpha) = I(K)$, this means that $S_I(\alpha) = I(K)$, and that $\hat{S}_I(K) = \max_{\alpha \in K} S_I(\alpha) = S_I(\alpha)$. So $\hat{S}_I(K) = I(K) > 0$. For the only if direction suppose that $\hat{S}_I(K) = I(K)$, that means that $\max_{\alpha \in K} S_I(\alpha) = I(K)$. But, by the distribution property we know that $I(K) = \sum_{\alpha \in K} S_I(\alpha)$. So it means that $\max_{\alpha \in K} S_I(\alpha) = \sum_{\alpha \in K} S_I(\alpha) = I(K)$. There exists α such that $S_I(\alpha) = I(K)$ (consequence of the definition of the max), and if there exists a $\beta \neq \alpha$ such that

$S_I(\beta) > 0$, then $\sum_{\alpha \in K} S_I(\alpha) > I(K)$. Contradiction. So it means that there is α such that $S_I(\alpha) = I(K)$ and for every $\beta \neq \alpha$, $S_I(\beta) = 0$. That means that every β is a free formula, and that α is inconsistent.

$\square$

References

Alchourrón, C.; Gärdenfors, P.; and Makinson, D. 1985. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic* 50:510–530.

Aumann, R., and Hart, S., eds. 2002. *Handbook of Game Theory with Economic Applications*, volume 3. North-Holland.

Barragáns-Martínez, B.; Pazos-Arias, J. J.; and Fernández-Vilas, A. 2004. On Measuring Levels of Inconsistency in Multi-Perspective Requirements Specifications. In *Proceedings of the 1st Conference on the Principles of Software Engineering (PRISE 2004)*, 21–30.

Barragáns-Martínez, B.; Pazos-Arias, J. J.; and Fernández-Vilas, A. 2005. Merging Requirements Views with Incompleteness and Inconsistency. In *Proceedings of the IEEE Australian Software Engineering Conference (ASWEC'05)*. Brisbane, Australia: IEEE Computer Society Press.

Benferhat, S.; Dubois, D.; and Prade, H. 1997. Some syntactic approaches to the handling of inconsistent knowledge bases: a comparative study, part 1: the flat case. *Studia Logica* 58:17–45.

Booth, R. 2001. A negotiation-style framework for non-prioritised revision. In *Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK'01)*, 137–150.

Booth, R. 2002. Social contraction and belief negotiation. In *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, 374–384.

Booth, R. 2006. Social contraction and belief negotiation. *Information Fusion* 7(1):19–34.

Chen, Q.; Zhang, C.; and Zhang, S. 2004. A verification model for electronic transaction protocols. In *Advanced Web Technologies and Applications*, volume 3007 of *LNCS*, 824–833. Springer.

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

Friedman, N., and Halpern, J. Y. 1996. Belief revision: a critique. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*, 421–431.

Gauwin, O.; Konieczny, S.; and Marquis, P. 2005. Conciliation and consensus in iterated belief merging. In *Proceedings of the Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, 514–526.

Grant, J., and Hunter, A. 2006. Measuring inconsistency in knowledgebases. *Journal of Intelligent Information Systems*. (in press).

Grant, J. 1978. Classifications for inconsistent theories. *Notre Dame Journal of Formal Logic* 19:435–444.

Hunter, A., and Konieczny, S. 2005. Approaches to measuring inconsistent information. In *Inconsistency Tolerance*, volume LNCS 3300, 189–234. Springer.

Hunter, A. 2002. Measuring inconsistency in knowledge via quasi-classical models. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI'2002)*, 68–73. MIT Press.

Hunter, A. 2004. Logical comparison of inconsistent perspectives using scoring functions. *Knowledge and Information Systems Journal* 6(5):528–543.

Hunter, A. 2006. How to act on inconsistent news: Ignore, resolve, or reject. *Data and Knowledge Engineering*. (in press).

Kemeny, J. 1953. A logical measure function. *Journal of Symbolic Logic* 18:289–308.

Knight, K. M. 2001. Measuring inconsistency. *Journal of Philosophical Logic* 31:77–98.

Knight, K. M. 2003. Two information measures for inconsistent sets. *Journal of Logic, Language and Information* 12:227–248.

Konieczny, S.; Lang, J.; and Marquis, P. 2003. Quantifying information and contradiction in propositional logic through epistemic tests. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, 106–111.

Konieczny, S. 2004. Belief base merging as a game. *Journal of Applied Non-Classical Logics* 14(3):275–294.

Lozinskii, E. 1994. Resolving contradictions: A plausible semantics for inconsistent systems. *Journal of Automated Reasoning* 12:1–31.

Manor, R., and Rescher, N. 1970. On inferences from inconsistent information. *Theory and Decision* 1:179–219.

Meyer, T.; Foo, N.; Zhang, D.; and Kwok, R. 2004. Logical foundations of negotiation: Strategies and preferences. In *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR'04)*, 311–318.

Mu, K.; Jin, Z.; Lu, R.; and Liu, W. 2005. Measuring inconsistency in requirements engineering. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU05)*, volume 3571 of *LNAI*, 440–451. Springer.

Nebel, B. 1991. Belief revision and default reasoning: syntax-based approaches. In *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning (KR'91)*, 417–428.

Paris, J. B. 1994. *The uncertain reasoner's companion: a mathematical perspective*. Number 39 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Priest, G. 1991. Minimally inconsistent LP. *Studia Logica* 50:321–331.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence* 32:57–95.

Shapley, L. S. 1953. A value for n-person games. In Kuhn, H. W., and Tucker, A. W., eds., *Contributions to the Theory of Games II (Annals of Mathematics Studies 28)*, 307–317. Princeton University Press.

Spohn, W. 1987. Ordinal conditional functions: a dynamic theory of epistemic states. In Harper, W. L., and Skyrms, B., eds., *Causation in Decision, Belief Change, and Statistics*, volume 2. Princeton University Press. 105–134.

Williams, M. 1995. Iterated theory base change: A computational model. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, 1541–1547.

Wong, P., and Besnard, P. 2001. Paraconsistent reasoning as an analytic tool. *Journal of the Interest Group in Propositional Logic* 9:233–246.

Zhang, D.; Foo, N.; Meyer, T.; and Kwok, R. 2004. Negotiation as mutual belief revision. In *Proceedings of the Nineteenth American National Conference on Artificial Intelligence (AAAI'04)*, 317–322.

MEASURING INCONSISTENCY THROUGH MINIMAL INCONSISTENT SETS

Anthony Hunter, Sébastien Konieczny.
Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08).
pages 358-366.
2008.

Measuring Inconsistency through Minimal Inconsistent Sets

Anthony Hunter
Department of Computer Science
University College London, UK
a.hunter@cs.ucl.ac.uk

Sébastien Konieczny
CRIL - CNRS
Université d'Artois, Lens, France
konieczny@cril.fr

Abstract

In this paper, we explore the links between measures of inconsistency for a belief base and the minimal inconsistent subsets of that belief base. The minimal inconsistent subsets can be considered as the relevant part of the base to take into account to evaluate the amount of inconsistency. We define a very natural inconsistency value from these minimal inconsistent sets. Then we show that the inconsistency value we obtain is a particular Shapley Inconsistency Value, and we provide a complete axiomatization of this value in terms of five simple and intuitive axioms. Defining this Shapley Inconsistency Value using the notion of minimal inconsistent subsets allows us to look forward to a viable implementation of this value using SAT solvers.

Introduction

The need to develop robust, but principled, logic-based techniques for analysing inconsistent information is increasingly recognized as an important research area for artificial intelligence in particular, and for computer science in general (Bertossi, Hunter, & Schaub 2004). This interest stems from the recognition that the dichotomy between consistent and inconsistent sets of formulae that comes from classical logics is not sufficient for describing inconsistent information.

A number of proposals have been made for measuring the degree of information of a belief base in the presence of inconsistency (Lozinskii 1994; Wong & Besnard 2001; Knight 2003; Konieczny, Lang, & Marquis 2003), and for measuring the degree of inconsistency of a belief base (Grant 1978; Knight 2001; Hunter 2002; Knight 2003; Konieczny, Lang, & Marquis 2003; Hunter 2004; 2003; Grant & Hunter 2006; Hunter & Konieczny 2006; Grant & Hunter 2008). For a review see (Hunter & Konieczny 2004).

These measures are potentially important in diverse applications in artificial intelligence, such as belief revision, belief merging, negotiation, multi-agent systems, decision-support, and software engineering tools. Already, measuring inconsistency has been seen to be a useful tool in analysing a diverse range of information types including news reports (Hunter 2006), integrity constraints (Grant & Hunter 2006), information merging (Qi, Liu, & Bell 2005), databases (Martinez *et al.* 2007), ontologies (Ma *et al.* 2007), software specifications (Barragáns-Martńez, Pazos-Arias, & Fernández-Vilas 2004; Mu *et al.* 2005), and ecommerce protocols (Chen, Zhang, & Zhang 2004).

Each of the current proposals for measuring inconsistency can be described as being one of the following two approaches.

The first approach involves "counting" the minimal number of formulae needed to produce the inconsistency in a set of formulae. The more formulae needed to produce the inconsistency, the less inconsistent the set (Knight 2001). This idea is an interesting one, but it rejects the possibility of a more fine-grained inspection of the (content of the) formulae. In particular, if one looks to singleton sets only, one is back to the initial problem, with only two values: consistent or inconsistent.

The second approach involves looking at the proportion of the language that is touched by the inconsistency in a set of formulae. This allows us to look *inside* the formulae (Hunter 2002; Konieczny, Lang, & Marquis 2003; Grant & Hunter 2006; 2008). This means that two formulae (singleton sets) can have different inconsistency measures. In these proposals one can identify the set of formulae with its conjunction (i.e. the set $\{\varphi, \varphi'\}$ has the same inconsistency measure as the set $\{\varphi \wedge \varphi'\}$). Whilst the lack of syntax sensitivity may be appropriate for some applications, it does mean that the distribution of the contradiction among the formulae is not taken into account.

It seems difficult to build measures that take these two dimensions into account. But, in (Hunter & Konieczny 2006) a unified framework was proposed with this aim. The main point was to define inconsistency values that give the inconsistency of each formula of the base, in constrast to the above inconsistency measures that give the inconsistency of the whole base. This allows us to draw a more precise picture of the inconsistencies of the base. The idea is to start from one of the measures considered in the two approaches above and use it to assign a measure of inconsistency to a set of formulae, and then to use a technique based on cooperative game theory: the Shapley value (Shapley 1953). This then allows us to identify the blame/responsibility of each formula in the inconsistency of the belief base (Hunter & Konieczny 2006). This means for example that we can use a measure from the second approach which considers the proportion of the language touched by inconsistency, and then using the Shapley value, apportion the blame for the inconsistency in the set to the individual formulae in a principled way.

Against this background, it is interesting to note that the use of minimal inconsistent subsets of a belief base has received much less attention as the basis for defining inconsistency measures. So in this paper, we explore the nature of some interesting measures of inconsistency based on minimal inconsistent subsets of the belief base, and we consider how these new measures relate to the family of Shapley Inconsistency Values.

It has been known for a long time that minimal inconsistent subsets of the base are a cornerstone of analysing inconsistencies. For instance, to recover consistency, one has just to remove one formula from each minimal inconsistent subset (Reiter 1987). For conflict resolution, where the syntactic representation of the information is important, measuring inconsistency in terms of the minimal inconsistent subsets is intuitive and it is informative for deciding how to change the set of formulae through a process such as negotiation, compromise, or resolution. So it should be natural to study how these minimal inconsistent subsets could be used to define measures of inconsistency.

The idea to analyse minimal inconsistent subsets of the belief base was followed in order to define scoring functions that, for each subset K' of a set of formulae K, gives a score that is the number of minimal inconsistent subsets of K that would be eliminated if K' were removed from K. Then this score is used to compare different subsets K' (Hunter 2004).

Obviously, these approaches are syntax sensitive, which for some applications, is necessary. Consider capturing requirements for a new corporate computer system in a process where a user may present his or her requirements in the form of a set of propositional formulae (Hunter & Nuseibeh 1998). Here presenting the set of requirements $\{\alpha, \beta\}$ should be treated differently to the set of requirements $\{\alpha \wedge \beta\}$ since the first set says that there are two requirements, the first being α and the second β, whereas in the second set says that is one requirement, namely $\alpha \wedge \beta$.

In the rest of this paper, we study how to use minimal inconsistent subsets in order to define inconsistency values, that will allow us to define the inconsistency of each formula of the base (like with Shapley Inconsistency Values (Hunter & Konieczny 2006)). To this end we introduce the family of MIV (for MinInc Inconsistency Value), and focus on a very intuitive one: MIV_C. As these values have the same aim as the family of Shapley Inconsistency Values, it is interesting to study the relationship between these two families. Moreover, we show a surprising result: the value MIV_C is in fact a Shapley Inconsistency Value. This result is very interesting as it allows us to state interesting logical properties, and to find a way to implement the Shapley Inconsistency Values using existing automated reasoning technology. The additional interest of this value is that its intuitive logical properties have led us to a complete axiomatization through five intuitive axioms.

Preliminaries

We consider a propositional language $\mathcal{L}$ built from a finite set of propositional symbols $\mathcal{P}$. We use $a, b, c, \dots$ to denote the propositional variables, and Greek letters $\alpha, \beta, \varphi, \dots$ to denote the formulae. An interpretation is a total function from $\mathcal{P}$ to $\{0, 1\}$. The set of all interpretations is denoted $\mathcal{W}$. An interpretation ω is a model of a formula φ, denoted $\omega \models \varphi$, if and only if it makes φ true in the usual truth-functional way. $Mod(\varphi)$ denotes the set of models of the formula φ, i.e. $Mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$. We will use $\subseteq$ to denote the set inclusion, and we will use $\subset$ to denote the strict set inclusion, i.e. $A \subset B$ iff $A \subseteq B$ and $B \not\subseteq A$. We will denote the set of natural numbers by $\mathbb{N}$ and the set of real numbers by $\mathbb{R}$.

Let A and B be two subsets of C, we note $C = A \oplus B$ if A and B form a partition of C, i.e. $C = A \oplus B$ iff $C = A \cup B$ and $A \cap B = \emptyset$.

A *belief base* K is a finite set of propositional formulae. More exactly, as we will need to identify the different formulae of a belief base in order to associate them with their inconsistency value, we will consider belief bases K as vectors of formulae. For logical properties we will need to use the set corresponding to each vector, so we suppose that we have a function such that for each vector $K = (\alpha_1, \dots, \alpha_n)$, $\overline{K}$ is the set $\{\alpha_1, \dots, \alpha_n\}$. As it will never be ambiguous, in the following we will omit the $\overline{}$ and write K as both the vector and the set. We use $\mathcal{K}_{\mathcal{L}}$ to denote the set of belief bases definable from formulae of the language $\mathcal{L}$.

A belief base is consistent if there is at least one interpretation that satisfies all its formulae. If a belief base K is not consistent, then one can define the minimal inconsistent subsets of K as:

$$\mathsf{MI}(K) = \{K' \subseteq K \mid K' \vdash \bot \text{ and } \forall K'' \subset K', K'' \nvdash \bot\}$$

If one wants to recover consistency from an inconsistent base K, then the minimal inconsistent subsets can be considered as the purest form of inconsistency, since to recover consistency, one has to remove at least one formula from each minimal inconsistent subset (Reiter 1987).

A *free formula* of a belief base K is a formula of K that does not belong to any minimal inconsistent subset of the belief base K. This means that this formula has nothing to do with the conflicts of the base.

Inconsistency values defined from minimal inconsistent sets

Apart from (Hunter & Konieczny 2006), existing inconsistency measures allow us to evaluate the amount of inconsistency of a whole base, but not to evaluate the amount of inconsistency of each formula of the base. In other words, existing inconsistency measures do not allow us to evaluate the responsability of each formula in the inconsistency of the base. Yet it is possible to define some such measures using minimal inconsistent sets. Furthermore, as these minimal inconsistent sets are the parts of the base where inconsistencies lie, it should be natural to use only the minimal inconsistent sets for evaluating the amount of inconsistency of the bases. This is the motivation for the following definition where f

is some function that takes as input a formula α and the set of minimal inconsistent subsets for a belief base K.

Definition 1 *A MinInc Inconsistency Value (MIV) is a function $MIV : \mathcal{K}_{\mathcal{L}} \times \mathcal{L} \rightarrow \mathbb{R}$ such that $MIV(K,\alpha) = f(\alpha, \mathsf{MI}(K))$ where f is a function of α and $\mathsf{MI}(K)$.*

Instances of a MIV (such as those given in Definitions 2 and 4) depend on the choice of function f.

So this definition states exactly the fact that the inconsistency value only takes into account the minimal inconsistent subsets of the base. In particular two different bases K and K' with exactly the same minimal inconsistent sets will have the same amount of inconsistency.

The simplest types of MIV one can define are the following ones:

Definition 2 *MIV_D and $MIV_{\#}$ are defined as follows:*

- $MIV_D(K,\alpha) = \begin{cases} 1 & \textit{if } \exists M \in \mathsf{MI}(K)\ \alpha \in M \\ 0 & \textit{otherwise} \end{cases}$
- $MIV_{\#}(K,\alpha) = |\{M \in \mathsf{MI}(K) \mid \alpha \in M\}|$

The first value is the drastic one, that takes value one if the formula belongs to a minimal inconsistent subset, and zero otherwise. The second one is a cardinality value, that counts the number of minimal inconsistent subsets the formula belongs to.

The first value is of little interest, since it allows us just to make a distinction between free formulas and the other ones. The second one is more useful, and allows us to find more interesting results. Let us check this on the following example.

Example 1 *Let $K_1 = \{a, \neg a, \neg a \wedge c, a \vee d, \neg d, b \wedge \neg b, e\}$, so the minimal inconsistent subsets of K_1 are*

$$\mathsf{MI}(K_1) = \{\{b \wedge \neg b\}, \{a, \neg a\}, \{a, \neg a \wedge c\}, \{\neg a, a \vee d, \neg d\}\}$$

and the $MIV_{\#}$ value gives as a result:

$MIV_{\#}(K_1, a) = 2$	$MIV_{\#}(K_1, \neg d) = 1$
$MIV_{\#}(K_1, \neg a) = 2$	$MIV_{\#}(K_1, b \wedge \neg b) = 1$
$MIV_{\#}(K_1, \neg a \wedge c) = 1$	$MIV_{\#}(K_1, e) = 0$
$MIV_{\#}(K_1, a \vee d) = 1$	

It is easy to check that $MIV_{\#}$ is just a scoring function (Hunter 2004) applied uniquely on formulae:

Definition 3 *Let $K \in \mathcal{K}_{\mathcal{L}}$. Let S be the* **scoring function** *for K defined as follows, where $S : \wp(K) \mapsto \mathbb{N}$ and $K' \in \wp(K)$*

$$S(K') = |\mathsf{MI}(K)| - |\mathsf{MI}(K - K')|$$

For a belief base K, a scoring function S gives the number of minimal inconsistent subsets of K that would be eliminated if the subset K' was removed from K. See (Hunter 2004) for more details on the use of these scoring functions. So if $\alpha \in K$, it is straighforward to see that we have $MIV_{\#}(\alpha, K) = S(\{\alpha\})$.

The evaluation of the inconsistency value of each formula given by $MIV_{\#}$ is still very rough. In particular, it does not take into account the cardinalities of the minimal inconsistent subsets the formula belongs to. But, as explained in several works (see for example (Knight 2001)), the size of the minimal inconsistent subset can have an impact on the evaluation of the inconsistency. The idea is that the smaller the value of the minimal inconsistent subset, the bigger is the inconsistency. To illustrate this we use the prototypical example of the lottery paradox given by Knight to motivate his approach.

Example 2 *There are a number of lottery tickets with one of them being the winning ticket. Suppose w_i denotes ticket i will win, then we have the assumption $w_1 \vee \ldots \vee w_n$. In addition, for each ticket i, we may pessimistically (or probabilistically if the number of tickets is important) assume that it will not win, and this is represented by the assumption $\neg w_i$. So the base K_L is:*

$$K_L = \{\neg w_1, \ldots, \neg w_n, w_1 \vee \ldots \vee w_n\}$$

Clearly if there are three or two (or one!) tickets in the lottery, then this base is highly inconsistent. But if there are millions of tickets there is intuitively (nearly) no conflict in the base.

So it could prove better not to simply take the number of minimal inconsistent subsets a formula belongs to, but to take into account their cardinalities. This idea gives a third MIV value:

Definition 4 *MIV_C is defined as follows:*

$$MIV_C(K,\alpha) = \sum_{M \in \mathsf{MI}(K)\, s.t.\, \alpha \in M} \frac{1}{|M|}$$

This allows us to define a much more precise view of the inconsistency, as illustrated in the following example.

Example 3 *Let $K_1 = \{a, \neg a, \neg a \wedge c, a \vee d, \neg d, b \wedge \neg b, e\}$ and the MIV_C value gives as a result:*

$MIV_C(K_1, a) = 1$	$MIV_C(K_1, \neg d) = \frac{1}{3}$
$MIV_C(K_1, \neg a) = \frac{5}{6}$	$MIV_C(K_1, b \wedge \neg b) = 1$
$MIV_C(K_1, \neg a \wedge c) = \frac{1}{2}$	$MIV_C(K_1, e) = 0$
$MIV_C(K_1, a \vee d) = \frac{1}{3}$	

We can compare these obtained results with the ones of Example 1, where less distinction was possible, with only three different levels. We can notice that now we can make a distinction between $\neg a \wedge c$, and $a \vee d$, that both belong to only one minimal inconsistent subset, but the one of $a \vee d$ is bigger. We also note that an inconsistent formula has a high degree of inconsistency according to the MIV_C value. One can remark that with MIV_C the formula $b \wedge \neg b$ is evaluated as more conflicting than the formula $\neg a$ which belongs to two larger minimal inconsistent subsets, whereas the evaluation is the converse for $MIV_{\#}$.

We can see more clearly on a dedicated example why MIV_C gives a more precise view of the conflict brought by each fomula than $MIV_{\#}$.

Example 4 *Consider $K_2 = \{a \wedge \neg a, b \wedge \neg b\}$ and $K_3 = \{a \wedge \neg b, \neg a \wedge b\}$. Here we see that the $MIV_{\#}$ value assigns the same value to each formula, even though for instance $a \wedge \neg a$ is entirely responsible for an inconsistency, whereas $a \wedge \neg b$ is only partially responsible for an inconsistency.*

$MIV_{\#}(K_2, a \wedge \neg a) = 1$	$MIV_{\#}(K_3, a \wedge \neg b) = 1$
$MIV_{\#}(K_2, b \wedge \neg b) = 1$	$MIV_{\#}(K_3, \neg a \wedge b) = 1$

In contrast, the MIV_C value is more discriminating and so for instance $a \wedge \neg a$, which is entirely responsible for an inconsistency, has the maximum value of 1, whereas $a \wedge \neg b$, which is only half of the cause of an inconsistency, has a value of 1/2.

$$MIV_C(K_2, a \wedge \neg a) = 1 \quad MIV_C(K_3, a \wedge \neg b) = \frac{1}{2}$$
$$MIV_C(K_2, b \wedge \neg b) = 1 \quad MIV_C(K_3, \neg a \wedge b) = \frac{1}{2}$$

More generally, we see that the MIV_C value is affected by the size of each minimal inconsistent subset: Returning to Example 2, we see that for K_L and for some $\alpha \in K_L$, the value of $MIV_C(K_L, \alpha)$ decreases as the cardinality of K_L increases.

We now give a few observations regarding the MIV_C definition. Other properties will be also derivable from later results of the paper.

Proposition 1

- *If α is a free formula in K, then $MIV_C(K, \alpha) = 0$*
- $MIV_C(K \cup K', \alpha) \geq MIV_C(K, \alpha)$
- *If $\alpha \equiv \perp$, then $MIV_C(K, \alpha) = 1$*
- *If $\phi \vdash \psi$ and $\phi \not\vdash \perp$ then*

$$MIV_C(K \cup \{\phi\}, \alpha) \geq MIV_C(K \cup \{\psi\}, \alpha)$$

So from the examples and observations in this section, it seems that MIV_C is an appealing and informative measure of inconsistency.

Shapley Inconsistency Values

Shapley Inconsistency Values were introduced in (Hunter & Konieczny 2006) in order to be able to define a measure of inconsistency for each formula, from a measure of inconsistency on belief bases.

The idea is to start from one basic measure of inconsistency from the literature, that allows us to evaluate the inconsistency of a belief base, to use this measure as the definition of a coalitional game, and to use a notion from cooperative game theory, the Shapley value (Shapley 1953), that allows us to define the merits of one individual in a given game. In our setting, with the scale reversal, this amounts to define the blame/responsability of one formula in a given base for the inconsistencies.

So the idea is similar to the one that drove the definition of MIV in the last section. Therefore it is natural to wonder if there are some links between the two approaches.

Let us first give the background on Shapley Inconsistency Values (SIV). We will just give here the definitions needed for this paper and for the proofs, for more details see (Hunter & Konieczny 2006).

First we recall the standard definitions of games in coalitional form and of the Shapley value (Aumann & Hart 2002).

Definition 5 *Let $N = \{1, \ldots, n\}$ be a set of n players. A game in coalitional form is given by a function $v : 2^N \to \mathbb{R}$, with $v(\emptyset) = 0$.*

This framework defines games in a very abstract way, focusing on the possible coalition formations. A coalition is just a subset of N. This function gives what payoff can be achieved by each coalition in the game v when all its members act together as a unit.

A natural notion of solution for this kind of game is to try to define the payoff that can be expected by each player i for the game v. This is what is called a *value*.

Definition 6 *A* value *is a function that assigns to each game v a vector of payoff $S(v) = (S_1, \ldots, S_n)$ in $\mathbb{R}^n$, where S_i is the payoff for player i.*

Despite these very abstract definitions of the game and of the value, it is possible to define a notion of solution, i.e. a value, that gives for any game the expected payoff (merits) of each player. The first (and main) such value has been defined by Shapley (Shapley 1953).

Basically the idea can be explained as follows: considering that the coalitions form according to some order (a first player enters the coalition, then another one, then a third one, etc), and that the payoff attached to a player is its marginal utility (i.e. the utility that it brings to the existing coalition), so if C is a coalition (subset of N) not containing i, player's i marginal utility is $v(C \cup \{i\}) - v(C)$. As one can not make any hypothesis on which order is the correct one, we may suppose that each order is equally probable. This leads to the following formula:

Let σ be a permutation on N, with σ_n denoting all the possible permutations on N. We need the following notation:

$$p^i_\sigma = \{j \in N \mid \sigma(j) < \sigma(i)\}$$

That means that p^i_σ represents all the players that precede player i for a given order σ.

Definition 7 *Let $i \in N$ be a player, and n be the number of players. The Shapley value of a game v is defined as.*

$$S_i(v) = \frac{1}{n!} \sum_{\sigma \in \sigma_n} v(p^i_\sigma \cup \{i\}) - v(p^i_\sigma)$$

The Shapley value can be directly computed from the possible coalitions (without looking at the permutations) using the following expression:

$$S_i(v) = \sum_{C \subseteq N} \frac{(c-1)!(n-c)!}{n!} (v(C) - v(C \setminus \{i\}))$$

where c is the cardinality of C.

Besides the fact that this definition gives very sensible results, its legitimacy is also given by a nice characterization result:

Proposition 2 (Shapley 1953) *The Shapley value is the only value that satisfies all of Efficiency, Symmetry, Dummy and Additivity.*

- $\sum_{i \in N} S_i(v) = v(N)$ **(Efficiency)**
- *If i and j are such that for all C s.t. $i, j \notin C$, $v(C \cup \{i\}) = v(C \cup \{j\})$, then $S_i(v) = S_j(v)$* **(Symmetry)**

• *If i is such that $\forall C\ v(C \cup \{i\}) = v(C)$, then $S_i(v) = 0$* **(Dummy)**

• $S_i(v + w) = S_i(v) + S_i(w)$ **(Additivity)**

This result supports several variations: there are other equivalent axiomatizations of the Shapley value, and there are some different values that can be defined by relaxing some of the above axioms. See (Aumann & Hart 2002).

So the idea is to consider an inconsistency measure (that allows us to evaluate the inconsistency of a belief base) as a game in coalitional form, and to compute the corresponding Shapley value, in order to be able to define the inconsistency value of each formula of the base.

We ask some properties for the underlying inconsistency measure:

Definition 8 *An inconsistency measure I is called a* **basic inconsistency measure** *if it satisfies the following properties, $\forall K, K' \in \mathcal{K}_{\mathcal{L}}$, $\forall \alpha, \beta \in \mathcal{L}$:*

• *$I(K) = 0$ iff K is consistent* **(Consistency)**
• *$I(K \cup K') \geq I(K)$* **(Monotony)**
• *If α is a free formula of $K \cup \{\alpha\}$, then $I(K \cup \{\alpha\}) = I(K)$* **(Free Formula Independence)**
• *If $\alpha \vdash \beta$ and $\alpha \nvdash \perp$, then $I(K \cup \{\alpha\}) \geq I(K \cup \{\beta\})$* **(Dominance)**

In (Hunter & Konieczny 2006), a Normalization property was also presented as an optional property. We do not consider that property in this paper.

Definition 9 *Let I be a basic inconsistency measure. We define the corresponding* **Shapley Inconsistency Value** *(SIV), noted S_I, as the Shapley value of the coalitional game defined by the function I, i.e. let $\alpha \in K$:*

$$S^I_\alpha(K) = \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$$

where n is the cardinality of K and c is the cardinality of C.

Note that this SIV gives a value for each formula of the base K. This definition allows us to define to what extent a formula inside a belief base is concerned with the inconsistencies of the base. It allows us to draw a precise picture of the contradiction of the base.

So, from a SIV, one can define an inconsistency measure for the whole belief base:

Definition 10 *Let K be a belief base,*

$$\hat{S}^I(K) = \max_{\alpha \in K} S^I_\alpha(K)$$

There are alternatives to Definition 10, see the discussion in (Hunter & Konieczny 2006).

MI Shapley Inconsistency Value

Since minimal inconsistent subsets of a base can be considered as fundamental features in charaterizing inconsistency, we use the notion here as the basic inconsistency measure.

Definition 11 *The* **MI inconsistency measure** *is defined as the number of minimal inconsistent sets of K, i.e. :*

$$I_{MI}(K) = |\mathsf{MI}(K)|$$

Example 5 $K = \{a, \neg a, \neg a \wedge c, a \vee d, \neg d, b \wedge \neg b, e\}$
Hence, we get the following:
$I_{MI}(K) = 4$ $\quad$ $I_{MI}(\{a, \neg a, \neg a \wedge c\}) = 2$
$I_{MI}(\{b \wedge \neg b, e\}) = 1$ $\quad$ $I_{MI}(\{\neg a, \neg a \wedge c\}) = 0$

Proposition 3 *The MI inconsistency measure I_{MI} is a basic inconsistency measure, i.e. it satisfies the properties of Consistency, Monotonicity, Free Formula Independence, and Dominance.*

And in fact the following result shows that this MI Shapley Inconsistency Value is exactly the MIV_C measure of Definition 4:

Proposition 4 $S^{I_{MI}}_\alpha(K) = MIV_C(K, \alpha)$

Proof: Let us first show the following lemma that will be useful in the proof.

Lemma 1 *If a simple game in coalitional form on a set of players $N = \{1, \dots, n\}$ is defined by a single winning coalition $C' \subseteq N$, i.e:*

$$v(C) = \begin{cases} 1 & \textit{if } C' \subseteq C \\ 0 & \textit{otherwise} \end{cases}$$

Then the corresponding Shapley value is:

$$S_i(v) = \begin{cases} 0 & \textit{if } i \notin C' \\ \frac{1}{|C'|} & \textit{if } i \in C' \end{cases}$$

Proof of Lemma 1 : The proof is direct using the logical properties of the Shapley value given in Proposition 2. Since by (Dummy) we get that if $i \notin C'$, then $S_i(v) = 0$. By (Efficiency) we know that the outcome of the grand coalition N must be shared in the sum of the Shapley values of the players: $\sum_{i \in N} S_i(v) = 1$. Since for players $i \notin C'$ we know that $S_i(v) = 0$, it means that it has to be split between members of C'. So $\sum_{i \in C'} S_i(v) = 1$. Now by (Symmetry) we get that for all $i, j \in C'$, we have $S_i(v) = S_j(v)$. So this implies that if $i \in C'$, then $S_i(v) = \frac{1}{|C'|}$. $\square$

Let us now state the result. First suppose that α is a free formula of K, then we have immediately by (Minimality) that $S^{I_{MI}}_\alpha(K) = 0$. We also have immediately by definition that $MIV_C(K, \alpha) = 0$. So the equality is satisfied in this case.

Now suppose that α is not a free formula of K. First remark that I_{MI} can be decomposed in $I_{MI}(C) = \sum_{M \in \mathsf{MI}(K)} \hat{M}(C)$, where $\hat{M}$ is the following characteristic function

$$\hat{M}(C) = \begin{cases} 1 & \text{if } M \subseteq C \\ 0 & \text{otherwise} \end{cases}$$

Let us denote by $\hat{M}(K)$ the game in coalitional form defined from K and the characteristic function $\hat{M}$.

So now let us start from the MI Shapley Inconsistency Value:

$$
\begin{aligned}
& S_\alpha^{I_{MI}}(K) \\
&= \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!} (I_{MI}(C) - I_{MI}(C \setminus \{\alpha\})) \\
&= \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!} (\sum_{M \in \mathsf{MI}(K)} \hat{M}(C) \\
&\qquad - \textstyle\sum_{M \in \mathsf{MI}(K)} \hat{M}(C \setminus \{\alpha\})) \\
&= \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!} (\sum_{M \in \mathsf{MI}(K)} (\hat{M}(C) - \hat{M}(C \setminus \{\alpha\}))) \\
&= \sum_{C \subseteq K} \sum_{M \in \mathsf{MI}(K)} \frac{(c-1)!(n-c)!}{n!} (\hat{M}(C) - \hat{M}(C \setminus \{\alpha\})) \\
&= \sum_{M \in \mathsf{MI}(K)} \sum_{C \subseteq K} \frac{(c-1)!(n-c)!}{n!} (\hat{M}(C) - \hat{M}(C \setminus \{\alpha\})) \\
&= \sum_{M \in \mathsf{MI}(K)} S_\alpha(\hat{M}(K))
\end{aligned}
$$

Now note that by Lemma 1 we have $S_\alpha(\hat{M}(K)) = \frac{1}{|M|}$.

That gives $S_\alpha^{I_{MI}}(K) = \sum_{M \in \mathsf{MI}(K)} \frac{1}{|M|} = MIV_C(K, \alpha)$.

□

This proposition is interesting for several reasons. First, it confirms the appeal of Shapley Inconsistency Values, since the very natural measure MIV_C is a special case of these measures. Second, it gives a simpler definition of $S_\alpha^{I_{MI}}$ than the one using the Shapley value. In general, obtaining a Shapley value is computationally demanding (Deng & Papadimitriou 1994). However, in the case of $S_\alpha^{I_{MI}}$, the above proposition hints at the possibility of computationally viable implementations for calculating these values. Finally, this equality is useful to state the logical properties of this value, as done in the next Section.

Logical Properties

It is quite difficult to state logical properties about inconsistency handling (and measure of inconsistency) in a purely classical framework, i.e. without adding too many hypotheses, by using an (arbitrary) paraconsistent logic to do so.

In (Hunter & Konieczny 2006) some logical properties for inconsistency measures are defined, as well as specific ones for Shapley Inconsistency Values. But there was no characterization theorem in that paper. We provide such a theorem below.

Let us first strengthen the condition on the basic inconsistency measure:

Definition 12 *A MinInc Separable basic inconsistency measure (MSBIM) I is a basic inconsitency measure that satisfies this additional property:*

- *If* $\mathsf{MI}(K \cup K') = \mathsf{MI}(K) \oplus \mathsf{MI}(K')$*, then* $I(K \cup K') = I(K) + I(K')$ **(MinInc Separability)**

This property basically expresses the fact that the inconsistency measure depends on the minimal inconsistent subsets, so that if we can partition the belief base in two subbases without "breaking" any minimal inconsistent subset, then the global inconsistency measure is the sum of the inconsistency measure of the two subbases. Clearly, the MI inconsistency measure satisfies this property.

Let us now enumerate the properties that we expect inconsistency values to satisfy:

- $\sum_{\alpha \in K} S_\alpha^I(K) = I(K)$ **(Distribution)**
- If $\exists \alpha, \beta \in K$ s.t. for all $K' \subseteq K$ s.t. $\alpha, \beta \notin K'$, $I(K' \cup \{\alpha\}) = I(K' \cup \{\beta\})$, then $S_\alpha^I(K) = S_\beta^I(K)$ **(Symmetry)**
- If α is a free formula of K, then $S_\alpha^I(K) = 0$ **(Minimality)**
- If $\alpha \vdash \beta$ and $\alpha \nvdash \bot$, then $S_\alpha^I(K) \geq S_\beta^I(K)$ **(Dominance)**
- If $\mathsf{MI}(K \cup K') = \mathsf{MI}(K) \oplus \mathsf{MI}(K')$, then $S_\alpha^I(K \cup K') = S_\alpha^I(K) + S_\alpha^I(K')$ **(Decomposability)**

The first four properties were already discussed in (Hunter & Konieczny 2006). The first three of these are closely related to original Shapley's properties. The distribution property states that the inconsistency values of the formulae sum to the total amount of inconsistency in the base ($I(K)$). The symmetry property ensures that only the amount of inconsistency brought by a formula matters for computing the inconsistency value. As one could expect, a formula that is not embedded in any contradiction (i.e. does not belong to any minimal inconsistent subset) will not be blamed by the inconsistency value. This is what is expressed in the minimality property. The dominance property states that logically stronger formulae bring (potentially) more conflicts. It was shown in (Hunter & Konieczny 2006) that every Shapley Inconsistency Value satisfies these four properties.

The Decomposability property is related to Shapley's Additivity property. We explain in (Hunter & Konieczny 2006) that a direct translation of this Additivity property makes little sense because it is not meaningful to add different (basic) inconsistency measures. But one can consider another translation of the additivity property, by looking to the "addition" of two different bases: the set union. So direct translation of that meaning leads to

$$S_\alpha^I(K \cup K') = S_\alpha^I(K) + S_\alpha^I(K')$$

This formulation is not satisfactory because it forgets the fact that new conflicts can appear when making the union of the two bases. So we want this property to hold only when joining two bases does not create any new inconsistencies. That is ensured by the condition of the Decomposability property. Note that this possibility of interaction between the two subgames that is not taken into account in the usual Additivity condition, is one of the criticisms about this condition. Let us quote for instance the following paragraph from (Luce & Raiffa 1957):

> *The last condition is not nearly so innocent as the other two. For although $v + w$ is a game composed from v and w, we cannot in general expect it to be played as if it were the two separate games. It will have its own structure which will determine a set of equilibrium outcomes which may be different from those for v and w.*

Therefore, one might very well argue that its a priori value should not necessarily be the sum of the values of the two component games. This strikes us as a flaw in the concept value, but we have no alternative to suggest.

In our framework the interaction between the two bases is simply the new logical conflicts that appears when joining the bases, that allows us to say when this addition can hold, and when it is not sensible.

For setting the characterization result we have to ask one additional property that states that each minimal inconsistent subset brings the same amount of conflict:

- If $M \in \mathsf{MI}(K)$, then $I(M) = 1$ **(MinInc)**

So now we reach the wanted characterization result

Proposition 5 *An inconsistency value satisfies Distribution, Symmetry, Minimality, Decomposability and MinInc if and only if it is the MI Shapley Inconsistency Value $S_\alpha^{I_{MI}}$.*

Proof: To prove that the MI Shapley Inconsistency Value satisfy the logical properties is easy. **(Distribution)**, **(Symmetry)**, **(Minimality)** are satisfied by all Shapley Inconsistency Values (Proposition 3 of (Hunter & Konieczny 2006)).

So it remains to show **(Decomposability)** and **(MinInc)**. **(MinInc)** is satisfied by definition since $I_{MI}(M) = |\mathsf{MI}(M)| = 1$ for any $M \in \mathsf{MI}(K)$.

For **(Decomposability)**, by definition $S_\alpha^I(K \cup K') = \sum_{C \subseteq K \cup K'} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\}))$. Now split C on K and K', i.e. define $H = C \cap K$ and $H' = C \cap K'$. It is easy to check that $C = H \cup H'$, and from the hypothesis that $\mathsf{MI}(K \cup K') = \mathsf{MI}(K) \oplus \mathsf{MI}(K')$ we deduce that $\mathsf{MI}(H \cup H') = \mathsf{MI}(H) \oplus \mathsf{MI}(H')$, so as $S^{I_{MI}} = MIV_C$ satisfies **(MinInc Separability)** we have that $I(C) = I(H) + I(H')$. So using this in the definition we have

$$\begin{aligned}
&S_\alpha^I(K \cup K') \\
&= \sum_{C \subseteq K \cup K'} \frac{(c-1)!(n-c)!}{n!}(I(C) - I(C \setminus \{\alpha\})) \\
&= \sum_{C \subseteq K \cup K'} \frac{(c-1)!(n-c)!}{n!}(I(H) + I(H') \\
&\qquad\qquad -I(H \setminus \{\alpha\}) - I(H' \setminus \{\alpha\})) \\
&= \sum_{C \subseteq K \cup K'} \frac{(c-1)!(n-c)!}{n!}(I(H) - I(H \setminus \{\alpha\}) \\
&\qquad + \sum_{C \subseteq K \cup K'} \frac{(c-1)!(n-c)!}{n!}(I(H') - I(H' \setminus \{\alpha\})) \\
&= \sum_{H \subseteq K} \frac{(c-1)!(n-c)!}{n!}(I(H) - I(H \setminus \{\alpha\}) \\
&\qquad + \sum_{H' \subseteq K'} \frac{(c-1)!(n-c)!}{n!}(I(H') - I(H' \setminus \{\alpha\})) \\
&= S_\alpha^I(K) + S_\alpha^I(K')
\end{aligned}$$

For the converse implication suppose that we have an inconsistency value that satisfies **(Distribution)**, **(Symmetry)**, **(Minimality)**, **(Decomposability)** and **(MinInc)**. We want to show that it is the MI Shapley Inconsistency Value.

First note that for any K such that $\mathsf{MI}(K) = \{M_1, \ldots, M_n\}$, if one chooses a sequence $M_1, \ldots, M_n$, then for all i where $1 \leq i < n$, the following holds:

$$\mathsf{MI}(M_1 \cup \ldots \cup M_i \cup M_{i+1}) = \mathsf{MI}(M_1 \cup \ldots \cup M_i) \oplus \mathsf{MI}(M_{i+1})$$

Hence, there is a sequence of the minimal inconsistent subsets of K, such that by use of **(Minimality)** and successive use of **(Decomposability)** we have that

$$S_\alpha^I(K) = \sum_{M \in \mathsf{MI}(K)} S_\alpha^I(M)$$

Now for each M if $\alpha \notin M$ we have by **(Minimality)** that $S_\alpha^I(M) = 0$. And if $\alpha \in M$ then we have by **(Distribution)** $\sum_{\alpha \in M} S_\alpha^I(M) = I(M)$. And by **(Symmetry)** we have that $\forall \alpha, \beta \in M$, $S_\alpha^I(M) = S_\beta^I(M)$. So we obtain that

$$\forall \alpha \in M,\ S_\alpha^I(M) = \frac{I(M)}{|M|}$$

and therefore

$$S_\alpha^I(K) = \sum_{M \in \mathsf{MI}(K) \text{ s.t. } \alpha \in M} \frac{I(M)}{|M|}$$

Now by **(MinInc)** we know that for all $M \in \mathsf{MI}(K)$, $I(M) = 1$. That gives

$$S_\alpha^I(K) = \sum_{M \in \mathsf{MI}(K) \text{ s.t. } \alpha \in M} \frac{1}{|M|}$$

That is the definition of MI Shapley Inconsistency Value. □

This result means that the Shapley Inconsistency Value $S_\alpha^{I_{MI}}$ is completely characterized by five simple and intuitive axioms.

Note that Dominance, although satisfied by SIV, is not required for stating this proposition.

Towards Implementation of Inconsistency Values

The development of SAT solvers has made impressive progress in recent years, allowing, despite the computational complexity of the problem, to practically solve a number of intractable problems (Kautz & Selman 2007).

Based on SAT solvers, some techniques have been aimed at the identification of minimal inconsistent subsets (called in these works Minimally Unsatisfiable Subformulas or MUS). Although the identification problem is computationally hard, since checking whether a set of clauses is a MUS or not is DP-complete, and checking whether a formula belongs to the set of MUSes of a base, is in Σ_2^P (Eiter & Gottlob 1992); it seems that finding each MUS can be practically feasible (Grégoire, Mazure, & Piette 2007; 2008).

Thanks to Proposition 4, we then can define an easy algorithm to compute the MI Shapley Inconsistency Value of the formulae of a base:

Input: A belief base $K = \{\alpha_1, \ldots, \alpha_n\}$
Output: A profile of values $(S^I_{\alpha_1}(K), \ldots, S^I_{\alpha_n}(K))$
1- **For** i **from** 1 **to** n
$S_i \leftarrow 0$
2- Compute $\mathsf{MI}(K)$
3- **For each** $C \in \mathsf{MI}(K)$
For each $\alpha_i \in C$
$S_i \leftarrow S_i + \frac{1}{|C|}$
4- **Return** $(S_1, \ldots, S_n)$

The hard step in the above algorithm is step 2. But if it can be viably computed, then the rest of the algorithm is just polynomial in the size of the $\mathsf{MI}(K)$ (but of course the size of $\mathsf{MI}(K)$ can be exponential in the size of K).

In future work, we plan to implement such an algorithm using an existing algorithm to identify MUS (Grégoire, Mazure, & Piette 2007; 2008). This would gives us a practical tool to measure inconsistency.

A possible approach to ameliorate the cost of entailment in finding minimal inconsistent subsets is to use approximate entailment: Proposed in (Levesque 1984), and developed in (Schaerf & Cadoli 1995), classical entailment is approximated by two sequences of entailment relations. The first is sound but not complete, and the second is complete but not sound. Both sequences converge to classical entailment. For a set of propositional formulae Δ, a formula α, and an approximate entailment relation $\models_i$, the decision of whether $\Delta \models_i \alpha$ holds or $\Delta \not\models_i \alpha$ holds can be computed in polynomial time. Approximate entailment has been developed for anytime coherence reasoning (Koriche 2001; 2002), and in furture work, we will investigate its potential for an approximate version of the MI Shapley Inconsistency Value.

By focussing on subsystems of classical logic, such as description logics, there appears to be much potential in harnessing existing specialized reasoning systems for finding minimal inconsistent subsets of a belief base (for example by using the Pellet reasoning system for description logics (Kalyanpur *et al.* 2005; Parsia, Sirin, & Kalyanpur 2005)). Furthermore, measuring inconsistency in description logic ontologies offers a potentially interesting and worthwhile application problem (Qi & Hunter 2007).

Another application area for inconsistency measures is in supporting reasoning with inconsistent databases (for example when using maximally consistent subsets of the database (Bertossi & Bravo 2005)). Again, this is an application where language restrictions and specialized reasoning systems offer the potential for viable means for finding minimal inconsistent subsets of a belief base, and thereby finding the MI Shapley Inconsistency Value.

Conclusion

As discussed in the introduction, there are a number of proposals for measures of inconsistency. The main novel contributions provided by this paper are :

- A first discussion on the definition of inconsistency values based on minimal inconsistent subsets of belief bases, and this leads to the definition of the family of Minimal Inconsistent Values;
- An equivalence between a particular Shapley Inconsistency Value (the MI Shapley Inconsistency Value) and a simple Minimal Inconsistent Value which gives an additional argument to support the Shapley Inconsistency Value definition since it captures this very natural value as particular case;
- The first (as far as we know) axiomatization of an inconsistency value;
- And finally, the characterization of the MI Shapley Inconsistency Value in terms of the MIV_C measure opens the possibility for computationally viable calculation of inconsistency values.

In future work, we would like to analyse the computational complexity of using the MI Shapley Inconsistency Value, develop algorithms and implementations (possibly based on approximation techniques), and undertake case studies of applications of this value.

Aknowledgements

This work has been partly supported by a bilateral project CNRS - Royal Society.

References

Aumann, R., and Hart, S., eds. 2002. *Handbook of Game Theory with Economic Applications*, volume 3. North-Holland.

Barragáns-Martńez, B.; Pazos-Arias, J.; and Fernández-Vilas, A. 2004. On measuring levels of inconsistency in multi-perspective requirements specifications. In *Proceedings of the 1st Conference on the Principles of Software Engineering (PRISE'04)*, 21–30.

Bertossi, L., and Bravo, L. 2005. Consistent query answers in virtual data integration systems. In *Inconsistency Tolerance*, volume 3300 of *LNCS*, 42–83. Springer.

Bertossi, L.; Hunter, A.; and Schaub, T. 2004. Introduction to inconsistency tolerance. In *Inconsistency Tolerance*, volume 3300 of *LNCS*, 1–16. Springer.

Chen, Q.; Zhang, C.; and Zhang, S. 2004. A verification model for electronic transaction protocols. In *Advanced Web Technologies and Applications*, volume 3007 of *LNCS*, 824–833. Springer.

Deng, X., and Papadimitriou, C. H. 1994. On the complexity of cooperative solution concepts. *Mathematics of Operations Research* 19(2):257–255.

Eiter, T., and Gottlob, G. 1992. On the complexity of propositional knowledge base revision , updates, and counterfactuals. *Artificial Intelligence* 57(2-3):227.

Grant, J., and Hunter, A. 2006. Measuring inconsistency in knowledgebases. *Journal of Intelligent Information Systems* 27:159–184.

Grant, J., and Hunter, A. 2008. Analysing inconsistent first-order knowledge bases. *Artificial Intelligence* 172:1064–1093.

Grant, J. 1978. Classifications for inconsistent theories. *Notre Dame Journal of Formal Logic* 19:435–444.

Grégoire, E.; Mazure, B.; and Piette, C. 2007. Boosting a complete technique to find mss and mus thanks to a local search oracle. In *International Joint Conference on Artificial Intelligence(IJCAI'07)*, 2300–2305.

Grégoire, E.; Mazure, B.; and Piette, C. 2008. Using local search to find msses and muses. *European Journal of Operational Research(EJOR)*. to appear.

Hunter, A., and Konieczny, S. 2004. Approaches to measuring inconsistent information. In *Inconsistency Tolerance*, volume 3300 of *LNCS*. Springer. 189–234.

Hunter, A., and Konieczny, S. 2006. Shapley inconsistency values. In *Proceedings of the 10th International Conference on Knowledge Representation (KR'06)*, 249–259. AAAI Press.

Hunter, A., and Nuseibeh, B. 1998. Managing inconsistent specifications: Reasoning, analysis and action. *ACM Transactions on Software Engineering and Methodology* 7:335–367.

Hunter, A. 2002. Measuring inconsistency in knowledge via quasi-classical models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'02)*, 68–73. MIT Press.

Hunter, A. 2003. Evaluating significance of inconsistencies. In *Proceedings of the 18th International Joint Conference on Artificial Intellignce (IJCAI'03)*, 468–473.

Hunter, A. 2004. Logical comparison of inconsistent perspectives using scoring functions. *Knowledge and Information Systems Journal* 6:528–543.

Hunter, A. 2006. How to act on inconsistent news: Ignore, resolve, or reject. *Data and Knowledge Engineering* 57:221–239.

Kalyanpur, A.; Parsia, B.; Sirin, E.; and Hendler, J. 2005. Debugging unsatisfiable classes in owl ontologies. *Journal of Web Semantics* 3(4):268–293.

Kautz, H., and Selman, B. 2007. The state of SAT. *Discrete Applied Mathematics* 155(12):1514–1524.

Knight, K. 2001. Measuring inconsistency. *Journal of Philosophical Logic* 31:77–98.

Knight, K. 2003. Two information measures for inconsistent sets. *Journal of Logic, Language and Information* 12:227–248.

Konieczny, S.; Lang, J.; and Marquis, P. 2003. Quantifying information and contradiction in propositional logic through epistemic actions. In *Proceedings of the 18th International Joint Conference on Artificial Intellignce (IJCAI'03)*, 106–111.

Koriche, F. 2001. On anytime coherence-based reasoning. In *Quantitative and Qualitative Approaches to Reasoning with Uncertainty*, volume 2143 of *LNCS*, 556–567. Springer.

Koriche, F. 2002. Approximate coherence-based reasoning. *Journal of Applied Non-classical Logics* 12(2):239–258.

Levesque, H. 1984. A logic of implicit and explicit belief. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'84)*, 198–202.

Lozinskii, E. 1994. Information and evidence in logic systems. *Journal of Experimental and Theoretical Artificial Intelligence* 6:163–193.

Luce, R. D., and Raiffa, H. 1957. *Games and Decision*. Wiley.

Ma, Y.; Qi, G.; Hitzler, P.; and Lin, Z. 2007. Measuring inconsistency for description logics based on paraconsistent semantics. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'07)*, volume 4724 of *LNCS*, 30–41. Springer.

Martinez, M.; Pugliese, A.; Simari, G.; Subrahmanian, V.; and Prade, H. 2007. How dirty is your relational database? an axiomatic approach. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'07)*, volume 4724 of *Lecture Notes in Computer Science*, 103 – 114. Springer.

Mu, K.; Jin, Z.; Lu, R.; and Liu, W. 2005. Measuring inconsistency in requirements engineering. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, volume 3571 of *LNAI*, 440–451. Springer.

Parsia, B.; Sirin, E.; and Kalyanpur, A. 2005. Debugging owl ontologies. In *Proceedings of the 14th international conference on World Wide Web (WWW'05)*, 633–640. ACM.

Qi, G., and Hunter, A. 2007. Measuring incoherence in description logic-based ontologies. In *The Semantic Web, 6th International Semantic Web Conference (ISWC'07)*, volume 4825 of *LNCS*, 381–394. Springer.

Qi, G.; Liu, W.; and Bell, D. 2005. Measuring conflict and agreement between two prioritized belief bases. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*, 552–557.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence* 32:57–95.

Schaerf, M., and Cadoli, M. 1995. Tractable reasoning via approximation. *Artificial Intelligence* 74:249–310.

Shapley, L. S. 1953. A value for n-person games. In Kuhn, H. W., and Tucker, A. W., eds., *Contributions to the Theory of Games II (Annals of Mathematics Studies 28)*, 307–317. Princeton University Press.

Wong, P., and Besnard, P. 2001. Paraconsistent reasoning as an analytic tool. *Journal of the Interest Group in Propositional Logic* 9:233–246.

IMPROVEMENT OPERATORS

Sébastien Konieczny, Ramón Pino Pérez.
Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08).
pages 177-186.
2008.

Improvement Operators

Sébastien Konieczny
CRIL - CNRS
Université d'Artois, Lens, France
konieczny@cril.fr

Ramón Pino Pérez
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Los Andes
Mérida, Venezuela
pino@ula.ve

Abstract

We introduce a new class of change operators. They are a generalization of usual iterated belief revision operators. The idea is to relax the success property, so the new information is not necessarily believed after the improvement. But its plausibility has increased in the epistemic state. So, iterating the process sufficiently many times, the new information will be finally believed. We give syntactical and semantical characterizations of these operators.

Introduction

Modelling belief change is a central topic in artificial intelligence, psychology and databases. One of the predominant approaches was proposed by Alchourrón, Gärdenfors and Makinson and is known as the AGM belief revision framework (Alchourrón, Gärdenfors, & Makinson 1985; Gärdenfors 1988; Katsuno & Mendelzon 1991). The main requirements imposed by AGM postulates are the principle of *coherence* asking to maintain consistency as far as possible, the so called principle of *minimal change* saying that we have to keep as much of the old information as possible, and the last important requirement is the principle of *primacy of update* (also called success property) that demands the new information to be true in the new belief base. The postulates proposed to characterize belief revision operators (Alchourrón, Gärdenfors, & Makinson 1985; Gärdenfors 1988; Katsuno & Mendelzon 1991; Hansson 1999) just aimed at capturing logically these principles.

A drawback of AGM definition of revision is that the conditions for the iteration of the process are very weak, and this is caused by the lack of expressive power of logical belief bases (Herzig, Konieczny, & Perussel 2003). In order to ensure good properties for the iteration of the revision process, one needs a more complex structure. So shifting from logical belief bases to epistemic states was proposed in (Darwiche & Pearl 1997). In this framework, one can define interesting iterated revision operators (Darwiche & Pearl 1997; Booth & Meyer 2006; Jin & Thielscher 2007; Konieczny & Pino Pérez 2000). Let us call these operators DP belief revision operators.

Another framework that allows to define interesting iterated change operator is the one of Ordinal Conditional Functions (OCF), also named kappa-rankings, that was proposed by Spohn in (Spohn 1988), and further developed in (Williams 1994). An OCF can be represented by a function that associates an ordinal to every interpretation, with at least one interpretation taking the value 0. The ordinal associated to the interpretation represents the degree of disbelief of the interpretation. This notion can be used to define a *degree of acceptance* of a formula.

So a change operator in this framework, called a transmutation (Williams 1994), is a function that changes the degree of acceptance of a formula. This means that it requires more information than AGM/DP belief revision operators, since, in addition to the new information, one needs to give its new degree of acceptance. This has one important drawback, since one has to find this new degree somewhere! It it is not a problem if it is given by the application, but if the only received input is the new information, justifying an "arbitrary" degree of acceptance can be problematic. On the other hand, this more general framework allows to define interesting operators. It allows to define revision and contraction operators, by choosing the right degree of acceptance. In particular most of the works on DP iterared revision operators use OCF operators as examples (see e.g. (Darwiche & Pearl 1997; Jin & Thielscher 2007)). But it also allows to define restructuring operators (Williams 1994; Spohn 1988), that modify the OCF, without changing the believed formulae. Such operators do not exist in the classical DP belief revision framework, that obey the success property, that asks the new information to be believed after the change.

The aim of this paper is to define such restructuring-like operators in the standard AGM/DP framework. We want to define change operators on epistemic states that do not (necessarily) satisfy the success property, although still improving the plausibility of the new information. We call these operators improvement operators. This idea is quite intuitive since usual AGM/DP belief revision operators can be considered as too strong: after revising by a new information, this information will be believed. Most of the time this is the wanted behaviour for the revision operators. But in some cases it may be sensible to take into account the new information more cautiously. Maybe because we have some confidence in the source of the new information, but not enough for accepting unconditionally this new information. This can be seen as a kind of learning/reinforcement process: each time the agent receives a new information α, this formula will gain in plausibility in the epistemic state of the agent. And if the agent receives the same new information many times, then he will finally believe it.

Our operators are close in spirit to the bad day/good day approach of Booth and Meyer (also called abstract interval orders revision) (Booth & Meyer 2007; Booth, Meyer, & Wong 2006). Unlike their operators that need an extra information, our operators are defined in the usual DP framework. We give more details on this relationship at the end of the paper.

The rest of the paper is organized as follows: we give the preliminaries in the first section. The second section is devoted to the introduction of improvement operators. The third section is devoted to a discussion of the irrelevance of syntax property. In the fourth and fifth sections we state the main results concerning syntactical and semantical characterizations, namely two representation theorems. The fifth section shows an example and how to encode improvement using OCF. The sixth section contains some interesting properties of improvement operators. The last section is the conclusion. There is also an appendix containing the proofs of the main results.

Preliminaries

We consider a propositional language $\mathcal{L}$ defined from a finite set of propositional variables $\mathcal{P}$ and the standard connectives. Let $\mathcal{L}^*$ denote the set of consistent formulae of $\mathcal{L}$.

An interpretation ω is a total function from $\mathcal{P}$ to $\{0,1\}$. The set of all interpretations is denoted $\mathcal{W}$. An interpretation ω is a model of a formula $\phi \in \mathcal{L}$ if and only if it makes it true in the usual truth functional way. $[\![\alpha]\!]$ denotes the set of models of the formula α, i.e., $[\![\alpha]\!] = \{\omega \in \mathcal{W} \mid \omega \models \alpha\}$. When $\{w_1, .., w_n\}$ is a set of models we denote by $\varphi_{w_1,..,w_n}$ a formula such that $[\![\varphi_{w_1,..,w_n}]\!] = \{w_1, .., w_n\}$.

We will use epistemic states to represent the beliefs of the agent, as usual in iterated belief revision (Darwiche & Pearl 1997). An epistemic state Ψ represents the current beliefs of the agent, but also additional conditional information guiding the revision process (usually represented by a pre-order on interpretations, a set of conditionals, a sequence of formulae, etc). Let $\mathcal{E}$ denote the set of all epistemic states. A projection function $B : \mathcal{E} \longrightarrow \mathcal{L}^*$ associates to each epistemic state Ψ a consistent formula $B(\Psi)$, that represents the current beliefs of the agent in the epistemic state Ψ.

For simplicity purpose we will only consider in this paper consistent epistemic states and consistent new information. Thus, we consider change operators as functions $\circ$ mapping an epistemic state and a consistent formula into a new epistemic state, *i.e.* in symbols, $\circ : \mathcal{E} \times \mathcal{L}^* \longrightarrow \mathcal{E}$. The image of a pair (Ψ, α) under $\circ$ will be denoted by $\Psi \circ \alpha$.

We adopt the following notations:

- $\Psi \circ^n \alpha$ defined as: $\begin{array}{lcl} \Psi \circ^1 \alpha & = & \Psi \circ \alpha \\ \Psi \circ^{n+1} \alpha & = & (\Psi \circ^n \alpha) \circ \alpha \end{array}$
- $\Psi \star \alpha = \Psi \circ^n \alpha$, where n is the first integer such that $B(\Psi \circ^n \alpha) \vdash \alpha$.

Note that $\star$ is undefined if there is no n such that $B(\Psi \circ^n \alpha) \vdash \alpha$, but for all operators $\circ$ considered in this work, the associated operator $\star$ will be total, that is for any pair Ψ, α there will exist n such that $B(\Psi \circ^n \alpha) \vdash \alpha$ (see postulate **(I1)**) below.

Finally, let $\leq$ be a a total pre-order, i.e a reflexive ($x \leq x$), transitive ($(x \leq y \wedge y \leq z) \rightarrow x \leq z$) and total ($x \leq y \vee y \leq x$) relation over $\mathcal{W}$. Then the corresponding strict relation $<$ is defined as $x < y$ iff $x \leq y$ and $y \not\leq x$, and the corresponding equivalence relation $\simeq$ is defined as $x \simeq y$ iff $x \leq y$ and $y \leq x$. We denote $w \ll w'$ when $w < w'$ and there is no w'' such that $w < w'' < w'$. We also use the notation $min(A, \leq) = \{w \in A \mid \nexists w' \in A \; w' < w\}$.

When a set $\mathcal{W}$ is equipped with a total pre-order $\leq$, then this set can be splitted in different levels, that gives the ordered sequence of its equivalence classes $\mathcal{W} = \langle S_0, \ldots S_n \rangle$. So $\forall x, y \in S_i \; x \simeq y$. We say in that case that x and y are at the same level of the pre-order. And $\forall x \in S_i \; \forall y \in S_j \; i < j$ implies $x < y$. We say in this case that x is in a lower level than y. We extend straightforwardly these definitions to compare subsets of equivalence classes, i.e if $A \subseteq S_i$ and $B \subseteq S_j$ then we say that A is in a lower level than B if $i < j$.

Improvement operators

First, let us state the basic logical properties that are asked for improvement operators.

Definition 1 *An operator $\circ$ is said to be a weak improvement operator if it satisfies (I1) to (I6):*
(I1) *There exists n such that $B(\Psi \circ^n \alpha) \vdash \alpha$*
(I2) *If $B(\Psi) \wedge \alpha \nvdash \bot$, then $B(\Psi \star \alpha) \equiv B(\Psi) \wedge \alpha$*
(I3) *If $\alpha \nvdash \bot$, then $B(\Psi \circ \alpha) \nvdash \bot$*
(I4) *For any positive integer n if $\alpha_i \equiv \beta_i$ for all $i \leq n$ then $B(\Psi \circ \alpha_1 \circ \cdots \circ \alpha_n) \equiv B(\Psi \circ \beta_1 \circ \cdots \circ \beta_n)$*
(I5) $B(\Psi \star \alpha) \wedge \beta \vdash B(\Psi \star (\alpha \wedge \beta))$
(I6) *If $B(\Psi\star\alpha)\wedge\beta \nvdash \bot$, then $B(\Psi\star(\alpha\wedge\beta)) \vdash B(\Psi\star\alpha)\wedge\beta$*

We have put together these properties because they allow to obtain a first basic representation theorem (see Theorem 1). These properties are very close to the usual ones for iterated belief revision (Darwiche & Pearl 1997). Note nevertheless that there is a real difference since in usual formulation $\star$ is a revision operator, whereas here it denotes a sequence of improvements.

Remark that (I3) is a straightforward consequence of the definition of the operator $\circ$, since we ask the new information and the epistemic states to be consistent. Although (I3) is redundant in our framework, we have chosen to put it explicitly to remain close to the usual DP postulates.

The main difference with usual belief revision operators is that we do not ask the fundamental success property $B(\Psi \circ \alpha) \vdash \alpha$. We ask instead the weaker (I1), that just requires that after a sequence of improvements, we will finally imply the new information. So this means that the (revision) operator $\star$ defined as a sequence of improvements $\circ$ is always defined.

Postulate (I4) is also stonger than the usual version of (Darwiche & Pearl 1997). We discuss it in the next Section.

Before establishing more specific postulates concerning the iteration by different formulas, we have to define new notions that help us to keep light notations.

Definition 2 *Let $\circ$ be a change operator satisfying (I1). Let α, β and Ψ be two formulae and an epistemic state respectively. We say that α is below β with respect to Ψ, given $\circ$, denoted $\alpha \prec^{\circ}_{\Psi} \beta$ (or simply $\alpha \prec_{\Psi} \beta$ if there is no ambiguity about $\circ$) if and only if $B(\Psi \star \alpha) \vdash B(\Psi \star (\alpha \vee \beta))$ and $B(\Psi \star \beta) \not\vdash B(\Psi \star (\alpha \vee \beta))$.*

The pair (α, β) is Ψ-consecutive, denoted $\alpha \prec\!\!\prec^{\circ}_{\Psi} \beta$ (or simply $\alpha \prec\!\!\prec_{\Psi} \beta$ if there is no ambiguity about $\circ$) if and only if $\alpha \prec_{\Psi} \beta$ and there is no formula γ such that $\alpha \prec_{\Psi} \gamma \prec_{\Psi} \beta$.

The idea of these two definitions is that $\alpha \prec^{\circ}_{\Psi} \beta$ denotes that α is more entrenched (plausible) than β in the epistemic state Ψ. And $\alpha \prec\!\!\prec^{\circ}_{\Psi} \beta$ denotes the fact that α is a formula immediately more entrenched (plausible) than β.

Now we are ready to state the postulates concerning more specific properties of iteration:

Definition 3 *A weak improvement operator is said to be an improvement operator if it satisfies I7 to I11*

(I7) *If $\alpha \vdash \mu$ then $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$*

(I8) *If $\alpha \vdash \neg\mu$ then $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$*

(I9) *If $B(\Psi \star \alpha) \not\vdash \neg\mu$ then $B((\Psi \circ \mu) \star \alpha) \vdash \mu$*

(I10) *If $B(\Psi \star \alpha) \vdash \neg\mu$ then $B((\Psi \circ \mu) \star \alpha) \not\vdash \mu$*

(I11) *If $B(\Psi \star \alpha) \vdash \neg\mu$, $\alpha \wedge \mu \not\vdash \bot$ and $\alpha \prec\!\!\prec_{\Psi} \alpha \wedge \mu$ then $B((\Psi \circ \mu) \star \alpha) \not\vdash \neg\mu$*

A first observation about these postulates is that they are expressed in terms of both $\circ$ and $\star$. And that it is thanks to these several iterations until revision[1], modeled by $\star$, that we can define powerful properties on $\circ$. Postulates (I7), (I8) are close to the properties (C1) and (C2) of (Darwiche & Pearl 1997), but translated for weak improvement operators. Postulate (I9) is also close to the property of Independence in (Jin & Thielscher 2007) (called also property (P) in (Booth & Meyer 2006)), but also translated for weak improvement operators. Postulates (I9) and (I11) deals with the improvement of the new information, i.e. the increase of its plausibility in the epistemic state. Postulates (I10) and (I11) deals with the the cautiousness of the approach, i.e. they express the fact that the increase of the plausibility of the new information is limited. Postulate (I10) says that if after a sequence of improvements by α, the obtained epistemic state imply $\neg\mu$, then, if before the sequence of improvements by α we improve by μ, then it will not be enough to imply μ after the sequence of improvements. This means that it is not possible to go directly by an improvement from an epistemic state where a formula is believed to one where its negation is believed. And postulate (I11) captures some of the ideas behind improvement operators as "small change" operators. Basically it says that if $\neg\mu$ is believed when revising by α, but μ is quite plausible given α, then improving by μ before starting the sequence of improvements needed to revise by α will be enough to ensure the result to be consistent with μ.

[1]Note that the $\star$ operator satisfies the success property, so it can be called *revision* operator. We will use this term in the following. The fact that $\star$ is a true AGM/DP revision operator will be proved in Corollary 1.

Irrelevance of syntax

As it has been pointed out by many authors (see for instance (Darwiche & Pearl 1997; Booth & Meyer 2006)) the postulate of independence (or irrelevance) of syntax is a delicate matter for epistemic states. Actually a basic translation of Darwiche and Pearl (R*4) in our framework would lead to:

$$\text{If } \alpha \equiv \beta, \text{then } B(\Psi \circ \alpha) \equiv B(\Psi \circ \beta) \qquad (1)$$

But even adding this postulate is not sufficient. Booth and Meyer have well illustrated this idea in (Booth & Meyer 2006). Actually, it is not enough that (1) holds in order to have a good iterative behavior with respect to revision by sequences of equivalent formulae. Consider:

$$\text{If } (\alpha \equiv \beta \;\&\; \gamma \equiv \theta), \text{then } B(\Psi \circ \alpha \circ \gamma) \equiv B(\Psi \circ \beta \circ \theta) \quad (2)$$

Postulate (1) doesn't entail postulate (2). So the good behaviour with respect to equivalent formulae is not guaranteed on two iterations. This is why Booth and Meyer have proposed in (Booth & Meyer 2006) to replace the usual postulate (1) by (2) in the usual DP framework.

We agree with Booth and Meyer that postulate (1) is not enough. But we think that (2) does not go far enough. In fact the example they give for showing that Postulate (1) does not avoid the problem at the second iteration can be easily extended to show that Postulate (2) does not avoid the problem at the third iteration. So one has to specify this for every number of iterations. That leads to the postulate **(I4)**.

The following example, which follows the same lines of the Example 1 in (Booth & Meyer 2006), shows that replacing the postulate (1) by the postulate (2) in the basic RAGM[2] framework is not enough to get the postulate **(I4)**.

Example 1 *Take a language with only two propositional letters p and q (in this order when we consider the interpretations). Let $\varphi_1 = p \vee \neg p$ and $\varphi_2 = q \vee \neg q$. Let Φ such that $\Phi \circ \varphi_1 \neq \Phi \circ \varphi_2$. Note that this is compatible with RAGM plus* (2)*, because the only constraint imposed by* (2) *is that $\leq_{\Phi\circ\varphi_1} = \leq_{\Phi\circ\varphi_2}$ but not that $\Phi \circ \varphi_1 = \Phi \circ \varphi_2$. Moreover we can take $B(\Phi \circ \varphi_1) = p \wedge q = B(\Phi \circ \varphi_2)$. Let Φ_1, Φ_2 be two epistemic states such that $B(\Phi_1) = p = B(\Phi_2)$, $00 <_{\Phi_1} 01$ and $01 <_{\Phi_2} 00$. Now, it is compatible with RAGM plus* (2) *to put $\leq_{\Phi\circ\varphi_1\circ p} = \leq_{\Phi_1}$ and $\leq_{\Phi\circ\varphi_2\circ\neg\neg p} = \leq_{\Phi_2}$. But then, by the representation, we have $B(\Phi \circ \varphi_1 \circ p \circ \neg p) = \neg p \wedge \neg q$ and $B(\Phi \circ \varphi_2 \circ \neg\neg p \circ \neg p) = \neg p \wedge q$, what clearly is a counter-example to* **(I4)***.*

Nevertheless, it worths noticing that all the well-known iterated revision operators, as natural revision (Boutilier 1996), Darwiche and Pearl $\bullet$ operator (Darwiche & Pearl 1997), Nayak's lexicographic revision (Nayak 1994;

[2]RAGM is the name that Booth and Meyer give to AGM/DP belief revision operators satisfying (R*1)-(R*6) (Darwiche & Pearl 1997).

Konieczny & Pino Pérez 2000) for instance, satisfy **(I4)**. The reason behind this phenomenon is that all these operators satisfy the following property

$$\text{If } \alpha \equiv \beta \text{ and } \leq_\Psi = \leq_\Phi \text{ then } \leq_{\Psi \circ \alpha} = \leq_{\Phi \circ \beta} \tag{3}$$

In the presence of the other RAGM properties, the previous property entails the property (2). But the converse is not true as we can see via the example 1.

Remark also that usual belief revision operators satisfy all other weak improvement properties (with $n = 1$ in (I1)), so weak improvements operators are a generalization of usual DP belief revision operators (Darwiche & Pearl 1997).

Representation theorem

Let us first define strong faithful assignements.

Definition 4 *A function $\Psi \mapsto \leq_\Psi$ that maps each epistemic state Ψ to a total pre-order on interpretations $\leq_\Psi$ is said to be a* strong faithful assignment *if and only if:*

1. *If $w \models B(\Psi)$ and $w' \models B(\Psi)$, then $w \simeq_\Psi w'$*
2. *If $w \models B(\Psi)$ and $w' \not\models B(\Psi)$, then $w <_\Psi w'$*
3. *For any positive integer n if $\alpha_i \equiv \beta_i$ for any $i \leq n$ then $\leq_{\Psi \circ \alpha_1 \circ \cdots \circ \alpha_n} = \leq_{\Psi \circ \beta_1 \circ \cdots \circ \beta_n}$*

Note that conditions 1 and 2 are equivalent to $[\![B(\Psi)]\!] = \min(\mathcal{W}, \leq_\Psi)$, and are the usual ones for faithful assignment (Darwiche & Pearl 1997). Condition 3 is a very natural condition that links pre-orders associated to iteration of improvements: two sequences of improvements of the same pre-order by equivalent formulae lead to the same pre-order.

Let us first show a first representation theorem on weak improvement operators, before turning on the more interesting iteration properties.

Theorem 1 *A change operator $\circ$ is a weak improvement operator if and only if there exists a strong faithful assignment that maps each epistemic state Ψ to a total pre-order on interpretations $\leq_\Psi$ such that*

$$[\![B(\Psi \star \alpha)]\!] = \min([\![\alpha]\!], \leq_\Psi) \tag{4}$$

It is easy to check that the faithful assignment representing $\circ$ in the previous theorem is unique.

An obvious corrolary of the previous Theorem and its proof is the following one:

Corollary 1 *If $\circ$ is a weak improvement operator, then $\star$ is an AGM/DP revision operator, i.e. it satisfies (R*1)-(R*6) of (Darwiche & Pearl 1997).*

As a consequence of the previous theorem we have also the following trichotomy property:

Proposition 1 *Let $\circ$ be a weak improvement operator. Then*

$$B(\Psi \star (\alpha \vee \beta)) = \begin{cases} B(\Psi \star \alpha) \text{ or} \\ B(\Psi \star \beta) \text{ or} \\ B(\Psi \star \alpha) \vee B(\Psi \star \beta) \end{cases}$$

Let us now give two corollaries of these results, that are useful to understand the definitions of $\prec_\Psi$ and $\prec\!\!\prec_\Psi$, and that will be useful in the proof of the main Theorem (Theorem 2).

Corollary 2 *Let $\circ$ be a weak improvement operator. Then $\alpha \prec_\Psi \beta$ if and only if there exist w, w' such that $w \in [\![B(\Psi \star \alpha)]\!]$, $w' \in [\![B(\Psi \star \beta)]\!]$, $w <_\Psi w'$.*

Corollary 3 *Let $\circ$ be a weak improvement operator. Then $\alpha \prec\!\!\prec_\Psi \beta$ if and only if there exist w, w' such that $w \in [\![B(\Psi \star \alpha)]\!]$, $w' \in [\![B(\Psi \star \beta)]\!]$, $w <_\Psi w'$ and there is no w'' such that $w <_\Psi w'' <_\Psi w'$.*

Main result

Let us turn now to the main representation result about improvement operators.

Definition 5 *Let $\circ$ be a weak improvement operator and $\Psi \mapsto \leq_\Psi$ its corresponding strong faithful assignment. The assignment will be called a gradual assignment if the properties S1, S2, S3, S4 and S5 are satisfied*

(S1) *If $w, w' \in [\![\alpha]\!]$ then $w \leq_\Psi w' \Leftrightarrow w \leq_{\Psi \circ \alpha} w'$*
(S2) *If $w, w' \in [\![\neg\alpha]\!]$ then $w \leq_\Psi w' \Leftrightarrow w \leq_{\Psi \circ \alpha} w'$*
(S3) *If $w \in [\![\alpha]\!], w' \in [\![\neg\alpha]\!]$ then $w \leq_\Psi w' \Rightarrow w <_{\Psi \circ \alpha} w'$*
(S4) *If $w \in [\![\alpha]\!], w' \in [\![\neg\alpha]\!]$ then $w' <_\Psi w \Rightarrow w' \leq_{\Psi \circ \alpha} w$*
(S5) *If $w \in [\![\alpha]\!], w' \in [\![\neg\alpha]\!]$ then $w' \ll_\Psi w \Rightarrow w \leq_{\Psi \circ \alpha} w'$*

Properties (S1) and (S2) correspond to usual properties (CR1) and (CR2) for DP iterated revision operators (Darwiche & Pearl 1997). Property (S3) is the new property proposed in (Jin & Thielscher 2007; Booth & Meyer 2006), and that forces to increase the plausibility of the models of the new information. Property (S4) shows how the increase of plausibility of the models of the new information is limited by improvement operators. This is an important difference with usual DP iterated revision operators (Darwiche & Pearl 1997; Jin & Thielscher 2007; Booth & Meyer 2006). Property (S5) asks (together with (S4)) that if a model of $\neg\alpha$ is just a little more plausible than a model of α, then after improvement the two models will have the same plausibility.

Theorem 2 *A change operator $\circ$ is an improvement operator if and only if there exists a gradual assignment such that*

$$[\![B(\Psi \star \alpha)]\!] = \min([\![\alpha]\!], \leq_\Psi)$$

This theorem has important consequences. In particular the relationship between $\leq_\Psi$ and $\leq_{\Psi \circ \alpha}$ imposed by Definition 5 is very tight. Actually, the total pre-order $\leq_{\Psi \circ \alpha}$ is completely determined by $\leq_\Psi$ and α as it will be stated in Proposition 2. We first need the following lemma:

Lemma 1 *Let $\circ$ be an improvement operator and $\Psi \mapsto \leq_\Psi$ its gradual assignment. If $w <_\Psi w'$, $w \in [\![\neg\alpha]\!]$, $w' \in [\![\alpha]\!]$ and $w \not\ll_\Psi w'$ then $w <_{\Psi \circ \alpha} w'$.*

This Lemma is interesting since it gives the missing relation between $\leq_{\Psi \circ \alpha}$ and $\leq_\Psi$, since all other relations are given by the properties of Definition 5.

Proposition 2 *Let $\circ$ be an improvement operator and $\Psi \mapsto \leq_\Psi$ its gradual assignment. Then for every formula α, the pre-order $\leq_{\Psi \circ \alpha}$ is completely determined by $\leq_\Psi$ and $[\![\alpha]\!]$.*

$w \in [\![\alpha]\!]$	$w' \in [\![\alpha]\!]$	$w \leq_\Psi w' \Leftrightarrow w \leq_{\Psi\circ\alpha} w'$	**(S1)**
$w \in [\![\neg\alpha]\!]$	$w' \in [\![\neg\alpha]\!]$	$w \leq_\Psi w' \Leftrightarrow w \leq_{\Psi\circ\alpha} w'$	**(S2)**
$w \in [\![\alpha]\!]$	$w' \in [\![\neg\alpha]\!]$	$w <_\Psi w' \Leftrightarrow w <_{\Psi\circ\alpha} w'$ $w \simeq_\Psi w' \Rightarrow w <_{\Psi\circ\alpha} w'$ $w' \ll_\Psi w \Rightarrow w \simeq_{\Psi\circ\alpha} w'$ $w' <_\Psi w \wedge w' \not\ll_\Psi w \Rightarrow w <_{\Psi\circ\alpha} w'$	**(S3)** **(S3)** **(S4) & (S5)** **(Lemma 1)**

Table 1: From $\leq_\Psi$ to $\leq_{\Psi\circ\alpha}$

This proposition is a very important one since it says, in a sense, that there is a unique improvement operator. In fact one can define different improvement operators by assigning to the initial epistemic state of the sequence different pre-orders. Once this pre-order is known, Proposition 2 tells us that there is no more freedom on the choice of subsequent pre-orders.

So clearly if ones considers pre-orders on interpretations as epistemic states (recall that the representation theorem just says that we can associate a pre-order on interpretation to each epistemic state, it does not presume anything on the exact nature of these epistemic states), then there is a unique improvement operator.

The exact construction of $\leq_{\Psi\circ\alpha}$ from $\leq_\Psi$ is given in Table 1. Roughly speaking, $\leq_{\Psi\circ\alpha}$ is obtained by shifting down one level the models of α in the total pre-order $\leq_\Psi$. This will be stated more formally in the next section.

Concrete example: Improvement via OCF

Let us now show how to implement an improvement operator via OCF. Let us denote Ord the class of ordinals.

Definition 6 *An Ordinal Conditional Function (OCF) κ is a function from the set of interpretations $\mathcal{W}$ to the set of ordinals such that at least one interpretation is assigned* 0*.*
A function from the set of interpretations $\mathcal{W}$ to the set of ordinals will be called a free OCF.
The set of OCF will be denoted $\mathcal{K}$.

Let us now state how to implement improvement using the framework of OCF. More precisely we will give two results: first we will show how to simply compute the resulting pre-order after an improvement, using a translation through free OCFs. Then, we will see how to define an improvement operator in the OCF framework.

So, what we do first is the following: giving $\leq_\Psi$ and α we describe $\leq_{\Psi\circ\alpha}$ using the machinery of OCF. At this point let us recall the equivalent view of a total pre-order $\leq$ introduced in the preliminaries: a total pre-order over $\mathcal{W}$ can be seen as the splitting of the set $\mathcal{W}$ in different levels (the equivalent classes), $\langle S_0, \ldots S_n\rangle$m the ordered sequence of its equivalence classes. Thus, $\forall x, y \in S_i\ x \simeq y$ and $\forall x \in S_i$ $\forall y \in S_j$ $i < j$ implies $x < y$.

Let κ be the canonical representative of $\leq_\Psi$, *i.e.* if $\leq_\Psi$ has n levels and w is in the level i, $\kappa(w) = i$.

Consider now the following free OCF:

$$\kappa_\alpha(w) = \begin{cases} \kappa(w) & \text{if } w \models \alpha \\ \kappa(w)+1 & \text{if } w \models \neg\alpha \end{cases}$$

It is not hard to see that this free OCF represents $\leq_{\Psi\circ\alpha}$, that is, the total pre-order associated to this function in the natural way ($w \leq_{\kappa_\alpha} w'$ iff $\kappa_\alpha(w) \leq \kappa_\alpha(w')$) satisfies all the properties of Table 1. So this result just aims at illustrating simply the behaviour of improvement operators in terms of total pre-orders via free OCF.

Note however that the free OCF used above is very particular, since it is build from the pre-order $\leq_\Psi$. In order to be able to define an improvement operator on any given OCF, this requires much more difficult definitions in order to modelize the smooth increase of plausibility of improvement operators.

So now we turn to a plain representation of an improvement operator $\circ$ in the full OCF framework. Thus, we assume that epistemic states are indeed OCF's and $\circ : \mathcal{K} \times \mathcal{L}^* \longrightarrow \mathcal{K}$. In this framework, we define the function B by putting $[\![B(\kappa)]\!] = \{w : \kappa(w) = 0\}$.

Remember that $\kappa(\alpha) = min\{\kappa(w) : w \in [\![\alpha]\!]\}$. Given κ and α, an OCF and a consistent formula respectively, we are going to define the new OCF $\kappa \circ \alpha$ by cases according to $\kappa(\alpha) > 0$ or $\kappa(\alpha) = 0$. In the first case ($\kappa(\alpha) > 0$) we perform the sliding down the models of α via an auxiliary function called $f^\kappa_{\alpha\downarrow}$ defined below. In the second case ($\kappa(\alpha) = 0$) we simulate the sliding down the models of α via an auxiliary function called $f^\kappa_{\alpha\uparrow}$ defined below that performs the sliding up the models of $\neg\alpha$. The functions $f^\kappa_{\alpha\downarrow} : [\![\alpha]\!] \longrightarrow Ord$ and $f^\kappa_{\alpha\uparrow} : [\![\neg\alpha]\!] \longrightarrow Ord$ are defined by putting

$$f^\kappa_{\alpha\downarrow}(w) = \begin{cases} max\{\rho : \exists w' \in [\![\neg\alpha]\!]\ \kappa(w') = \rho\ \&\ \rho < \kappa(w) \\ \quad \&\ \nexists w'' \in [\![\alpha]\!]\ \rho < \kappa(w'') < \kappa(w)\} \\ \quad \text{if this set is nonempty} \\ \kappa(w) - 1 \ \text{ otherwise} \end{cases}$$

$f^\kappa_{\alpha\downarrow}$ maps w, a model of α, into the first rank below $\kappa(w)$ where there is a model of $\neg\alpha$ in the case that there is no models of α strictly in between this two levels. Otherwise $f^\kappa_{\alpha\downarrow}$ maps w into $\kappa(w) - 1$.

$$f^\kappa_{\alpha\uparrow}(w) = \begin{cases} min\{\rho : \exists w' \in [\![\alpha]\!]\ \kappa(w') = \rho\ \&\ \kappa(w) < \rho \\ \quad \&\ \nexists w'' \in [\![\neg\alpha]\!]\ \kappa(w) < \kappa(w'') < \rho\} \\ \quad \text{if this set is nonempty} \\ \kappa(w) + 1 \ \text{ otherwise} \end{cases}$$

$f^{\kappa}_{\alpha\uparrow}$ maps w, a model of $\neg\alpha$, into the first rank above $\kappa(w)$ where there is a model of α in the case that there is no models of $\neg\alpha$ strictly in between this two ranks. Otherwise $f^{\kappa}_{\alpha\uparrow}$ maps w into $\kappa(w) + 1$.

Again, we define two functions mapping worlds into ordinals according to whether or not $\kappa(\alpha) = 0$. When $\kappa(\alpha) = 0$ we put

$$\kappa \uparrow \neg\alpha(w) = \begin{cases} \kappa(w) & \text{if } w \models \alpha \\ f^{\kappa}_{\alpha\uparrow}(w) & \text{if } w \models \neg\alpha \end{cases}$$

and when $\kappa(\alpha) > 0$ we put

$$\kappa \downarrow \alpha(w) = \begin{cases} \kappa(w) & \text{if } w \models \neg\alpha \\ f^{\kappa}_{\alpha\downarrow}(w) & \text{if } w \models \alpha \end{cases}$$

Finally we define $\kappa \circ \alpha$ by putting

$$\kappa \circ \alpha = \begin{cases} \kappa \uparrow \neg\alpha & \text{if } \kappa(\alpha) = 0 \\ \kappa \downarrow \alpha & \text{if } \kappa(\alpha) > 0 \end{cases}$$

Let us now take an example in order to see how it works.

Example 2 *Consider a language with propositional variables p, q and r in this order. Let κ be the OCF with image $\{0, 1, 2, 4, 5\}$ described in the diagram below and α a formula such that $\alpha = \neg p$. The following diagrams shows κ, $\kappa \circ \alpha$, $\kappa \circ \alpha \circ \alpha$ and $\kappa \circ \alpha \circ \alpha \circ \alpha$ (the models of α are in boldface):*

5	110	110	5
4	**011**	− − − − − −	4
3	− − − − − −	**011**	3
2	**010 000 001**	− − − − − −	2
1	101 100	101 100 **010 000 001**	1
0	111	111	0
	κ	$\kappa \circ \alpha$	

6		110	6
5	110	− − − − − −	5
4	− − − − − −	− − − − − −	4
3	− − − − − −	− − − − − −	3
2	− − − − − −	101 100	2
1	101 100 **011**	111 **011**	1
0	111 **010 000 001**	**010 000 001**	0
	$\kappa \circ \alpha \circ \alpha$	$\kappa \circ \alpha \circ \alpha \circ \alpha$	

Properties of improvement operators

Let us give now some additional properties on improvement operators, that illustrate how it relates with existing operators.

Proposition 3 *Improvement operators can not be represented as Spohn's Conditionalisation nor Williams' Adjustment.*

This is quite an intuitive result since Conditionalisation and Adjustement operate the same "global" change on the interpretation ranks, whereas, as sum up in Table 1, improvement requires a more adaptative behaviour (that depends more on the ranks of the other interpretations).

Proposition 4 • *There exists n such that $\Psi \circ^n \alpha$ is Nayak's lexicographic revision (Nayak 1994; Konieczny & Pino Pérez 2000). Let us note $\Psi \star_{lex} \alpha = \Psi \circ^n \alpha$.*

• *Actually, the first n such that $\Psi \star_{lex} \alpha = \Psi \circ^n \alpha$ is a fixed point for improvement by α, in the sense that $\leq_{\Psi \star_{lex} \alpha} = \leq_{\Psi \star_{lex} \alpha \circ \alpha}$.*

This can be shown easily with the help of Proposition 2 or with the representation of $\leq_{\Psi \circ \alpha}$ via the free OCF. In fact it requires at most k iterations where k is the level in $\leq_\Psi$ of the worst world of α (i.e. the model of α at the highest level) to reach this fixed point.

This last proposition is interesting since it illustrate the fact that the process of improvement does not stop as soon as the new information is believed. So in particular:

Proposition 5 *$B(\Psi) \vdash \alpha$ does not imply that $\leq_{\Psi \circ \alpha} = \leq_\Psi$ and therefore does not imply $\Psi \circ \alpha = \Psi$.*

It is worth noticing that even if we have a fixed point in the sense of Proposition 4, *i.e.* $\leq_\Psi = \leq_{\Psi \circ \alpha}$ we can have $\Psi \neq \Psi \circ \alpha$. The operator defined via the OCF is an example of such a situation.

Remark 1 *Improvements operators can be used to define contractions operators. Actualy, define $\Psi \odot \alpha = \Psi \circ^n \neg\alpha$ where n is the smallest integer such that $\Psi \circ^n \neg\alpha \nvdash \alpha$. Then, $\odot$ is a contraction operator.*

Let us now elaborate on the links between improvement operators and the bad day/good day approach of Booth et al. (Booth & Meyer 2007; Booth, Meyer, & Wong 2006) (also called abstract interval orders revision). In both cases the change is small, in the sense that the increase of plausibility of the models of the new information is limited. A first difference is that their operators are defined as revision of total pre-orders, whereas improvements are defined on general DP epistemic states. A second, more important difference between Booth et al. approach and ours is that they need an extra-logical information in order to guide the process, whereas our operators are completely defined in the usual DP framework. This is an important improvement, that allows for instance to easily iterate the process.

Actually, given $\leq_\Psi$, it is possible to define $\preceq$, a $\leq_\Psi$-*faithful tpo* (see (Booth, Meyer, & Wong 2006)), such that the revision of $\preceq$ by α, in the sense of Booth-Meyer-Wong, is exactly $\leq_{\Psi \circ \alpha}$. So, according to this link, improvement operators could be considered in a sense as a special case of Booth and Meyer operators.

Finally there is a very interesting behaviour of improvement operator with respect to long term behaviour. When working in a finite framework, no existing iterated belief revision operator escapes one of the following limit cases after a long course of revisions: maxichoice revision, or full meet revision.

Full meet revision means that the beliefs of the new epistemic state is either the conjunction of the new information with the beliefs of the old epistemic state it if is consistent, or just the new information otherwise. This is problematic, since it means that after a long course of revision the agent has lost all his beliefs. But for instance Lehmann's operators (Lehmann 1995) lead to this limit case when working on finite frameworks.

Maxichoice revision means that the revision leads to an epistemic states whose beliefs are a complete formula. This is also problematic, since it means that when this situation is reached, any revision by any formula will allow to be completely determined about each issue. It can be argued that this is due to the long revision history, that allows the agent to have a very precise view of the world. But still it seems sensible to be able to be uncertain about some issues, and to lose some certainties sometimes. Note that most of DP-like operators lead to this limit case (Darwiche & Pearl 1997; Nayak 1994; Boutilier 1996; Konieczny & Pino Pérez 2000; Booth & Meyer 2006; Jin & Thielscher 2007).

Note that it is not the case in the framework of OCFs, after any sequence of conditionalization, or adjustment, it is possible to reach any other OCF by some sequence. This is one of the advantage of using a more quantitative framework (using a degree of acceptance for each input formula), compared to the fully qualitative one that is the DP iterated belief revision framework.

It is interesting to note that improvement operators have no limit case. Actually, after any sequence of improvements, it is possible to reach any formula (as beliefs of an epistemic state) by an adequate sequence of improvements (the same result holds for associated pre-orders: any pre-order can be reached after an adequate sequence of improvements starting from any other pre-order). Whereas for DP iterated belief revision operators it is not the case: after some sequences of revisions, some formulae (or pre-orders) are not reachable anymore.

The following proposition summarizes this property of improvement operators:

Proposition 6 *Let $\leq$ be any pre-order on interpretations and $\leq_\Psi$ the pre-order associated to Ψ then there exists a sequence of formulae $\alpha_1, \ldots, \alpha_n$ such that $\leq_{\Psi \circ \alpha_1 \circ \ldots \circ \alpha_n} = \leq$.*

Improvement operators are, as far as we know, the first change operators defined in the DP framework that allows to avoid these limit cases.

Conclusion

We have introduced a new family of change operators called improvement operators. These operators have a more cautious behaviour than usual DP iterated revision operators. The main iterated revision operators of the literature satisfy all the properties of weak improvement operators. In that respect weak improvement operators can be considered as a generalization of iterated revision operators.

An essential point for being able to state logical properties and theorems on improvement operators is the interesting relationship between $\circ$ and its corresponding revision operator $\star$.

(I11), the last postulate required for improvement operators is very strong, in the sense that it determines in a unique way the pre-orders associated to the improvement. Thus, there are room to explore some variants of **(I11)** leading to other interesting weak improvements operators. We keep this as future work.

Acknowledgements

The authors would like to thank the reviewers of the paper for their helpful comments.

The second author was partially supported by a research grant of the Mairie de Paris and by the project CDCHT-ULA N° C-1451-07-05-A. Part of this work was done when the second author was a visiting professor at CRIL (CNRS UMR 8188) from September to December 2007 and a visiting researcher at TSI Department of Telecom ParisTech (CNRS UMR 5141 LTCI) from January to April 2008. The second author thanks to CRIL and TSI Department for the excellent working conditions.

Appendix

Proof of Theorem 1: (*only if*) Let $\circ$ be a weak improvement operator. We define an assignment $\Psi \mapsto \leq_\Psi$ by putting

$$w \leq_\Psi w' \text{ if and only if } w \models B(\Psi \star \varphi_{w,w'})$$

By I4, the relation $\leq_\Psi$ is well defined, *i.e.* it does not depend on the choice of the formula $\varphi_{w,w'}$. We prove now that $\leq_\Psi$ is a total pre-order.

Totality: Let w, w' be any interpretations (eventually $w = w'$). By I1, $[\![\Psi \star \varphi_{w,w'}]\!] \subseteq \{w, w'\}$ and by definition $[\![\Psi \star \varphi_{w,w'}]\!] \neq \emptyset$, so $w \in [\![\Psi \star \varphi_{w,w'}]\!]$ or $w' \in [\![\Psi \star \varphi_{w,w'}]\!]$ (or both), *i.e.* $w \leq_\Psi w'$ or $w' \leq_\Psi w$.

Transitivity: Suppose $w \leq_\Psi w'$ and $w' \leq_\Psi w''$. We want to show that $w \leq_\Psi w''$, that is to say $w \in [\![B(\Psi \star \varphi_{w,w''})]\!]$. Suppose towards a contradiction that it is not the case, so $w \notin [\![B(\Psi \star \varphi_{w,w''})]\!]$. As by definition $[\![B(\Psi \star \varphi_{w,w''})]\!] \neq \emptyset$, this means that $[\![B(\Psi \star \varphi_{w,w''})]\!] = \{w''\}$. Let us consider two cases:

1) If $B(\Psi \star \varphi_{w,w',w''}) \wedge \varphi_{w,w''}$ is not consistent. Then, as by definition $B(\Psi \star \varphi_{w,w',w''}) \nvdash \bot$, we have $[\![B(\Psi \star \varphi_{w,w',w''})]\!] = \{w'\}$. In this case as $B(\Psi \star \varphi_{w,w',w''}) \wedge \varphi_{w,w'}$ is consistent, by I5 and I6 and I4 we get that $[\![B(\Psi \star \varphi_{w,w'})]\!] = [\![B(\Psi \star \varphi_{w,w',w''}) \wedge \varphi_{w,w'}]\!] = \{w'\}$. This means by definition that $w' <_\Psi w$. Contradiction.

2) If $B(\Psi \star \varphi_{w,w',w''}) \wedge \varphi_{w,w''}$ is consistent. Then by I5 and I6 and I4 we get that $[\![B(\Psi \star \varphi_{w,w''})]\!] = [\![B(\Psi \star \varphi_{w,w',w''}) \wedge \varphi_{w,w''}]\!] = \{w''\}$. This means by definition that $w'' <_\Psi w$. Contradiction.

Let us prove now equation (4). First we will prove that $[\![B(\Psi \star \alpha)]\!] \subseteq \min([\![\alpha]\!], \leq_\Psi)$. Take w in $[\![B(\Psi \star \alpha)]\!]$. Thus, $B(\Psi \star \alpha) \wedge \varphi_{w,w'} \nvdash \bot$ for any $w' \in [\![\alpha]\!]$. Then, by I5 and I6 and I4, $B(\Psi \star \alpha) \wedge \varphi_{w,w'} \equiv B(\Psi \star \varphi_{w,w'})$. Therefore $w \in [\![B(\Psi \star \varphi_{w,w'})]\!]$, that is $w \leq_\Psi w'$ for any $w' \in [\![\alpha]\!]$ what exactly means $w \in \min([\![\alpha]\!], \leq_\Psi)$.

Now we will prove the converse inclusion that is $\min([\![\alpha]\!], \leq_\Psi) \subseteq [\![B(\Psi \star \alpha)]\!]$. Suppose that $w \in \min([\![\alpha]\!], \leq_\Psi)$. We want to show that $w \in [\![B(\Psi \star \alpha)]\!]$. Towards a contradiction suppose that $w \notin [\![B(\Psi \star \alpha)]\!]$. Let w' be a model of $B(\Psi \star \alpha)$.Then, by I5 and I6 and I4, $B(\Psi \star \alpha) \wedge \varphi_{w,w'} \equiv B(\Psi \star \varphi_{w,w'})$. By assumption $w \notin [\![B(\Psi \star \alpha)]\!]$ so $[\![B(\Psi \star \varphi_{w,w'})]\!] = \{w'\}$. Therefore $w' <_\Psi w$, contradicting the minimality of w in $[\![\alpha]\!]$ with respect to $\leq_\Psi$.

Now we prove the conditions of the strong faithful assignment. First to show conditions 1 and 2 it is equivalent to show that $[\![B(\Psi)]\!] = \min(\mathcal{W}, \leq_\Psi)$. Suppose that $w \models$

$B(\Psi)$. We want to see that $w \leq_\Psi w'$ for any interpretation w'. In order to do that, let w' be a interpretation. Note that $w \models B(\Psi) \wedge \varphi_{w,w'}$, so $B(\Psi) \wedge \varphi_{w,w'} \not\vdash \bot$. Then, by I2, $B(\Psi \star \varphi_{w,w'}) \equiv B(\Psi) \wedge \varphi_{w,w'}$. Therefore, $w \models B(\Psi \star \varphi_{w,w'})$, *i.e.* $w \leq_\Psi w'$. This proves that $[[B(\Psi)]] \subseteq \min(\mathcal{W}, \leq_\Psi)$. For the converse inclusion take $w \in \min(\mathcal{W}, \leq_\Psi)$. Towards a contradiction, suppose that $w \notin [[B(\Psi)]]$. Let w' be a model of $B(\Psi)$. Then $[[B(\Psi) \wedge \varphi_{w,w'}]] = \{w'\}$. Thus, by I2, $B(\Psi \star \varphi_{w,w'}) \equiv B(\Psi) \wedge \varphi_{w,w'}$ and therefore $[[B(\Psi \star \varphi_{w,w'})]] = \{w'\}$, *i.e.* $w' <_\Psi w$, contradicting the minimality of w with respect to $\leq_\Psi$.

Now for condition 3 suppose $\alpha_i \equiv \beta_i$ for any $i \leq n$ we want to show $\leq_{\Psi\circ\alpha_1\circ\cdots\circ\alpha_n} = \leq_{\Psi\circ\beta_1\circ\cdots\circ\beta_n}$. We proceed by induction on $k = 0, \ldots, n$. For $k = 0$ is trivial because $\leq_\Psi = \leq_\Psi$. For shortening the notation put $\Theta_k = \Psi \circ \alpha_1 \circ \cdots \circ \alpha_k$ and $\Gamma_k = \Psi \circ \beta_1 \circ \cdots \circ \beta_k$. Thus our induction hypothesis is $\leq_{\Theta_k} = \leq_{\Gamma_k}$. We want to show that $\leq_{\Theta_k\circ\alpha_{k+1}} = \leq_{\Gamma_k\circ\beta_{k+1}}$. In order to do that, we prove that each level of $\leq_{\Theta_k\circ\alpha_{k+1}}$ is equal to the corresponding level of $\leq_{\Gamma_k\circ\beta_{k+1}}$. This is done by induction on the number of levels of $\leq_{\Theta_k\circ\alpha_{k+1}}$. We sketch the proof. For the level 0: we want to see that $min(\mathcal{W}, \leq_{\Theta_k\circ\alpha_{k+1}}) = min(\mathcal{W}, \leq_{\Gamma_k\circ\beta_{k+1}})$. By equation (4), we have $min(\mathcal{W}, \leq_{\Theta_k\circ\alpha_{k+1}}) = [[B(\Gamma_k \circ \alpha_{k+1})]]$ and $min(\mathcal{W}, \leq_{\Gamma_k\circ\beta_{k+1}}) = [[B(\Gamma_k \circ \beta_{k+1})]]$. By I4$^+$, $[[B(\Theta_k \circ \alpha_{k+1})]] = [[B(\Gamma_k \circ \beta_{k+1})]]$. Therefore, $min(\mathcal{W}, \leq_{\Theta_k\circ\alpha_{k+1}}) = min(\mathcal{W}, \leq_{\Gamma_k\circ\beta_{k+1}})$. Now suppose that the first i levels of $\leq_{\Theta_k\circ\alpha_{k+1}}$ correspond exactly to the first i levels of $\leq_{\Gamma_k\circ\beta_{k+1}}$. We will prove that the level $i+1$ of $\leq_{\Theta_k\circ\alpha_{k+1}}$ is contained in the level $i+1$ of $\leq_{\Gamma_k\circ\beta_{k+1}}$ (and with a symmetrical argument we will prove the inverse inclusion). Towards a contradiction suppose that w is in the level $i+1$ of $\leq_{\Theta_k\circ\alpha_{k+1}}$ and w is not in the level $i+1$ of $\leq_{\Gamma_k\circ\beta_{k+1}}$. Take w' in the level $i+1$ of $\leq_{\Gamma_k\circ\beta_{k+1}}$. As the first i levels of $\leq_{\Theta_k\circ\alpha_{k+1}}$ and $\leq_{\Gamma_k\circ\beta_{k+1}}$ are equal, w' is in a level j, with $j > i$ for the pre-order $\leq_{\Theta_k\circ\alpha_{k+1}}$. Consider now the formula $\varphi_{w,w'}$. Then it is clear that $w \in min([[\varphi_{w,w'}]], \leq_{\Theta_k\circ\alpha_{k+1}})$ and $w \notin min([[\varphi_{w,w'}]], \leq_{\Gamma_k\circ\beta_{k+1}})$. From this, by equation (4), follows $[[B(\Theta_k \circ \alpha_{k+1} \circ \varphi_{w,w'})]] \neq [[B(\Gamma_k \circ \beta_{k+1} \circ \varphi_{w,w'})]]$, contradicting I4$^+$.

(*if*) Suppose that we have a strong faithful assignment $\Psi \mapsto \leq_\Psi$ such that equation (4) holds. We want to check that I1-I6 hold.

(I1) Follows from equation 4.

(I2) Let us first show that $B(\Psi) \wedge \alpha \vdash B(\Psi \star \alpha)$. If $w \models B(\Psi) \wedge \alpha$ this means that $w \in \min(\mathcal{W}, \leq_\Psi)$. So for any $w' \in \mathcal{W}$, we have $w \leq_\Psi w'$. This is in particular true for all the models of α, so $w \in \min(\alpha, \leq_\Psi)$, that is, by definition, $w \models B(\Psi\star\alpha)$. Let us now show that $B(\Psi\star\alpha) \vdash B(\Psi)\wedge\alpha$. By definition $w \models B(\Psi\star\alpha)$ means $w \in \min([[\alpha]], \leq_\Psi)$. So $w \models \alpha$. Let us show that $w \models B(\Psi)$. Suppose that it is not the case. In this case, and since by hypothesis $B(\Psi)\wedge\alpha \nvdash \bot$ we can choose a $w' \in [[B(\Psi) \wedge \alpha]]$. So, as $w' \in [[B(\Psi)]]$ and $w \notin [[B(\Psi)]]$, we have $w' <_\Psi w$. But as $w', w \models \alpha$, this implies that $w \notin \min([[\alpha]], \leq_\Psi)$. Contradiction.

(I4) Suppose $\alpha_i \equiv \beta_i$ for any $i \leq n$ we want to show $B(\Psi \circ \alpha_1 \circ \cdots \circ \alpha_n) \equiv B(\Psi \circ \beta_1 \circ \cdots \circ \beta_n)$. By equation (4), this is equivalent to prove $min([[\alpha_n]], \leq_{\Psi\circ\alpha_1\circ\cdots\circ\alpha_{n-1}}) = min([[\beta_n]], \leq_{\Psi\circ\beta_1\circ\cdots\circ\beta_{n-1}})$. But this is clear because, by S6, we have $\leq_{\Psi\circ\alpha_1\circ\cdots\circ\alpha_{n-1}} = \leq_{\Psi\circ\beta_1\circ\cdots\circ\beta_{n-1}}$ and by hypothesis $[[\alpha_n]] = [[\beta_n]]$.

(I5 and I6) By equation (4) we have $[[B(\Psi \star (\alpha \wedge \beta))]] = \min([[\alpha \wedge \beta]], \leq_\Psi)$ and $[[B(\Psi \star \alpha) \wedge \beta]] = \min([[\alpha]], \leq_\Psi) \cap [[\beta]]$. Thus, it is enough to see that

$$\min([[\alpha]], \leq_\Psi) \cap [[\beta]] = \min([[\alpha \wedge \beta]], \leq_\Psi)$$

under the hypothesis $\min([[\alpha]], \leq_\Psi) \cap [[\beta]] \neq \emptyset$. It is quite clear that $\min([[\alpha]], \leq_\Psi) \cap [[\beta]] \subseteq \min([[\alpha \wedge \beta]], \leq_\Psi)$. For the other inclusion take $w \in \min([[\alpha \wedge \beta]], \leq_\Psi)$. As w is in $[[\beta]]$ it remains to see that $w \in \min([[\alpha]], \leq_\Psi)$. We know that $w \in [[\alpha]]$. We claim that it is minimal in $[[\alpha]]$ with respect to $\leq_\Psi$. Towards a contradiction, suppose that ir is not the case. As $\leq_\Psi$ is a total pre-order there exists $w' \in \min([[\alpha]], \leq_\Psi)$ such that $w' <_\Psi w$. By hypothesis, there exists $w'' \in \min([[\alpha]], \leq_\Psi) \cap [[\beta]]$. Again as $\leq_\Psi$ is a total pre-order, $w' \sim_\Psi w''$, therefore $w'' <_\Psi w$ contradicting the minimality of w in $[[\alpha \wedge \beta]]$ with respect to $\leq_\Psi$. ∎

Proof of Proposition 1: If $\min([[\alpha]], \leq_\Psi)$ and $\min([[\beta]], \leq_\Psi)$ are in the same level with respect to $\leq_\Psi$ then $\min([[(]]\alpha \vee \beta), \leq_\Psi) = \min([[\alpha]], \leq_\Psi) \cup \min([[\beta]], \leq_\Psi)$. Thus, by Theorem 1, $[[B(\Psi \star (\alpha \vee \beta))]] = [[B(\Psi \star \alpha)]] \cup [[B(\Psi \star \beta)]]$. Otherwise, $\min([[\alpha]], \leq_\Psi)$ is in a lower level than $\min([[\beta]], \leq_\Psi)$ or $\min([[\beta]], \leq_\Psi)$ is in a lower level than $\min([[\alpha]], \leq_\Psi)$. In the first case, $\min([[(]]\alpha \vee \beta), \leq_\Psi) = \min([[\alpha]], \leq_\Psi)$. Thus, by the Theorem 1, $[[B(\Psi \star (\alpha \vee \beta))]] = [[B(\Psi \star \alpha)]]$. In the second case, $\min([[(]]\alpha \vee \beta), \leq_\Psi) = \min([[\beta]], \leq_\Psi)$. Thus, by the Theorem 1, $[[B(\Psi \star (\alpha \vee \beta))]] = [[B(\Psi \star \beta)]]$. ∎

Proof of Corollary 2: (*only if*) Assume that $\alpha \prec_\Psi \beta$, that is $B(\Psi\star\alpha) \vdash B(\Psi\star(\alpha\vee\beta))$ and $B(\Psi\star\beta) \not\vdash B(\Psi\star(\alpha\vee\beta))$. By Proposition 1 and its proof, necessarily $\min([[\alpha]], \leq_\Psi)$ is in a lower level than $\min([[\beta]], \leq_\Psi)$. Thus, by Theorem 1, it is enough to take $w \in \min([[\alpha]], \leq_\Psi)$ and $w' \in \min([[\beta]], \leq_\Psi)$ to get $w \in [[B(\Psi \star \alpha)]]$, $w' \in [[B(\Psi \star \beta)]]$, $w <_\Psi w'$.
(*if*) Assume that there exist w, w' such that $w \in [[B(\Psi \star \alpha)]]$, $w' \in [[B(\Psi \star \beta)]]$, $w <_\Psi w'$. Then, by Theorem 1, $\min([[\alpha]], \leq_\Psi)$ is in a lower level than $\min([[\beta]], \leq_\Psi)$. Then, by Proposition 1 and its proof, $B(\Psi \star (\alpha \vee \beta)) \equiv B(\Psi \star \alpha)$. On the other hand $B(\Psi \star \beta) \not\vdash B(\Psi \star (\alpha \vee \beta))$ because $\min([[\alpha]], \leq_\Psi)$ and $\min([[\beta]], \leq_\Psi)$ are not in the same level. Therefore $\alpha \prec_\Psi \beta$

∎

Proof of Corollary 3: (*only if*) Assume $\alpha \prec\!\!\prec_\Psi \beta$. By Corollary 2, we get w, w' such that $w \in [[B(\Psi \star \alpha)]]$, $w' \in [[B(\Psi \star \beta)]]$, $w <_\Psi w'$. Towards a contradiction, suppose that there exists w'' such that $w <_\Psi w'' <_\Psi w'$. But it is clear, using Corollary 2, that $\alpha \prec_\Psi \varphi_{w''} \prec_\Psi \beta$ contradicting the fact $\alpha \prec\!\!\prec_\Psi \beta$.
(*if*) Assume there exist w, w' such that $w \in [[B(\Psi \star \alpha)]]$,

$w' \in [\![B(\Psi \star \beta)]\!]$, $w <_\Psi w'$ and there is no w'' such that $w <_\Psi w'' <_\Psi w'$. By Corollary 2, $\alpha \prec_\Psi \beta$. Thus, the only possibility for $\alpha \not\ll_\Psi \beta$, is the existence of γ such that $\alpha \prec_\Psi \gamma \prec_\Psi \beta$. Again, by Corollary 2, taking $w'' \in [\![B(\Psi \star \gamma)]\!]$ we have $w <_\Psi w'' <_\Psi w'$, a contradiction. ∎

Proof of Theorem 2: (*only if*) By Theorem 1 we know that there exists an epistemic assignment $\Psi \mapsto \leq_\Psi$ such that the equation (4) holds. Thus, it remains to prove that the assignment is indeed a gradual assignment, *i.e.* it satisfies S1, S2, S3, S4 and S5.

(S1) Suppose $w, w' \in [\![\alpha]\!]$. Thus, $\varphi_{w,w'} \vdash \alpha$. By I7, $B((\Psi \circ \alpha) \star \varphi_{w,w'}) \equiv B(\Psi \star \varphi_{w,w'})$. Then by equation (4) we have

$$
\begin{aligned}
w \leq_{\Psi\circ\alpha} w' &\Leftrightarrow w \in \min(\{w, w'\}, \leq_{\Psi\circ\alpha}) \\
&\Leftrightarrow w \in [\![B((\Psi \circ \alpha) \star \varphi_{w,w'})]\!] \\
&\Leftrightarrow w \in [\![B(\Psi \star \varphi_{w,w'})]\!] \\
&\Leftrightarrow w \in \min(\{w, w'\}, \leq_\Psi) \\
&\Leftrightarrow w \leq_\Psi w'
\end{aligned}
$$

(S2) The proof is analogous to the one of S1 but using I8 instead of I7.

(S3) Suppose that $w \in [\![\alpha]\!]$, $w' \in [\![\neg\alpha]\!]$ and $w \leq_\Psi w'$. We want to show that $w <_{\Psi\circ\alpha} w'$. As $w \leq_\Psi w'$, necessarily $w \in \min(\{w, w'\}, \leq_\Psi)$ what, by Equation (4), means $w \in [\![B(\Psi \star \varphi_{w,w'})]\!]$. Then $B(\Psi \star \varphi_{w,w'}) \nvdash \neg\alpha$, so by I9, $B((\Psi \circ \alpha) \star \varphi_{w,w'}) \vdash \alpha$. Then, by I1 and I3, $[\![B((\Psi \circ \alpha) \star \varphi_{w,w'})]\!] = \{w\}$. From this, using Equation (4), we get $w <_{\Psi\circ\alpha} w'$.

(S4) Suppose that $w \in [\![\alpha]\!]$, $w' \in [\![\neg\alpha]\!]$ and $w' <_\Psi w$. We want to show that $w' \leq_{\Psi\circ\alpha} w$. From the hypothesis $w' <_\Psi w$ we get $\min(\{w, w'\}, \leq_\Psi) = \{w'\}$. Then, by Equation (4), $[\![B(\Psi \star \varphi_{w,w'})]\!] = \{w'\}$. Therefore $B(\Psi \star \varphi_{w,w'}) \vdash \neg\alpha$. Thus, by I10, $B((\Psi \circ \alpha) \star \varphi_{w,w'}) \nvdash \alpha$. Then $w' \in [\![B((\Psi \circ \alpha) \star \varphi_{w,w'})]\!]$, and by Equation (4), $w' \in \min(\{w, w'\}, \leq_{\Psi\circ\alpha})$, *i.e.* $w' \leq_{\Psi\circ\alpha} w$.

(S5) Suppose that $w \in [\![\alpha]\!]$, $w' \in [\![\neg\alpha]\!]$, $w' <_\Psi w$ and that there is no w'' such that $w' <_\Psi w'' <_\Psi w$. We want to show that $w \leq_{\Psi\circ\alpha} w'$. From $w' <_\Psi w$ we have $\min(\{w, w'\}, \leq_\Psi) = \{w'\}$, so from Theorem 1 we have $[\![B(\Psi \star \varphi_{w,w'})]\!] = \{w'\}$. So $B(\Psi \star \varphi_{w,w'}) \vdash \neg\alpha$. On the other hand, the assumptions with the Corollary 3 gives us $\varphi_{w,w'} \ll_\Psi \varphi_{w,w'} \wedge \alpha$. Then, by I11, $B((\Psi \circ \alpha) \star \varphi_{w,w'}) \nvdash \neg\alpha$, that means by Theorem 1, $w \leq_{\Psi\circ\alpha} w'$.

(*if*) By Theorem 1 we know that $\circ$ is a weak improvement operator. Thus, it remains to check that I7-I11 hold.

(I7) Suppose that $\alpha \vdash \mu$, *i.e.* $[\![\alpha]\!] \subseteq [\![\mu]\!]$. We want to show $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$. By Equation (4) this is equivalent to prove that $\min([\![\alpha]\!], \leq_{\Psi\circ\mu}) = \min([\![\alpha]\!], \leq_\Psi)$. That is a straightforward consequence of S1 that gives (since $[\![\alpha]\!] \subseteq [\![\mu]\!]$) $\forall w, w' \models \alpha, w \leq_\Psi w'$ iff $w \leq_{\Psi\circ\mu} w'$.

(I8) Suppose that $\alpha \vdash \neg\mu$, *i.e.* $[\![\alpha]\!] \subseteq [\![\neg\mu]\!]$. We want to show $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$. By Equation (4) this is equivalent to prove that $\min([\![\alpha]\!], \leq_{\Psi\circ\mu}) = \min([\![\alpha]\!], \leq_\Psi)$. Like for (I7), this is a straightforward consequence of S2 that gives (since $[\![\alpha]\!] \subseteq [\![\neg\mu]\!]$) $\forall w, w' \models \alpha, w \leq_\Psi w'$ iff $w \leq_{\Psi\circ\mu} w'$.

(I9) Let us remark from the fact that $\leq_\Psi$ and $\leq_{\Psi\circ\mu}$ are total pre-orders, that postulate S3 is equivalent to the following one:

(S3') If $w \in [\![\mu]\!], w' \in [\![\neg\mu]\!]$ then $w' \leq_{\Psi\circ\mu} w \Rightarrow w' <_\Psi w$

Now suppose that $B(\Psi \star \alpha) \nvdash \neg\mu$. We want to show $B((\Psi \circ \mu) \star \alpha) \vdash \mu$. Towards a contradiction, suppose that $B((\Psi \circ \mu) \star \alpha) \nvdash \mu$, *i.e.* there exists $w \in [\![B((\Psi \circ \mu) \star \alpha)]\!]$ such that $w \notin [\![\mu]\!]$. By Equation (4) $w \in \min([\![\alpha]\!], \leq_{\Psi\circ\mu})$, so for any $w' \in [\![\alpha]\!]$, $w \leq_{\Psi\circ\mu} w'$. By the assumption, there exists $w'' \in [\![B(\Psi \star \alpha)]\!] \cap [\![\mu]\!]$. In particular, by I1, $w'' \in [\![\alpha]\!]$. Thus, $w \leq_{\Psi\circ\mu} w''$. On the other hand, by Equation (4) $w'' \in \min([\![\alpha]\!], \leq_\Psi)$. As $w \in [\![\neg\mu]\!]$ and $w'' \in [\![\mu]\!]$, and $w \leq_{\Psi\circ\mu} w''$, by S3', $w <_\Psi w''$. But, since $w \in [\![\alpha]\!]$, this contradicts the minimality of w'' in $[\![\alpha]\!]$ with respect to $\leq_\Psi$.

(I10) Let us remark from the fact that $\leq_\Psi$ and $\leq_{\Psi\circ\mu}$ are total pre-orders, the postulate S4 is equivalent to the following one:

(S4') If $w \in [\![\mu]\!], w' \in [\![\neg\mu]\!]$ then $w <_{\Psi\circ\mu} w' \Rightarrow w \leq_\Psi w'$

Suppose that $B(\Psi \star \alpha) \vdash \neg\mu$. We want to show $B((\Psi \circ \mu) \star \alpha) \nvdash \mu$. Towards a contradiction suppose that $B((\Psi \circ \mu) \star \alpha) \vdash \mu$. Let w, w' be such that $w \models B((\Psi \circ \mu) \star \alpha)$ and $w' \models B(\Psi \star \alpha)$. By the assumptions $w \in [\![\mu]\!]$ and $w' \in [\![\neg\mu]\!]$. By Equation (4), $w \in \min([\![\alpha]\!], \leq_{\Psi\circ\mu})$ and $w' \in \min([\![\alpha]\!], \leq_\Psi)$. By the assumptions, $w \not\simeq_{\Psi\circ\mu} w'$ and $w \not\simeq_\Psi w'$, because if not $w' \in \min([\![\alpha]\!], \leq_{\Psi\circ\mu})$ or $w \in \min([\![\alpha]\!], \leq_\Psi)$. But this is impossible because in the first case $w' \in [\![\mu]\!]$, a contradiction and in he second case $w \in [\![\neg\mu]\!]$, a contradiction. Thus, necessarily $w <_{\Psi\circ\mu} w'$ and $w' <_\Psi w$. As we have $w \in [\![\mu]\!]$, $w' \in [\![\neg\mu]\!]$ and $w <_{\Psi\circ\mu} w'$, by S4', $w \leq_\Psi w'$, a contradiction.

(I11) Assume $B(\Psi \star \alpha) \vdash \neg\mu, \alpha \wedge \mu \nvdash \perp$ and $\alpha \ll_\Psi \alpha \wedge \mu$. We want to show that $B((\Psi \circ \mu) \star \alpha) \nvdash \neg\mu$. Towards a contradiction, suppose that $B((\Psi \circ \mu) \star \alpha) \vdash \neg\mu$. Let w, w' such that $w' \in [\![B(\Psi \star \alpha)]\!]$ and $w \in [\![B(\Psi \star (\alpha \wedge \mu))]\!]$. By the assumptions we have $w' \in [\![\neg\mu]\!]$ and $w \in [\![\mu]\!]$. By Corollary 3, $w' <_\Psi w$ and there is no w'' such that $w' <_\Psi w'' <_\Psi w$. By S5, $w \leq_{\Psi\circ\mu} w'$. By S4, $w' \leq_{\Psi\circ\mu} w$. Therefore, $w \simeq_{\Psi\circ\mu} w'$. That means that $[\![B(\Psi \star \alpha)]\!]$ and $[\![B(\Psi \star \alpha \wedge \mu)]\!]$ are in the same level with respect to $\leq_{\Psi\circ\mu}$. We claim that this level is the level of $\min([\![\alpha]\!], \leq_{\Psi\circ\mu})$. But this is a contradiction because we have $w \in \min(\min([\![\alpha]\!], \leq_{\Psi\circ\mu}))$ and therefore $w \models \neg\mu$ which contradicts the fact that $w \models \mu$. Now we turn to the proof of our claim. Towards a contradiction, suppose the claim is not true. Then, necessarily there is $w'' \in \min([\![\alpha]\!], \leq_{\Psi\circ\mu})$ such that $w'' <_{\Psi\circ\mu} w$. We consider two cases: $w'' \in [\![\mu]\!]$ and $w'' \in [\![\neg\mu]\!]$. In the case $w'' \in [\![\mu]\!]$, we don't have $w'' <_\Psi w$ because $w \in \min([\![\alpha \wedge \mu]\!], \leq_\Psi)$. Therefore $w \leq_\Psi w''$. Then, by S1, $w \leq_{\Psi\circ\mu} w''$, a contradiction. In the case $w'' \in [\![\neg\mu]\!]$, we don't have $w'' <_\Psi w'$ because $w' \in \min([\![\alpha]\!], \leq_\Psi)$. Therefore $w' \leq_\Psi w''$. Then, by S2, $w' \leq_{\Psi\circ\mu} w''$, that is $w \leq_{\Psi\circ\mu} w''$, a contradiction. ∎

Proof of Lemma 1: Define $A = \{w'' \in \mathcal{W} : w <_\Psi w'' <_\Psi w'\}$. By the assumptions $w <_\Psi w'$ and $w \not\ll_\Psi w'$, the set A is nonempty. Thus $A \cap [[\neg\alpha]] \neq \emptyset$ or $A \cap [[\alpha]] \neq \emptyset$. We consider first the case $A \cap [[\neg\alpha]] \neq \emptyset$. Take $w'' \in max(A \cap [[\neg\alpha]], \leq_\Psi)$. By definition of A, $w <_\Psi w''$ and $w'' <_\Psi w'$. We consider two subcases:

- $w'' \ll_\Psi w'$. In this situation, we conclude by S4 and S5 $w'' \sim_{\Psi\circ\alpha} w'$. By S2, $w <_{\Psi\circ\alpha} w''$. Therefore by transitivity $w <_{\Psi\circ\alpha} w'$.
- $w'' \not\ll_\Psi w'$. In this situation we take w''' such that $w'' \ll_\Psi w'''$. Its clear that $w''' <_\Psi w'$ and by definition of w'', $w''' \in [[\alpha]]$. By S4 and S5, $w'' \simeq_{\Psi\circ\alpha} w'''$. Thus $w <_{\Psi\circ\alpha} w'''$. By S1, $w''' <_{\Psi\circ\alpha} w'$. Then, by transitivity, $w <_{\Psi\circ\alpha} w'$.

For the second case, $A \cap [[\alpha]] \neq \emptyset$, we proceed with an analogous reasoning, but this time taking $w'' \in min(A \cap [[\alpha]], \leq_\Psi)$. ∎

Proof of Proposition 2: Towards a contradiction, suppose that we have $\leq^1_{\Psi\circ\alpha} \neq \leq^2_{\Psi\circ\alpha}$ and both pre-orders obey to (S1-S5). Let w, w' be witness of this inequality. Thus, without lost of generality, we can suppose $w <^1_{\Psi\circ\alpha} w'$ and $w' \leq^2_{\Psi\circ\alpha} w$. By S1, it is not the case $w, w' \in [[\alpha]]$, since otherwise by $w <^1_{\Psi\circ\alpha} w'$ we obtain $w <_\Psi w'$ and by $w' \leq^2_{\Psi\circ\alpha} w$ we obtain $w' \leq_\Psi w$ and a contradiction. Similarly by S2, it is not the case $w, w' \in [[\neg\alpha]]$. Thus, the only possibilities are $w \in [[\alpha]]$ and $w' \in [[\neg\alpha]]$ or $w \in [[\neg\alpha]]$ and $w' \in [[\alpha]]$.
We consider the first case, *i.e.* $w \in [[\alpha]]$ and $w' \in [[\neg\alpha]]$. As $w' \leq^2_{\Psi\circ\alpha} w$, by S3, $w \not\leq_\Psi w'$, *i.e.* $w' <_\Psi w$. If $w' \not\ll_\Psi w$ then, by the Lemma 1 $w' <^1_{\Psi\circ\alpha} w$, a contradiction. If $w' \ll_\Psi w$, by S4 and S5, $w' \simeq^1_{\Psi\circ\alpha} w$, again a contradiction. Now, we consider the second case, *i.e.* $w \in [[\neg\alpha]]$ and $w' \in [[\alpha]]$. As $w <^1_{\Psi\circ\alpha} w'$, by S3, $w' \not\leq_\Psi w$, *i.e.* $w <_\Psi w'$. Suppose $w \ll_\Psi w'$. Then, by S5, $w' \leq^1_{\Psi\circ\alpha} w$ a contradiction. So $w \not\ll_\Psi w'$, and by Lemma 1, $w <^2_{\Psi\circ\alpha} w'$, a contradiction. ∎

References

Alchourrón, C. E.; Gärdenfors, P.; and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50:510–530.

Booth, R., and Meyer, T. 2006. Admissible and restrained revision. *Journal of Artificial Intelligence Research* 26:127–151.

Booth, R., and Meyer, T. 2007. On the dynamics of total preorders: Revising abstract interval orders. In Mellouli, K., ed., *ECSQARU*, volume 4724 of *Lecture Notes in Computer Science*, 42–53. Springer.

Booth, R.; Meyer, T.; and Wong, K.-S. 2006. A bad day surfing is better than a good day working: How to revise a total preorder. In Doherty, P.; Mylopoulos, J.; and Welty, C. A., eds., *KR*, 230–238. AAAI Press.

Boutilier, C. 1996. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic* 25(3):262–305.

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

Gärdenfors, P. 1988. *Knowledge in flux*. MIT Press.

Hansson, S. O. 1999. *A Textbook of Belief Dynamics. Theory Change and Database Updating*. Kluwer.

Herzig, A.; Konieczny, S.; and Perussel, L. 2003. On iterated revision in the agm framework. In *Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03)*, 477–488.

Jin, Y., and Thielscher, M. 2007. Iterated belief revision, revised. *Artificial Intelligence* 171:1–18.

Katsuno, H., and Mendelzon, A. O. 1991. Propositional knowledge base revision and minimal change. *Artificial Intelligence* 52:263–294.

Konieczny, S., and Pino Pérez, R. 2000. A framework for iterated revision. *Journal of Applied Non-Classical Logics* 10(3-4):339–367.

Lehmann, D. 1995. Belief revision, revised. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, 1534–1540.

Nayak, A. C. 1994. Iterated belief change based on epistemic entrenchment. *Erkenntnis* 41:353–390.

Spohn, W. 1988. Ordinal conditional functions: A dynamic theory of epistemic states. In Harper, W. L., and skyrms, B., eds., *Causation in Decision: Belief Change and Statistics*. Kluwer. 105–134.

Williams, M. A. 1994. Transmutations of knowledge systems. In *Proceedings of the Fourth International Conference on the Principles of Knowledge Representation and Reasoning (KR'94)*, 619–629.

ON ITERATED REVISION IN THE AGM FRAMEWORK

Andreas Herzig, Sébastien Konieczny, Laurent Perussel.
Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03).
pages 477-488.
2003.

On iterated revision in the AGM framework

Andreas Herzig, Sébastien Konieczny, Laurent Perrussel

Institut de Recherche en Informatique de Toulouse
118 route de Narbonne - 31062 Toulouse - France
{herzig,konieczny,perrussel}@irit.fr

Abstract. While AGM belief revision identifies belief states with sets of formulas, proposals for iterated revision are usually based on more complex belief states. In this paper we investigate within the AGM framework several postulates embodying some aspects of iterated revision. Our main results are negative: when added to the AGM postulates, our postulates force revision to be maxichoice (whenever the new piece of information is inconsistent with the current beliefs the resulting belief set is maximal). We also compare our results to revision operators with memory and we investigate some postulates proposed in this framework.

1 Introduction

While AGM belief revision identifies belief states with sets of formulas, proposals for iterated revision are usually based on more complex belief states. Following the work of [7], they are usually represented by total pre-orders on interpretations. In fact in [6], Darwiche and Pearl first stated their postulates (C1-C4) in the classical AGM framework. But it has been shown in [8, 15] that (C2) is inconsistent with AGM, and that under the AGM postulates (C1) implies (C3) and (C4). To remove these contradictions, Darwiche and Pearl rephrased their and the AGM postulates in terms of epistemic states [7]. This has lead to a widely accepted framework for iterated revision, and most of the work on iterated belief revision now uses this more complex framework.

So an interesting question investigated in this paper is which requirements on iteration one can consistently add to the usual AGM framework. We focus on the status of old information, and formulate several postulates embodying that aspect of iterated revision. They all express that old information about A determine in some way the current status of A.

In particular, the first postulate says that if the agent was informed about A before revision (in the sense that either A or $\neg A$ was accepted) then the agent should remain informed about A after revision.

Our second postulate is motivated by the following basic algorithm for the revision of a belief set B by a new piece of information A [11, 19]: first put A in the new belief set, then add as many old beliefs from B as possible. So the second postulate expresses that the corresponding operator is idempotent with respect to B. We also study a family of postulates that generalizes this idea.

We also review other postulates coming from the iterated revision literature, in the classical belief set framework.

Our results are mainly negative: when added to the AGM postulates, our postulates lead to extreme revision operators. In particular the first two postulates force revision to be maxichoice: whenever the new piece of information is inconsistent with the current beliefs then the resulting belief set is maximal.

These "impossibility results" about iterated revision in the usual AGM framework can be seen as a justification for the increase in representational complexity that shows up when one goes from AGM to iterated belief revision frameworks (see e.g. [7, 15, 18, 13, 17, 4]). Instead of "flat" belief sets (alias sets of interpretations), the latter work with epistemic states, that can be represented by pre-orders on interpretations.

The paper is organized as follows. In section 2 we give some definitions and notations. In section 3 we consider the Darwiche and Pearl postulates in the AGM framework. More specifically, we focus on their first postulate. In section 4 we investigate the implications of trying to retain old information as much as possible. In section 5 we explore a family of postulates, saying that re-introducing old pieces of information is harmless. In section 6 we compare our results to revision operators with memory [13, 14] and we investigate the implications of some postulates coming from this work. We conclude in section 7.

2 Preliminaries

We work with a propositional language built from a set of atomic variables, denoted by $p, q, \ldots$ Formulas are denoted by $A, B, C, \ldots$ We identify finite sets of formulas (that we call belief sets) with the conjunction of their elements. A belief set B is *informed about* a formula C if $B \vdash C$ or $B \vdash \neg C$. A belief set B is *maximal* (or complete) if B is informed about every C.

The set of all interpretations is denoted $\mathcal{W}$, and the set of all belief sets is denoted $\mathcal{B}$. For a formula B, $Mod(B)$ denotes the set of models of B, i.e. $Mod(B) = \{\omega \in \mathcal{W} : \ \omega \models B\}$. For a set of interpretations $M \subseteq \mathcal{W}$, $Form(W)$ denotes the formula (up to logical equivalence) whose set of models is M, i.e. $Form(W) = \{B : \omega \models B \ \text{ iff } \omega \in M\}$.

A pre-order $\leq$ is a reflexive and transitive relation. $<$ is its strict counterpart: $\omega < \omega'$ if and only if $\omega \leq \omega'$ and $\omega' \not\leq \omega$. And $\simeq$ is defined by $\omega \simeq \omega'$ iff $\omega \leq \omega'$ and $\omega' \leq \omega$. A pre-order is total is for all ω, ω' we have $\omega \leq \omega'$ or $\omega' \leq \omega$. $\min(M, \leq)$ denotes the set $\{\omega \in M | \nexists \omega' \in M : \omega' < \omega\}$.

Definition 1 (AGM belief revision). *An AGM belief revision operator $\star$ is a function that maps a belief set B and a formula A to a belief set $B \star A$ such that :*

(R1) $B \star A \vdash A$
(R2) *If* $B \wedge A \nvdash \bot$, *then* $B \star A \equiv B \wedge A$
(R3) *If* $A \nvdash \bot$, *then* $B \star A \nvdash \bot$
(R4) *If* $B_1 \equiv B_2$ *and* $A_1 \equiv A_2$, *then* $B_1 \star A_1 \equiv B_2 \star A_2$

(R5) $(B \star A) \wedge C \vdash B \star (A \wedge C)$
(R6) *If* $(B \star A) \wedge C \nvdash \bot$, *then* $B \star (A \wedge C) \vdash (B \star A) \wedge C$

The postulates (R1-R4) are often called the *basic* AGM postulates, and the set (R1-R6) the *extended* AGM postulates, indicating that people consider the former to be more fundamental. Notice however that they do not put very hard constraints on $\star$. It is the two last ones (R5) and (R6) that allow to state the below representation theorem, which says that a revision operator corresponds to a family of pre-orders on interpretations. (The theorem is due to Katsuno and Mendelzon, but the idea can be directly traced back to Grove [10].) But first we need the following:

Definition 2 (Faithful assignment). *A function that maps each belief set B to a pre-order $\leq_B$ on interpretations is called a* faithful assignment *if and only if the following holds:*

1. *If $\omega \models B$ and $\omega' \models B$, then $\omega \simeq_B \omega'$*
2. *If $\omega \models B$ and $\omega' \not\models B$, then $\omega <_B \omega'$*
3. *If $B_1 = B_2$, then $\leq_{B_1} = \leq_{B_2}$*

Theorem 1. *A revision operator $\star$ satisfies postulates (R1-R6) if and only if there exists a faithful assignment that maps each belief set B to a total pre-order $\leq_B$ such that:*

$$Mod(B \star A) = \min(Mod(A), \leq_B)$$

We say that the assignment is the faithful assignment *corresponding to* the revision operator.

Let us now introduce a special family of revision operators, called maxichoice revision operators [1, 9].

Definition 3 (maxichoice revision). *A belief revision operator $\star$ is a* maxichoice *revision operator if for every B and A, if $B \vdash \neg A$ then $B \star A$ is maximal.*

Maxichoice revision operators are not very satisfactory, since they are too precise and have a too drastic behaviour. In fact, with those operators, learning any piece of information that conflicts with the current beliefs, however incomplete they are, causes the agent to have beliefs on any formula: for any formula A, either the agent believes that A holds or he believes that $\neg A$ holds. They are considered as an upper-bound for revision operators (the lower-bound being full-meet revision operators [1, 9]).

We will use a characterization of maxichoice operators on the semantical level. First we define:

Definition 4. *A* linear faithful assignment *is a faithful assignment that satisfies*

4. *If $\omega \not\models B$ and $\omega' \not\models B$, then $\omega <_B \omega'$ or $\omega' <_B \omega$*

The following result is is part of the folklore in the literature on revision:

Theorem 2. *A revision operator $\star$ is a maxichoice operator if and only if its corresponding assignment is a linear faithful assignment.*

The proof is straightforward.

3 Darwiche and Pearl postulates in the AGM framework

In [6], Darwiche and Pearl first stated their well-known postulates (C1-C4) in the classical AGM framework.

(C1) If $A \vdash C$, then $(B \star C) \star A \equiv B \star A$
(C2) If $A \vdash \neg C$, then $(B \star C) \star A \equiv B \star A$
(C3) If $B \star A \vdash C$, then $(B \star C) \star A \vdash C$
(C4) If $B \star A \nvdash \neg C$, then $(B \star C) \star A \nvdash \neg C$

But it has been shown in [8, 15] that (C2) is inconsistent with AGM, and that under the AGM postulates (C1) implies (C3) and (C4). To remove these contradictions, Darwiche and Pearl rephrased their and the AGM postulates in terms of epistemic states [7].

As (C1) is consistent with the AGM postulates, one might wonder what the constraints imposed by this postulate on the revision operators are like. This question has not been investigated as far as we know. The consistency of (C1) with AGM is easily established by noticing that the full meet revision operator satisfies (C1) [15]. But is this the only AGM operator satisfying (C1), or do we face a wider family?

Let us define another particular family of revision operators.

Definition 5. *Let $\leq$ be a total pre-order on interpretations. A revision operator $\star$ is said to be* imposed *by $\leq$ if its corresponding faithful assignment satisfies the following property:*

i. If $\omega \not\models B$ and $\omega' \not\models B$, then ($\omega \leq_B \omega'$ iff $\omega \leq \omega'$).

As far as we know, this family of operators has not been studied yet. Such operators are not satisfactory since the result of a revision does not depend of the belief set, but merely of the new piece of information (see theorem 3). This seems to be counter-intuitive and to go against the basic ideas behind revision. Nevertheless, such operators fulfill all AGM postulates, and the full meet revision operator is a particular case (when $\leq$ is a flat pre-order, i.e. $\omega \simeq \omega', \forall \omega, \omega' \in \mathcal{W}$).

Theorem 3. *Let $\star$ be an AGM revision operator, and let f be any function mapping formulas to formulas such that $f(A) \vdash A$ and if $A_1 \equiv A_2$ then $f(A_1) \equiv f(A_2)$. $\star$ is imposed if and only if for any belief set B and formula A, the following holds:*

(IMP) *If $B \vdash \neg A$ then $B \star A \equiv f(A)$.*

Proof. The only if part is straightforward: define $f(A)$ as $\min(Mod(A), \leq)$.

For the if part we need to build the imposed pre-order $\leq$ from $f(A)$. This can be established by noting that if we take a formula A that has exactly two (distinct) models ω and ω', then by (IMP) for every B such that $A \wedge B \vdash \bot$, we have $B \star A \equiv f(A)$. By (R1) and (R3), $Mod(f(A)) = \{\omega\}$ or $Mod(f(A)) =$

$\{\omega'\}$ or $Mod(f(A)) = \{\omega, \omega'\}$. Since $\star$ is an AGM operator, the faithful assignment gives us, for *every* B inconsistent with A, that $\omega <_B \omega'$ whenever $Mod(f(A)) = \{\omega\}$, $\omega' <_B \omega$ whenever $Mod(f(A)) = \{\omega'\}$, and $\omega \simeq_B \omega'$ whenever $Mod(f(A)) = \{\omega, \omega'\}$. That means that there exists a pre-order $\leq$ defined as $\omega \leq \omega'$ iff $\omega \in Mod(f(Form(\omega, \omega')))$ and such that for all B such that $\omega, \omega' \not\models B$, $\omega \leq_B \omega'$ iff $\omega \leq \omega'$.

This result states that for any revision that is not an expansion the old belief set is not taken into account in the result of the revision.

Now let us return to the case of the (C1) postulate and state the following result:

Theorem 4. *An AGM revision operator satisfies (C1) if and only if it is imposed.*

Proof. The if part is straighforward, since either $B \wedge A$ is consistent and then (C1) is a consequence of (R2), or $B \wedge A$ is not consistent, and then (C1) is a consequence of theorem 3.

For the only if part, suppose that the operator $\star$ satisfies (R1-R6) and (C1). We will show that the operator is imposed and there exists an f such that (IMP) is satisfied. If $\star$ satisfies (C1) then (IMP) holds, since for every A and B such that $A \wedge B$ is not consistent, by (R2) we have that $B \star A \equiv (\neg A \star (A \vee B)) \star A$. Thus by (C1) we get that $(\neg A \star (A \vee B)) \star A = \neg A \star A$, consequently we get $B \star A \equiv \neg A \star A$. Thus f can be defined by stipulating that $f(A) = \neg A \star A$. This means that the result of the revision depends only on the input A.

This result casts serious doubts on the (C1) postulate in the AGM framework.

4 "Keep on being informed about A"

When an agent receives new information she has to modify her current set of beliefs B in order to take it into account. One major requirement of AGM theory is the principle of *minimal change*, that means that when one revises a belief set by a new piece of information, one has to keep *"as much as possible"* of the old belief set.

The following property tries to capture this intuition, by saying that revising by A can not induce a loss of information: if B is informed about C, then learning A can not lead to loose this information.

(Compl) If $B \vdash C$ then $B \star A \vdash C$ or $B \star A \vdash \neg C$

Unfortunately it can be proved that :

Theorem 5. *If $\star$ satisfies (R1-R6) and (Compl), then $\star$ is a maxichoice revision operator.*

Proof. This can be proved straightforwardly: suppose $B \vdash A$. If $\vdash A$ then the theorem holds. Else we have $B \vdash A \vee C$ and $B \vdash A \vee \neg C$. By (Compl), $B \star \neg A \vdash A \vee C$ or $B \star \neg A \vdash \neg A \wedge \neg C$, and $B \star \neg A \vdash A \vee \neg C$ or $B \star \neg A \vdash \neg A \wedge C$. Among the four cases, the one where $B \star \neg A \vdash (A \vee C) \wedge (A \vee \neg C)$ is impossible because $B \star \neg A \vdash A$ by (R4) and $\nvdash A$. The one where $B \star \neg A \vdash (\neg A \wedge \neg C) \wedge (\neg A \wedge C)$ is impossible because $B \star \neg A \vdash \bot$. It follows that $B \star \neg A \vdash \neg C$ or $B \star \neg A \vdash C$.

It is straightforward to show that every maxichoice revision operator satisfies (Compl). Together with the preceding theorem it follows that (Compl) characterizes maxichoice revision.

Remark 1. Formula (3.17) in [9] is just (COMPL) (modulo a typo). There, proposition (3.19) says that "$B \star A$ is maximal for any sentence A such that $\neg A \in B$", i.e. (3.17) entails maxichoice revision. The proof refers to observation 3.2 of [2], but the latter presupposes already that $\star$ is a maxichoice operator, and establishes that this entails maximality.

So this postulate puts too strong a requirement on classical AGM revision operators.

In the next section we will investigate another requirement also based on the assumption that we can keep as much as possible of the old information.

5 "Re-introducing old information doesn't harm"

Another way of ensuring that one does not forget previous information is to suppose that we can re-introduce the old belief set without changing the current one. It can be seen as some kind of left-idempotency of the revision operator. This idea is very close to the one used for defining revision with memory operators [14, 13, 3].

First we need the following abbreviations.

Definition 6. *Given a set of beliefs B and pieces of information A_i, then for $1 \leq i \leq n$ we define B_i by:*

$$B_i = (...((B \star A_1) \star A_2) \star \ldots) \star A_i$$

Thus $B_0 = B$, $B_1 = B \star A_1$, and $B_2 = (B \star A_1) \star A_2$.

Our abbreviation enables us to concisely formulate the following family of postulates:

(Mem$_i$) $B_i \equiv B \star B_i$, for $i \geq 0$

Hence:

(Mem$_0$) says $B_0 \equiv B \star B_0$, i.e. $B \equiv B \star B$,
(Mem$_1$) says $B_1 \equiv B \star B_1$, i.e. $B \star A_1 \equiv B \star (B \star A_1)$, and
(Mem$_2$) says $B_2 \equiv B \star B_2$, i.e. $(B \star A_1) \star A_2 \equiv B \star ((B \star A_1) \star A_2)$.

$\vdots$

Let us see now what is the relation of the postulates (Mem_i) with the AGM postulates.

Theorem 6. *(Mem_0) is derivable from the basic AGM postulates.*

The proof only uses the postulate (R2).

Theorem 7. *(Mem_1) is derivable from the extended AGM postulates.*

Proof. From (R1) we know that $(B \star A) \wedge A \equiv B \star A$. Now using (R5) and (R6) with $C = B \star A$, we have $B \star (A \wedge (B \star A)) \equiv (B \star A) \wedge (B \star A)$. That is directly $B \star (B \star A) \equiv B \star A$.

Theorem 8. *(Mem_2), (Mem_3), etc. cannot be derived from the AGM postulates.*

Proof. This can be established e.g. by considering Dalal's revision operator [5], which is known to satisfy the AGM postulates [12] and showing that is does not satisfy the (Mem_i) postulates. Indeed, consider $B = \neg p$, $A_1 = \neg q$, $A_2 = p \vee q$. Then $B_2 = (\neg p \star \neg q) \star (p \vee q) = (\neg p \wedge \neg q) \star (p \vee q) = p \oplus q$ where $\oplus$ is the exclusive or. But this is different from $B \star B_2 = \neg p \star ((\neg p \star \neg q) \star (p \vee q))$ $= \neg p \star ((\neg p \wedge \neg q) \star (p \vee q)) = \neg p \star (p \oplus q) = \neg p \wedge q$.

We can easily find revision operators satisfying these additional postulates :

Theorem 9. *If $\star$ is a maxichoice revision operator then $\star$ satisfies every postulate (Mem_i).*

The postulates of this family are ordered by strength, as shows the following result:

Theorem 10. *If $\star$ satisfies postulate (Mem_{i+1}) then $\star$ satisfies postulate (Mem_i).*

The other way round, (Mem_i) does not always imply (Mem_{i+1}): this is immediate for $i = 0$.

So is those families of operators, defined from the (Mem_i) postulates, are wide ones ? It is not the case. We show that, once again, only maxichoice revision operators satisfy our postulates.

Theorem 11. *If $\star$ satisfies (R1-R6) and (Mem_2), then $\star$ is a maxichoice revision operator.*

Proof. Suppose that A is consistent and that $B \vdash \neg A$. We want to show that $B \star A$ is maximal, i.e. for an arbitrary C we have that either $B \star A \vdash C$, or $B \star A \vdash \neg C$.

First, (Mem_2) tells us that $(\neg A \vee C) \star B \star A = (\neg A \vee C) \star ((\neg A \vee C) \star B \star A)$, and similarly $(\neg A \vee \neg C) \star B \star A = (\neg A \vee \neg C) \star ((\neg A \vee \neg C) \star B \star A)$. As $B \vdash \neg A$ we have $B = (\neg A \vee C) \star B$ by (R2), and similarly $B = (\neg A \vee \neg C) \star B$. Hence $(\neg A \vee C) \star B \star A = (\neg A \vee C) \star ((\neg A \vee C) \star B \star A) = (\neg A \vee C) \star (B \star A)$, and similarly

$(\neg A \vee \neg C) \star B \star A = (\neg A \vee \neg C) \star ((\neg A \vee \neg C) \star B \star A) = (\neg A \vee \neg C) \star (B \star A)$. Now suppose that not(either $B \star A \vdash C$, or $B \star A \vdash \neg C$), i.e. $B \star A$ is consistent with C, and $B \star A$ consistent with $\neg C$. Then we must have $(\neg A \vee C) \star B \star A = (\neg A \vee C) \star ((\neg A \vee C) \star B \star A) = (\neg A \vee C) \star (B \star A) = (\neg A \vee C) \wedge (B \star A)$, and $(\neg A \vee \neg C) \star B \star A = (\neg A \vee \neg C) \star ((\neg A \vee \neg C) \star B \star A) = (\neg A \vee \neg C) \star (B \star A) = (\neg A \vee \neg C) \wedge (B \star A)$. As $B \star A \vdash A$, we would have that $(\neg A \vee C) \wedge (B \star A) \vdash C$, and $(\neg A \vee \neg C) \wedge (B \star A) \vdash \neg C$. But by AGM $(\neg A \vee C) \star (B \star A)$ must be consistent.

A corollary of the theorems 10 and 11 is that a revision operator satisfies a (Mem_i) postulate if and only if it is a maxichoice revision operator. So each postulate of this family is a characterisation of maxichoice operators.

As explained at the beginning of this section, the idea of this family of postulates seems very close to the one behind the definition of revision with memory operators. In the next section we will investigate more deeply the links between revision with memory operators and the requirements on classical AGM revision operators.

6 The relation with revision with memory operators

Belief revision operators with memory [14, 13] keep trace of the history of beliefs in order to be able to use them whenever further revisions make this possible. They are based on a notion of belief state that is more complex than the flat set of beliefs of the AGM framework.

Basically, if we represent epistemic states Φ by a pre-order on interpretations, noted $\leq_\Phi$, we can extract the associated belief set with the projection operator $Bel(\Phi) = \min(\mathcal{W}, \leq_\Phi)$. The pre-order $\leq_\Phi$ represents the agent's relative confidence in interpretations. For example $\omega <_\Phi \omega'$ means that for the agent in the epistemic state Φ the interpretation ω seems (strictly) more plausible than the interpretation ω'.

The usual logical notations extend straightforwardly to epistemic states (they in fact denote conditions on the associated belief sets). For example $\Phi \vdash C$, $\Phi \wedge C$ and $\omega \models \Phi$ respectively mean $Bel(\Phi) \vdash C$, $Bel(\Phi) \wedge C$ and $\omega \models Bel(\Phi)$.

Now let us define revision with memory operators. This family of operators is parametrized by a classical AGM operator. It can be seen as a tool to change a classical AGM operator with bad iteration properties into an operator that has good ones.

Definition 7 (Revision with memory). *Suppose that we dispose of a classical AGM operator $\star$. (We will use its corresponding faithful assignment $C \rightarrow \leq_C$.) Then we define the epistemic state (the pre-order) $\Phi \circ C$ that results from the revision with memory of Φ by the new information C as:*

$$\begin{aligned} \omega \leq_{\Phi \circ C} \omega' \text{ iff } & \omega <_C \omega' \text{ or} \\ & \omega \simeq_C \omega' \text{ and } \omega \leq_\Phi \omega' \end{aligned}$$

This definition means that each incoming piece of information induces some credibility ordering. (The exact ordering induced depends on the classical AGM operator that has been chosen.[1]) And the new epistemic state is built by listening first to this incoming piece of information, and then to the old epistemic state (this is the well known *primacy of update* principle).

In fact, it is shown in [13], that an epistemic state for revision with memory operators can be encoded as the history of the new pieces of information acquired by the agent since its "birth". So we can suppose that the agent starts from an "empty" epistemic state Ξ, that is represented by a flat pre-order[2], and successively accommodates all the pieces of information. So if we suppose that all revision sequences start from Ξ, it can be shown that all revision with memory operators satisfy the (Mem_i) postulates, since they all take the history of the revisions into account.

Theorem 12. *A revision operator with memory satisfies (Mem_i), $\forall i$.*

In fact, a logical characterization for revision with memory operators has been given in [13]. Most of the postulates are generalizations of AGM postulates in the epistemic states framework, but there are also some specific postulates characterizing revision with memory. We will examine now their status in the classical belief set framework. Those postulates have been written for epistemic states, but we can translate them for belief sets (with some simplifications) as follows :

(Hist1) $(B \star A) \star C \equiv B \star (A \star C)$
(Hist2) If $C \star A \equiv A$, then $(B \star C) \star A \equiv B \star A$
(Hist3) If $C \star A \vdash D$, then $(B \star C) \star A \vdash D$

The first postulate expresses some kind of associativity and aims at expressing the strong influence of the new piece of information. The second one says that if a formula C does not distinguish between the models of A, then learning C before A is without effect on the resulting belief set. The third one says that the consequences of a revision also holds if we first learn another piece of information.

The counterpart of (Hist1), (Hist2) and (Hist3) for epistemic states are respectively named (H7), (H'7) and (H'8) in [13]. It is shown there that in the presence of the other postulates (H1-H6) (that are mainly a generalisation of AGM postulates in the epistemic state framework), (H7) is equivalent to (H'7-H'8).

This equivalence no longer holds in the belief set framework. Let us see now the implications of these three postulates in this framework.

Theorem 13. *There is no operator that satisfies (R1-R6) and (Hist1).*

[1] Note that one of the possibilities is a two level pre-order with the models of the formula at the lowest level, and the counter-models at the top level. That gives the more "classical" operator of the family [18, 16, 20, 3].

[2] that is $\forall \omega, \omega' \; \omega \simeq_{\Xi} \omega'$

Proof. Let $\omega_0, \omega_1, \omega_2, \omega_3$ be 4 distinct interpretations. Now take four formulas A, B, C, D such that $Mod(A) = \{\omega_1, \omega_2\}$, $Mod(B) = \{\omega_0, \omega_1\}$, $Mod(C) = \{\omega_2, \omega_3\}$ and $Mod(D) = \{\omega_1, \omega_3\}$. From (Hist1) we have that $(B \star A) \star C = B \star (A \star C)$, that is from (R2) $(B \wedge A) \star C = B \star (A \wedge C)$. As $Mod(A \wedge C) = \{\omega_2\}$, from (R1) and (R3) it follows that $Mod(B \star (A \wedge C)) = \{\omega_2\}$, hence $Mod((B \wedge A) \star C) = \{\omega_2\}$. On the other side, starting from (Hist1) with $(B \star D) \star C = B \star (D \star C)$, we obtain similarly $Mod(B \star (D \wedge C)) = Mod((B \wedge D) \star C) = \{\omega_3\}$. Now notice that $B \wedge D \equiv B \wedge A$, so (R4) says that $(B \wedge A) \star C \equiv (B \wedge D) \star C$. Contradiction.

Note that (Hist2) is stronger that the postulate (C1) proposed by Darwiche and Pearl. As (C1) is consistent with the AGM postulates we will consider a weakening of the (Hist2) postulate, that accounts for the case when $A \not\vdash C$:

(StrictHist2) If $C \star A \equiv A$ and $A \nvdash C$, then $(B \star C) \star A \equiv B \star A$

Theorem 14. *If an operator $\star$ satisfies (R1-R6) and (StrictHist2), then $\star$ is a maxichoice revision operator.*

Proof. We show that if $\star$ satisfies (StrictHist2), then $\star$ is maxichoice. If $\star$ is not maxichoice, then there exists a formula C such that $\leq_C$ is not linear, that means that we can find a formula A and two distinct interpretations ω, ω', with $Mod(A) = \{\omega, \omega'\}$ (with $\omega \neq \omega'$) such that $C \wedge A$ is not consistent[3] and $\omega \simeq_C \omega'$, ie $C \star A = A$. (StrictHist2) then says that for all B $(B \star C) \star A = B \star A$. In particular if we take B such that $Mod(B) = Mod(C) \cup \{\omega\}$, that means that $C \star A = B \star A = A$. But from (R2) we get that $B \star A = B \wedge A$, so $Mod(B \star A) = \{\omega\}$. Contradiction.

So, as a corollary of theorems 14 et 4, every operator satisfying (R1-R6) and (Hist2) must be an imposed maxichoice operator.

Theorem 15. *There is no operator that satisfies (R1-R6) and (Hist3).*

Proof. Let $\omega_0, \omega_1, \omega_2$ be 3 distinct interpretations. Now take four formulas A, B, C, D such that $Mod(A) = \{\omega_1, \omega_2\}$, $Mod(B) = \{\omega_0\}$, $Mod(C) = \{\omega_0, \omega_1\}$, and $Mod(D) = \{\omega_0, \omega_2\}$. As from (R2) $C \star A = C \wedge A$, then $Mod(C \star A) = \{\omega_1\}$, so from (Hist3) and (R3), that means that $Mod((B \star C) \star A) = \{\omega_1\}$. On the other side, starting from $D \star A$, we find similarly that $Mod((B \star D) \star A) = \{\omega_2\}$. Finally, as from (R2) we find easily that $(B \star C) \equiv (B \star D)$, from (R4) we have that $(B \star C) \star A \equiv (B \star D) \star A$. Contradiction.

These three results show, once again, that it is hard to try to formulate iteration postulates in the AGM framework. Whereas those properties are meaningful in the epistemic state framework, two of them, (Hist1) and (Hist3), are not consistent with AGM postulates for belief set revision, and the last one, (StrictHist2), implies the maxichoice property.

[3] When $\star$ is an AGM revision operator and $C \star A \equiv A$, then $C \wedge A \nvdash \perp$ is equivalent to $A \vdash C$.

7 Conclusion

Studies in iterated belief revision have been stated in the epistemic state framework mainly because of the influence of Darwiche and Pearl's proposal [6, 7] and its incompatibility with the AGM belief set framework. But since, few work has been done to see if some properties on iteration can be stated in the classical framework.

We have addressed this issue in this paper by looking at some candidates postulates. In different ways, all of them express that the result of a revision must keep as much as possible of the old information.

Our results are mainly negative. When the proposed postulates are not inconsistent with classical AGM ones, they inexorably lead to the maxichoice property, which is far from satisfactory for a sensible revision operator. So the results obtained in this paper can be seen as "impossibility results" about iteration in the classical AGM framework.

This study is then important to justify the gap, both in terms of knowledge representation and in terms of computational complexity, induced by all the iterated revision approaches that abandon the classical framework and work with more complex objects, viz. epistemic states.

References

1. 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.
2. C. E. Alchourrón and D. Makinson. The logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.
3. S. Benferhat, D. Dubois, and O. Papini. A sequential reversible belief revision method based on polynomials. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI'99)*, pages 733–738, 1999.
4. R. Booth. On the logic of iterated non-prioritised revision. In *Workshop on Conditionals, Information and Inference. Hagen, Germany*, 2002.
5. M. Dalal. Investigations into a theory of knowledge base revision: preliminary report. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'88)*, pages 475–479, 1988.
6. A. Darwiche and J. Pearl. On the logic of iterated belief revision. In Morgan Kaufmann, editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1994 Conference (TARK'94)*, pages 5–23, 1994.
7. A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.
8. M. Freund and D. Lehmann. Belief revision and rational inference. Technical Report TR-94-16, Institute of Computer Science, The Hebrew University of Jerusalem, 1994.
9. P. Gärdenfors. *Knowledge in flux*. MIT Press, 1988.
10. A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17(157-180), 1988.
11. Sven Ove Hansson. *A textbook of belief dynamics*. Kluwer Academic Press, 1999.

12. H. Katsuno and A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.
13. S. Konieczny and R. Pino Pérez. A framework for iterated revision. *Journal of Applied Non-Classical Logics*, 10(3-4):339–367, 2000.
14. S. Konieczny and R. Pino Pérez. Some operators for iterated revision. In *Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01)*, pages 498–509, 2001.
15. D. Lehmann. Belief revision, revised. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, pages 1534–1540, 1995.
16. P. Liberatore. The complexity of iterated belief revision. In *Proceedings of the Sixth International Conference on Database Theory (ICDT'97)*, pages 276–290, 1997.
17. T. Meyer. Basic infobase change. *Studia Logica*, 67:215–242, 2001.
18. A. C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.
19. B. Nebel. Syntax-based approaches to belief revision. In P. Gärdenfors, editor, *Belief revision*, volume 29 of *Journal of Cambridge Tracts in Theoretical Computer Science*, pages 52–88. Cambridge University Press, 1992.
20. O. Papini. Iterated revision operations stemming from the history of an agent's observations. In H. Rott and M. A. Williams, editors, *Frontiers in Belief revision*, pages 279–301. Kluwer, 1999.

DA^2 MERGING OPERATORS

Sébastien Konieczny, Jérôme Lang, Pierre Marquis.
Artificial Intelligence. 157(1-2).
pages 49-79.
2004.

DA^2 Merging Operators ⋆

S. Konieczny [a] J. Lang [a] P. Marquis [b]

[a] *IRIT - CNRS / Université Paul Sabatier*
118 route de Narbonne - 31062 Toulouse - France
{konieczny,lang}@irit.fr

[b] *CRIL - CNRS / Université d'Artois*
rue de l'Université - S.P. 16 - 62307 Lens - France
marquis@cril.univ-artois.fr

Abstract

A new framework for propositional merging is presented. *DA^2 merging operators*, parameterized by a distance between interpretations and two aggregation functions, are introduced. Many distances and aggregation functions can be used and many merging operators already defined in the literature (including both model-based ones and syntax-based ones) can be encoded as specific DA^2 operators. Both logical and complexity properties of those operators are studied. An important result is that (under very weak assumptions) query entailment from merged bases is "only" at the first level of the polynomial hierarchy when any of the DA^2 operators is used. As a by-product, complexity results for several existing merging operators are derived as well.

Key words: Knowledge Representation, Belief Merging, Computational Complexity.

1 Introduction

Belief merging is an important issue of many AI fields (see (1) for a panorama of applications of data and belief fusion.) Although particular requirements can be asked for each application, several pieces of information are usually brought into play when propositional base merging is concerned. In the following:

⋆ This paper is an extended and revised version of the paper entitled "*Distance-based merging: a general framework and some complexity results*", that appeared in the proceedings of the *Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 97-108, Toulouse, 2002.

- A *belief profile* $E = \{K_1, \dots, K_n\}$ is a finite multi-set of belief bases, where each *belief base* K_i represents the set of beliefs from source i. Each K_i is a finite set of consistent propositional formulas $\varphi_{i,j}$ encoding the explicit beliefs from source i.
- IC is a propositional formula encoding some *integrity constraints*. IC represents some information the result of the merging has to obey (e.g. some physical constraints, norms, etc..)

The purpose of merging E is to characterize a formula (or a set of formulas) $\triangle_{IC}(E)$, considered as the overall belief from the n sources given the integrity constraints IC. Recently, several families of such merging operators have been defined and characterized in a logical way (2; 3; 4; 5; 6). Among them are the so-called *model-based* merging operators (2; 3; 4; 5) where the models of $\triangle_{IC}(E)$ are defined as the models of IC which are preferred according to some criterion depending on E. Often, such preference information takes the form of a total pre-order over interpretations, induced by a notion of distance $d(\omega, E)$ between an interpretation ω and the belief profile E. The distance $d(\omega, E)$ is typically defined by aggregating the distances $d(\omega, K_i)$ for every K_i. Usually, model-based merging operators take only into account consistent belief bases K_i. Other merging operators are so-called *syntax-based* ones (7; 8; 9). They are based on the selection of some consistent subsets of the set-theoretic union $\bigcup_{i=1}^{n} K_i$ of the belief bases. This allows for taking inconsistent belief bases K_i into account and to incorporate some additional preference information into the merging process. Indeed, as in belief revision, relying on the syntax of K_i is a way to specify (implicitly but in a cheap way with respect to representation) that explicit beliefs are preferred to implicit beliefs (10; 11). But the price to be paid is the introduction of an additional connective ",", which is not truth functional. Moreover, since they are based on the set-theoretic union $\bigcup_{i=1}^{n} K_i$ of the bases, such operators usually do not take into account the frequency of each explicit piece of belief into the merging process (the fact that $\varphi_{i,j}$ is believed in one source only or in the n sources under consideration is not considered relevant, which is often counter-intuitive.)

In this paper, a new framework for defining propositional merging operators is provided. A family of merging operators parameterized by a distance d between interpretations and two aggregation functions $\oplus$ and $\odot$ is presented. Accordingly, DA^2 merging operator is a short for Distance-based merging operator, obtained through 2 Aggregation steps. The parameters d, $\oplus$, $\odot$ are used to define a notion of distance between an interpretation and a belief profile E in a two-step fashion. Like in existing model-based approaches to merging, the models of the merging $\triangle_{IC}^{d,\oplus,\odot}(E)$ of E given some integrity constraints IC are exactly the models of IC that are as close as possible to E with respect to the distance. Moreover, the first aggregation step allows to take into account the syntax of belief bases within the merging process (and to handle inconsistent ones in a satisfying way.)

The contribution of this work is many fold. First, our framework is general enough to encompass many model-based merging operators as specific cases, especially those given in (2; 3; 4; 5; 12; 13; 6). In addition, despite the model-theoretic ground of our approach, several syntax-based merging operators provided so far in the literature can be captured as well (7; 8; 9). We show that, by imposing few conditions on the parameters, several logical properties that are expected when merging operators are considered, are satisfied by DA^2 operators.

Another very strong feature offered by our framework is that query entailment from $\triangle_{IC}^{d,\oplus,\odot}(E)$ is guaranteed to lay at the first level of the polynomial hierarchy provided that d, $\oplus$ and $\odot$ can be computed in polynomial time. Accordingly, improving the generality of the model-based merging operators framework through an additional aggregation step does not result in a complexity shift.

We specifically focus on some simple families of distances and aggregation functions. By letting the parameters d, $\oplus$ and $\odot$ vary in these respective sets, several merging operators are obtained; some of them were already known and are thus encoded as specific cases in our framework, and others are new operators. In any case, we investigate the logical properties and identify the complexity of each operator under consideration. As a by-product, the complexity of several model-based merging operators already pointed out so far is also identified.

The remaining of the paper is as follows. In Section 2 we give some formal preliminaries, and we recall some notions of computational complexity and some axiomatic properties for belief merging. In Section 3 we give a glimpse at the two main families of merging methods: model-based merging operators and syntax-based ones. In Section 4 we introduce DA^2 merging operators and give some examples. In Section 5 we study the computational complexity of this class of operators. This section also gives complexity results for some specific operators from the class. In Section 6 we address the logical properties of the operators. Finally Section 7 concludes the paper and presents some directions for future work.

2 Formal Preliminaries

We consider a propositional language $PROP_{PS}$ built up from a finite set PS of propositional symbols in the usual way. $\top$ (resp. $\bot$) denotes the Boolean constant interpreted to 1 (true) (resp. 0 (false)). An interpretation is a total function from PS to $BOOL = \{0, 1\}$. It is denoted by a tuple of literals over PS (or a tuple of truth values 0, 1 when a total ordering over PS is given).

The set of all interpretations is denoted by $\mathcal{W}$. An interpretation ω is a model of a formula if it makes it true in the usual classical truth functional way.

Provided that φ is a formula from $PROP_{PS}$, $Mod(\varphi)$ denotes the set of models of φ, i.e., $Mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$. Conversely, let M be a set of interpretations, $form(M)$ denotes the logical formula (unique up to logical equivalence) whose models are M.

Two belief bases K_1 and K_2 are said to be logically equivalent ($K_1 \equiv K_2$) if $\bigwedge K_1 \equiv \bigwedge K_2$, and two belief profiles E_1 and E_2 are said to be equivalent ($E_1 \equiv E_2$) if and only if there is a bijection between E_1 and E_2 such that each belief base of E_1 is logically equivalent to its image in E_2. A belief base K_i is said to be consistent if and only if the conjunction $\bigwedge K_i$ of its formulas is consistent. Similarly, a belief profile E is said to be consistent if the conjunction of its belief bases $\bigwedge E = \bigwedge_{K_i \in E} \bigwedge_{\varphi_{i,j} \in K_i} \varphi_{i,j}$ is consistent. $\sqcup$ denotes the multi-set union. For every belief profile E and for every integer n, E^n denotes the multi-set containing E n times.

For any set A, let $\leq$ be any binary relation over $A \times A$. $\leq$ is said to be *total* if $\forall a, b \in A$, $a \leq b$ or $b \leq a$; *reflexive* if $\forall a \in A$, $a \leq a$; *transitive* if $\forall a, b, c \in A$, ($a \leq b$ and $b \leq c$) implies $a \leq c$. Let $\leq$ be any binary relation, $<$ is its strict counterpart, i.e., $a < b$ if and only if $a \leq b$ and $b \not\leq a$, and $\simeq$ is its indifference relation, i.e. $a \simeq b$ if and only if $a \leq b$ and $b \leq a$. We denote $\min(A, \leq)$ the set $\{a \in A \mid \nexists b \in A \; b < a\}$.

2.1 Computational Complexity

The complexity results we give in this paper refer to some complexity classes which we now briefly recall (see (14) for more details), especially the classes Δ_2^p and Θ_2^p (15; 16) from the polynomial hierarchy PH, as well as the class BH_2 from the Boolean hierarchy. We assume the reader familiar with the classes P, NP et coNP and we now introduce the following three classes located at the first level of the polynomial hierarchy:

- BH_2 (also known as DP) is the class of all languages L such that $L = L_1 \cap L_2$, where L_1 is in NP and L_2 in coNP. The canonical BH_2-complete problem is SAT–UNSAT: given two propositional formulas φ and ψ, $\langle \varphi, \psi \rangle$ is in SAT–UNSAT if and only if φ is consistent and ψ is inconsistent.
- $\Delta_2^p = \mathsf{P}^{\mathsf{NP}}$ is the class of all languages that can be recognized in polynomial time by a deterministic Turing machine equipped with an NP oracle, where an NP oracle solves whatever instance of a problem from NP in unit time.
- $\Theta_2^p = \Delta_2^p[\mathcal{O}(log\ n)]$ is the class of all languages that can be recognized in polynomial time by a deterministic Turing machine using a number of calls to an NP oracle bounded by a logarithmic function of the size of the input

data.

Note that the following inclusions hold:

$$\mathsf{NP} \cup \mathsf{coNP} \subseteq \mathsf{BH}_2 \subseteq \Theta_2^p \subseteq \Delta_2^p \subseteq \mathsf{PH}.$$

Finally, $\mathsf{F}\Delta_2^p$ is the class of function problems associated wih Δ_2^p, i.e. those that can be solved in deterministic polynomial time on a Turing machine equipped with an NP oracle.

2.2 Logical Properties for Belief Merging

Some work in belief merging aims at finding sets of axiomatic properties operators may exhibit the expected behaviour (17; 2; 4; 18; 5; 12). We focus here on the characterization of Integrity Constraints (IC) merging operators (5; 13).

Definition 1 (IC merging operators) *Let E, E_1, E_2 be belief profiles, K_1, K_2 be consistent belief bases, and IC, IC_1, IC_2 be formulas from $PROP_{PS}$. $\triangle$ is an* IC merging operator *if and only if it satisfies the following postulates:*

(IC0) $\triangle_{IC}(E) \models IC$.
(IC1) *If IC is consistent, then $\triangle_{IC}(E)$ is consistent.*
(IC2) *If $\bigwedge E$ is consistent with IC, then $\triangle_{IC}(E) \equiv \bigwedge E \wedge IC$.*
(IC3) *If $E_1 \equiv E_2$ and $IC_1 \equiv IC_2$, then $\triangle_{IC_1}(E_1) \equiv \triangle_{IC_2}(E_2)$.*
(IC4) *If $K_1 \models IC$ and $K_2 \models IC$, then $\triangle_{IC}(\{K_1, K_2\}) \wedge K_1$ is consistent if and only if $\triangle_{IC}(\{K_1, K_2\}) \wedge K_2$ is consistent.*
(IC5) $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2) \models \triangle_{IC}(E_1 \sqcup E_2)$.
(IC6) *If $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$ is consistent , then $\triangle_{IC}(E_1 \sqcup E_2) \models \triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$.*
(IC7) $\triangle_{IC_1}(E) \wedge IC_2 \models \triangle_{IC_1 \wedge IC_2}(E)$.
(IC8) *If $\triangle_{IC_1}(E) \wedge IC_2$ is consistent, then $\triangle_{IC_1 \wedge IC_2}(E) \models \triangle_{IC_1}(E)$.*

The intuitive meaning of the properties is the following: (IC0) ensures that the result of merging satisfies the integrity constraints. (IC1) states that, if the integrity constraints are consistent, then the result of merging will be consistent. (IC2) states that if possible, the result of merging is simply the conjunction of the belief bases with the integrity constraints. (IC3) is the principle of irrelevance of syntax : the result of merging has to depend only on the expressed opinions and not on their syntactical presentation. (IC4) is a fairness postulate meaning that the result of merging of *two* belief bases should not give preference to one of them (if it is consistent with one of both, it has to be consistent with the other one.) It is a symmetry condition, that aims to rule out operators that can give priority to one of the bases. Note that (IC4) is a strong impartiality requirement and may appear very strong in some cases, but nevertheless it is satisfied by many interesting merging operators. Note that stating this property makes sense only because the belief bases K_i are required to be consistent. (IC5) expresses the following idea: if belief profiles are viewed as expressing the beliefs of the members of a group, then if E_1 (corresponding to a first group) compromises on a set of alternatives which A belongs to, and E_2 (corresponding to a second group) compromises on another set of alternatives which contains A too, then A has to be in the chosen alternatives if we join the two groups. (IC5) and (IC6) together state that if one could find two subgroups which agree on at least one alternative, then the result of the global merging will be exactly those alternatives the two groups agree on. (IC7) and (IC8) state that the notion of closeness is well-behaved, i.e., that an alternative that is preferred among the possible alternatives (IC_1), will remain preferred if one restricts the possible choices ($IC_1 \wedge IC_2$).

Two sub-classes of IC merging operators have been defined. *IC Majority operators* aim at resolving conflicts by adhering to the majority wishes, while *IC arbitration operators* have a more consensual behaviour:

Definition 2 (majority and arbitration) *An* IC majority operator *is an IC merging operator that satisfies the following majority postulate:*

(Maj) $\exists n \;\; \triangle_{IC}\,(E_1 \sqcup {E_2}^n) \models \triangle_{IC}(E_2)$.

An IC arbitration operator *is an IC merging operator that satisfies the following arbitration postulate:*

(Arb) $$\left.\begin{array}{l} \triangle_{IC_1}(K_1) \equiv \triangle_{IC_2}(K_2) \\ \triangle_{IC_1 \Leftrightarrow \neg IC_2}(\{K_1, K_2\}) \equiv (IC_1 \Leftrightarrow \neg IC_2) \\ IC_1 \not\models IC_2 \\ IC_2 \not\models IC_1 \end{array}\right\} \Rightarrow \triangle_{IC_1 \vee IC_2}(\{K_1, K_2\}) \equiv \triangle_{IC_1}(K_1).$$

See (5; 12) for explanations about those two postulates and the behaviour of the two corresponding classes of merging operators. For the sake of simplicity, we simply refer to such operators in the following as majority (resp. arbitration) ones, omitting IC.

3 Model-based Merging vs Syntax-based Merging

In this section, we recall the two main families of belief merging operators: the model-based ones and the syntax-based ones.

3.1 Model-based Merging

The idea here is that the result of the merging process is a belief base (up to logical equivalence) whose models are the *best* ones for the given belief profile E. Formally, provided that $\leq_E$ denotes an arbitrary binary relation (usually $\leq_E$ is required to be total, reflexive and transitive) on $\mathcal{W}$:

$$Mod(\triangle_{IC}(E)) = \min(Mod(IC), \leq_E)$$

Accordingly, in order to define a model-based merging operator, one just has to point out a function that maps each belief profile E to a binary relation $\leq_E$ (see (5) for conditions on this function.)

A compact way to characterize $\leq_E$ consists in deriving it from a notion of distance between an interpretation ω and a belief profile E (in this case $\leq_E$ is a total pre-order):

$$\omega \leq_E \omega' \text{ if and only if } d(\omega, E) \leq d(\omega', E).$$

$d(\omega, E)$ is usually defined by choosing a distance between interpretations aiming at building "individual" evaluations of each interpretation for each belief base, and then by aggregating those evaluations in a "social" evaluation of each interpretation. Indeed, assume that we have a distance d between interpretations (cf. Definition 5) that fits our particular application. Then one can define an (individual) belief base evaluation of each interpretation as the minimal distance between this interpretation and the models of the belief base:

$$d(\omega, K) = \min_{\omega' \models K} d(\omega, \omega').$$

Then it remains to compute a (social) belief profile evaluation of the interpretations using some aggregation function $*$:

$$d(\omega, E) = *_{K \in E}\, d(\omega, K).$$

In the first works on model-based merging, the distance used was Dalal's distance (19), namely, the Hamming distance between interpretations, and the aggregation function was the *sum* or the *max* (2; 3). In (5; 12) it has been shown that one can take any distance between interpretations without changing the logical properties of the operators and a *leximax* aggregation function was proposed as an example of arbitration operator.

Syntax-based merging (also called formula-based merging) operators work from preferred consistent subsets of formulas. The differences between the operators of this family lie in the definition of the preference relation (maximality with respect to set inclusion for instance.)

Let us briefly present the operators given in (7; 8).

Definition 3 *Let* $\textsc{maxcons}(K, IC)$ *be the set of the* maxcons *of* $K \cup \{IC\}$ *that contain* IC*, i.e., the maximal (with respect to set inclusion) consistent subsets of* $K \cup \{IC\}$ *that contain* IC*. Formally,* $\textsc{maxcons}(K, IC)$ *is the set of all* M *such that:*

- $M \subseteq K \cup \{IC\}$*, and*
- $IC \subseteq M$*, and*
- *if* $M \subset M' \subseteq K \cup \{IC\}$*, then* $M' \models \bot$.

Let $\textsc{maxcons}(E, IC) = \textsc{maxcons}(\bigcup_{K_i \in E} K_i, IC)$*. When the maximality of the sets is defined in terms of cardinality, we will use the subscript "card", i.e. we will note the set* $\textsc{maxcons}_{card}(E, IC)$*.*

Let us define the following operators:

Definition 4 *Let* E *be a belief profile and* IC *be a belief base:*

$\triangle^{C1}{}_{IC}(E) = \bigvee \textsc{maxcons}(E, IC)$.
$\triangle^{C3}{}_{IC}(E) = \bigvee\{M : M \in \textsc{maxcons}(E, \top) \textit{ and } M \cup \{IC\} \textit{ consistent}\}$.
$\triangle^{C4}{}_{IC}(E) = \bigvee \textsc{maxcons}_{card}(E, IC)$.
$\triangle^{C5}{}_{IC}(E) = \bigvee\{M \cup \{IC\} : M \in \textsc{maxcons}(E, \top) \textit{ and } M \cup \{IC\} \textit{ consistent}\}$
if this set is non empty and IC *otherwise.*

The $\triangle^{C1}$ operator takes as result of the combination the set of the maximal consistent subsets of $E \cup \{IC\}$ that contain the constraints IC. The $\triangle^{C3}$ operator computes first the set of the maximal consistent subsets of E, and then selects those that are consistent with the constraints. The $\triangle^{C4}$ operator selects the set of consistent subsets of $E \cup \{IC\}$ that contain the constraints IC and that are maximal with respect to cardinality.

$\triangle^{C1}{}_{IC}(E)$, $\triangle^{C3}{}_{IC}(E)$ and $\triangle^{C4}{}_{IC}(E)$ correspond respectively to $Comb1(E, IC)$, $Comb3(E, IC)$ and $Comb4(E, IC)$ as defined in (8) (there is no actual need to to consider the $Comb2$ operator since it is equivalent to $Comb1$ (8).) The $\triangle^{C5}$ operator is a slight modification of $\triangle^{C3}$ in order to get more logical properties (9).

Once the union of the belief bases is performed, the problem is to extract some coherent piece of information from it. Thus, such an approach is very close to Rescher and Manor's inference (20), Brewka's preferred subtheories (21), to the work by Benferhat *et al.* on entailment from inconsistent databases (22; 23; 24), as well as to several approaches to belief revision (25; 10; 26) and to reasoning with counterfactuals (27).

A drawback of this approach is that the distribution of information is not taken into account in the consistency restoration process. To deal with this drawback, it has been proposed in (9) to select only the maxcons that best fit a merging criterion. Those selection functions are related to those used in the AGM belief revision framework for *partial meet revision functions* (28). In both cases the selection functions aim at selecting only some of the maxcons (the "best" ones.) The idea for belief merging is to use the selection function to incorporate a "social" evaluation of maxcons.

In (9) three particular criteria have been proposed and studied. The first one selects the maxcons that are consistent with as many belief bases as possible. The second one takes the maxcons that have the smallest symmetrical difference (with respect to cardinality) with the belief bases and the last one takes the maxcons that have the largest intersection (with respect to cardinality) with the belief bases.

4 DA^2 Merging

4.1 The General Framework

Defining a merging operator in our framework simply consists in setting three parameters: a distance d and two aggregation functions $\oplus$ and $\odot$. Let us first make it precise what such notions mean in this paper:

Definition 5 (distances) *A distance between interpretations*[1] *is a total function d from $\mathcal{W} \times \mathcal{W}$ to $I\!N$ such that for every ω_1, $\omega_2 \in \mathcal{W}$*

- $d(\omega_1, \omega_2) = d(\omega_2, \omega_1)$, *and*
- $d(\omega_1, \omega_2) = 0$ *if and only if* $\omega_1 = \omega_2$.

Any distance between interpretations d induces a distance between an inter-

[1] We slightly abuse words here, since d is only a pseudo-distance (triangular inequality is not required.)

pretation ω and a formula φ given by

$$d(\omega, \varphi) = \min_{\omega' \models \varphi} d(\omega, \omega').$$

Definition 6 (aggregation functions) *An aggregation function is a total function $\oplus$ associating a nonnegative integer to every finite tuple of nonnegative integers and verifying* (non-decreasingness), (minimality) *and* (identity).

- *if $x \leq y$, then $\oplus(x_1, \dots, x, \dots, x_n) \leq \oplus(x_1, \dots, y, \dots, x_n)$.* (non-decreasingness)
- *$\oplus(x_1, \dots, x_n) = 0$ if and only if $x_1 = \dots = x_n = 0$.* (minimality)
- *for every nonnegative integer x, $\oplus(x) = x$.* (identity)

We are now in position to define DA^2 merging operators. Basically the distance gives the closeness between an interpretation and each formula of a belief base. Then a first aggregation function $\oplus$ evaluates the plausibility (resp. desirability) of the interpretation for an agent (belief base) K_i from those closeness degrees when formulas are interpreted as information items (resp. preference items). And finally the second aggregation function $\odot$ evaluates the plausibility (resp. desirability) of the interpretation for the whole group (belief profile.)

Definition 7 (DA^2 merging operators) *Let d be a distance between interpretations and $\oplus$ and $\odot$ be two aggregation functions. For every belief profile $E = \{K_1, \dots, K_n\}$ and every integrity constraint IC, $\triangle_{IC}^{d,\oplus,\odot}(E)$ is defined in a model-theoretical way by:*

$$Mod(\triangle_{IC}^{d,\oplus,\odot}(E)) = \min(IC, \leq_E^{d,\oplus,\odot}).$$

$\leq_E^{d,\oplus,\odot}$ is defined as $\omega \leq_E^{d,\oplus,\odot} \omega'$ if and only if $d(\omega, E) \leq d(\omega', E)$, where

$$d(\omega, E) = \odot(d(\omega, K_1), \dots, d(\omega, K_n)),$$

and for every $K_i = \{\varphi_{i,1}, \dots, \varphi_{i,n_i}\}$,

$$d(\omega, K_i) = \oplus(d(\omega, \varphi_{i,1}), \dots, d(\omega, \varphi_{i,n_i})).$$

Defining two separate aggregation steps is not a theoretical fantasy that is only motivated by a struggle for generalization; rather, it formalizes the different nature of belief bases and belief profiles:

- a belief base is the set of elementary data reported by a given entity. The precise meaning of this rather vague formulation ("entity") depends on the context of the merging problem:

  - when merging several pieces of belief stemming from different "sources" (in practice, a source may be a sensor, an expert, a database...), the formulas inside a belief base K_i are the *pieces of information provided by source* i;
  - when evaluating alternatives with respect to different criteria, the formulas inside a belief base K_i are the *pieces of information pertaining to criterion* i;
  - when aggregating individual preferences in a group decision making context, the formulas inside a "belief base" K_i are the *elementary goals expressed by agent* i. In this case, the formulas $\varphi_{i,j}$ are no longer *beliefs* but *preferences* (which does not prevent one from using the same merging operators.) In this case, still calling these formulas "beliefs" is no longer appropriate, but, for the sake of simplicity, we nevertheless use the terminology "belief", rather than systematically writing "beliefs or preferences", which would be rather awkward.
- a belief profile E consists of the collection of all belief bases K_i corresponding to the different sources, criteria or agents involved in the problem.

Now, since the relationship between a belief base and its elementary pieces of information and the relationship between a belief profile and its belief bases are of different nature, there is no reason for not using *two* (generally distinct) aggregation functions $\oplus$ and $\odot$. In other words, both aggregation steps corresponds to different processes. The first step is an *intra-source* (more generally, intra-entity) aggregation: $\oplus$ aggregates scores with respect to the elementary (explicit) pieces of information contained in each K_i (it allows, in particular, to take inconsistent belief bases into account.) The second step is an *inter-source* (more generally, inter-entity) aggregation: $\odot$ aggregates the "$\oplus$-aggregated scores" pertaining to the different sources. Such a two-step approach is used in a group decision context by (29).

Interestingly, few conditions are imposed on d, $\oplus$, and $\odot$. As we will see in the next section, many distances and aggregation functions can be used. Often, the aggregation functions $\oplus$ and $\odot$ are required to be symmetric (i.e., no priority is given to some explicit beliefs in a belief base, and no priority is given to some belief base in a belief profile.) However, this condition is not mandatory here and this is important when some preference information are available, especially when all sources i are not equally reliable. For instance, the *weighted sum* aggregation function gives rise to (non-symmetric) merging operators.

Let us stress that, contrarily to usual model-based operators, our definition allows for inconsistent belief bases to take (a non-trivial) part in the merging process.

Example 1 *Assume that we want to merge* $E = \{K_1, K_2, K_3, K_4\}$ *under the integrity constraints* $IC = \top$, *where*

- $K_1 = \{a, b, c, a \Rightarrow \neg b\}$,
- $K_2 = \{a, b\}$,
- $K_3 = \{\neg a, \neg b\}$,
- $K_4 = \{a, a \Rightarrow b\}$.

In this example, K_1 believes that c holds. Since this piece of information is not involved in any contradiction, it seems sensible to be confident in K_1 about the truth of c. Model-based merging operators can not handle this situation: inconsistent belief bases can not be taken into account. Thus, provided that the Hamming distance d_H between interpretations is considered, the operator $\triangle^{d_H,\Sigma}$ (2; 3; 5; 13) gives a merged base whose models (over $\{a, b, c\}$) are: $(a, b, \neg c)$ and (a, b, c); the operator $\triangle^{d_H,Gmax}$ (5; 13) gives a merged base whose models are: $(\neg a, b, \neg c)$, $(\neg a, b, c)$, $(a, \neg b, \neg c)$, and $(a, \neg b, c)$. In any of these two cases, nothing can be said about the truth of c in the merged base, which is often counter-intuitive since no argument against it can be found in the input data.

Syntax-based operators render possible the exploitation of inconsistent belief bases. Thus, on the previous example, c holds in the merged base, whatever the syntax-based operator at work (among those considered in the paper.) Obviously, this would not be the case, would the inconsistent base K_1 be replaced by an equivalent one, as $\{a, \neg a\}$. However, syntax-based operators are not affected by how the formulas are distributed among the belief bases. Consider the two standard syntax-based operators $\triangle^{C1}$ and $\triangle^{C4}$, selecting the maximal subsets of E with respect to set inclusion and to cardinality, respectively. On the previous example, $\triangle^{C1}$ returns a merged base equivalent to c and $\triangle^{C4}$ to $c \wedge \neg a$. So, a is in the result for none of these two operators, whereas a holds in three of four input bases.

Our DA^2 operators achieve a compromise between model-based operators and syntax-based operators, by taking into account the way information is distributed and by taking advantage of the information stemming from inconsistent belief bases. For instance, our operator $\triangle^{d_D,sum,sum}$ (cf. Section 4.2) gives a merged base whose single model is (a, b, c), and $\triangle^{d_D,sum,lex}$ returns a merged base whose models are $(\neg a, b, c)$ and $(a, \neg b, c)$. So, using any of these two operators, we can conclude that c holds from the merged base.

DA^2 merging operators can be viewed as a generalization of model-based merging operators, with an additional aggregation step. One can then ask why we restrict the approach to two aggregation steps instead of characterizing DA^3 merging operators and so on... Actually, it can be sensible to use those additional aggregation steps to characterize the common belief of an organization structured in a hierarchical way. For example if an organization is composed of several departments, which are divided in services, that group several teams, etc., we can figure out an aggregation step for the team level, a second one for the service level, etc. At each step it is possible to use a different aggregation scheme. A detailed study of such operators is left to further research. In the light of our results, we can nevertheless make some important remarks concerning DA^n operators. On the one hand, the first aggregation step has a specific role since it allows to take inconsistent belief bases into account in the merging process. This underlies a main difference between DA^n operators (with n ≥ 2) and DA^1 operators, the usual model-based merging operators. The latter are not suited to use inconsistent belief bases in a valuable way. The differences induced by the second and the third aggregation steps are in some sense less significant. On the other hand, it is easy to show that all our complexity results pertaining to DA^2 operators can be extended to DA^n operators (the complexity does not change provided that the number of aggregation steps is bounded a priori.) Finally, as a tool for modelizing corporation merging, we think that DA^n operators are not fully adequate. Indeed, they would suppose that the number of hierarchical divisions is the same in all the branches, and that all the groups at a given level use the same aggregation method; this is a strong, unrealistic assumption.

4.2 Instantiating our Framework

Let us now instantiate our framework and focus on some simple families of distances and aggregation functions.

Definition 8 (some distances) *Let ω_1, $\omega_2 \in \mathcal{W}$ be two interpretations.*

- *The* drastic distance d_D *is defined by*

$$d_D(\omega_1, \omega_2) = \begin{cases} 0 \text{ if } \omega_1 = \omega_2, \\ 1 \text{ otherwise.} \end{cases}$$

- *The* Hamming distance d_H *is defined by*

$$d_H(\omega_1, \omega_2) = |\{x \in PS \mid \omega_1(x) \neq \omega_2(x)\}|.$$

- *Let q be a total function from PS to $\mathbb{N}^*$. The* weighted Hamming distance d_{H_q} *induced by q is defined by*

$$d_{H_q}(\omega_1, \omega_2) = \textstyle\sum_{\{x \in PS \mid \omega_1(x) \neq \omega_2(x)\}} q(x).$$

These distances satisfy the requirements imposed in Definition 7. The Hamming distance is the distance most commonly considered in model-based merging. It is very simple to express, but it is very sensitive to the representation language of the problem (i.e., the choice of propositional symbols.) Interestingly, many other distances can be used. For instance, weighted Hamming distances are relevant when some propositional symbols are known as more important than others. [2]

As to aggregation functions, many choices are possible. We just give here two well-known classes of such functions.

Definition 9 (weighted sums) *Let q be a total function from $\{1,\ldots,n\}$ to $\mathbb{N}^*$ such that $q(1)=1$ whenever $n=1$. The weighted sum WS_q induced by q is defined by*

$$WS_q(e_1,\ldots,e_n)=\sum_{i=1}^{n} q(i)e_i.$$

q is a *weight function*, that gives to each formula (resp. belief base) φ_i (resp. K_i) of index i its weight $q(i)$ denoting the formula (resp. belief base) reliability. The requirement $q(1)=1$ whenever $n=1$ ensures that when we merge a singleton, the aggregation function has no impact.

Definition 10 (ordered weighed sums) *Let q be a total function from $\{1,\ldots,n\}$ to $\mathbb{N}$ such that $q(1)=1$ whenever $n=1$, and $q(1)\neq 0$ in any case. The ordered weighted sum OWS_q induced by q is defined by*

$$OWS_q(e_1,\ldots,e_n)=\sum_{i=1}^{n} q(i)e_{\sigma(i)}$$

where σ is a permutation of $\{1,\ldots,n\}$ such that $e_{\sigma(1)}\geq e_{\sigma(2)}\geq\ldots\geq e_{\sigma(n)}$.

The requirement $q(1)\neq 0$ is needed to meet the minimality condition (definition 6.) When using q with OWS_q, $q(i)$ reflects the importance given to the i^{th} largest value. With the slight difference that q is normalized (but without requiring that $q(1)=1$ whenever $n=1$), the latter family is well-known in multi-criteria decision making under the terminology "Ordered Weighted Averages" (OWAs) (30).

When $q(i)=1$ for every $i\in 1,\ldots,n$, WS_q and OWS_q are the usual *sum*. When $q(1)=1$ and $q(2)=\ldots=q(n)=0$, we have $OWS_q(e_1,\ldots,e_n)=max(e_1,\ldots,e_n)$. Lastly, let M be a upper bound of the scores, i.e., for any possible $(e_1,\ldots,e_n)$ we have $e_i<M$, and let $q(i)=M^{n-i}$ for all i. Then the rank order on vectors of scores induced by OWS_q is exactly the *leximax* (abbreviated by *lex*) ordering $\leq_{lex}$. Namely, we have $(e_1,\ldots,e_n)<_{lex}(e'_1,\ldots,e'_n)$

[2] Consider this example where information items about a murder coming from different witnesses; let a stand for "the murderer is a male" and b stand for "the murderer had an umbrella". Attaching a larger weight to a than to b means that the interpretation (a,b) is closer to $(a,\neg b)$ than to $(\neg a,b)$, reflecting that a mistake about b is more plausible than a mistake about a.

if and only if there exists k in $1, \ldots, n$ such that for all $i < k$, $e_{\sigma(i)} = e'_{\sigma'(i)}$ and $e_{\sigma(k)} < e'_{\sigma'(k)}$ if and only if $OWS_q(e_1, \ldots, e_n) < OWS_q(e'_1, \ldots, e'_n)$.

All these functions satisfy the requirements imposed in Definition 7; all of them are symmetric but *weighted sum* when q is not uniform. [3]

Many other possible choices for $\oplus$ and $\odot$ can be found in the literature of multi-criteria decision making (31). Noticeable examples of such aggregation functions are the *Choquet integral*, which generalizes both the weighted sum and the ordered weighted sum, and its ordinal counterpart, the *Sugeno integral* (32). These aggregation functions are still polynomially computable, which makes the following complexity results applicable when instantiating $\oplus$ and $\odot$ with such functions.

Note that functions such as the purely utilitarian *sum* or *weighted sum* allow for compensation between scores (and lead to majority-like operators), while the egalitarian functions *max* and *lex* do not.

By letting the parameters d, $\oplus$ and $\odot$ vary, several merging operators are obtained; some of them were already known and are thus encoded as specific cases in our framework, while others are new operators. For example, $\triangle^{d_D,max,max}$ is the *basic* merging operator (5), giving $\bigwedge E \wedge IC$ if consistent and IC otherwise. $\triangle^{d_D,max,sum}$ is the *drastic* merging operator which amounts to select the models of IC satisfying the greatest number of belief bases from E. It is equivalent to the drastic majority operator as defined in (9) when working with deductively closed belief bases. $\triangle^{d_D,sum,sum}$ corresponds to the intersection operator of (9). $\triangle^{d_D,WS_q,max}$ corresponds to an operator used in (29) in a group decision context. When singleton belief bases are considered [4] (in this case $\oplus$ is irrelevant) every $\triangle^{d,\oplus,max}$ operator is a $\triangle^{Max}$ operator (2; 13), every $\triangle^{d,\oplus,sum}$ operator is a $\triangle^{\Sigma}$ operator (2; 3; 5), and every $\triangle^{d,\oplus,lex}$ operator is a $\triangle^{GMax}$ operator (5; 13). Still with singleton belief bases, $\triangle^{d_D,\oplus,WS_q}$ is a penalty-based merging operator (where one minimizes the sum of the penalties $q(i)$ attached to the K_i's) (33), and taking $d = d_D$ and $\oplus = WMAX_q$ (defined by $WMAX_q(x_1, \ldots, x_n) = \max_{i=1,\ldots,n} \min(q(i), x_i)$) we get a possibilistic merging operator (6) (the scales used for scores are different but it is easy to show that this difference has no impact, i.e., the induced orderings over interpretations coincide.) Finally, the operators $\triangle^{d_D,sum,\odot}$, with $\odot \in \{sum, WS_q, max, lex\}$ have been proposed in (34) as a compromise between model-based and syntax-based approaches and a way to take into account inconsistent belief bases in the merging process.

We will now illustrate the behaviour of these different operators on an example.

[3] q is uniform when $\forall i, j \in 1, \ldots, n,\ q(i) = q(j)$.

[4] Or when each K_i is replaced by $\{\bigwedge K_i\}$ before merging.

$$
\begin{aligned}
\triangle_{IC}^{d_D,max,max}(E) &\equiv \top. \\
\triangle_{IC}^{d_D,max,sum}(E),\ \triangle_{IC}^{d_D,max,lex}(E),\ \triangle_{IC}^{d_H,max,sum}(E) &\equiv a \wedge b\ . \\
\triangle_{IC}^{d_D,sum,max}(E) &\equiv \neg b\ . \\
\triangle_{IC}^{d_D,sum,sum}(E) &\equiv (\neg a \wedge \neg b) \vee (a \wedge b \wedge c)\ . \\
\triangle_{IC}^{d_D,sum,lex}(E) &\equiv \neg a \wedge \neg b\ . \\
\triangle_{IC}^{d_H,sum,max}(E),\ \triangle_{IC}^{d_H,sum,lex}(E) &\equiv a \wedge \neg b \wedge c. \\
\triangle_{IC}^{d_H,max,max}(E), \triangle_{IC}^{d_H,max,lex}(E) &= (\neg a \wedge b \wedge c) \vee (a \wedge \neg b \wedge c)\ . \\
\triangle_{IC}^{d_H,sum,sum} &\equiv a \wedge c\ .
\end{aligned}
$$

Fig. 1. Result of merging for the operators of Example 2

Example 2 *Consider the following belief profile* $E = \{K_1, K_2, K_3, K_4\}$ *that we want to merge under the integrity constraints* $IC = \top$.

- $K_1 = \{a \wedge b \wedge c, a \Rightarrow \neg b\}$,
- $K_2 = \{a \wedge b\}$,
- $K_3 = \{\neg a \wedge \neg b, \neg b\}$,
- $K_4 = \{a, a \Rightarrow b\}$.

The result of merging E *according to the different operators with* $d \in \{d_D, d_H\}$, $\oplus \in \{max, sum\}$ *and* $\odot \in \{max, sum, lex\}$ *under no constraints (i.e., IC=* $\top$*) is given on Figure 1.*

Table 1 gives an example of computation with the $\triangle^{d_H,sum,lex}$ *operator. In the leftmost column of this table, every interpretation* (x, y, z) *with* $x, y, z \in \{0, 1\}$ *is the one mapping* a *to* x, b *to* y *and* c *to* z. *Each cell except those of the extreme columns gives the (Hamming) distance between the interpretation indexing its row and the formula or the belief base indexing its column. Each cell of the rightmost column contains the vector (ordered in a decreasing way) of distances from the interpretation* ω *indexing the corresponding row and each belief base* K_i $(i = 1, \ldots, 4)$. *As explained before, each such vector can be encoded as an integer using an* OWS_q *function, and such a score can be interpreted as the distance between* ω *and* E. *The main point is that the natural ordering over such scores representing vectors coincides with the leximax one over the corresponding vectors.*

For the example, the result of merging process with $d = d_H, \oplus = sum, \odot = lex$ *is* $Mod(\triangle_{\top}^{d_H,sum,lex}(E)) = \{(1, 0, 1)\}$ *since* $I = (1, 0, 1)$ *is the unique interpretation leading to the minimal vector* 1111 *(corresponding to a minimal distance to E.)*

	$a \wedge b \wedge c$	$a \Rightarrow \neg b$	$a \wedge b$	$\neg a \wedge \neg b$	$\neg b$	a	$a \Rightarrow b$	$\mathbf{K_1}$	$\mathbf{K_2}$	$\mathbf{K_3}$	$\mathbf{K_4}$	**E**
$(0,0,0)$	3	0	2	0	0	1	0	3	2	0	1	3210
$(0,0,1)$	2	0	2	0	0	1	0	2	2	0	1	2210
$(0,1,0)$	2	0	1	1	1	1	0	2	1	2	1	2211
$(0,1,1)$	1	0	1	1	1	1	0	1	1	2	1	2111
$(1,0,0)$	2	0	1	1	0	0	1	2	1	1	1	2111
$(1,0,1)$	1	0	1	1	0	0	1	1	1	1	1	1111
$(1,1,0)$	1	1	0	2	1	0	0	2	0	3	0	3200
$(1,1,1)$	0	1	0	2	1	0	0	1	0	3	0	3100

Table 1
$\triangle^{d_H,sum,lex}$ operator

The wide variety of the results we obtained shows the degree of flexibility achieved by our framework. The example illustrates several aspects of merging operators: the belief base K_1 is not consistent, but it is the only base that gives an information about c, so it can be sensible to take c as true in the result of merging. DA^2 operators can encode merging operators that are syntax-dependent: for example, K_3 is logically equivalent to $\neg a \wedge \neg b$, but replacing K_3 by this formula would lead to different results of the merging operation. Syntax is relevant for DA^2 merging operators since one has to consider that different formulas of a same base are distinct reasons to believe in the same information. Taking syntax into account is important from the point of view of representation of beliefs (or goals). In our framework, unlike with the classical model-based merging operators, the symbol "," can be taken to be a connective that is interpreted differently from "$\wedge$".

We do not consider the case $\oplus = lex$, since this choice induces some specific difficulties as to the second aggregation step. The first one is definitional: what does it mean to aggregate vectors (instead of atomic values) using an OWS_q function, especially when the vectors have different sizes? Several conflicting intuitions may exist. But would the induced operators exhibit the expected behavioural properties of merging? One may argue that, as shown above, it is possible to find out an OWS_q function the total pre-order induced by it coincides with leximax; however, this leads to another problem, namely, a representational problem: how to encode in a faithful way an aggregation function over vectors using some aggregation function over (atomic) values, so that the induced pre-orders coincide? A solution to both problems may come from a systematic study of more general aggregation functions than those used in this paper. This idea is of interest, but has not been considered here. It can be considered as an open question of this paper (however see (35) for a related issue.)

5 Computational Complexity

Let us now turn to the complexity issue. First of all, we can get a general hardness result that holds for *any* merging operator satisfying (IC1) and (IC2); this result, extremely close to a similar BH_2-hardness result for belief revision in (36), gives us a general lower bound of the complexity of inference from merging.

Proposition 1 *For any merging operator Δ satisfying properties (IC1) and (IC2), the complexity of inference from a merged base is BH_2-hard.*

Proof : Let $\langle \varphi, \psi \rangle$ be a pair of propositional formulas; without loss of generality, assume that φ and ψ do not share any propositional symbols. Then with $\langle \varphi, \psi \rangle$ we associate the following instance of INFERENCE-FROM-MERGING: $IC = \top$, $E = \{\{\varphi \vee x\}, \{\varphi \vee \neg x\}\}$, where x is a new symbol (appearing neither in φ nor in ψ), and $\alpha = \varphi \wedge \neg\psi$. Then we have $\Delta(E) \models \alpha$ if and only if φ is satisfiable and ψ is unsatisfiable, that is, if and only if $\langle \varphi, \psi \rangle$ is a positive instance of SAT-UNSAT. Indeed: consider first the case φ is satisfiable; in this case, (IC2) implies that $\Delta(E) \equiv (\varphi \vee x) \wedge (\varphi \vee \neg x) \equiv \varphi$; now, $\Delta(E) \models \varphi \wedge \neg\psi$ if and only if $\varphi \models \varphi \wedge \neg\psi$, which, since φ and ψ do not share any symbol, holds if and only if ψ is unsatisfiable. Consider now the case φ is unsatisfiable. Then $\alpha = \varphi \wedge \neg\psi$ is unsatisfiable, and property (IC1) tells that it cannot be the case that $\Delta(E) \models \alpha$. Therefore, we have $\Delta(E) \models \alpha$ if and only if $\langle \varphi, \psi \rangle$ is a positive instance of SAT-UNSAT.

□

Now, as to finding an upper bound, we obtain a fairly general membership result which states that provided that d, $\oplus$ and $\odot$ can be computed in polynomial time, determining whether a given formula is entailed by the merging of a belief profile is in Δ_2^p; in addition to this, if d, $\oplus$ and $\odot$ are bounded by polynomial functions, then the above problem falls in Θ_2^p. Let us now state this more formally:

Proposition 2 *Let $\triangle^{d,\oplus,\odot}$ be a DA[2] merging operator. Given a belief profile E and two formulas IC and α:*

- *If d, $\oplus$ and $\odot$ are computable in polynomial time, then determining whether $\triangle_{IC}^{d,\oplus,\odot}(E) \models \alpha$ holds is in Δ_2^p.*
- *If d, $\oplus$ and $\odot$ are computable in polynomial time and are polynomially bounded,[5] then determining whether $\triangle_{IC}^{d,\oplus,\odot}(E) \models \alpha$ holds is in Θ_2^p.*

[5] A function $f : \mathbb{N}^n \to \mathbb{N}$ is polynomially bounded if and only if it is bounded by a polynomial function; more formally, when f is a function with a variable number of arguments, such as our aggregation functions, f is polynomially bounded if and

Proof : These results are consequences of the two following lemmata:

Lemma 1 *Let k be an integer; if d, $\oplus$ and $\odot$ are computable in polynomial time, then the problem of determining whether* $\min_{\omega \models IC} d(\omega, E) \leq k$ *given IC, E and k is in* NP*.*

Proof : It is sufficient to consider the following nondeterministic algorithm:

i) guess an interpretation ω and N interpretations $\omega_{i,j}$ ($i = 1, \ldots, n$, $j = 1, \ldots, n_i$) over $Var(E \cup \{IC\})$, where $N = \Sigma_{i=1,\ldots,n} n_i$ is the total number of formulas $\varphi_{i,j}$ in E;
ii) check that $\omega \models IC$ and that $\omega_{i,j} \models \varphi_{i,j}$ for all $i = 1, \ldots, n$ and all $j = 1, \ldots, n_i$;
iii) compute $d(\omega, \omega_{i,j})$ for all $i = 1, \ldots, n$ and all $j = 1, \ldots, n_i$;
iv) compute $d(\omega, K_i)$ for all $i = 1, \ldots, n$;
v) compute $d(\omega, E)$ and check that $d(\omega, E) \leq k$.

This algorithm runs in polynomial time in the size of the input (E, IC, and k represented in binary notation) since d, $\oplus$, $\odot$ are computable in polynomial time.

$\square$

Lemma 2 *If for any $\omega \in \mathcal{W}$ the value of $d(\omega, E)$ is bounded by the value $h(|E| + |IC|)$ (where h is a function with values in $\mathbb{N}$), then* $\min_{\omega \models IC} d(\omega, E)$ *can be computed using $\lceil \log_2 h(|E| + |IC|) \rceil$ calls to an* NP *oracle.*

Proof : $min = \min_{\omega \models IC} d(\omega, E)$ can be computed using binary search on $\{0, \ldots, h(|E| + |IC|)\}$ with at each step a call to an NP oracle to check whether $\min_{\omega \models IC} d(\omega, E) \leq k$ (that is in NP from Lemma 1.) Since a binary search on $\{0, \ldots, h(|E| + |IC|)\}$ needs at most $\lceil \log_2 h(|E| + |IC|) \rceil$ steps, the result follows.

$\square$

- *Point 1. of Proposition 2*

If d, $\oplus$ and $\odot$ are computable in polynomial time, then for every belief profile E and every $\omega \in \mathcal{W}$, the binary representation of $d(\omega, E)$ is bounded by $p(|E| + |IC|)$, where p is a polynomial. Hence, the value of $d(\omega, E)$ is bounded by $2^{p(|E|+|IC|)}$. From Lemma 2, we can conclude that $min = \min_{\omega \models IC} d(\omega, E)$ can be computed using a polynomial number of calls to an NP oracle. Now, let E be a belief profile, IC be a formula, k be an integer and α be a formula, the problem of determining whether there exists a model ω of IC such that $d(\omega, E) = k$ and such that $\omega \not\models \alpha$ is in NP (note the similarity between this proof and the one of Lemma 1):

only if there exists a collection of polynomial functions $\{pol_i \mid i \geq 1\}$ such that $f(x_1, ..., x_n) \leq pol_n(x_1, ..., x_n)$ for every n and for all $x_1, ..., x_n$.

i) guess an interpretation ω and N interpretations $\omega_{i,j}$ ($i = 1, \ldots, n$, $j = 1, \ldots, n_i$) over $Var(E \cup \{IC, \alpha\})$, where $N = \Sigma_{i=1,\ldots,n} n_i$ is the total number of formulas $\varphi_{i,j}$ in E;
ii) check that $\omega \models IC \land \neg\alpha$ and that $\omega_{i,j} \models \varphi_{i,j}$ for all $i = 1, \ldots, n$ and all $j = 1, \ldots, n_i$;
iii) compute $d(\omega, \omega_{i,j})$ for all $i = 1, \ldots, n$ and all $j = 1, \ldots, n_i$;
iv) compute $d(\omega, K_i)$ for all $i = 1, \ldots, n$;
v) compute $d(\omega, E)$ and check that $d(\omega, E) = k$.

So we can show that $\triangle_{IC}^{d,\oplus,\odot}(E) \not\models \alpha$ using first a polynomial number of calls to an NP oracle in order to compute min, and then using an additional call to an NP oracle in order to determine whether there exists a model ω of IC such that $d(\omega, E) = min$ and $\omega \not\models \alpha$. Hence the membership to Δ_2^p for this problem. The fact that Δ_2^p is closed for the complement concludes the proof.

• *Point 2. of Proposition 2*

When d, $\oplus$ and $\odot$ are polynomially bounded, the proof is similar to the one of point 1., but the computation of $\min_{\omega \models IC} d(\omega, E)$ needs only a logarithmic number of steps since h is polynomially bounded, hence the membership to Θ_2^p.

□

As shown by the previous proposition, improving the generality of the model-based merging operators framework through an additional aggregation step does not result in a complexity shift : the decision problem for query entailment is still at the first level of PH.

Importantly, our results rely on the assumption that distances and aggregation functions can be computed in polynomial time. First, it should be remarked that all Δ_2^p membership results would still hold provided that distances and aggregation functions are in $\mathsf{F}\Delta_2^p$. Second, let us discuss the reasonableness of this assumption. On the one hand, all "usual" distance and aggregation functions used in the Knowledge Representation and in the Multicriteria Decision Making communities are consistent with it. On the other hand, there do exist interesting non polynomially-computable distances (and maybe also aggregation functions, although this is less clear). [6]

[6] Here is an example. Consider a set of deterministic events (or actions) E, where the dynamics of each event is described by a STRIPS list, and let us define the distance d_E by $d_E(\omega, \omega') = \min(L_E(\omega, \omega'), L_E(\omega', \omega))$ where $L_E(\omega, \omega')$ is the length of the shortest event sequence (or the shortest plan, if E is a set of actions) leading from ω to ω'. Then, using well-known results about the complexity of propositional STRIPS planning (37) imply that unless $\mathsf{P} = \mathsf{PSPACE}$, d_E is not polynomially computable.

We have also identified the complexity of query entailment from a merged base for the following DA^2 merging operators. Due to some similarity in the proofs of the three following Propositions, their proofs are written in one block. For these three Propositions, when X is a complexity class, X-c means X-complete.

Proposition 3 (complexity results for $d = d_D$)
Given a belief profile E and two formulas IC and α from $PROP_{PS}$, the complexity of $\triangle_{IC}^{d_D,\oplus,\odot}(E) \models^? \alpha$ is reported in the following table.

$\oplus/\odot$	max	sum	lex	WS_q	OWS_q
max	BH_2-c	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Θ_2^p-c
sum	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

Proposition 4 (complexity results for $d = d_H$)
Given a belief profile E and two formulas IC and α from $PROP_{PS}$, the complexity of $\triangle_{IC}^{d_H,\oplus,\odot}(E) \models^? \alpha$ is reported in the following table

$\oplus/\odot$	max	sum	lex	WS_q	OWS_q
max	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
sum	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

Proposition 5 (complexity results for $d = d_{H_q}$)
Given a belief profile E and two formulas IC and α from $PROP_{PS}$, the complexity of $\triangle_{IC}^{d_{H_q},\oplus,\odot}(E) \models^? \alpha$ is reported in the following table.

$\oplus/\odot$	max	sum	lex	WS_q	OWS_q
max	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
sum	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

Proof :

- Membership: All the membership results (for Propositions 3, 4 and 5) are direct consequences of Proposition 2, except to what concerns the basic merging operator ($d = d_D, \oplus = \odot = max$) and the membership-to-$\Theta_2^p$ results reported in Proposition 3 in the situation one of the two aggregation functions is a OWS_q function while the other one is max. Indeed, in all the remaining cases, all the distances and aggregation functions considered in the three tables can be computed in polynomial time. In addition, the distances d_D and d_H and the aggregation functions sum and max are polynomially bounded. As a consequence, we obtain immediately the membership to Θ_2^p of the inference problem with $\triangle^{d,\oplus,\odot}$ with $d \in \{d_D, d_H\}$, $\oplus \in \{sum, max\}$, $\odot \in \{sum, max\}$.

Now, focusing on the situation $d = d_D$, let us consider the case $\oplus$ is max and $\odot$ is a OWS_q function. For every interpretation $\omega \in \mathcal{W}$, let us note $k_E(\omega)$ the number of belief bases K_i ($i \in 1, \ldots, n$) from E such that $\omega \models K_i$ holds. Then we have $d(\omega, E) = \sum_{i=1}^{n-k_E(\omega)} q_i$. For any $j \in 0, \ldots, n$, it is easy to determine in nondeterministic polynomial time whether there exists a model ω of IC such that $d(\omega, E) \leq \sum_{i=1}^{n-j} q_i$. Now, since $d(\omega, E)$ can only take at most $n + 1$ different values, its minimal value min over $Mod(IC)$ can be computed through binary search using at most $\lceil \log_2 n \rceil$ calls to an NP oracle which implements the nondeterministic algorithm above (starting with $j = 0$.) Once min has been computed, a final call to an NP oracle can be used to determine whether there exists a model ω of IC such that $d(\omega, E) = min$ and $\omega \not\models \alpha$. The fact that Θ_2^p is closed for the complement concludes the proof. The case $\oplus$ is a OWS_q function and $\odot$ is max can be handled in a similar way. The main difference is that $d(\omega, E)$ can only take at most $max_{i=1,\ldots,n} card(K_i)$ different values.
Finally, as to the basic merging operator, determining whether a formula α is a logical consequence of the merged base E given IC can be achieved using the following algorithm:

if $sat(E \cup \{IC\})$
then return($unsat(E \cup \{IC, \neg\alpha\})$)
else return($unsat(\{IC, \neg\alpha\})$).

Since only one satisfiability test (sat) and one unsatisfiability test ($unsat$) are required, the decision problem is in BH_2.

- Hardness:
  - Proposition 3: The Θ_2^p-hardness results are direct consequences of hardness results for cardinality-maximizing base revision $\circ_C$ (Theorem 5.14 from (36)) since we have $\Delta_{IC}^{d_D,sum,max}(\{\{\varphi_1, \ldots, \varphi_n\}\}) \equiv \Delta_{IC}^{d_D,sum,sum}(\{\{\varphi_1, \ldots, \varphi_n\}\}) \equiv \Delta_{IC}^{d_D,max,sum}(\{\{\varphi_1\}, \ldots, \{\varphi_n\}\}) \equiv \Delta_{IC}^{d_D,max,lex}(\{\{\varphi_1\}, \ldots, \{\varphi_n\}\}) \equiv \{\varphi_1, \ldots, \varphi_n\} \circ_C IC$.

    Indeed, whenever a single aggregation step is done and the drastic distance d_D is considered, lex gives the same ordering as sum. Since lex is a

specific OWS_q function, the corresponding Θ_2^p-hardness results still hold in the case $\oplus$ is a OWS_q function and $\odot = max$, as well as in the case $\oplus = max$ and $\odot$ is a OWS_q function.
As to the case where $\oplus$ is a OWS_q function and $\odot = sum$, the Δ_2^p-hardness result can be established by considering the following polynomial reduction from the Δ_2^p-complete problem MAX-SAT-ASG$_{odd}$ (16). MAX-SAT-ASG$_{odd}$ is the following decision problem:

> **Input:** Σ, a propositional formula such that $Var(\Sigma) = \{x_1, \ldots, x_n\}$.
> **Question:** Is the greatest model ω of Σ (over $Var(\Sigma)$) with respect to the lexicographic ordering $\preccurlyeq$ induced by $x_1 < x_2 < \ldots < x_n$ such that $\omega(x_n) = 1$?

To every formula Σ such that $Var(\Sigma) = \{x_1, \ldots, x_n\}$, we associate in polynomial time the tuple $M(\Sigma) = \langle E, IC, \alpha\rangle$, where $E = \{K_i \mid i \in 1, \ldots, n\}$, $IC = \Sigma$, $\alpha = x_n$ and for each $i \in 1 \ldots n$, $K_i = \{\bigwedge_{k=1}^{n+2-j} x_i \mid j \in 1 \ldots n+2-i\}$. Accordingly, each K_i contains $n+2-i$ formulas that are syntactically distinct but all equivalent to x_i. We consider now the OWS_q function $\oplus$ induced by q such that $q(1) = 1$ and for every $j > 1$, $q(j) = 2^{i-2}$. By construction, for any $\omega \in \mathcal{W}$ and any $i \in 1, \ldots, n$, we have $d_D(\omega, K_i) = 0$ if $\omega \models x_i$ and $d_D(\omega, K_i) = 2^{n-i+1}$ if $\omega \not\models x_i$. Accordingly, $d_D(\omega, E) = \sum_{i=1}^{n} d_D(\omega, K_i) = \sum_{i=1,\ldots,n \mid \omega \not\models x_i} 2^{n-i+1}$. We immediately get that ω is a model of IC that minimizes $d_D(\omega, E)$ if and only if ω is the (unique) greatest model of Σ w.r.t. $\preccurlyeq$, which leads easily to the result.
The Δ_2^p-hardness result in the case $\oplus = sum$ and $\odot = lex$ can be easily derived by taking advantage of the Δ_2^p-hardness result in the case each K_i is a singleton reduced to a conjunction of atoms (hence $\oplus$ is irrelevant), $\odot$ is lex and the Hamming distance d_H is considered (the proof is given in the following.) Indeed, to each $K_i = \bigwedge_{j=1}^{n_i} x_{i,j}$, we can associate the *set* of formulas $K_i = \{x_{i,j} \mid j \in 1,\ldots,n_i\}$ and for every interpretation $\omega \in \mathcal{W}$, we have $d_H(\omega, K_i) = \sum_{j=1}^{n_i} d_D(\omega, x_{i,j})$. Roughly, the Hamming distance is encoded here through a first aggregation step (using $\oplus = sum$) based on the drastic distance. Since sum is a specific WS_q function and lex is a specific OWS_q function, this hardness result can be extended to the rest of the table, except for the case ($\oplus$ is a WS_q function and $\odot = max$ or $\odot = sum$) or $\odot$ is a WS_q function.
As to these cases, the Δ_2^p-hardness of linear base revision $\circ_L$ (Theorem 5.9 from (36)) can be used to obtain the desired result. Indeed, it is sufficient to consider belief bases K_i reduced to singletons (hence the first aggregation step using $\oplus$ is irrelevant) or similarly a belief profile E consisting of a singleton (so that $\odot$ is irrelevant) since we have $\triangle_{IC}^{d_D,\oplus,\odot}(\{K_1, \ldots, K_n\}) \equiv \{K_1, \ldots, K_n\} \circ_L IC$, where $\odot$ is the weighted sum induced by q such that $q(i) = 2^{n-i}$, and each K_i is viewed as the unique formula it contains. Here, the preference ordering over $\{K_1, \ldots, K_n\}$ is such that $K_1 < K_2 < \ldots < K_n$.
Finally, as to the basic merging operator, the BH_2-hardness result is a direct consequence of Propositions 1 and 6.

– Proposition 4: The Θ_2^p-hardness results still hold in the situation E contains only one belief base K, and K itself contains only one formula that is a conjunction of atoms. This merely shows that our hardness result is independent from the aggregation functions $\oplus$ and $\odot$ under consideration (since they are irrelevant whenever E and K are singletons) but is a consequence of the distance that is used (Hamming). Indeed, in this restricted case, $\triangle_{IC}^{d_H,\oplus,\odot}(\{K\})$ is equivalent to $K \circ_D IC$ where $\circ_D$ is Dalal's revision operator (19). The fact that the inference problem from $K \circ_D IC$ is Θ_2^p-hard (even in the restricted case where K is a conjunction of atoms) concludes the proof (see Theorem 6.9 from (15)).

We now show that the Δ_2^p-hardness results hold in the restricted case each K_i is a singleton, reduced to a conjunction of literals (which means that the $\oplus$-aggregation step is irrelevant), whenever $\odot = lex$. Since lex can be viewed as a specific OWS_q, the hardness result holds for OWS_q functions as well. We consider the following polynomial reduction M from MAX-SAT-ASG$_{odd}$ to the inference problem from a merged base. Let Σ be a propositional formula such that $Var(\Sigma) = \{x_1, \ldots, x_n\}$. Let $M(\Sigma) =$

$$\langle E = \{K_i = \{x_i \wedge \bigwedge_{j=i+1}^{2n-i+1} new_j\} \mid i \in 1, \ldots, n\}, IC = \Sigma \wedge \bigwedge_{j=2}^{2n} \neg new_j, \alpha = x_n\rangle$$

where each new_j ($j \in 2, \ldots, 2n$) is a new variable (not occurring in Σ.) Now, for every model ω of IC and for every $i \in 1, \ldots, n-1$, we have

$$d_H(\omega, K_i = \{x_i \wedge \bigwedge_{j=i+1}^{2n-i+1} new_j\}) > d_H(\omega, K_{i+1} = \{x_{i+1} \wedge \bigwedge_{j=i+2}^{2n-i+2} new_j\}).$$

This shows that the vectors L_ω^E obtained by sorting the set $\{d_H(\omega, K_i) \mid i \in 1, \ldots, n\}$ in decreasing lexicographic order are always sorted in the same way (independently of ω): the first element is $d_H(\omega, K_1)$, the second one is $d_H(\omega, K_2)$, etc. Furthermore, whenever a model ω_1 of IC is strictly smaller than a model ω_2 of IC with respect to the lexicographic ordering $\preccurlyeq$ induced by $x_1 < x_2 < \ldots < x_n$, then $L_{\omega_1}^E$ is strictly greater than $L_{\omega_2}^E$ (with respect to the lexicographic ordering over vectors of integers.) Since the models of IC are totally ordered with respect to $\preccurlyeq$, exactly one model of IC is minimal with respect to the preference ordering induced by E: this is the model of IC that is maximal with respect to $\preccurlyeq$. Accordingly, x_n is true in this model if and only if $\triangle_{IC}^{d_H,\oplus,lex}(E) \models \alpha$ holds. This concludes the proof.

Finally, we show that the remaining Δ_2^p-hardness results hold in the case one of the aggregation function is a WS_q function, i.e., whenever each K_i is a singleton (even reduced to an atom) or E is a singleton. In the first case, this merely shows that our hardness result is independent from the aggregation function $\oplus$ under consideration but holds in the case $\odot$ is a WS_q function and $d = d_H$ is the Hamming distance. Let us consider the following polynomial reduction M from MAX-SAT-ASG$_{odd}$ to the inference problem from a merged base. Let Σ be a propositional formula such that $Var(\Sigma) = \{x_1, \ldots, x_n\}$. Let

$$M(\Sigma) = \langle E = \{K_i = \{x_i\} \mid i \in 1, \ldots, n\}, IC = \Sigma, \alpha = x_n\rangle$$

and the aggregation function $\odot$ is the weighted sum operator induced by $q(i) = 2^{n-i}$. Accordingly, for any interpretation $\omega \in \mathcal{W}$, we have $d(\omega, E) = \sum_{i=1}^{n} q(i) d_H(\omega, K_i)$. By construction, for any interpretations ω_1, $\omega_2 \in \mathcal{W}$, we have $d_H(\omega_1, E) \leq d_H(\omega_2, E)$ if and only if $\omega_2 \preccurlyeq \omega_1$ where $\preccurlyeq$ is the lexicographic ordering induced by $x_1 < x_2 < \ldots < x_n$. Accordingly, the greatest model ω of Σ with respect to $\preccurlyeq$ is the unique model of $\triangle_{IC}^{d_H,\oplus,\odot}(E)$. As a consequence, the greatest model ω of Σ with respect to $\preccurlyeq$ is such that $\omega(x_n) = 1$ if and only if $\triangle_{IC}^{d_H,\oplus,\odot}(E) \models \alpha$. This concludes the proof.

- Proposition 5: We show that Δ_2^p-hardness holds in the very restricted case E contains only one belief base K, and K itself contains only one formula that is a conjunction of atoms. This merely shows that our hardness result is independent from the aggregation functions $\oplus$ and $\odot$ under consideration (since they are irrelevant whenever E and K are singletons) but is a consequence of the family of distances that is used (weighted Hamming). Let us consider the following polynomial reduction M from MAX-SAT-ASG$_{odd}$ to the inference problem from a merged base. Let Σ be a propositional formula such that $Var(\Sigma) = \{x_1, \ldots, x_n\}$. Let

$$M(\Sigma) = \langle E = \{\{\bigwedge_{i=1}^{n} x_i\}\}, IC = \Sigma, \alpha = x_n\rangle$$

and the weighted Hamming distance d_{H_q} induced by q such that $\forall i \in 1, \ldots, n$, $q(x_i) = 2^{n-i}$. By construction, for any interpretations ω_1, $\omega_2 \in \mathcal{W}$, we have $d_{H_q}(\omega_1, \bigwedge_{i=1}^{n} x_i) \leq d_{H_q}(\omega_2, \bigwedge_{i=1}^{n} x_i)$ if and only if $\omega_2 \preccurlyeq \omega_1$ where $\preccurlyeq$ is the lexicographic ordering induced by $x_1 < x_2 < \ldots < x_n$. Accordingly, the greatest model ω of Σ with respect to $\preccurlyeq$ is the unique model of $\triangle_{IC}^{d_{H_q},\oplus,\odot}(E)$. As a consequence, the greatest model ω of Σ with respect to $\preccurlyeq$ is such that $\omega(x_n) = 1$ if and only if $\triangle_{IC}^{d_{H_q},\oplus,\odot}(E) \models \alpha$. This concludes the proof.

□

Looking at the tables above, we can observe that the choice of the distance d has a great influence on the complexity results. Thus, whenever $d = d_H$ or $d = d_{H_q}$, the complexity results for inference from a merged base coincide whenever $\oplus$ (or $\odot$) is a WS_q function or a OWS_q function. This is no longer the case when $d = d_D$ is considered.

Together with Proposition 2, the complexity of many model-based merging operators already pointed out in the literature are derived as a by-product of the previous complexity results. To the best of our knowledge, the complexity of such operators has not been identified up to now[7], hence this is an additional contribution of this work. We can also note that, while the complexity of our DA^2 operators is not very high (first level of PH, at most), finding out significant tractable restrictions seems a hard task since intractability is still the case in many restricted situations (see the proofs.) Finally, our results show that some syntax-based merging operators (the ones based on set inclusion instead of cardinality and "located" at the second level of PH) cannot be encoded in polynomial time as DA^2 operators (unless PH collapses.)

6 Logical properties

Let first see what are the logical properties of DA^2 merging operators in the general case.

Proposition 6 *Let d be any distance, and let f and g be two aggregation functions. $\triangle^{d,\oplus,\odot}$ satisfies (IC0), (IC1), (IC2), (IC7), (IC8). The other postulates are not satisfied in the general case.*

Proof :

(IC0) By definition $Mod(\triangle_{IC}(E)) \subseteq Mod(IC)$.

(IC1) $\oplus$ and $\odot$ are functions with values in $I\!N$, so if $Mod(IC) \neq \emptyset$, there is always a minimal model ω of IC such that for every model ω' of IC $d(\omega, E) \leq d(\omega', E)$. So $\omega \models \triangle_{IC}(E)$ and $\triangle_{IC}(E) \not\models \bot$.

(IC2) By assumption, $\bigwedge E$ is consistent, i.e., there exists ω such that $\omega \models (\varphi_{11} \wedge \ldots \wedge \varphi_{1n_1}) \wedge \ldots \wedge (\varphi_{n1} \wedge \ldots \wedge \varphi_{nn_n})$. By definition of the distance, $d(\omega, \varphi) = 0$ if $\omega \models \varphi$, so by (minimality) of $\oplus$ we get $\oplus(d(\omega, \varphi_{i1}), \ldots, d(\omega, \varphi_{in_i})) = d(\omega, K_i) = 0$ if and only if $\omega \models \varphi_{i1} \wedge \ldots \wedge \varphi_{in_i}$. By (minimality) of $\odot$ we have that $\odot(d(\omega, K_1), \ldots, d(\omega,\ K_n)) = d(\omega, E) = 0$ if and only if $\omega \models K_1 \wedge \ldots \wedge K_n$. So $\omega \models \triangle_{IC}(E)$ if and only if $\omega \models \bigwedge E \wedge IC$.

(IC7) Suppose $\omega \models \triangle_{IC_1}(E) \wedge IC_2$. For any $\omega' \models IC_1$, we have $d(\omega, E) \leq d(\omega', E)$. Hence $\omega' \models IC_1 \wedge IC_2$, $d(\omega, E) \leq d(\omega', E)$. Subsequently $\omega \models \triangle_{IC_1 \wedge IC_2}(E)$.

(IC8) Suppose that $\triangle_{IC_1}(E) \wedge IC_2$ is consistent. Then there exists a model ω' of $\triangle_{IC_1}(E) \wedge IC_2$. Consider a model ω of $\triangle_{IC_1 \wedge IC_2}(E)$ and suppose that $\omega \not\models \triangle_{IC_1}(E)$. We have $d(\omega', E) < d(\omega, E)$, and since $\omega' \models IC_1 \wedge IC_2$, we have $\omega \notin \min(Mod(IC_1 \wedge IC_2), \leq_E^{d,\oplus,\odot})$, hence $\omega \not\models \triangle_{IC_1 \wedge IC_2}(E)$. Contradiction.

[7] However, $(\triangle_{IC}^{d_H,sum,sum}(E) \overset{?}{\models} \alpha) \in \Delta_2^p$ can be recovered from a complexity results given in (38), page 151.

$\square$

Clearly enough, it is not the case that every DA^2 merging operator is an IC merging operator (not satisfying some postulates is motivated by the need to give some importance to the syntax in order to take inconsistent belief bases into account.)

Concerning the operators examined in the previous section, we have identified the following properties:

Proposition 7 $\triangle^{d,\oplus,\odot}$ *satisfies the logical properties stated in Tables 2 and 3. Since all these operators are already known to satisfy (IC0), (IC1), (IC2), (IC7) and (IC8) (cf. Proposition 6), we refrain from repeating such postulates here. For the sake of readability, postulate (ICi) is noted i and M (resp. A) stands for (Maj) (resp. (Arb)).*

<table>
<tr><th>$\oplus/\odot$</th><th>max</th><th>sum</th><th>lex</th><th>WS_q</th><th>OWS_q</th></tr>
<tr><td>max</td><td>3,4,5,A</td><td colspan="2">3,4,5,6,M,A</td><td>5,6,M</td><td>3,4</td></tr>
<tr><td>sum</td><td>5,A</td><td>5,6,M</td><td>5,6,A</td><td>5,6,M</td><td></td></tr>
<tr><td>$WS_q - OWS_q$</td><td>5,A</td><td>5,6,M</td><td>5,6,A</td><td>5,6,M</td><td></td></tr>
</table>

Table 2
Logical properties ($d = d_D$)

$\oplus/\odot$	max	sum	lex	WS_q	OWS_q
max	5,A	5,6,M	5,6,A	5,6,M	
sum	5,A	5,6,M	5,6,A	5,6,M	
$WS_q - OWS_q$	5,A	5,6,M	5,6,A	5,6,M	

Table 3
Logical properties ($d = d_H$ or $d = d_{H_q}$)

Proof :

(IC3) Most operators of the table do not satisfy (IC3). For the operators with $\oplus = WS_q$, this is because (IC3) refers to the equivalence of belief profiles, and the definition of this equivalence does not take weights into account. A counter-example for operators with $d = d_D$ is $K_1 = \{a, b\}$, $K_2 = \{a \wedge b\}$, $K_3 = \{\neg b\}$ giving $\triangle_\top(\{K_1, K_2\}) \not\equiv \triangle_\top(\{K_1, K_3\})$. A counter-example for operators with $d = d_H$ is $K_1 = \{a, b\}$, $K_2 = \{a, b, b\}$, $K_3 = \{\neg b\}$. Nevertheless (IC3) holds for $d = d_D$, $\oplus = max$, and $\odot \in \{max, sum, lex, OWS_q\}$. It is because each $d(\omega, E)$ is a vector of 0 and 1 (0 is set whenever $\omega \models K_i$ and 1 otherwise.) It is not the case for $d = d_D$, $\oplus = max$, $\odot = WS_q$, since in this case the result is sensible to permutations (because of the weights.)

(IC4) For most operators of the table, (IC4) is not satisfied, since those operators are sensible to the syntax of the base (in particular to the number of formulas.) Let us take as counter-example $K_1 = \{a, b, a \wedge b\}$ and $K_2 = \{\neg a\}$. Nevertheless (IC4) holds for $d = d_D, \oplus = max, \odot \in \{max, sum, lex, OWS_q\}$. Since if $K_1 \wedge K_2 \not\models \bot$, (IC4) holds trivially by (IC2), and if $K_1 \wedge K_2 \models \bot$, then if $\omega \models K_1$, then $d(\omega, \{K_1, K_2\}) = \odot(0, 1)$ and if $\omega \models K_2$, then $d(\omega, \{K_1, K_2\}) = \odot(1, 0)$. It is then sufficient to remark that every $\odot \in \{max, sum, lex,\ OWS_q\}$ is a symmetrical operator, so $\odot(0, 1) = \odot(1, 0)$.

(IC5) To show that the operators satisfy (IC5), it is enough to show that the following property holds: if $d(\omega, E_1) \leq d(\omega', E_1)$ and $d(\omega, E_2) \leq d(\omega', E_2)$, then $d(\omega, E_1 \sqcup E_2) \leq d(\omega', E_1 \sqcup E_2)$. This property depends only on $\odot$ and it is satisfied for $\odot \in \{max, sum, lex, WS_q\}$.

(IC6) To show that the operators satisfy (IC6), it is enough to show that the following property holds: if $d(\omega, E_1) < d(\omega', E_1)$ and $d(\omega, E_2) \leq d(\omega', E_2)$, then $d(\omega, E_1 \sqcup E_2) < d(\omega', E_1 \sqcup E_2)$. This property depends only on $\odot$ and it is satisfied for $\odot \in \{sum, lex, WS_q\}$.

(Maj) Showing that all operators with $\odot \in \{sum,\ WS_q\}$ satisfy (Maj) is easy from the properties of sum. It is also easy to show that operators with $\odot \in \{max, lex\}$ do not satisfy (Maj) since one can find a counter-example where the repetition of one base does not change the result [8] . Consider the following counter-examples: ($E_1 = \{K_1\} = \{\{a, b\}\}$ and $E_2 = \{K_2\} = \{\{\neg a, \neg b\}\}$), or ($E_1 = \{K_1\} = \{\{a \wedge b\}\}$ and $E_2 = \{K_2\} = \{\{\neg a \wedge \neg b\}\}$).

(Arb) It is easy to show that (Arb) holds for all operators with $\odot = max$ since the stronger following property holds: if $\triangle_{IC_1}(K_1) \equiv \triangle_{IC_2}(K_2)$, then $\triangle_{IC_1 \vee IC_2}(\{K_1, K_2\}) \equiv \triangle_{IC_1}(K_1)$.

To show that (Arb) holds for $\odot = lex$ operators, assume that $\triangle_{IC_1}(K_1) \equiv \triangle_{IC_2}(K_2)$, that is there exists a model ω of $IC_1 \wedge IC_2$ such that for every model ω' of IC_1, $d(\omega, K_1) \leq d(\omega', K_1)$ and for every model ω'' of IC_2, $d(\omega, K_2) \leq d(\omega'', K_2)$. W.l.o.g let us suppose that $d(\omega, K_1) \leq d(\omega, K_2)$. To show that $\triangle_{IC_1 \vee IC_2}(\{K_1, K_2\}) \equiv \triangle_{IC_1}(K_1)$, we show that if $\omega' \models IC_1 \vee IC_2$ and $\omega' \not\models \triangle_{IC_1}(K_1)$, then $\omega' \not\models \triangle_{IC_1 \vee IC_2}(\{K_1, K_2\})$. Consider the following three cases :

1. $\omega' \models IC_1 \wedge IC_2$. Then we have $d(\omega', K_1) > d(\omega, K_1)$ and $d(\omega', K_2) > d(\omega, K_2)$. As a consequence $d(\omega', \{K_1, K_2\}) > d(\omega, \{K_1, K_2\})$, hence $\omega' \not\models \triangle_{IC_1 \vee IC_2}(\{K_1, K_2\})$.
2. $\omega' \models IC_2 \wedge \neg IC_1$. Since $\omega' \models IC_2$ we know that $d(\omega', K_2) > d(\omega, K_2)$, by transitivity $d(\omega', K_2) > d(\omega, K_1)$. Then we have $d(\omega', \{K_1, K_2\}) > d(\omega, \{K_1, K_2\})$, hence $\omega' \not\models \triangle_{IC_1 \vee IC_2}(\{K_1, K_2\})$.
3. $\omega' \models IC_1 \wedge \neg IC_2$. Suppose that $\omega' \models \triangle_{IC_1 \vee IC_2}(\{K_1, K_2\})$. This implies that $d(\omega', \{K_1, K_2\}) \leq d(\omega, \{K_1, K_2\})$. This requires one of the following cases to hold (recall that we assume $d(\omega, K_1) \leq d(\omega, K_2)$):
   i. $d(\omega', K_1) < d(\omega, K_1)$ and $d(\omega', K_2) < d(\omega, K_1)$.

[8] Except for $d = d_D$, $\oplus = max$, $\odot = lex$, since in this case the lex operator induces the same ordering as the one induced by sum.

ii. $d(\omega', K_1) = d(\omega, K_1)$ and $d(\omega', K_2) \leq d(\omega, K_2)$.
iii. $d(\omega', K_2) = d(\omega, K_1)$ and $d(\omega', K_1) \leq d(\omega, K_2)$.

The first two cases are not possible since, as $\omega' \models IC_1$ and $\omega' \not\models \triangle_{IC_1}(K_1)$, we have $d(\omega', K_1) > d(\omega, K_1)$. So let us consider the last case and note that we have $d(\omega', K_1) \leq d(\omega, K_2)$ and $d(\omega', K_2) = d(\omega, K_1) \leq d(\omega, K_2)$. (Arb) requires that for every model ω'' of $IC_2 \wedge \neg IC_1$, $d(\omega'', \{K_1, K_2\}) = d(\omega', \{K_1, K_2\})$. So for any $\omega'' \models IC_2 \wedge \neg IC_1$ $d(\omega'', K_2) \leq d(\omega, K_2)$, hence $\omega'' \models \triangle_{IC_2}(K_2)$. But, by hypothesis, $\triangle_{IC_1}(K_1) \equiv \triangle_{IC_2}(K_2)$, hence $\omega'' \models IC_2$. Contradiction.

To show that operators with $\odot \in \{sum, WS_q\}$ do not satisfy (Arb), consider the following counter-example: $K_1 = \{a \wedge b\}$, $K_2 = \{\neg a \wedge \neg b\}$, $IC_1 = \neg(a \wedge b)$ and $IC_2 = a \wedge b$.

□

The tables above show that our DA^2 operators exhibit different properties. We remark that only $\triangle^{d_D, max, sum}$ satisfies all listed properties. Failing to satisfy (IC3) (*irrelevance to the syntax*) in many cases is not surprising, since we want to allow our operators to take syntax into account. (IC4) imposes that, when merging two belief bases, if the result is consistent with one belief base, it has to be consistent with the other one – such fairness postulate is not expected when working with nonsymmetric operators (so, unsurprisingly, it is not satisfied for $\odot = WS_q$.) This postulate is not satisfied by any operator for which d is Hamming distance since cardinalities of the belief bases have an influence on $\oplus$, and more generally, it is hardly satisfiable when working with syntax-dependent operators. (IC5) and (IC6) correspond to Pareto dominance in social choice theory and are really important; so it is worth noting that almost all operators satisfy them (only operators for which $\odot = max$ or OWS_q do not satisfy (IC6).) As shown before, OWS_q gathers many aggregation functions; not surprisingly, the price to be paid is the lack of many logical properties in the general case.

We saw through the previous results that DA^2 merging operators do not (and aim not at) satisfy all IC merging operators properties. Hence this is natural to look for additional requirements under which all those properties would be satisfied.

Let us first define some natural additional properties on aggregation functions:

1) If $\varphi_1 \wedge \ldots \wedge \varphi_n$ is consistent,
then $\oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)) = \oplus(d(\omega, \varphi_1 \wedge \ldots \wedge \varphi_n))$.[9] *(and)*

[9] Since $\oplus$ is an aggregation function, we have $\oplus(d(\omega, \varphi_1 \wedge \ldots \wedge \varphi_n)) = d(\omega, \varphi_1 \wedge$

2) For any permutation σ, $\oplus(x_1, \ldots, x_n) = \oplus(\sigma(x_1, \ldots, x_n))$ *(symmetry)*

3) If $\oplus(x_1, \ldots, x_n) \leq \oplus(y_1, \ldots, y_n)$, then $\oplus(x_1, \ldots, x_n, z) \leq \oplus(y_1, \ldots, y_n, z)$. *(composition)*

4) If $\oplus(x_1, \ldots, x_n, z) \leq \oplus(y_1, \ldots, y_n, z)$, then $\oplus(x_1, \ldots, x_n) \leq \oplus(y_1, \ldots, y_n)$. *(decomposition)*

We have obtained the following representation theorem for DA^2 merging operators:

Proposition 8 *A DA^2 merging operator $\triangle^{d,\oplus,\odot}$ satisfies (IC0)-(IC8) if and only if the function $\oplus$ satisfies (and), and the function $\odot$ satisfies (symmetry), (composition) and (decomposition).*

Proof :

(If) We know that (IC0), (IC1), (IC2), (IC7) and (IC8) are directly satisfied (*cf.* Proposition 6). Let us consider the other properties. Let $E_1 = \{K_1, \ldots, K_n\}$ and $E_2 = \{K'_1, \ldots K'_n\}$.

(IC3) Assume that $E_1 \equiv E_2$. Hence we can find a permutation σ such that for every $i \in 1, \ldots, n$, $K_{\sigma(i)} \equiv K'_i$. Now, since $\oplus$ satisfies (and) and is non-decreasing in each argument, we have $d(\omega, K_{\sigma(i)}) = d(\omega, K'_i)$, so, as $\odot$ satisfies (symmetry) one gets $d(\omega, E_1) = \odot(d(\omega, K'_1), \ldots d(\omega, K'_n)) = d(\omega, E_2)$. Consequently $\triangle_{IC}(E_1) \equiv \triangle_{IC}(E_2)$. The result for $IC_1 \equiv IC_2$ is obvious from the definition of the operators.

(IC4) Suppose that $\triangle_{IC}(\{K_1, K_2\}) \wedge K_1 \not\models \bot$ and that $\triangle_{IC}(\{K_1, K_2\}) \wedge K_2 \models \bot$. As a consequence, we have $\min_{\omega \models K_1} \odot(d(\omega, K_1), d(\omega, K_2)) < \min_{\omega \models K_2} \odot(d(\omega, K_1), d(\omega, K_2))$. Since $\oplus$ satisfies (and), this is equivalent to $\min_{\omega \models K_1} \odot(0, d(\omega, K_2)) < \min_{\omega \models K_2} \odot(d(\omega, K_1), 0)$. Then by (symmetry), this is equivalent to $\min_{\omega \models K_1} \odot(d(\omega, K_2), 0) < \min_{\omega \models K_2} \odot(d(\omega, K_1), 0)$. Hence, since $\odot$ is non-decreasing in each argument, we get $\min_{\omega \models K_1} d(\omega, K_2) < \min_{\omega \models K_2} d(\omega, K_1)$. Now, let us take $K'_{j=1,2} = \bigwedge_{\varphi_i \in K_j} \varphi_i$. Since $\oplus$ satisfies (and), we have $d(\omega, K_j) = d(\omega, K'_j)$ for every interpretation ω. So we get $\min_{\omega \models K'_1} d(\omega, K'_2) < \min_{\omega \models K'_2} d(\omega, K'_1)$. Now, by definition of the distance $d(\omega, \omega') = d(\omega', \omega)$ for every pair of interpretations ω, ω'; from the definition of $d(\omega, \varphi)$, we have $\min_{\omega \models \varphi} d(\omega, \varphi') = \min_{\omega \models \varphi'} d(\omega, \varphi)$ for every pair of formulas φ, φ'. Since $\oplus$ is non-decreasing in each argument, we obtain $\min_{\omega \models \varphi} \oplus(d(\omega, \varphi')) = \min_{\omega \models \varphi'} \oplus(d(\omega, \varphi))$. But, taking $\varphi = K'_1$ and $\varphi' = K'_2$, this contradicts $\min_{\omega \models K'_1} d(\omega, K'_2) < \min_{\omega \models K'_2} d(\omega, K'_1)$.

(IC5) Consider $E_1 = \{K_1, \ldots K_n\}$ and $E_2 = \{K'_1, \ldots K'_n\}$. Suppose that ω is a model of $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$. The, for every model ω' of IC we have both:

$$\odot(d(\omega, K_1), \ldots, d(\omega, K_n)) \leq \odot(d(\omega', K_1), \ldots, d(\omega', K_n)), \text{ and}$$

$\ldots \wedge \varphi_n)$.

$$\odot(d(\omega, K'_1), \ldots, d(\omega, K'_{n'})) \leq \odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'})).$$

Since we have $\odot(d(\omega, K_1), \ldots, d(\omega, K_n)) \leq \odot(d(\omega', K_1), \ldots, d(\omega', K_n))$, using (composition) several times we obtain that:

$$\odot\,(d(\omega, K_1), \ldots, d(\omega, K_n), d(\omega, K'_1), \ldots, d(\omega, K'_{n'})) \\ \leq \odot(d(\omega', K_1), \ldots, d(\omega', K_n), d(\omega, K'_1), \ldots, d(\omega, K'_{n'})). \quad (1)$$

Similarly, since $\odot(d(\omega, K'_1), \ldots, d(\omega, K'_{n'})) \leq \odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'}))$, using (composition) several times gives:

$$\odot\,(d(\omega, K'_1), \ldots, d(\omega, K'_{n'}), d(\omega', K_1), \ldots, d(\omega', K_n)) \\ \leq \odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'}), d(\omega', K_1), \ldots, d(\omega', K_n)). \quad (2)$$

By (symmetry), we have that:

$$\odot\,(d(\omega, K'_1), \ldots, d(\omega, K'_{n'}), d(\omega', K_1), \ldots, d(\omega', K_n)) \\ = \odot(d(\omega', K_1), \ldots, d(\omega', K_n), d(\omega, K'_1), \ldots, d(\omega, K'_{n'})). \quad (3)$$

By transitivity, using (1), (2) and (3), we have for every model ω' of IC:

$$\odot\,(d(\omega, K_1), \ldots, d(\omega, K_n), d(\omega, K'_1), \ldots, d(\omega, K'_{n'})) \\ \leq \odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'}), d(\omega', K_1), \ldots, d(\omega', K_n)).$$

This exactly means that $\omega \models \triangle_{IC}(E_1 \sqcup E_2)$.

(IC6) Suppose that $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2) \not\models \perp$ and that $\triangle_{IC}(E_1 \sqcup E_2) \not\models \triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$. There exists ω such that $\omega \models \triangle_{IC}(E_1 \sqcup E_2)$ and $\omega \not\models \triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$. Let us assume w.l.o.g. that $\omega \not\models \triangle_{IC}(E_1)$. Since $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2) \not\models \perp$, let us consider any $\omega' \models \triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$. Since $\omega' \models \triangle_{IC}(E_1)$ and $\omega \not\models \triangle_{IC}(E_1)$, we obtain:

$$\odot(d(\omega', K_1), \ldots, d(\omega', K_n)) < \odot(d(\omega, K_1), \ldots, d(\omega, K_n)).$$

Since $\omega' \models \triangle_{IC}(E_2)$, we have:

$$\odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'})) \leq \odot(d(\omega, K'_1), \ldots, d(\omega, K'_{n'})).$$

Using (decomposition) several times, we get:

$$\odot\,(d(\omega', K_1), \ldots, d(\omega', K_n), d(\omega', K'_1), \ldots, d(\omega', K'_{n'})) \\ < \odot(d(\omega, K_1), \ldots, d(\omega, K_n), d(\omega', K'_1), \ldots, d(\omega', K'_{n'})).$$

Using (composition) several times, we get:

$$\odot\,(d(\omega', K'_1), \ldots, d(\omega', K'_{n'}), d(\omega, K_1), \ldots, d(\omega, K_n)) \\ \leq \odot(d(\omega, K'_1), \ldots, d(\omega, K'_{n'}), d(\omega, K_1), \ldots, d(\omega, K_n)).$$

By (symmetry), we have:

$$\odot\,(d(\omega, K_1), \ldots, d(\omega, K_n), d(\omega', K'_1), \ldots, d(\omega', K'_{n'})) \\ = \odot(d(\omega', K'_1), \ldots, d(\omega', K'_{n'}), d(\omega, K_1), \ldots, d(\omega, K_n)).$$

Now by transitivity:

$$\odot\,(d(\omega', K_1), \ldots, d(\omega', K_n), d(\omega', K'_1), \ldots, d(\omega', K'_{n'})) \\ < \odot(d(\omega, K_1), \ldots, d(\omega, K_n), d(\omega, K'_1), \ldots, d(\omega, K'_{n'})).$$

That is $d(\omega', E_1 \sqcup E_2) < d(\omega, E_1 \sqcup E_2)$. This means $\omega \not\models \triangle_{IC}(E_1 \sqcup E_2)$. Contradiction.

(Only if)

(symmetry) $\odot(x_1, \ldots, x_n) = \odot(\sigma(x_1, \ldots, x_n))$. Direct from (IC3).

(composition) If $\odot(x_1, \ldots, x_n) \leq \odot(y_1, \ldots, y_n)$, then let us consider two interpretations ω, ω' such that for every $i \in 1, \ldots, n$, $d(\omega, K_i) = x_i$ and $d(\omega', K_i) = y_i$. From the definition of the DA^2 operators, we have $\omega \models \triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n\})$. Now let us take a belief base K', such that $d(\omega, K') = d(\omega', K') = z$, we have both $\omega \models \triangle_{form(\{\omega,\omega'\})}(K')$ and $\omega' \models \triangle_{form(\{\omega,\omega'\})}(K')$. Now, from (IC5), we conclude that $\omega \models \triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n, K\})$, or equivalently (from the definition of the operators) $\odot(x_1, \ldots, x_n, z) \leq \odot(y_1, \ldots, y_n, z)$.

(decomposition) We will show the equivalent condition:

$$\text{if } \odot(x_1, \ldots, x_n) < \odot(y_1, \ldots, y_n), \text{ then } \odot(x_1, \ldots, x_n, w) < \odot(y_1, \ldots, y_n, w).$$

Suppose $\odot(x_1, \ldots, x_n) < \odot(y_1, \ldots, y_n)$. Let us consider two interpretations ω, ω' s. t. for every $i \in 1, \ldots, n$, we have $d(\omega, K_i) = x_i$ and $d(\omega', K_i) = y_i$. From the definition of DA^2 operators, we get $\omega \models \triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n\})$ and $\omega' \not\models \triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n\})$. Now let us consider a base K', such that $d(\omega, K') = d(\omega', K') = z$; we have $\omega \models \triangle_{form(\{\omega,\omega'\})}(K')$ and $\omega' \models \triangle_{form(\{\omega,\omega'\})}(K')$. Since $\omega \models \triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n\}) \wedge \triangle_{form(\{\omega,\omega'\})}(K')$, the conjunction is consistent and from (IC6) we obtain $Mod(\triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n, K'\})) \subseteq Mod(\triangle_{form(\{\omega,\omega'\})}(\{K_1, \ldots, K_n\})) = \{\omega\}$. So, by definition of the operator, we have $\odot(x_1, \ldots, x_n, z) < \odot(y_1, \ldots, y_n, z)$.

(and) Suppose that $\varphi = \varphi_1 \wedge \ldots \wedge \varphi_n$ is consistent. We want to show that for every interpretation ω, $\oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)) = \oplus(d(\omega, \varphi))$. There are 2 cases:

case 1: $\omega \models \varphi$.

By definition of the distances, we have $d(\omega, \varphi) = d(\omega, \varphi_i) = 0$; by (minimality) of $\oplus$, $\oplus(d(\omega, \varphi)) = \oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)) = 0$.

case 2: $\omega \not\models \varphi$.

Consider the result of $\triangle_{form(\{\omega\})\vee\varphi}(\{\{form(\{\omega\})\}, \{\varphi_1, \ldots, \varphi_n\}\})$, by (IC0) and (IC1) this base has to be consistent, so it has to pick some models in

$\{\omega\} \cup Mod(\varphi)$. Furthermore (IC4) states that ω and some models of φ have to be in the result. Let us consider one such model ω' of φ. Then we have $d(\omega, \{\{form(\{\omega\})\}, \{\varphi_1, \ldots, \varphi_n\}\}) = d(\omega', \{\{form(\{\omega\})\}, \{\varphi_1, \ldots, \varphi_n\}\})$. Now, by definition of DA^2 merging operators:

$$d(\omega, E) = \odot(\oplus(d(\omega, form(\{\omega\}))), \oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n))),$$

i.e., $\odot(0, \oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)))$. We have also:

$$d(\omega', E) = \odot(\oplus(d(\omega', form(\{\omega\}))), \oplus(d(\omega', \varphi_1), \ldots, d(\omega', \varphi_n))),$$

or equivalently $\odot(\oplus(d(\omega', \omega)), 0)$. Now by (symmetry) and (non-decreasingness) of $\odot$, we get that $\oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)) = \oplus(d(\omega', \omega))$. By the definition of the distance, this is equivalent to $\oplus(d(\omega, \varphi_1), \ldots, d(\omega, \varphi_n)) = \oplus(d(\omega, \varphi))$.

$\square$

7 Conclusion

The major contribution of this paper is a new framework for propositional merging. It is general enough to encompass many existing operators (both model-based ones and syntax-based ones) and to allow the definition of many new operators (symmetric or not.) Both the logical properties and the computational properties of the merging operators pertaining to our framework have been investigated. Some of our results are large-scope ones in the sense that they make sense under very weak conditions on the three parameters that must be set to define an operator in our framework. By instantiating our framework and considering several distances and aggregation functions, more refined results have also been obtained. Finally, a representation theorem for characterizing the "fully rational" DA^2 merging operators has been given.

This work calls for the investigation of several other perspectives. One of them consists in analyzing the properties of the DA^2 operators that are achieved when some other aggregation functions or some other distances are considered. For instance, suppose that a collection of formulas of interest (topics) is available. In this situation, the distance between ω_1 and ω_2 can be defined as the number of relevant formulas on which ω_1 and ω_2 differs (i.e., such that one of them satisfies the formula and the other one violates it.) Several additional distances could also be defined and investigated (see e.g., (39) for distances based on Choquet integrals.)

Finally, it would be interesting to extend our study to non-uniform DA^2 operators, i.e., those obtained by associating a specific aggregation function to each belief base K_i (instead of considering the same one for each K_i.)

Acknowledgments

The authors want to thank Laurence Cholvy for several comments on a first draft of this paper. They also want to thank the anonymous referees for their helpful comments.

The third author has been partly supported by the IUT de Lens, the Université d'Artois, the Région Nord/Pas-de-Calais under the TACT-TIC project, and by the European Community FEDER Program.

References

[1] I. Bloch, A. Hunter, Fusion: General concepts and characteristics, International Journal of Intelligent Systems 16 (10) (2001) 1107–1134, special Issue on Data and Knowledge Fusion.

[2] P. Z. Revesz, On the semantics of arbitration, International Journal of Algebra and Computation 7 (2) (1997) 133–160.

[3] J. Lin, A. O. Mendelzon, Knowledge base merging by majority, in: Dynamic Worlds: From the Frame Problem to Knowledge Management, Kluwer, 1999.

[4] P. Liberatore, M. Schaerf, Arbitration (or how to merge knowledge bases), IEEE Transactions on Knowledge and Data Engineering 10 (1) (1998) 76–90.

[5] S. Konieczny, R. Pino Pérez, Merging with integrity constraints, in: Proceedings of the Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99), LNAI 1638, 1999, pp. 233–244.

[6] S. Benferhat, D. Dubois, S. Kaci, H. Prade, Encoding information fusion in possibilistic logic: a general framework for rational syntactic merging, in: Proceedings of the Fourteenth European Conference on Artificial Intelligence (ECAI'00), 2000, pp. 3–7.

[7] C. Baral, S. Kraus, J. Minker, Combining multiple knowledge bases, IEEE Transactions on Knowledge and Data Engineering 3 (2) (1991) 208–220.

[8] C. Baral, S. Kraus, J. Minker, V. S. Subrahmanian, Combining knowledge bases consisting of first-order theories, Computational Intelligence 8 (1) (1992) 45–71.

[9] S. Konieczny, On the difference between merging knowledge bases and

combining them, in: Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR'00), 2000, pp. 135–144.

[10] B. Nebel, A knowledge level analysis of belief revision, in: Proceedings of the First International Conference on the Principles of Knowledge Representation and Reasoning (KR'89), Toronto, 1989, pp. 301–311.

[11] S. O. Hansson, Revision of belief sets and belief bases, Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3 (1998) 17–75.

[12] S. Konieczny, R. Pino Pérez, On the frontier between arbitration and majority, in: Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02), 2002, pp. 109–118.

[13] S. Konieczny, R. Pino Pérez, Merging information under constraints: a qualitative framework, Journal of Logic and Computation 12 (5) (2002) 773–808.

[14] C. H. Papadimitriou, Computational Complexity, Addison-Wesley, 1994.

[15] T. Eiter, G. Gottlob, On the complexity of propositional knowledge base revision, updates, and counterfactuals, Artificial Intelligence 57 (2-3) (1992) 227–270.

[16] K. W. Wagner, More complicated questions about maxima and minima, and some closures of NP, Theoretical Computer Science 51 (1987) 53–80.

[17] P. Z. Revesz, On the semantics of theory change: arbitration between old and new information, in: Proceedings of the Twelfth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Databases, 1993, pp. 71–92.

[18] S. Konieczny, R. Pino Pérez, On the logic of merging, in: Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98), 1998, pp. 488–498.

[19] M. Dalal, Investigations into a theory of knowledge base revision: preliminary report, in: Proceedings of the Seventh American National Conference on Artificial Intelligence (AAAI'88), 1988, pp. 475–479.

[20] N. Rescher, R. Manor, On inference from inconsistent premises, Theory and Decision 1 (1970) 179–219.

[21] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning, in: Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI'89), 1989, pp. 1043–1048.

[22] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, H. Prade, Inconsistency management and prioritized syntax-based entailment, in: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93), 1993, pp. 640–645.

[23] S. Benferhat, D. Dubois, H. Prade, Some syntactic approaches to the handling of inconsistent knowledge bases: a comparative study, part 1: the flat case, Studia Logica 58 (1997) 17–45.

[24] S. Benferhat, D. Dubois, J. Lang, H. Prade, A. Saffioti, P. Smets, A

general approach for inconsistency handling and merging information in prioritized knowledge bases, in: Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98), 1998, pp. 466–477.

[25] R. Fagin, J. Ullman, M. Vardi, On the semantics of updates in databases, in: Proceedings of the Second ACM SIGACT SIGMOD Symposium on Principles of Database Systems (PODS'83), Atlanta (GE), 1983, pp. 352–365.

[26] B. Nebel, Belief revision and default reasoning: syntax-based approaches, in: Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning (KR'91), Cambridge (MA), 1991, pp. 417–428.

[27] M. Ginsberg, Counterfactuals, Artificial Intelligence 30 (1986) 35–79.

[28] C. E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: Partial meet contraction and revision functions, Journal of Symbolic Logic 50 (1985) 510–530.

[29] C. Lafage, J. Lang, Logical representation of preferences for group decision theory, in: Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR'00), 2000, pp. 457–468.

[30] R. R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Transactions on Systems, Man and Cybernetics 18 (1998) 183–190.

[31] J. L. Marichal, Aggregation operators for multicriteria decision aid, Ph.D. thesis, Université de Liege (1999).

[32] T. Murofushi, M. Sugeno, An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure, Fuzzy Sets and Systems 29 (1989) 201–227.

[33] G. Pinkas, Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge, Artificial Intelligence 77 (1995) 203–247.

[34] E. Grégoire, A. Sofiane, Fusing syntax and semantics in knowledge fusion, in: Proc. of the Eusflat Conference (Eusflat'2001), 2001, pp. 414–417.

[35] H. Fargier, R. Sabbadin, Qualitative decision under uncertainty: Back to expected utility, in: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03), 2003, pp. 303–308.

[36] B. Nebel, How hard is it to revise a belief base?, Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3: Belief Change (1998) 77–145.

[37] T. Bylander, The computational complexity of propositional STRIPS planning, Artificial Intelligence 69 (1-2) (1994) 165–204.

[38] P. Liberatore, M. Schaerf, Brels: a system for the integration of knowledge bases, in: Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR'00), 2000, pp. 145–152.

[39] C. Lafage, J. Lang, Propositional distances and preference representation, in: Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01), 2001, pp. 48–59.

On the merging of Dung's argumentation systems

Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, Pierre Marquis.
Artificial Intelligence. 171.
pages 740-753.
2007.

On the Merging of Dung's Argumentation Systems

Sylvie Coste-Marquis [a], Caroline Devred [a],
Sébastien Konieczny [a], Marie-Christine Lagasquie-Schiex [b],
Pierre Marquis [a]

[a] *CRIL-CNRS, Université d'Artois, Lens, France*
[b] *IRIT-CNRS, UPS, Toulouse, France*

Abstract

In this paper, the problem of deriving sensible information from a collection of argumentation systems coming from different agents is addressed. The underlying argumentation theory is Dung's one: each argumentation system gives both a set of arguments and the way they interact (*i.e.*, attack or non-attack) according to the corresponding agent. The inadequacy of the simple, yet appealing, method which consists in voting on the agents' selected extensions calls for a new approach. To this purpose, a general framework for merging argumentation systems from Dung's theory of argumentation is presented. The objective is achieved through a three-step process: first, each argumentation system is expanded into a partial system over the set of all arguments considered by the group of agents (reflecting that some agents may easily ignore arguments pointed out by other agents, as well as how such arguments interact with her own ones); then, merging is used on the expanded systems as a way to solve the possible conflicts between them, and a set of argumentation systems which are as close as possible to the whole profile is generated; finally, voting is used on the selected extensions of the resulting systems so as to characterize the acceptable arguments at the group level.

Key words: Argumentation frameworks, Argument in agent system

---

1 Introduction

Argumentation is based on the exchange and the evaluation of interacting arguments which may represent information of various kinds, especially beliefs

---

⋆ This paper is an extended and revised version of a paper entitled "Merging Argumentation Systems" that appeared in the Proceedings of AAAI'05, pages 614–619.

Preprint submitted to Artificial Intelligence *2 April 2007*

or goals. Argumentation can be used for modelling some aspects of reasoning, decision making, and dialogue; as such, it has been applied to several domains, including law. For instance, when an agent has conflicting beliefs (viewed as arguments), a (nontrivial) set of plausible consequences can be derived through argumentation from the most acceptable arguments for the agent (additional information like a plausibility ordering are often taken into account in the evaluation phase). Much work has been devoted to this reasoning issue (see for example (13; 21; 26; 25; 1; 27)).

Several theories of argumentation exist; many of them make explicit the nature of arguments, the way arguments are generated, how they interact and how to evaluate them, and finally a characterization of the most acceptable arguments. A key issue is the interaction between arguments which is typically based on a notion of attack; for example, when an argument takes the form of a logical proof, arguments for a statement and arguments against it can be put forward. In that case, the attack relation relies on logical inconsistency.

Dung's theory of argumentation includes several formal systems developed so far for commonsense reasoning or logic programming (13). It is abstract enough to manage without any assumptions on the nature of arguments or the attack relation. Indeed, an argumentation system *à la* Dung consists of a set of (abstract) arguments, together with a binary relation on it (the attack relation). Several semantics can be used for defining interesting sets of arguments (so-called extensions) from which acceptable sets of arguments (*i.e.*, the derivable sets) can be characterized.

In a multi-agent setting, argumentation can also be used to represent (part of) some information exchange processes, like negotiation, or persuasion (see for example (22; 28; 18; 24; 3; 4; 5)). For instance, a negotiation process between two agents about whether some belief must be considered as true given some evidence can be modelled as a two-player game where each move consists in reporting an argument which attacks arguments given by the opponent.

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

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

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

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

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

Like the input of the overall merging process, the result of the merging step is a set of argumentation systems. However, while the first one reflects different points of view (since each system is provided by a specific agent), the second set expresses some uncertainty on the merging due to the presence of conflicts.

The last step of the process consists in defining the acceptable arguments for the group under the uncertainty provided by this set of argumentation systems. Once again, several sensible definitions are given. We show that the sets of arguments considered acceptable when the input is the set of argumentation systems primarily furnished by the agents may drastically differ from the sets of arguments considered acceptable after the merging step, and by means of example, we show that the latter ones are more in accordance with the intuition.

The rest of the paper is organized as follows. After a refresher on Dung's theory of argumentation (in which our approach takes place), we give a simple motivating example (Section 3) which shows that voting on the arguments accepted by each agent is not adequate for defining the arguments accepted by the group. Then we introduce a notion of partial argumentation system (Section 4) which extends the notion of argumentation system and enables to handle the case when agents do not share the same set of arguments. On this ground, we define a family of merging operators for argumentation systems (Section 5) and we study the properties of some of them (especially, those based on the edit distance) (Section 6). Then, we focus on acceptability for partial argumentation systems (Section 7). Finally, we conclude the paper and give a short presentation of some possible refinements of our framework (Section 8).

2 Dung's Theory of Argumentation

Let us present some basic definitions at work in Dung's theory of argumentation (13). We restrict them to finite argumentation frameworks.

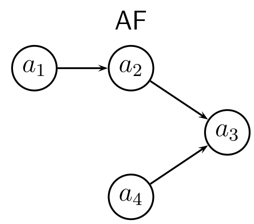
Definition 1 (Argumentation system (AF))
A (finite) argumentation system $\mathsf{AF} = \langle A, R \rangle$ *over* A *is given by a finite set* A *of arguments and a binary relation* R *on* A *called an* attack relation. $a_i R a_j$ *means that* a_i *attacks* a_j *(also denoted by* $(a_i, a_j) \in R$*).*

For our study, we are not interested in the structure of arguments and we consider an arbitrary attack relation.

$\langle A, R \rangle$ defines a directed graph $\mathcal{G}$ called the *attack graph*.

Example 2 *The argumentation system* $\mathsf{AF} = \langle A = \{a_1, a_2, a_3, a_4\},\ R = \{(a_2, a_3), (a_4, a_3), (a_1,\ a_2)\}\rangle$ *defines the following graph* $\mathcal{G}$*:*

Acceptability is about the selection of the most acceptable arguments. Two mainstream approaches exist:



- *Individual acceptability*: acceptability of an *argument* depends only on its properties (see (16; 2));
- *Collective acceptability*: an argument can be defended by other arguments; in this case, the acceptability of a *set of arguments* is considered (see (13)).

Dung's theory is concerned with the second approach. Whether an argument can be accepted depends on the way arguments interact. Collective acceptability is based on two key notions: lack of conflict between arguments and collective defense.

Definition 3 ((13)) *Let $\langle A, R\rangle$ be an argumentation system.*

Conflict-free set *A set $E \subseteq A$ is* conflict-free *if and only if $\nexists a, b \in E$ such that aRb.*

Collective defense *Consider $E \subseteq A$, $a \in A$. E* (collectively) defends *a if and only if $\forall b \in A$, if bRa, then $\exists c \in E$ such that cRb (a is said acceptable w.r.t. E). E* defends all its elements *if and only if $\forall a \in E$, E collectively defends a.*

Dung defines several semantics for collective acceptability based on those two notions (13). Among them the *admissible semantics*, the *preferred semantics*, the *stable semantics* and the *grounded semantics*.

Definition 4 ((13)) *Let $\langle A, R\rangle$ be an argumentation system.*

Admissible semantics *A set $E \subseteq A$ is* admissible *if and only if E is conflict-free and E defends all its elements.*

Preferred semantics *A set $E \subseteq A$ is a* preferred extension *if and only if E is maximal for set inclusion among the admissible sets.*

Stable semantics *A set $E \subseteq A$ is a* stable extension *if and only if E is conflict-free and every $a \in A \setminus E$ is attacked by an element of E.*

Grounded semantics *The* grounded extension *of $\langle A, R\rangle$ is the smallest subset of A with respect to set inclusion among the subsets of A which are admissible and coincide with the set of arguments acceptable w.r.t. itself.*

Note that in all the above definitions, *each attacker* of a given argument is considered independently of the other attackers (there is no way to represent synergetic effects and the possibility to quantify all attackers as a whole is not considered – there exist other works which are concerned with this aspect, see (19; 8; 6; 9; 10; 23)).

Definition 5 (Well-founded argumentation system (13))
An argumentation framework $AF = \langle A, R\rangle$ is well-founded if and only if there does not exist an infinite sequence $a_0, a_1, \ldots, a_n \ldots$ of arguments from A, such that for each i, $a_{i+1} R a_i$.

Among other things, It is shown in (13) that:

- Any admissible set of $\langle A, R\rangle$ is included in a preferred extension of $\langle A, R\rangle$.
- Each $\langle A, R\rangle$ has at least one preferred extension.
- Each $\langle A, R\rangle$ has exactly one grounded extension of $\langle A, R\rangle$ and this extension is included in each preferred extension.
- If $\langle A, R\rangle$ is well-founded then it has a unique preferred extension which is also the only stable extension and the grounded extension.
- Any stable extension of $\langle A, R\rangle$ is also a preferred extension (the converse is false).
- Some $\langle A, R\rangle$ do not have a stable extension.

The acceptability status of each subset of arguments can now be defined by the following relation:

Definition 6 (Acceptability relation) *An* acceptability relation*, denoted by Acc_{AF}, for a given argumentation system $\mathsf{AF} = \langle A, R\rangle$, is a total function from 2^A to $\{true, false\}$ which associates each subset E of A with true if E is an acceptable set for AF and with false otherwise.*

Usually, an acceptability relation is based on a specific semantics (plus a selection principle). For instance, a set of arguments can be considered acceptable if and only if it is included in one extension (credulous selection) or in every extension (skeptical selection). Alternatively, a set of arguments can be considered acceptable if and only if it coincides with one extension for the chosen semantics. Whatever the way it is defined, an acceptability relation can be viewed as a choice function among the elements of 2^A. In this context, the "acceptability of an argument" a can correspond either to the acceptability of the singleton $\{a\}$, or to the membership of a to an acceptable set (see (12)).

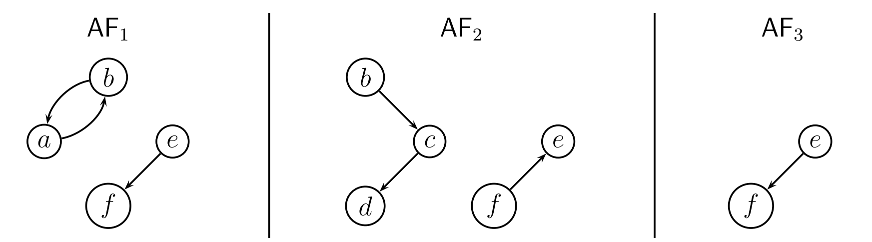
3 Simple is not so Beautiful

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

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

Example 7 *Consider the three following argumentation systems:*

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



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

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

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

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

3.2 Union is not merging (in general)

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

Example 7 (continued) *The resulting AF is* $\bigcup_{i=1}^{3} \mathsf{AF}_i = \langle \{a, b, c, d, e, f\}, \{(a, b), (b, a), (b, c), (c, d), (e, f), (f, e)\} \rangle$.

Example 7 shows that the union approach to merging argumentation systems suffers from a major problem: it solves conflicts by giving to the explicit attack information some undue prominence to implicit non-attack information. Thus, when a pair of arguments (like, say (f, e)) does not belong to the attack relation furnished by an agent (say, Agent 1) while both arguments (f and e) belong to the set of arguments she points out, the meaning is that for Agent 1, argument f *does not attack* argument e. Imagine now that in the considered profile of argumentation systems, 999 agents report the same system as Agent 1, and the 1000^{th} agent is Agent 2. In the resulting argumentation system considered at the group level, assuming that union is used as a merging operator, it will be the case that f attacks e while 999 agents over 1000 believes that it is not the case!

4 Partial Argumentation Systems

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

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

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

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

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

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

Handling these different kinds of information within a uniform setting calls for an extension[1] of the notion of argumentation systems, that we call *partial argumentation systems.*

Definition 8 (Partial argumentation system (PAF)) *A* (finite) partial argumentation system *over A is a quadruple $\mathsf{PAF} = \langle A, R, I, N\rangle$ where*

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

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

Each AF is a particular PAF for which the set I is empty (we say that such an AF is equivalent to the associated PAF). In an AF, the N relation also

[1] In (11), a new binary relation on the arguments is also introduced in Dung's argumentation framework : however, this new relation represents a notion of support between arguments. Clearly enough, this is unrelated with the relation introduced here representing the ignorance about the attack between arguments.

exists even if it is not given explicitly ($I = \varnothing$ and $N = A \times A \setminus R$). So, an AF could also be denoted by $\langle A, R, N \rangle$.

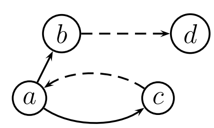
Each PAF over A can be viewed as a compact representation of a set of AFs over A, called its *completions*:

Definition 9 (Completion of a PAF) *Let* $\mathsf{PAF} = \langle A, R, I \rangle$*. Let* $\mathsf{AF} = \langle A, S \rangle$*.* AF *is a* completion *of* PAF *if and only if* $R \subseteq S \subseteq R \cup I$*.*

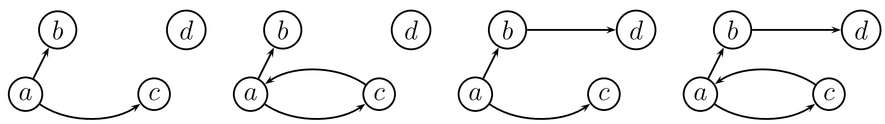
The set of all completions of PAF *is denoted* $\mathcal{C}(\mathsf{PAF})$*.*

Example 10 *The partial argumentation system* $\mathsf{PAF} = \langle A = \{a, b, c, d\},\ R = \{(a, b),\ (a, c)\},\ I = \{(c, a), (b, d)\},\ N = \{(a, a), (b, b), (c, c), (d, d), (b, a), (b, c), (c, b), (a, d), (d, a), (d, b), (c, d), (d, c)\}\rangle$ *is illustrated on the following figure (solid arrows represent the attack relation and dotted arrows represent the ignorance relation; non-attack relations are not represented explicitly as in the AF case):*

PAF



The completions of this PAF are:



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

Definition 11 (Expansion of an AF) *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of* n *AFs such that* $\mathsf{AF}_i = \langle A_i, R_i, N_i \rangle$*. Let* $\mathsf{AF} = \langle A, R \rangle$ *be an argumentation system. An* expansion *of* AF *given* $\mathcal{P}$ *is any PAF* $\mathtt{exp}(\mathsf{AF}, \mathcal{P})$ *defined by* $\langle A \cup \bigcup_i A_i, R', I',\ N' \rangle$ *such that* $R \subseteq R'$ *and* $(A \times A) \setminus R \subseteq N'$*.* $\mathtt{exp}$ *is referred to as an* expansion function*.*

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

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

Each agent may also compute the exact interaction between a and b when the attack relation is not primitive but defined from more basic notions (as in the approach by Elvang-Gøransson et al., see e.g., (15; 16; 17)). Note that if she has limited computational resources, Agent i can compute exact interactions as far as she can, then express ignorance for the remaining ones.

In the following, we specifically focus on *consensual expansions*. Intuitively, the consensual expansion of an argumentation system $\mathsf{AF} = \langle A, R \rangle$ given a profile of such systems is obtained by adding a pair of arguments (a, b) (where at least one of a, b is not in A) into the attack (resp. the non-attack relation) *provided that all other agents of the profile who know the two arguments agree on the existence of the attack*[2] (resp. *the non-attack*); otherwise, it is added to the ignorance relation.

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

Definition 12 (Consensual expansion) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of n AFs such that $\mathsf{AF}_i = \langle A_i, R_i \rangle$. Let $\mathsf{AF} = \langle A, R, N \rangle$ be an argumentation system. Let $conf(\mathcal{P}) = (\bigcup_i R_i) \cap (\bigcup_i N_i)$ be the set of interactions for which a conflict exists within the profile. The* consensual expansion *of AF over $\mathcal{P}$ is the tuple denoted by $\mathtt{exp}_C = \langle A', R', I', N' \rangle$ with:*

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

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

Proposition 13 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of n AFs such that $\mathsf{AF}_i = \langle A_i, R_i \rangle$. Let $\mathsf{AF} = \langle A, R, N \rangle$ be an argumentation system. The consensual expansion $\mathtt{exp}_C$ of AF over $\mathcal{P}$ is an expansion of AF over $\mathcal{P}$ in the sense of Definition 11.*

[2] *i.e.*, if $a, b \in A_i$, then $(a, b) \in R_i$.

Proof : Consider $(a, b) \in A' \times A'$. There are several cases:

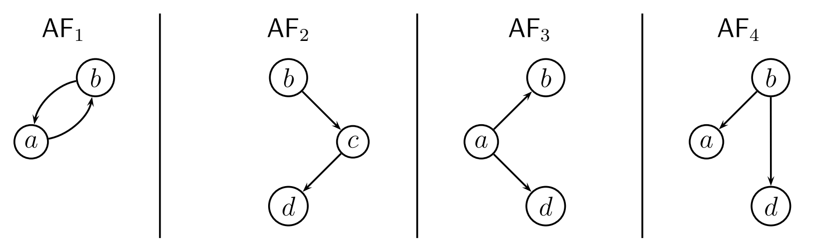
- if $(a, b) \in R$ then $(a, b) \in R'$ and $(a, b) \notin I' \cup N'$ (so, $R \subseteq R'$);
- if $(a, b) \notin R$ and $(a, b) \in N$ then $(a, b) \notin I' \cup R'$ and $(a, b) \in N'$ (so, $N \subseteq N'$);
- if $(a, b) \notin R \cup N$ then there are two cases:
  - · if $\nexists \mathsf{AF}_i \in \mathcal{P}$ such that $a, b \in A_i$ then $(a, b) \notin conf(\mathcal{P})$; so, $(a, b) \in N'$ and $(a, b) \notin R' \cup I'$;
  - · if $\exists \mathsf{AF}_i \in \mathcal{P}$ such that $a, b \in A_i$ then we have 4 possible cases:
    if $(a, b) \in R_i$ and $\nexists AF_j \in \mathcal{P}$ such that $(a, b) \in N_j$ then $(a, b) \notin conf(\mathcal{P})$; so, $(a, b) \in R'$ and $(a, b) \notin N' \cup I'$;
    if $(a, b) \in R_i$ and $\exists AF_j \in \mathcal{P}$ such that $(a, b) \in N_j$ then $(a, b) \in conf(\mathcal{P})$; so, $(a, b) \in I'$ and $(a, b) \notin R' \cup N'$;
    if $(a, b) \in N_i$ and $\nexists AF_j \in \mathcal{P}$ such that $(a, b) \in R_j$ then $(a, b) \notin conf(\mathcal{P})$; so, $(a, b) \in N'$ and $(a, b) \notin R' \cup I'$;
    if $(a, b) \in N_i$ and $\exists AF_j \in \mathcal{P}$ such that $(a, b) \in R_j$ then $(a, b) \in conf(\mathcal{P})$; so, $(a, b) \in I'$ and $(a, b) \notin R' \cup N'$.

So, R', I' and N' form a partition of $A' \times A'$ which satisfies $R \subseteq R'$ and $N \subseteq N'$. $\square$

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

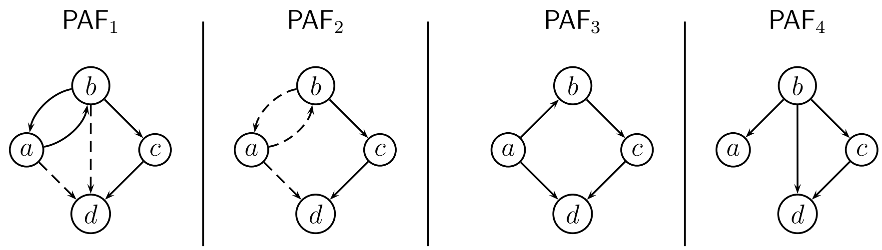
Example 14 *Consider the profile consisting of the following four argumentation systems:*

- $\mathsf{AF}_1 = \langle A_1 = \{a, b\}, R_1 = \{(a, b), (b, a)\}\rangle$,
- $\mathsf{AF}_2 = \langle A_2 = \{b, c, d\}, R_2 = \{(b, c), (c, d)\}\rangle$,
- $\mathsf{AF}_3 = \langle A_3 = \{a, b, d\}, R_3 = \{(a, b), (a, d)\}\rangle$,
- $\mathsf{AF}_4 = \langle A_4 = \{a, b, d\}, R_4 = \{(b, d), (b, a)\}\rangle$.



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

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



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

5 Merging Operators

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

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

Definition 15 (Pseudo-distance) *A* pseudo-distance *d between PAFs over A is a mapping which associates a non-negative real number to each pair of PAFs over A and satisfies the properties of* symmetry *($d(x, y) = d(y, x)$) and* minimality *($d(x, y) = 0$ if and only if $x = y$).*

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

Definition 16 (Aggregation function) *An* aggregation function *is a mapping $\otimes$ from $(\mathbb{R}+)^n$ to $(\mathbb{R}+)$ (strictly speaking, it is a family of mappings, one for each n), that satisfies*

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

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

Definition 17 (Merging of n AFs) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of n AFs. Let d be any pseudo-distance between PAFs, let $\otimes$ be an aggregation function, and let $\mathtt{exp}_1, \ldots, \mathtt{exp}_n$ be n expansion functions. The* merging *of $\mathcal{P}$ is the set of AFs*

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

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

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

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

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

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

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

Note that it would be possible to refine Definition 17 so as to include integrity constraints into the picture. This can be useful if there exists some (unquestionable) knowledge about the expected result (some attacks between arguments which have to hold for the group). It is then enough to look only to the AFs which satisfy the constraints, similarly to what is done in propositional belief base merging (see e.g., (20)). In contrast to the belief base merging scenario, constraints on the *structure* of the candidate AFs can also be set. In particular, considering only acyclic AFs can prove valuable since (1) such AFs are well-founded, (which implies that only one extension has to be considered whatever the underlying semantics – among Dung's ones), and (2) this extension (which turns out to be the grounded one, see (13)) can be computed in time polynomial in the size of the AF (while computing a single extension is intractable for the other semantics in the general case – under the standard assumptions of complexity theory – see (14)).

Now, many pseudo-distances between PAFs and many aggregation functions can be used, giving rise to many merging operators. Usual aggregation functions include the sum Σ, the max $\mathcal{M}\mathtt{ax}$ and the leximax $\mathcal{L}\mathtt{eximax}$[3] but using

[3] When applied to a vector of n real numbers, the leximax function $\mathcal{L}\mathtt{eximax}$ gives

non-symmetric functions is also possible (this may be particularly valuable if some agents are more important than others). Some aggregation functions (like the sum) enable the merging process to take into account the number of agents believing that an argument attacks or not another argument:

Example 7 (continued) *Two agents over three agree with the fact that e attacks f and f does not attack e. It may prove sensible that the group agrees with the majority.*

The choice of the aggregation function is very important for tuning the operator behaviour with the expected one. For example, sum is a possible choice in order to solve conflicts using majority. Otherwise, the leximax function can prove more valuable if the aim is to behave in a more consensual way, trying to define a result close to the AF of each agent of the group. The distinction between majority and arbitration operators as considered in propositional belief base merging (20) also applies here.

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

Definition 18 (Edit distance) *Let* $\mathsf{PAF}_1 = \langle A, R_1, I_1, N_1\rangle$ *and* $\mathsf{PAF}_2 = \langle A, R_2, I_2, N_2\rangle$ *be two PAFs over A.*

- *Let a, b be two arguments $\in A$. The* edit distance between PAF_1 and PAF_2 over a, b *is the mapping $de_{a,b}$ such that:*
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 0$ *if and only if* $(a, b) \in R_1 \cap R_2$ *or* $I_1 \cap I_2$ *or* $N_1 \cap N_2$,
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 1$ *if and only if* $(a, b) \in R_1 \cap N_2$ *or* $N_1 \cap R_2$,
  - $de_{a,b}(\mathsf{PAF1}, \mathsf{PAF2}) = 0.5$ *otherwise.*
- *The* edit distance between PAF_1 and PAF_2 *is given by*

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

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

It is easy to show that:

Proposition 19 *The edit distance de between PAFs is a distance.*

Proof : We show that de and $de_{a,b}$, $\forall (a, b) \in A \times A$ are distances, *i.e.* they are (1) symmetric, they satisfy (2) the minimality requirement and (3) the triangular inequality:

(1) Obvious.

---

the list of those numbers sorted in a decreasing way. Such lists are compared w.r.t. the lexicographic ordering induced by the standard ordering on real numbers.

(2) ($\Rightarrow$) Consider $\mathsf{PAF}_1 = \langle A, R_1, I_1, N_1\rangle$ and $\mathsf{PAF}_2 = \langle A, R_2, I_2, N_2\rangle$ such that $\mathsf{PAF}_1 = \mathsf{PAF}_2$. For all $(a,b) \in A \times A$, if $\mathsf{PAF}_1 = \mathsf{PAF}_2$ then $(a,b) \in R_1 \cap R_2$ or $(a,b) \in I_1 \cap I_2$ or $(a,b) \in N_1 \cap N_2$. So, $\forall (a,b) \in A \times A$, $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_2) = 0$, and $de(\mathsf{PAF}_1, \mathsf{PAF}_2) = 0$.

($\Leftarrow$) Suppose $de(\mathsf{PAF}_1, \mathsf{PAF}_2) = 0$ and make a *reductio ad absurdum*: if $\mathsf{PAF}_1 \neq \mathsf{PAF}_2$ then $\exists (a,b) \in A \times A$ such that $(a,b) \notin R_1 \cap R_2$, $(a,b) \notin I_1 \cap I_2$ and $(a,b) \notin N_1 \cap N_2$; so, $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_2) \neq 0$; so, $de(\mathsf{PAF}_1, \mathsf{PAF}_2) \neq 0$ which is a contradiction with the hypothesis; so, $\mathsf{PAF}_1 = \mathsf{PAF}_2$. The same reasoning can be achieved with $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_2) = 0$ and the same result is obtained: $\mathsf{PAF}_1 = \mathsf{PAF}_2$.

(3) Consider $\mathsf{PAF}_1 = \langle A, R_1, I_1, N_1\rangle$, $\mathsf{PAF}_2 = \langle A, R_2, I_2, N_2\rangle$ and $\mathsf{PAF}_3 = \langle A, R_3, I_3, N_3\rangle$. $\forall (a,b) \in A \times A$, we compute and compare $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_2)$, $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_3)$ and $de_{a,b}$ $(\mathsf{PAF}_3, \mathsf{PAF}_2)$, respectively denoted by x, y, z. We have three possible cases:
- $x = 0$: $\forall y, z$, we have $x \leq y + z$;
- $x = 0.5$: $x \leq y + z$ is false if and only if $y = z = 0$; however, $y = z = 0$ implies that $(a,b) \in R_1 \cap R_2 \cap R_3$ or $(a,b) \in I_1 \cap I_2 \cap I_3$ or $(a,b) \in N_1 \cap N_2 \cap N_3$ which also implies $x = 0$ (contradiction with the hypothesis); so, $x \leq y + z$;
- $x = 1$: we have $(a,b) \in R_1 \cap N_2$ or $(a,b) \in N_1 \cap R_2$; suppose that $(a,b) \in R_1 \cap N_2$ then there are 3 possible cases:
  - $(a,b) \in R_3$: so, $y = 0$, $z = 1$ and we have $x \leq y + z$;
  - $(a,b) \in I_3$: so, $y = 0.5$, $z = 0.5$ and we have $x \leq y + z$;
  - $(a,b) \in N_3$: so, $y = 1$, $z = 0$ and we have $x \leq y + z$.

  The same reasoning can be achieved if $(a,b) \in N_1 \cap R_2$. So, $\forall (a,b) \in A \times A$: $de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_2) \leq de_{a,b}(\mathsf{PAF}_1, \mathsf{PAF}_3) + de_{a,b}(\mathsf{PAF}_3, \mathsf{PAF}_2)$;

  summing over all $(a,b) \in A \times A$, we get:
  $de(\mathsf{PAF}_1, \mathsf{PAF}_2) \leq de(\mathsf{PAF}_1, \mathsf{PAF}_3) + de(\mathsf{PAF}_3, \mathsf{PAF}_2)$.

□

Let us now illustrate the notion of edit distance as well some associated merging operators on Example 14.

Example 14 (continued) *We consider the following argumentation system* $\mathsf{AF}'_1 = \langle \{a, b, c, d\}, \{(a,b), (b,a), (b,c), (c,d)\}\rangle$.

The edit distance between AF'_1 *and each of the PAFs* PAF_1, PAF_2, PAF_3, PAF_4 *obtained by consensual expansion from the profile* $\langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4\rangle$ *is:*

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

Taking the sum as the aggregation function, we obtain:
$\Sigma_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 6.5$.

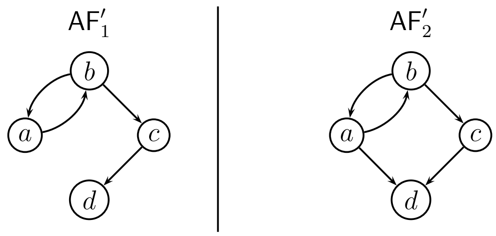
Taking the max, we obtain: $\mathcal{M}\mathtt{ax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = 2$.

Taking the leximax, we obtain: $\mathcal{L}\mathtt{eximax}_{i=1}^{4} de(\mathsf{AF}'_1, \mathsf{PAF}_i) = (2, 2, 1.5, 1)$.

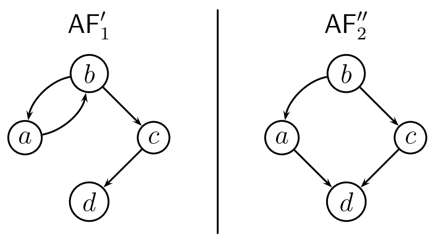
By computing such distances for all candidate AFs (i.e., all AFs over $\{a, b, c, d\}$*), we can compute the result of the merging:*

$\Delta_{de}^{\Sigma}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_4 \rangle)$ *is the set containing the two following AFs:*

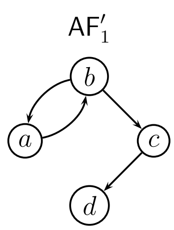
- $\mathsf{AF}'_1 = \langle \{a, b, c, d\}, \{(a, b), (b, a), (b, c), (c, d)\} \rangle$,
- $\mathsf{AF}'_2 = \langle \{a, b, c, d\}, \{(a, b), (b, a), (b, c), (a, d), (c, d)\} \rangle$.



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



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



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

- AF'_1 *is the most consensual AF obtained as it is almost equidistant from each PAF;*
- AF'_2 *is much closer to* PAF_1, PAF_2 *and* PAF_3 *than to* PAF_4, *thus it is selected with the sum as an aggregation operator but it is too far from* PAF_4 *for being selected with the* $\mathcal{M}\mathtt{ax}$ *or* $\mathcal{L}\mathtt{eximax}$ *operators.*

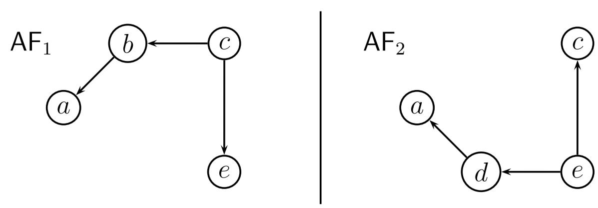
- AF''_2 *is nearly equidistant from all four PAFs of the profile but less consensual than* AF'_1*, thus it is selected neither with* Σ *nor with* $\mathcal{L}\mathtt{eximax}$ *but only with* $\mathcal{M}\mathtt{ax}$ *as it is not far from any of the given PAFs.*

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

Here is another simple example:

Example 20 *Consider the two following argumentation systems:*

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



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

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

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

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

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

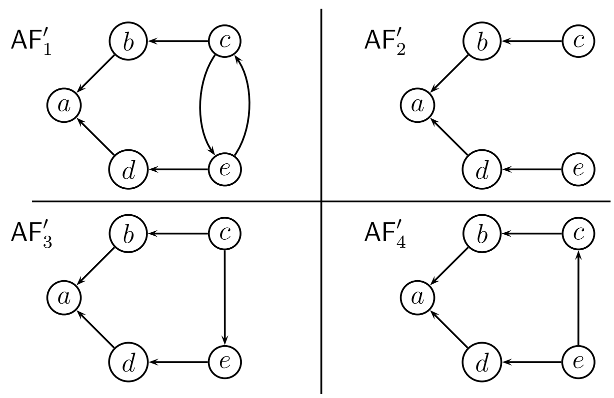
$$\Delta_{de}^{\mathcal{M}\mathtt{ax}}(\langle\mathsf{AF}_1, \mathsf{AF}_2\rangle) = \Delta_{de}^{\mathcal{L}\mathtt{eximax}}(\langle\mathsf{AF}_1, \mathsf{AF}_2\rangle) = \{\mathsf{AF}'_1, \mathsf{AF}'_2\} \text{ with}$$

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

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

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

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



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

6 Some Properties

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

6.1 Properties of PAFs and consensual expansions

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

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

Definition 21 (Clash-free part of a profile of PAFs)
Let $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *be a profile of PAFs. The* clash-free part *of* $\mathcal{P}$ *is denoted by* $CFP(\mathcal{P})$ *and is defined by:*

$$CFP(\mathcal{P}) = \langle \bigcup_i A_i, \bigcup_i R_i \setminus \bigcup_i N_i, I_{CFP}, \bigcup_i N_i \setminus \bigcup_i R_i \rangle$$

where $I_{CFP} = (\bigcup_i A_i \times \bigcup_i A_i) \setminus ((\bigcup_i R_i \setminus \bigcup_i N_i) \cup (\bigcup_i N_i \setminus \bigcup_i R_i))$.

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

Example 14 (continued) *With* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$, $CFP(\mathcal{P}) = \langle \{a, b, c, d\}, \{(b, c), (c, d)\}, \{(a, b), (b, a), (a, d), (b, d), (a, c), (c, a)\} \rangle$.

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

Definition 22 (Common part of a profile of PAFs)
Let $\mathcal{P} = \langle \mathsf{PAF}_1, \dots, \mathsf{PAF}_n \rangle$ *be a profile of PAFs. The* common part *of* $\mathcal{P}$ *is denoted by* $CP(\mathcal{P})$ *and is defined by:* $CP(\mathcal{P}) = \langle \bigcap_i A_i, \bigcap_i R_i, \bigcap_i I_i, \bigcap_i N_i \rangle$.

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

Example 14 (continued) *With* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$, $CP(\mathcal{P}) = \langle \{b\}, \varnothing, \varnothing, \{(b, b)\} \rangle$.

We have the following easy property:

Proposition 23 *Let* $\mathcal{P} = \langle \mathsf{PAF}_1, \dots, \mathsf{PAF}_n \rangle$ *be a profile of PAFs. The* common part *of* $\mathcal{P}$ *is pointwise included into the clash-free part of* $\mathcal{P}$, *i.e.:*

- $\bigcap_i R_i \subseteq \bigcup_i R_i \setminus \bigcup_i N_i$;
- $\bigcap_i I_i \subseteq I_{CFP}$;
- $\bigcap_i N_i \subseteq \bigcup_i N_i \setminus \bigcup_i R_i$.

Proof : The proof is straightforward:

- $\bigcap_i R_i \subseteq \bigcup_i R_i$ is obvious; and we also have $\forall (a, b) \in \bigcap_i R_i$, $(a, b) \notin N_j$ for all j (otherwise, $\exists \mathsf{PAF}_k$ such that $(a, b) \in R_k \cap N_k$ that is impossible by definition), so $\bigcap_i R_i \subseteq \bigcup_i R_i \setminus \bigcup_i N_i$.
- In the same way, we can prove $\bigcap_i N_i \subseteq \bigcup_i N_i \setminus \bigcup_i R_i$.
- if $\forall (a, b) \in \bigcap_i I_i$ then, by definition, $(a, b) \notin R_i$ and $(a, b) \notin N_i$ for all i ; so, $(a, b) \in (\bigcup_i A_i \times \bigcup_i A_i) \setminus ((\bigcup_i R_i \setminus \bigcup_i N_i) \cup (\bigcup_i N_i \setminus \bigcup_i R_i))$.

$\square$

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

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

Proposition 24 *Let $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ be a profile of n PAFs over the same set of arguments A. Consider the clash-free part of $\mathcal{P}$ denoted by $CFP(\mathcal{P})$ $= \langle A_{CFP}, R_{CFP}, I_{CFP}, N_{CFP} \rangle$ and the common part of $\mathcal{P}$ denoted by $CP(\mathcal{P})$ $= \langle A_{CP}, R_{CP}, I_{CP}, N_{CP} \rangle$. If $\bigcup_i I_i \subseteq conf(\mathcal{P}) = (\bigcup_i R_i) \cap (\bigcup_i N_i)$, we have:*

- $A_{CFP} = A_{CP}$,
- $R_{CFP} = R_{CP}$,
- $N_{CFP} = N_{CP}$.

Proof : All the PAFs are over the same set of arguments, so we have $A = \bigcup_i A_i = \bigcap_i A_i$ and $A_{CFP} = A_{CP}$.

First, we prove that $R_{CFP} = R_{CP}$.

- $R_{CFP} \subseteq R_{CP}$: consider $(a, b) \in R_{CFP}$; so $(a, b) \in \bigcup_i R_i \setminus \bigcup_i N_i$; suppose that $(a, b) \notin R_{CP}$; so $\exists \mathsf{PAF}_k$ such that $(a, b) \notin R_k$; so $(a, b) \in N_k$ or $(a, b) \in I_k$;
  In the first case, we have $(a, b) \notin \bigcup_i R_i \setminus \bigcup_i N_i$: contradiction with the hypothesis $(a, b) \in R_{CFP}$;
  In the second case, we retrieve the first case because $\bigcup_i I_i \subseteq conf(\mathcal{P}) = (\bigcup_i R_i) \cap (\bigcup_i N_i)$.
  Thus $(a, b) \in R_{CP}$.
- $R_{CFP} \supseteq R_{CP}$: given by Proposition 23.

$N_{CFP} = N_{CP}$ is proven in the same way. □

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

Example 14 (continued) *With $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3, \mathsf{AF}_4 \rangle$ and $\mathtt{exp}_C(\mathcal{P}) = \langle \mathtt{exp}_C\ (\mathsf{AF}_1,\ \mathcal{P}),\ \mathtt{exp}_C(\mathsf{AF}_2, \mathcal{P}),\ \mathtt{exp}_C(\mathsf{AF}_3, \mathcal{P}),\ \mathtt{exp}_C(\mathsf{AF}_4, \mathcal{P}) \rangle$, we have:*

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

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

Proposition 25 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. For each i, we have:*

- $A_{CFP(\mathcal{P})} = A_{\mathtt{exp}_C(\mathsf{AF}_{i,\mathcal{P}})}$,
- $R_{CFP(\mathcal{P})} \subseteq R_{\mathtt{exp}_C(\mathsf{AF}_{i,\mathcal{P}})}$,
- $N_{CFP(\mathcal{P})} \subseteq N_{\mathtt{exp}_C(\mathsf{AF}_{i,\mathcal{P}})}$.

Proof : Consider AF_i, denoted by $\langle A_i, R_i, N_i \rangle$, and the set $conf(\mathcal{P}) = (\bigcup_i R_i) \cap (\bigcup_i N_i)$. Each $\mathtt{exp}_C(\mathsf{AF}_i, \mathcal{P})$ is denoted by $\langle A'_i, R'_i, I'_i, N'_i \rangle$.

- By definition, the set of arguments is the same for $CFP(\mathsf{AF}_1, \ldots, \mathsf{AF}_n)$ and for each $\mathtt{exp}_C(\mathsf{AF}_i, \mathcal{P})$, $\forall \mathsf{AF}_i$: it is equal to $\bigcup_i A_i$.
- Consider $a, b \in \bigcup_i A_i$ such that $(a, b) \in R_{CFP(\mathcal{P})} = (\bigcup_i R_i) \setminus (\bigcup_i N_i)$; so, we have $(a, b) \notin conf(\mathcal{P})$ and $(a, b) \notin N_i$. So, $(a, b) \in R'_i = R_i \cup (((\bigcup_i R_i) \setminus conf(\mathcal{P})) \setminus N_i)$.
- Consider $a, b \in \bigcup_i A_i$ such that $(a, b) \in N_{CFP(\mathcal{P})} = (\bigcup_i N_i) \setminus (\bigcup_i R_i)$; so, we have $(a, b) \notin conf(\mathcal{P})$ and $(a, b) \notin \bigcup_i R_i$. So, $(a, b) \notin I'_i$, and $(a, b) \notin R'_i$. So, $(a, b) \in N'_i$.

□

Now, concordance between AFs can be defined as follows:

Definition 26 (Concordance) *Let $\mathsf{AF}_1 = \langle A_1, R_1 \rangle$, $\mathsf{AF}_2 = \langle A_2, R_2 \rangle$ be two AFs. AF_1, AF_2 are said to be* concordant *if and only if $\forall (a, b) \in (A_1 \cap A_2) \times (A_1 \cap A_2)$, $(a, b) \in R_1$ if and only if $(a, b) \in R_2$. Otherwise they are said to be* discordant.
Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. $\mathcal{P}$ is said to be concordant *if and only if all its AFs are pairwise concordant. Otherwise it is said to be* discordant.

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

Proposition 27 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of argumentation systems. $\mathcal{P}$ is concordant if and only if $conf(\mathcal{P}) = \bigcup_i R_i \cap \bigcup_i N_i$ is empty.*

Proof : $\mathcal{P}$ is concordant $\Leftrightarrow$
$\forall \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$, $\nexists a, b \in A_i \cap A_j$ such that $(a, b) \in (R_i \setminus R_j) \cup (R_j \setminus R_i) \Leftrightarrow$
$\forall \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$, $\nexists a, b \in A_i \cap A_j$ such that $(a, b) \in R_i$ and $(a, b) \in N_j \Leftrightarrow$
$\bigcup_i R_i \cap \bigcup_i N_i = \varnothing$. □

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

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

Proof : $CFP(\mathcal{P})$ is denoted by $\langle A_{CFP}, R_{CFP}, I_{CFP}, N_{CFP} \rangle$.

The proof for "$\mathcal{P}$ concordant $\Rightarrow CFP(\mathcal{P}) = \bigcup_i \mathsf{AF}_i$" is made using a *reductio ad absurdum*. We suppose that $CFP(\mathcal{P}) \neq \bigcup_i \mathsf{AF}_i$ and we have the following possibilities:

- $\exists (a,b) \in R_{CFP}$ and $(a,b) \notin \bigcup_i R_i$; this case is impossible because, by definition, $(a,b) \in (\bigcup_i R_i) \setminus (\bigcup_i N_i)$;
- $\exists (a,b) \in N_{CFP}$ and $(a,b) \notin \bigcup_i N_i$; this case is impossible because, by definition, $(a,b) \in (\bigcup_i N_i) \setminus (\bigcup_i R_i)$;
- $\exists (a,b) \notin R_{CFP}$ and $(a,b) \in \bigcup_i R_i$; so, by definition, $(a,b) \in (\bigcup_i N_i)$; so, $\exists \mathsf{AF}_k, \mathsf{AF}_j$ such that $(a,b) \in R_k$ and $(a,b) \in N_j$; so, $\exists \mathsf{AF}_k, \mathsf{AF}_j$ such that $(a,b) \in A_k \cap A_j$ and $(a,b) \in R_k \setminus R_j$; so, contradiction with the hypothesis $\mathcal{P}$ concordant;
- $\exists (a,b) \notin N_{CFP}$ and $(a,b) \in \bigcup_i N_i$; so, by definition, $(a,b) \in (\bigcup_i R_i)$; so, $\exists \mathsf{AF}_k, \mathsf{AF}_j$ such that $(a,b) \in R_k$ and $(a,b) \in N_j$; so, $\exists \mathsf{AF}_k, \mathsf{AF}_j$ such that $(a,b) \in A_k \cap A_j$ and $(a,b) \in R_k \setminus R_j$; so, contradiction with the hypothesis $\mathcal{P}$ concordant.

For each possibility, we obtain a contradiction. So, if $\mathcal{P}$ is concordant, then $CFP(\mathcal{P}) = \bigcup_i \mathsf{AF}_i$.

The proof for "$\mathcal{P}$ concordant $\Leftarrow CFP(\mathcal{P}) = \bigcup_i \mathsf{AF}_i$" is also made using a *reductio ad absurdum*. If $\mathcal{P}$ is discordant then $\exists \mathsf{AF}_i, \mathsf{AF}_j$ such that $\exists (a,b) \in A_i \cap A_j$ and $(a,b) \in (R_i \setminus R_j) \cup (R_j \setminus R_i)$. So, $a, b \in \bigcup_k A_k$, $(a,b) \in \bigcup_k R_k$ and $(a,b) \in \bigcup_k N_k$; so, (a,b) appears in the attack relation and in the non-attack relation of $\bigcup_i \mathsf{AF}_i$. However, by definition, (a,b) cannot appear in the same time in R_{CFP} and in N_{CFP}. So, contradiction with the hypothesis $CFP(\mathsf{AF}_1, \ldots, \mathsf{AF}_n) = \bigcup_i \mathsf{AF}_i$. $\square$

Proposition 29 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. $\mathcal{P}$ is concordant if and only if $\mathtt{exp}_C(\mathcal{P}) = \langle \mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P}) \rangle$ is reduced to $\langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle$ (i.e., each of the n elements of the vector is $\bigcup_i \mathsf{AF}_i$).*

Proof : Consider a concordant profile of AFs $\mathcal{P}$. $\forall \mathsf{AF}_i = \langle A_i, R_i, N_i \rangle$, let us consider $\mathtt{exp}_C(\mathsf{AF}_i, \mathcal{P}) = \langle A'_i, R'_i, I'_i, N'_i \rangle$. $\forall a, b \in \bigcup_i A_i$, there are several cases:

- if $(a,b) \in R_i$ then $(a,b) \in R'_i$;
- if $(a,b) \notin R_i$ and $(a,b) \in A_i \times A_i$ then $(a,b) \in N_i$, so $(a,b) \in N'_i$; with $\mathcal{P}$ concordant, we also know that $\nexists \mathsf{AF}_j \in \mathcal{P}$ such that $(a,b) \in R_j$;
- if $(a,b) \notin R_i$ and $(a,b) \notin A_i \times A_i$ then there are two cases:
  · either $\exists \mathsf{AF}_j \in \mathcal{P}$ such that $(a,b) \in R_j$: because $\mathcal{P}$ is concordant, $(a,b) \in$

R'_i;

· or $\nexists \mathsf{AF}_j \in \mathcal{P}$ such that $(a,b) \in R_j$: so, $(a,b) \in N'_i$.

In all the cases, if (a,b) is an attack interaction for one of the AF_i, (a,b) is also an attack interaction for the consensual PAFs. So, all the consensual PAFs are equal to $\bigcup_i \mathsf{AF}_i$.

For the second part of the proof, consider $\mathtt{exp}_C(\mathcal{P}) = \langle \bigcup_i \mathsf{AF}_i, , \ldots, \bigcup_i \mathsf{AF}_i \rangle$. We suppose that $\mathcal{P}$ is discordant. So, $\exists \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$ such that $\exists a, b \in A_i \cap A_j$ and $(a,b) \in (R_i \setminus R_j) \cup (R_j \setminus R_i)$. If we suppose that $(a,b) \in R_i$, then $\mathtt{exp}_C(\mathsf{AF}_j, \mathcal{P})$ cannot contain the attack (a,b); so, $\mathtt{exp}_C(\mathsf{AF}_j, \mathcal{P}) \neq \bigcup_i \mathsf{AF}_i$: contradiction. And the same problem appears when we suppose that $(a,b) \in R_j$. So, $\mathcal{P}$ is concordant. □

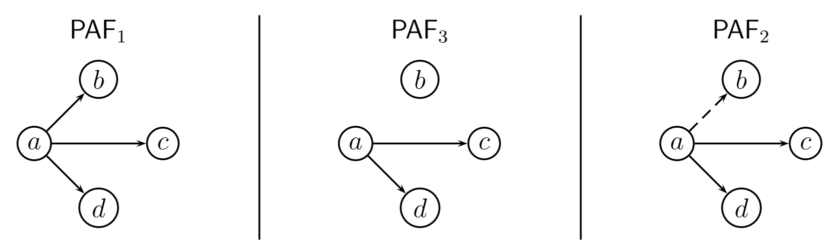
Note that $\bigcup_i \mathsf{AF}_i$ may appear into $\mathtt{exp}_C(\mathcal{P})$, even if $\mathcal{P}$ is discordant. This is illustrated by the following example:

Example 30 *Consider the profile* $\mathcal{P} = \langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *consisting of the following three AFs:*

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

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

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



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

Proposition 31 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs. Let* (a,b) *be a pair of arguments such that* $a, b \in \bigcup_i A_i$ *and* $\nexists \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$ *such that* $(a,b) \in (R_i \setminus R_j) \cup (R_j \setminus R_i)$.

$\exists \mathsf{AF}_l \in \mathcal{P}$ *such that* $(a,b) \in R_l$ *if and only if* $\forall \mathsf{AF}_k \in \mathcal{P}$, $(a,b) \in R'_k$ *with* R'_k *denoting the attack relation of the PAF* $\mathtt{exp}_C(\mathsf{AF}_k, \mathcal{P})$.

Proof : Consider $\mathsf{AF}_k \in \mathcal{P}$. Since $\exists \mathsf{AF}_l \in \mathcal{P}$ such that $(a,b) \in R_l$ and $\nexists \mathsf{AF}_i, \mathsf{AF}_j \in \mathcal{P}$ such that $(a,b) \in (R_i \backslash R_j) \cup (R_j \backslash R_i)$, $(a,b) \notin N_k$; so, $(a,b) \in R'_k$.

The second part of the proof is obvious with a *reductio ad absurdum*: if we suppose that $\nexists \mathsf{AF}_l \in \mathcal{P}$ such that $(a,b) \in R_l$ then we obtain $\forall \mathsf{AF}_k \in \mathcal{P}$, $(a,b) \in N'_k$ which is a contradiction with $\forall \mathsf{AF}_k \in \mathcal{P}$, $(a,b) \in R'_k$. $\square$

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

Definition 32 (Compatibility) *Let* $\mathcal{P} = \langle \mathsf{PAF}_1, \ldots, \mathsf{PAF}_n \rangle$ *be a profile of PAFs over a set of arguments* A. $\mathsf{PAF}_1, \ldots, \mathsf{PAF}_n$ *are said to be* compatible *if and only if they have at least one common completion. Otherwise they are said to be* incompatible.

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

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

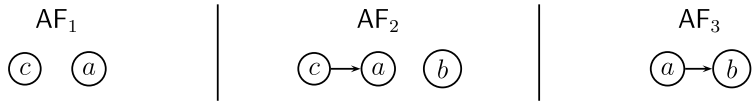
Proposition 33 *Let* $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ *be a profile of AFs.* $\mathcal{P}$ *is concordant if and only if* $\mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P})$ *are compatible.*

Proof : The first part of the proof is obvious: if $\mathcal{P}$ is concordant then the profile $\mathtt{exp}_C(\mathcal{P}) = \langle \mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P}) \rangle$ is reduced to $\langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle$ (see Proposition 29); so, $\mathtt{exp}_C(\mathsf{AF}_1, \mathcal{P}), \ldots, \mathtt{exp}_C(\mathsf{AF}_n, \mathcal{P})$ are equal and have a common completion.

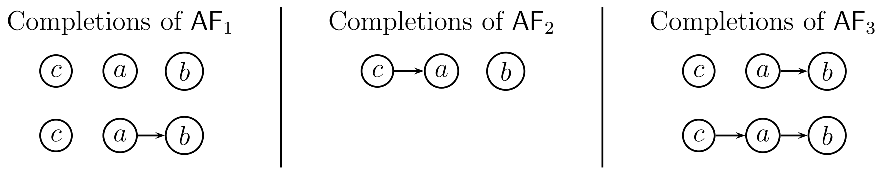
The second part of the proof uses a *reductio ad absurdum*: if we suppose that $\mathcal{P}$ is discordant then $\exists \mathsf{AF}_i, \mathsf{AF}_j$ such that $\exists (a,b) \in R_i \cap N_j$; so, $(a,b) \in R'_i$ with R'_i denoting the attack relation of $\mathtt{exp}_C(\mathsf{AF}_i, \mathcal{P})$ and $(a,b) \in N'_j$ with N'_j denoting the non-attack relation of $\mathtt{exp}_C(\mathsf{AF}_j, \mathcal{P})$; so, all the completions of $\mathtt{exp}_C(\mathsf{AF}_i, \mathcal{P})$ must contain the attack (a,b) and no completion of $\mathtt{exp}_C(\mathsf{AF}_j, \mathcal{P})$ can contain the attack (a,b); so, AF_i and AF_j do not have a common completion which is in contradiction with the hypothesis of compatibility. $\square$

Example 34 *Consider the following argumentation systems* AF_1, AF_2 *and* AF_3.

The completions of their respective consensual expansions PAF_1, PAF_2 *and*



PAF_3 *are:*



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

6.2 Properties of merging operators

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

Proposition 35 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs. Assume that the expansion function used for each agent is the consensual one. If $\mathcal{P}$ is concordant then $\Delta^{\otimes}_{de}(\mathcal{P}) = \{\bigcup_i \mathsf{AF}_i\}$.*

Proof : If $\mathcal{P}$ is concordant, then by Proposition 29, we have $\mathtt{exp}_C(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle) = \langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle$. It remains to show that $\Delta^{\otimes}_{de}(\langle \bigcup_i \mathsf{AF}_i, \ldots, \bigcup_i \mathsf{AF}_i \rangle) = \{\bigcup_i \mathsf{AF}_i\}$, which is obvious since de, as a distance, satisfies the minimality requirement ($\bigcup_i \mathsf{AF}_i$ is the unique PAF at edit distance 0 from itself). □

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

Proposition 36 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of argumentation systems. Assume that the expansion function used for each agent is the consensual one. For any aggregation function $\otimes$, we have that : $\forall \mathsf{AF} = \langle A, R, N \rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle)$:*

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

Proof : Let $CFP(\mathcal{P}) = \langle A_{CFP}, R_{CFP}, N_{CFP} \rangle$ (the ignorance relation does not appear here because argumentation systems (and not partial ones) are considered).

- $A_{CFP} = \bigcup_i A_i \subseteq A = \bigcup_i A_i$.
- By Proposition 25, we know that $CFP(\mathcal{P})$ is pointwise included in each $\mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})$. Let first consider the case $(a, b) \in R_{CFP}$, we have $(a, b) \in R_{\mathsf{exp}_C(\mathsf{AF}_i,\mathcal{P})}$, $\forall \mathsf{AF}_i$.
  Consider $\mathsf{AF} = \langle A, R\rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n\rangle)$. Suppose that $(a, b) \notin R$ and consider $\mathsf{AF}' = \langle A' = A, R' = R \cup \{(a, b)\}\rangle$.
  $\forall \mathsf{AF}_i$, $de(\mathsf{AF}', \mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})) = de(\mathsf{AF}, \mathsf{exp}_C(\mathsf{AF}_i, \mathcal{P})) - 1$, since $(a, b) \in R' \cap R_{\mathsf{exp}_C(\mathsf{AF}_i,\mathcal{P})}$ and $\notin R \cap R_{\mathsf{exp}_C(\mathsf{AF}_i,\mathcal{P})}$; so, since $\otimes$ respects monotonicity, we have $\otimes^n_{i=1} de(\mathsf{AF}', \mathsf{exp}_C(\mathcal{P})) < \otimes^n_{i=1} de(\mathsf{AF}, \mathsf{exp}_C(\mathcal{P}))$ and we obtain a contradiction with $\mathsf{AF} \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n\rangle)$; so, $(a, b) \in R$. Hence $R_{CFP(\mathcal{P})} \subseteq R$.
- In the same way, we can prove that if $(a, b) \in N_{CFP}$ then $(a, b) \in N$. So $N_{CFP(\mathcal{P})} \subseteq N$.

$\square$

As a direct corollary of Propositions 23 and 36, we get that:

Corollary 37 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n\rangle$ be a profile of argumentation systems. Assume that the expansion function used for each agent is the consensual one. For any aggregation function $\otimes$, we have that: $\forall \mathsf{AF} = \langle A, R, N\rangle \in \Delta^{\otimes}_{de}(\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n\rangle)$:*

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

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

Definition 38 (Majority PAF) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n\rangle$ be a profile of AFs over the same set A of arguments. The* majority PAF *$MP(\mathcal{P})$ of $\mathcal{P}$ is the triple $\langle R, N, I\rangle$ such that $\forall a, b \in A$:*[4]

- $(a, b) \in R$ *if and only if* $\#(\{i \in 1 \ldots n \mid (a, b) \in R_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in N_i\})$;
- $(a, b) \in N$ *if and only if* $\#(\{i \in 1 \ldots n \mid (a, b) \in N_i\}) > \#(\{i \in 1 \ldots n \mid (a, b) \in R_i\})$;

[4] For any set S, $\#(S)$ denotes the cardinality of S.

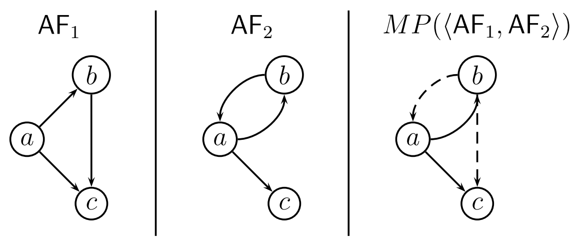
- $(a, b) \in I$ *otherwise.*

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

Proposition 39 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set A of arguments. The majority PAF $MP(\mathcal{P})$ of $\mathcal{P}$ is a PAF.*

Proof : Obvious since by construction, R and I are disjoint sets and N is the complement of $R \cup I$ into $A \times A$. □

Example 40 *Consider* $\mathsf{AF}_1 = \langle \{a, b, c\}, \{(a, b), (b, c), (a, c)\} \rangle$, $\mathsf{AF}_2 = \langle \{a, b, c\}, \{(a, b), (b, a), (a, c)\} \rangle$.



We have $MP(\langle \mathsf{AF}_1, \mathsf{AF}_2 \rangle) = \langle \{a, b, c\}, \{(a, b), (a, c)\}, \{(b, c), (b, a)\}, \{(a, a), (b, b), (c, c), (c, a), (c, b)\} \rangle$.

Proposition 41 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set A of arguments. $\Delta_{de}^{\Sigma}(\mathcal{P}) = \mathcal{C}(MP(\mathcal{P}))$.*

Proof : The key is that the edit distance between an AF denoted by AF and a profile of AFs over A when Σ is the aggregation operator is the sum over the AF_i of the profile of the sum over every ordered pair of arguments over A of the edit distances between AF and AF_i (this is a consequence of the associativity of the sum).

Let AF be an AF over A which minimizes $\sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$. Let $a, b \in A$. If $\#(\{i \in 1 \ldots n \mid (a, b) \in R_i\} > \#(\{i \in 1 \ldots n \mid (a, b) \in N_i\}$, then (a, b) must be in the attack relation of AF; otherwise, the AF AF' over A which coincides with AF except that (a, b) is in the attack relation of AF' would be such that $\sum_{i=1}^{n} de(\mathsf{AF}', \mathsf{AF}_i) < \sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$. Similarly, if $\#(\{i \in 1 \ldots n \mid (a, b) \in N_i\} > \#(\{i \in 1 \ldots n \mid (a, b) \in R_i\}$, then (a, b) must not be in the attack relation of AF.

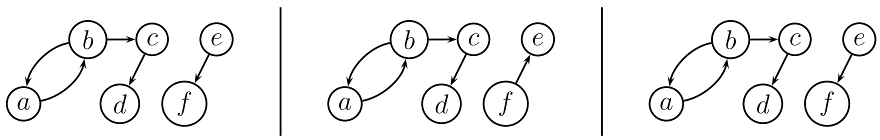
In the remaining case, *i.e.*, when $\#(\{i \in 1 \ldots n \mid (a, b) \in R_i\} = \#(\{i \in 1 \ldots n \mid (a, b) \in N_i\}$, let AF' be the AF over A which coincides with AF except that (a, b) is in the attack relation of AF' if and only if (a, b) is not in the attack relation of AF. Then $\sum_{i=1}^{n} de(\mathsf{AF}', \mathsf{AF}_i) = \sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$.

This shows that every AF over A which minimizes $\sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$ is a completion of the majority PAF $MP(\mathcal{P})$.

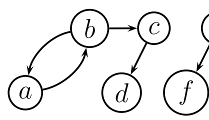
Conversely, since every completion AF' of $MP(\mathcal{P})$ is such that $\sum_{i=1}^{n} de(\mathsf{AF}', \mathsf{AF}_i) = \sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$ where AF minimizes $\sum_{i=1}^{n} de(\mathsf{AF}, \mathsf{AF}_i)$, the conclusion follows. $\square$

Let us illustrate the previous proposition on Example 7:

Example 7 (continued) *The consensual expansions of* AF_1*,* AF_2 *and* AF_3 *are respectively:*



So, the majority PAF of $\langle \mathsf{AF}_1, \mathsf{AF}_2, \mathsf{AF}_3 \rangle$ *is:*



Using the edit distance and sum as the aggregation function, this PAF also represents the result of the merging in the sense that the latter is the set of all completions of this PAF.

Computing the majority PAF of a profile of AFs over the same set of arguments amounts to *voting on the attack relations* associated to each AF. As explained in Section 3, this can prove more suited to our goal than the approach which consists in voting directly on the acceptable sets of arguments for each agent. The previous proposition shows that such a simple voting approach corresponds to a specific merging operator in our framework (but many other operators, especially arbitration ones, can also be used).

7 Acceptability for Merged AFs

Starting from a profile of AFs (over possibly different sets of arguments), a merging operator enables the computation of a set of AFs (this time, over the same set of arguments) which are the best candidates to represent the AFs of the group (a kind of "consensus").

There is an important epistemic difference between those two sets of AFs, the first one reflects different points of view given by different agents (and it can be the case that two distinct agents give the same AF), while the second set expresses some uncertainty on the merging due to the presence of conflicts.

Let us recall that the main goal of this paper is to characterize the sets of arguments acceptable by the whole group of agents. In order to achieve it, it remains to define some mechanisms for exploiting the resulting set of AFs. This calls for a notion of joint acceptability.

Definition 42 (Joint acceptability) *A* joint acceptability relation *for a profile $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ of AFs, denoted by $Acc_{\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle}$, is a total function from $2^{\bigcup_i A_i}$ to $\{true, false\}$ which associates each subset E of $\bigcup_i A_i$ with true if E is a jointly acceptable set for $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ and with false otherwise.*

For instance, a joint acceptability relation for a profile $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ can be defined by the acceptability relations Acc_{AF_i} (based themselves on some semantics and some selection principles), which can coincide for every AF_i (but this is not mandatory) and a voting method $V : \{true, false\}^n \mapsto \{true, false\}$:

$$Acc_{\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle}(E) = V(Acc_{\mathsf{AF}_1}(E), \ldots, Acc_{\mathsf{AF}_n}(E)).$$

Here are some instances of Definition 42 based on voting methods:

Definition 43 (Acceptabilities for profiles of AFs) *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of n AFs over the same set of arguments A. Let Acc_{AF_i} be the (local) acceptability relation associated with AF_i. If $n = 1$, then we define $Acc_{\langle \mathsf{AF}_1 \rangle} = Acc_{\mathsf{AF}_1}$. Otherwise, for any subset S of A, we say that:*

- *S is* skeptically jointly acceptable for *$\mathcal{P}$ if and only if S is included in at least one acceptable set for each AF_i:*
  $\forall \mathsf{AF}_i \in \mathcal{P}, \exists E_i$ *such that* $Acc_{\mathsf{AF}_i}(E_i)$ *is true and* $S \subseteq E_i$.
- *S is* credulously jointly acceptable for *$\mathcal{P}$ if and only if S is included in at least one acceptable set for at least one AF_i:*
  $\exists \mathsf{AF}_i \in \mathcal{P}, \exists E_i$ *such that* $Acc_{\mathsf{AF}_i}(E_i)$ *is true and* $S \subseteq E_i$.
- *S is* jointly acceptable by majority for *$\mathcal{P}$ if and only if S is included in at least one acceptable set for at least a weak majority of AF_i:*
  $\#(\{\mathsf{AF}_i \mid \exists E_i$ *such that* $Acc_{\mathsf{AF}_i}(E_i)$ *is true and* $S \subseteq E_i\}) \geq \frac{n}{2}$.

Obviously enough, when none of the local acceptabilities Acc_{AF_i} is trivial (*i.e.*, equivalent to the constant function $false$) for the profile under consideration, we have that any set of arguments which is skeptically jointly acceptable is also jointly acceptable by majority, and that any set of arguments which is jointly acceptable by majority is also credulously jointly acceptable.

Note that skeptical (resp. credulous) joint acceptability does not require that the skeptical (resp. credulous) inference principle is at work for defining local acceptabilities Acc_{AF_i}, which remain unconstrained.

Focusing on the preferred semantics together with credulous local acceptabilities, let us re-consider some previous examples:

Example 20 (continued) *Using the edit distance and* $\otimes = \mathcal{L}\texttt{eximax}$ *(or* $\mathcal{M}\mathsf{ax}$*) as the aggregation function, we get two AFs* AF'_1 *and* AF'_2 *in the merging.*

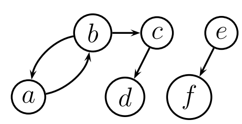
If the local acceptability relations are based on credulous inference from preferred extensions, we have:

- $Acc_{\mathsf{AF}'_1}(E) =$ *true if and only if* $E \subseteq \{c, d\}$ *or* $E \subseteq \{b, e\}$*;*
- $Acc_{\mathsf{AF}'_2}(E) =$ *true if and only if* $E \subseteq \{a, c, e\}$*.*

$\{c\}$ *and* $\{e\}$ *are skeptically jointly acceptable and* $\{b, e\}$,$\{c, d\}$ *and* $\{a, c, e\}$ *(and their subsets) are credulously (and by majority) jointly acceptable for the merging.*

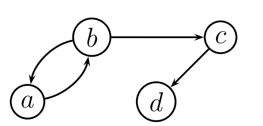
Using this method, the argument a can still be derived credulously, contrariwise to what happens when the union of the two AFs AF_1 *and* AF_2 *is considered.*

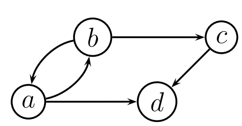
Example 7 (continued) *Using the edit distance and the sum as the aggregation function, we get one AF in the merging, denoted* AF*:*



AF *has two preferred extensions :* $\{a, c, e\}$ *and* $\{b, d, e\}$*. So,* $Acc_{\mathsf{AF}}(E) =$ *true if and only if* $E \subseteq \{a, c, e\}$ *or* $E \subseteq \{b, d, e\}$*. The three joint acceptability relations coincide here (as there is only one AF in the result). The sets* $\{a, c, e\}$ *and* $\{b, d, e\}$ *(and their subsets) are credulously, skeptically and by majority, jointly acceptable for the merging, which is a more sensible result that the one obtained using a voting method on the derived arguments of the initial AFs (as explained in Section 3).*

Example 14 (continued) *Using the edit distance and the sum as the aggregation function, we get two AFs in the merging:*





The preferred extensions for these 2 AFs coincide (they are $\{a, c\}$ *and* $\{b, d\}$*). As the preferred extensions for the 2 AFs are the same ones, the three relations* *of joint acceptability coincide here. Thus, the sets $\{a, c\}$ and $\{b, d\}$ (and their subsets) are skeptically, credulously and by majority jointly acceptable for the merging.*

It is interesting to compare the joint acceptability relation for the input profile $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ with the joint acceptability relation for the merging $\Delta_d^{\otimes}(\mathcal{P})$. Unsurprisingly, both predicates are not logically connected (*i.e.*, none of them implies the other one), even in the case when the two joint acceptability relations are based on the same notion of local acceptability (for instance, considering a set of arguments E as acceptable for an AF when it is included in at least one of its preferred extensions) and the same voting method (for instance, the simple majority rule).

Thus, it can be the case that new jointly acceptable sets are obtained after merging while they were not jointly acceptable at start:

Proposition 44 *Let $\mathcal{P} = \langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ be a profile of AFs over the same set of arguments A. The set of all jointly acceptable sets for the profile $\mathcal{P}$ is not necessarily equal to the set of all jointly acceptable sets for the merging of $\mathcal{P}$.*

A counter-example is given by Example 14.

When each local acceptability relation corresponds exactly to the collective acceptability proposed by Dung (for a given semantics and $\forall \mathsf{AF}_i$, $Acc_{\mathsf{AF}_i}(E) = true$ if and only if E is an extension of AF_i for this semantics), the following remarks can be done:

- If a set of arguments is included in *one* of the acceptable sets for an agent, it is not necessarily included into one of the acceptable sets of any AF from the merging (and it also holds for singletons). The converse is also true.
- More surprisingly, even if a set of arguments is included into *each* acceptable set for an agent, it is not guaranteed to be included into an acceptable set of an AF from the merging. Conversely, if a set of arguments is included into every acceptable set of the AFs from the merging, it is not guaranteed to be included into an acceptable set for one of the agents. Intuitively, this can be explained by the fact that if an argument is accepted by all agents *for bad reasons* (for instance, because they lack information about attacks on it), it can be rejected by the group after the merging. More formally, this is due to the fact that nothing ensures that one of the initial AFs will belong to the result of the merging and also to the fact that acceptability is nonmonotonic (in the sense that adding a single attack (a, b) in an AF may drastically change its extensions).

8 Conclusion and Perspectives

We have presented a framework for deriving sensible information from a collection of argumentation systems *à la* Dung. Our approach consists in merging such systems. The proposed framework is general enough to allow for the representation of many different scenarios. It is not assumed that all agents must share the same sets of arguments. No assumption is made concerning the meaning of the attack relations, so that such relations may differ not only because agents have different points of view on the way arguments interact but more generally may disagree on what an interaction is. Each agent may be associated to a specific expansion function, which enables for encoding many attitudes when facing a new argument. Many different distances between PAFs and many different aggregation functions can be used to define argumentation systems which best represent the whole group.

By means of example, we have shown that our merging-based approach leads to results which are much more expected than those furnished by a direct vote on the (sets of) arguments acceptable by each agent. We have also shown that union cannot be taken as a valuable merging operator in the general case. We have investigated formally some properties of the merging operators which we point out. Among other results, we have shown that merging operators based on the edit distance preserve all the information on which all the agents participating in the merging process agree, and more generally, all the information on which the agents participating in the merging process do not disagree. We have also shown that the merging operator based on the edit distance and the sum as aggregation function is closely related to the merging approach which consists in voting on the attack relations when the input profile gathers argumentation systems over the same set of arguments. Finally, we have proven that in the general case, the derivable sets of arguments when joint acceptability concerns the input profile may drastically differ from the the derivable sets of arguments when joint acceptability concerns the profile obtained after the merging step.

We plan to refine our framework in several directions:

Merging PAFs. Our framework can be extended to PAFs merging (instead of AFs). This enables us to take into account agents with incomplete belief states regarding the attack relation between arguments. Expansions of PAFs can be defined in a very similar way to expansions of AFs (what mainly changes is the way ignorance is handled). As PAFs are more expressive than AFs, an interesting issue for further research is to define acceptability for PAFs.

Attacks strengths. Assume that each attack believed by Agent i is associated to a numerical value reflecting the strength of the attack according to the agent, *i.e.*, the degree to which Agent i believes that a attacks b. It is easy to take into account those values by modifying slightly the definition of the edit distance over an ordered pair of arguments (for instance, viewing such values as weights once normalized within $[0, 1]$). Another possibility regarding attack strengths is, from unweighted attack relations, to generate a weighted one, representing different degrees of accordance in the group. For instance, each attack (a, b) in the majority PAF of a profile $\langle \mathsf{AF}_1, \ldots, \mathsf{AF}_n \rangle$ can be labelled by the ratio $\frac{\#(\{i \in 1 \ldots n | (a,b) \in R_i\})}{n}$ and similarly for the non-attack relation (this leads to consider both the attack and the non-attack relations of the majority PAF as fuzzy relations). Corresponding acceptability relations remain to be defined. This is another perspective of this work.

Merging audiences. In (7), an extension of the notion of AF, called valued AF — VAF for short —, has been proposed in order to take advantage of values representing the agent's preferences in the context of a given audience. A further perspective of our work concerns the merging of such VAFs.

Acknowledgments

The authors are grateful to the anonymous referees for their helpful comments. This work has been partly supported by the Région Nord/Pas-de-Calais, the IRCICA consortium and by the European Community FEDER Program.

References

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

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

[3] L. Amgoud, N. Maudet, and S. Parsons. Arguments, Dialogue and Negotiation. In *Proc. of ECAI'00*, pages 338–342, 2000.

[4] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In *Intelligent Agents VIII: Agent theories, architectures and languages*, pages 190–205, 2002.

[5] L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In *Proc. of KR'04*, pages 175–182, 2004.

[6] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[7] T. Bench-Capon, S. Doutre, and P.E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171:42–71, 2007.

[8] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128 (1-2):203–235, 2001.
[9] A. Bochman. Collective argumentation and disjunctive programming. *Journal of Logic and Computation*, 13 (3):405–428, 2003.
[10] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.
[11] C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proc. of ECSQARU'05 – LNAI 3571*, pages 378–389. Springer-Verlag, 2005.
[12] S. Coste-Marquis, C. Devred, and P. Marquis. Symmetric argumentation frameworks. In *Proc. of ECSQARU'05 – LNAI 3571*, pages 317–328, 2005.
[13] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
[14] P. Dunne and T. Bench-Capon. Coherence in finite argument systems. *Artificial Intelligence*, 141:187–203, 2002.
[15] 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.
[16] M. Elvang-Gøransson, J. Fox, and P. Krause. Dialectic reasoning with inconsistent information. In *Proc. of UAI'93*, pages 114–121, 1993.
[17] M. Elvang-Gøransson and A. Hunter. Argumentative logics: Reasoning with classically inconsistent information. *Data and Knowledge Engineering*, 16:125–145, 1995.
[18] T. F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice.* Kluwer Academic, 1995.
[19] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261, 1999.
[20] S. Konieczny and R. Pino Pérez. Merging information under constraints: a qualitative framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.
[21] P. Krause, S. Ambler, M. Elvang, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.
[22] J. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical logic*, 8:117–133, 1979.
[23] S. Nielsen and S. Parsons. Computing preferred extensions for argumentation systems with sets of attacking arguments. In *Proc. of COMMA'06*, pages 97–108. IOS Press, 2006.
[24] S. Parsons and N. R. Jennings. Negotiation through argumentation—a preliminary report. In *Proc. of ICMAS'96*, pages 267–274, 1996.
[25] J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133:233–282, 2001.
[26] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.
[27] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In *Handbook of Philosophical Logic*, volume 4, pages 218–319. Kluwer Academic, 2002.
[28] D. Walton and E. Krabbe. *Commitment in dialogue. Basic concepts of interpersonal reasoning.* State University of New York Press, 1995.

BELIEF BASE MERGING AS A GAME

Sébastien Konieczny.
Journal of Applied Non-Classical Logics. 14(3).
pages 275-294.
2004.

Belief base merging as a game

Sébastien Konieczny

CRIL - CNRS
Université d'Artois
62300 Lens (France)
konieczny@cril.univ-artois.fr

ABSTRACT. *We propose in this paper a new family of belief merging operators, that is based on a game between sources : until a coherent set of sources is reached, at each round a contest is organized to find out the weakest sources, then those sources has to concede (weaken their point of view). This idea leads to numerous new interesting operators (depending of the exact meaning of "weakest" and "concede", that gives the two parameters for this family) and opens new perspectives for belief merging. Some existing operators are also recovered as particular cases. Those operators can be seen as a special case of Booth's Belief Negotiation Models [BOO 02], but the achieved restriction forms a consistent family of merging operators that worths to be studied on its own.*

KEYWORDS: *belief merging, belief negotiation.*

1. Introduction

The problem of (propositional) belief merging [REV 97, LIN 99, LIB 98, KON 99, KON 02a, KON 04] can be summarized by the following question: given a set of sources (propositional belief bases) that are (typically) mutually inconsistent, how do we obtain a coherent belief base reflecting the beliefs of the set?

The idea here is that some/each sources has to concede on some points in order to solve the conflicts. If one has some notion of relative reliability between sources, it is enough and sensible to force the less reliable ones to give up first. There is a variety of different means to do that, which has provided a large literature, e.g. [CHO 93, CHO 95, CHO 98, BEN 98a, BEN 98b]. But often we do not have such information, and even if we get it, it remains the more fundamental problem of how to merge sources of equal reliability [KON 99, KON 02a].

In this paper we will investigate the merging methods based on a notion of game between the sources. The intuitive idea is simple: when trying to impose its wish, each source will try to form some coalition with the closest (more compatible) other sources. So the source that is the "furthest" from the other ones will certainly be the weakest one. And it will be that source that will have to concede first. In this work, we will not focus on how the coalitions form, we only take this idea to designate the *weakest* ones.

So the merging is based on the following game: until a coherent set of sources is reached, at each round a contest is organized to find out the weakest sources, then those sources have to concede (weaken their point of view).

We can state several intuitions and justifications for the use of such operators. We have already given the first one: coalition with near-minded sources. In a group decision process between rational sources, it can be sensible to expect the sources to look for near-minded sources in order to find help to defend their view, so the "furthest" source is the more likely to have to concede on its view.

A second intuition is the one given by a social pressure on the sources. When confronting several points of view, usually people that have the more exotic views try to change their opinion in order to be accepted by the other members of the group, so opinions that are defended by the least number of sources are usually given up more easily in the negotiation process.

A last intuition that gives the main rationale for that kind of operator is Condorcet's Jury theorem. This theorem states that if all the members of a jury are reliable (in the sense that they have greater than 50% chance to find the truth), then listening to the majority is the more rational choice.

After stating some useful definitions and notations in Section 2, we will define the new family of operators we propose in Section 3. The definition will use a notion of weakening and choice functions. We will explore these notions in Section 4. We will give some examples of specific operators in Section 5 in order to illustrate their behaviour. We will look at the logical properties of those operators in Section 6. In Section 7 we will look at the links between this work and related works (especially Booth's proposal [BOO 01, BOO 02]). We conclude in Section 8 with some open issues and perspectives of this work.

2. Definitions

We consider a propositional language $\mathcal{L}$ over a finite alphabet $\mathcal{P}$ of propositional symbols. An interpretation is a function from $\mathcal{P}$ to $\{0, 1\}$. The set of all the interpretations is denoted $\mathcal{W}$. An interpretation ω is a model of a formula φ, noted $\omega \models \varphi$, if and only if it makes it true in the usual classical truth functional way. Let φ be a formula, $mod(\varphi)$ denotes the set of models of φ, i.e. $mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$. Conversely, let X be a set of interpretations, $form(X)$ denotes the formula (up to logical equivalence) whose set of models is X.

A *belief base* φ is a consistent propositional formula (or, equivalently, a finite consistent set of propositional formulae considered conjunctively). Let us note $\mathcal{K}$ the set of all belief bases.

Let $\varphi_1, \ldots, \varphi_n$ be n belief bases (not necessarily different). We call *belief profile* the multi-set Ψ consisting of those n belief bases: $\Psi = (\varphi_1, \ldots, \varphi_n)$ (i.e. two sources can have the same belief base). We note $\bigwedge \Psi$ the conjunction of the belief bases of Ψ, i.e. $\bigwedge \Psi = \varphi_1 \wedge \cdots \wedge \varphi_n$. We say that a belief profile is consistent if $\bigwedge \Psi$ is consistent. The multi-set union will be noted $\sqcup$ and the multi-set inclusion will be noted $\sqsubseteq$. The cardinal of a finite (multi-)set Ψ is noted $\#(\Psi)$ (the cardinal of a finite multi-set is the sum of the numbers of occurrences of each of its elements).

Let us define those multi-set notions more formally. As a set Φ (from elements in a set A) can be defined from its characteristic function $\chi_\Phi : A \to \{0, 1\}$, a multi-set Ψ (from elements in a set A) can be defined from its characteristic function $\chi_\Psi : A \to \mathbb{N}$, where $\mathbb{N}$ is the set of nonnegative integers[1]. Then the multi-set union, noted $\sqcup$, is defined as $\chi_{\Psi \sqcup \Psi'} = \chi_\Psi + \chi_{\Psi'}$. The cardinality of a multi-set is defined as $\#(\Psi) = \sum_{a \in A} \chi_\Psi(a)$.

Indeed, all set notions used in this paper (subset, inclusion, union, etc.), are for multi-sets. For the sake of simplicity, and since it cannot lead to confusion since those notions are a generalization of the set ones, we will omit the "multi-".

Let $\mathcal{E}$ be the set of all finite non-empty belief profiles.

Two belief profiles Ψ_1 and Ψ_2 are said to be equivalent ($\Psi_1 \equiv \Psi_2$) if and only if there is a bijection between Ψ_1 and Ψ_2 such that each belief base of Ψ_1 is logically equivalent to its image in Ψ_2.

3. Belief game model

In [BOO 01, BOO 02] Richard Booth proposes a framework for merging sources of information incrementally. He named this framework *"Belief Negotiation Model"* (BNM). In this work we will use the name *"Belief Game Model"* (BGM) because in our framework there is no room for negotiation, so we find it more accurate and it allows us to make a distinction in this paper between Booth's proposal and ours. The BGM framework can be seen as a restriction of Booth's BNM framework: the main differences between Booth's proposal and our is that Booth's one take the sources as candidates to weakening, whereas we restrict ourselves to *"points of view"* (logical content of the sources). That means that in Booth's if one source has to weaken, it can be the case that another source with exactly the same beliefs do not have to weaken too (that is not allowed in our framework). Our proposal involves more anonymity by saying that only beliefs decide who has to weaken, not the identity of one source. Similarly, the choice functions are more "Markovian" in our framework than in Booth's one, that means that we only look at the current profile to choose the bases, whereas in

1. if $a \in A$, $\chi_\Psi(a) = n$ intuitively means that a appears n times in the multi-set Ψ.

Booth's work, one can use the whole history of profiles to make the choice. We think that those hypothesis are more realistic (and necessary) on a belief merging point of view, whereas Booth's framework allows to model more generalized negotiation schemes, where one can decide for example that each source has to weaken one after the other (see Section 7 for a deeper comparison of the two approaches).

DEFINITION 1. — *A* choice function *is a function $g : \mathcal{E} \to \mathcal{E}$ such that:*

– $g(\Psi) \sqsubseteq \Psi$

– *If* $\bigwedge \Psi \not\equiv \top$*, then* $\exists \varphi \in g(\Psi)$ *s.t.* $\varphi \not\equiv \top$

– *If* $\Psi \equiv \Psi'$*, then* $g(\Psi) \equiv g(\Psi')$

The choice function aims to find which are the sources that must weaken at a given round (see definition 3). As the weakening function aims to weaken the belief base, and as there is no weaker base than a tautological one, the second condition states that at least one non-tautological base must be selected. So it states that at each round at least one base will be weaken. This condition is necessary to ensures to always reach a result with Belief Game Model. Note that a consequence of this condition is that we have $g(\Psi) \neq \emptyset$ as soon as the profile contains at least one non-tautological base. Last condition is an irrelevance of syntax condition. It states that the selection of the bases to weaken does not depend on the particular form of the bases, but only on their informational content. Note that we also have an additional property: anonymity, that means that the result does not depend on the "name" of the source, but only on its point of view. This is due to the fact that we work with multi-sets, that are equivalent by permutation. If one works with another representation (ordered lists of sources for example), this anonymity property can be given by the last condition, provided that the equivalence between two belief profiles is rightly defined (as in Section 2).

DEFINITION 2. — *A* weakening function *is a function $\blacktriangledown : \mathcal{L} \to \mathcal{L}$ such that:*

– $\varphi \vdash \blacktriangledown(\varphi)$

– *If* $\varphi \equiv \blacktriangledown(\varphi)$*, then* $\varphi \equiv \top$

– *If* $\varphi \equiv \varphi'$*, then* $\blacktriangledown(\varphi) \equiv \blacktriangledown(\varphi')$

The weakening function aims to give the new beliefs of a source that have been chosen to be weakened. The two first conditions ensure that the base will be replaced by a strictly weaker one (unless the base is already a tautological one). The last condition is an irrelevance of syntax requirement : the result of the weakening must only depend on the information conveyed by the base, not on its syntactical form.

We extend the weakening functions on belief profiles as follows: let Ψ' be a subset of Ψ,

$$\blacktriangledown_{\Psi'}(\Psi) = \bigsqcup_{\varphi \in \Psi'} \blacktriangledown(\varphi) \sqcup \bigsqcup_{\varphi \in \Psi \setminus \Psi'} \varphi$$

This means that we only weaken the belief bases of Ψ that are in Ψ', and the other ones do not change.

DEFINITION 3. — *A Belief Game Model is a pair* $\mathcal{N} = \langle g, \blacktriangledown \rangle$ *where* g *is a choice function and* $\blacktriangledown$ *is a weakening function.*

The solution to a belief profile Ψ *for a Belief Game Model* $\mathcal{N} = \langle g, \blacktriangledown \rangle$*, noted* $\mathcal{N}(\Psi)$*, is the belief profile* $\Psi_{\mathcal{N}}$*, defined as:*

– $\Psi_0 = \Psi$

– $\Psi_{i+1} = \blacktriangledown_{g(\Psi_i)}(\Psi_i)$

– $\Psi_{\mathcal{N}}$ *is the first* Ψ_i *that is consistent*

So the solution to a belief profile is the result of a game on the beliefs of the sources. At each round there is a contest to find out the weakest bases (the losers), and the losers have to concede on their belief by weakening them.

It may prove reasonable that each source has its own weakening function, that denotes different conceding politics. After all, the point is that the source has to weaken its beliefs, not how she does so. So we can figure out a generalization of the belief game model, where there is no one weakening function, but one for each source[2]. But this is not the point in this paper, so we will suppose that there is a unique weakening function for all the sources.

In some cases, the result of the merging has to obey some constraints (physical constraints, norms, etc...). We will assume that these integrity constraints are encoded as a propositional formula (a belief base), and we will note this base μ. Then we introduce the following notion:

DEFINITION 4. — *The solution to a belief profile* Ψ *for a Belief Game Model* $\mathcal{N} = \langle g, \blacktriangledown \rangle$ *under the integrity constraints* μ*, noted* $\mathcal{N}_\mu(\Psi)$*, is the belief profile* $\Psi^{\mu}_{\mathcal{N}}$ *defined as:*

– $\Psi_0 = \Psi$

– $\Psi_{i+1} = \blacktriangledown_{g(\Psi_i)}(\Psi_i)$

– $\Psi^{\mu}_{\mathcal{N}}$ *is the first* Ψ_i *that is consistent with* μ

Often in the following in this paper we will call result of the merging operator (Belief Game Model), the belief base $\bigwedge \Psi^{\mu}_{\mathcal{N}} \wedge \mu$. This abuse of notation is not problematic, since this belief base denotes the consensus point obtained by the belief profile $\Psi^{\mu}_{\mathcal{N}}$ solution of the Belief Game Model process.

Note that the definition of the Belief Game Model and of the weakening and choice functions ensures that each belief profile Ψ has a solution as soon as the constraints μ are consistent.

2. Technically, it forces to drop out the weakening function from the belief game model $\mathcal{N}$ and to put it in the input, i.e. the input would be a list of sources i that are couples compound of a belief base and a weakening function: $\langle \varphi_i, \blacktriangledown_i \rangle$

THEOREM 5. — *Let Ψ be a belief profile, and μ be a belief base. If Ψ is non-empty and μ is consistent, then $\bigwedge \Psi_{\mathcal{N}}^{\mu} \wedge \mu$ is consistent and $\Psi_{\mathcal{N}}^{\mu}$ is reached for a finite number of rounds.*

To prove that $\Psi_{\mathcal{N}}^{\mu}$ is reached for a finite number of rounds, it is sufficient to note that, at a given round, either Ψ_i is consistent with μ, so $\Psi_{\mathcal{N}}^{\mu} = \Psi_i$, and the process end, so we are done. Or Ψ_i is not consistent with μ, in this case there is an other round. From the defintion of the weakening of a profile, the resulting profile Ψ_{i+1} is logically strictly weaker than Ψ_i, that means that each base of Ψ_{i+1} is either logically equivalent to the corresponding base in Ψ_i, or logically strictly weaker than this base (see definition 2). Now just note that the logically weakest profile, is the one where each base is equivalent to $\top$, and that this profile is achievable from every given belief profile by successive applications of the weakening function. Since we work in a finite propositional logic setting, this can be done by a finite number of rounds. Finally, note that this profile is consistent with every consistent integrity constraint μ. So if the process described in definition 4 has not stopped before, it is guaranteed to give a result with this belief profile.

4. Weakening and choice functions

In order to define a particular Belief Game Model, we have to choose a choice function and a weakening function. We will give in this section some natural choices for these functions and see what are the resulting BGM operators.

4.1. *Weakening functions*

Let us first turn out to weakening functions. Can we find a "natural" one ? In fact it is a difficult task, since the exact choice of a weakening function depends on the expected behaviour for the Belief Game Model and depends also on the existence of some "preferential" information. But if we have no such additional information, we have at least two natural candidates : drastic weakening and dilation.

DEFINITION 6. — *Let φ be a belief base. The* drastic weakening function *forget all the information about one source, i.e. :* $\blacktriangledown_{\top}(\varphi) = \top$.

This weakening function simply forget all the information in φ !

After this rough function, let us see a more fine grained one. Let us first recall what is the Hamming's distance between interpretations (also called Dalal's distance [DAL 88]) since we will use it several times in this paper.

DEFINITION 7. — *The Hamming distance between interpretations is the number of propositional symbols on which the two interpretations differ. Let ω and ω' be two interpretations, then*

$$d_H(\omega, \omega') = \#(\{a \in \mathcal{P} \mid \omega(a) \neq \omega'(a)\})$$

Then the dilation weakening function is defined as :

DEFINITION 8. — *Let* φ *be a belief base. The* dilation weakening function *is defined as :*

$$mod(\blacktriangledown_\delta(\varphi)) = \{\omega \in \mathcal{W} \mid \exists \omega' \models \varphi \; d_H(\omega, \omega') \leq 1\}$$

This weakening function takes as models of the weakened base $\blacktriangledown_\delta(\varphi)$, all the models that are at an Hamming distance less or equal to 1 from the models of φ, i.e. all the models that are in the base φ and all the models that are achievable by flipping the truth value of one propositional symbol in a model of φ. This is close to the dilation operator used in morpho-logics [BLO 00, BLO 04].

4.2. *Choice functions*

Let us now turn to choice functions. The aim of this function is to determine the "losers", that are the sources that have to concede by weakening their beliefs at a given round.

One of the simplest choice functions is identity (denoted g_{id}). It is not the expected behaviour for this function, but it can prove the rationality of our operators if, even in this case, we obtain a sensible merging.

We will focus on two families of choice functions. The first one is model-based, the second one is formula-based. We think that most of the sensible choice functions belong to one of those families.

4.2.1. *Model-based choice functions*

We will focus here on some modelizations of what can be called "social pressure", and can be viewed as a majority principle. Namely, at each round it is the "furthest" sources from the group that will concede. The exact choice of the meaning of "furthest" will fix the chosen operator for this family. Technically we will use a distance between belief bases and an aggregation function to evaluate the distance of a belief base with respect to the others.

We will start from the definition of the distance between two belief bases.

DEFINITION 9. — *A (pseudo)distance*[3] d *between two belief bases is a function* $d : \mathcal{L} \times \mathcal{L} \rightarrow \mathbb{N}$ *such that:*

– $d(\varphi, \varphi') = 0$ *iff* $\varphi \wedge \varphi' \nvdash \bot$

– $d(\varphi, \varphi') = d(\varphi', \varphi)$

3. Remark that we miss an important property of distances: we have only $d(\varphi, \varphi') = 0$ *if* $\varphi = \varphi'$, but not the *only if* part. Remark also that we do not require the triangular inequality.

Two examples of such distances are :

– $d_D(\varphi, \varphi') = \begin{cases} 0 & \text{if } \varphi \wedge \varphi' \nvdash \bot \\ 1 & \text{otherwise} \end{cases}$

– $d_H(\varphi, \varphi') = \min_{\omega \models \varphi, \omega' \models \varphi'} d_H(\omega, \omega')$

DEFINITION 10. — *An aggregation function is a total function f associating a nonnegative integer to every finite tuple of nonnegative integers and verifying (non-decreasingness), (minimality) and (identity).*

– *if $x \leq y$, then $f(x_1, \ldots, x, \ldots, x_n) \leq f(x_1, \ldots, y, \ldots, x_n)$.* *(non-decreasingness)*

– $f(x_1, \ldots, x_n) = 0$ *if and only if* $x_1 = \ldots = x_n = 0$. *(minimality)*

– *for every nonnegative integer x, $f(x) = x$.* *(identity)*

We say that an aggregation function is symmetric if it also satisfies :

– *For any permutation σ, $f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$* *(symmetry)*

DEFINITION 11. — *A model-based choice function $g^{d,h}$ is defined as :*

$$g^{d,h}(\Psi) = \{\varphi_i \in \Psi \mid h(d(\varphi_i, \varphi_1), \ldots, d(\varphi_i, \varphi_n)) \text{ is maximal } \}$$

where h is an aggregation function, and d is a distance between belief bases.

We say that the model-based choice function is symmetric if the aggregation function is symmetric.

We will focus on some specific aggregation functions in this paper, but we can use different aggregation functions here. In particular we will only focus on symmetrical aggregation functions in this paper (to fit with choice function requirements) but note that the definition allows non-symmetrical functions. This allows to define operators that are not anonymous, i.e. where each base has not the same importance. So one can use priorities (a weight or a pre-order on the sources) for denoting different level of reliability, different hierarchical importance, etc.

We will use in the following as examples of aggregation functions, two typical ones, the sum (noted Σ) and the maximum (noted max).

4.2.2. *Formula-based choice functions*

Not all interesting choice functions are captured in the definition given in the previous section. In particular, a lot of interesting choice functions can be defined by using maximal consistent subsets. Note, however that, conversely to usual formula-based merging operators [BAR 92, KON 00], we use multi-sets instead of simple sets.

DEFINITION 12. — *Let* MAXCONS(Ψ) *be the set of the* maxcons *of Ψ, i.e. the maximal (with respect to multi-set inclusion) consistent subsets of Ψ. Formally,* MAXCONS(Ψ) *is the set of all multi-sets M such that:*

– *$M \sqsubseteq \Psi$ and*

– *if* $M \sqsubset M' \sqsubseteq \Psi$, *then* $\bigwedge M' \models \bot$.

DEFINITION 13. — *A formula-based choice function* g^{mc} *is a function of the set of the maxcons of* Ψ *and the belief base, i.e. :*

$$g^{mc}(\Psi) = \{\varphi_i \in \Psi \mid h(\varphi_i, \text{MAXCONS}(\Psi)) \textit{ is minimal } \}$$

Examples of the use of maxcons are numerous, let us see two of them.

DEFINITION 14. —

– $h^{mc1}(\varphi, \text{MAXCONS}(\Psi)) = \#(\{M \mid M \in \text{MAXCONS}(\Psi) \textit{ and } \varphi \in M\})$

– $h^{mc2}(\varphi, \text{MAXCONS}(\Psi)) = \max(\{\#(M) \mid M \in \text{MAXCONS}(\Psi) \textit{ and } \varphi \in M\})$

The first function computes the number of maxcons the belief base belongs to. The second function computes the size of the biggest maxcons the belief base belongs to.

We will note g^{mc1} (respectively g^{mc2}) the formula-based choice function that use h^{mc1} (resp. h^{mc2}).

5. Instantiating the BGM framework

In this section we will try to illustrate how interesting the defined Belief Game Model framework is by giving several examples. We will first see some of the simplest operators that we can define with this framework. Then we will illustrate the behaviour of more complex operators on a typical merging example.

5.1. *Some simple examples*

Let us first see what operators are obtained with the simplest weakening and choice functions (that means that we will either choose the weakening function to be the drastic one, or the choice function to be identity).

– $\langle g_{id}, \blacktriangledown_\top \rangle$: In this case the belief base result of the BGM on Ψ under the constraint μ is the conjunction of all the bases of the profile with the integrity constraints ($\bigwedge \Psi \wedge \mu$) if this conjunction is consistent, and μ otherwise. This operator is called the *basic merging operator* [KON 99].

– $\langle g_{id}, \blacktriangledown_\delta \rangle$: In this case, at each step of the game, each source weakens using dilation. This gives the well known model-based merging operator $\triangle^{d_H, \max}$ defined in [REV 93, REV 97, KON 02a].

– $\langle g^{d_D, \Sigma}, \blacktriangledown_\top \rangle$: Here, the result is the cardinality-maximal consistent subset of Ψ if it is unique and consistent with the constraints μ, and it is simply μ otherwise. This operator is a new one. It is interesting since it can be viewed as a generalized conjunction : it gives the conjunction of all the bases and the constraints if it is consistent, but if it is not, it tries to find the result by doing the least number of repairs (forget one belief base) of the belief profile. If there is no ambiguity on the correction (i.e. a unique cardinality-maxcons), then it accepts it as the result.

– $\langle g^{d_D,\max}, \blacktriangledown_\top \rangle$: This operator gives as result the conjunction of all the formulas that belong to all maxcons (also called free formulas in [BEN 97, BEN 99]) and the integrity constraints if it is consistent, and μ otherwise.

– $\langle g^{mc1}, \blacktriangledown_\top \rangle$: This operator gives the conjunction of the formulas that belong to the maximum number of maxcons and the integrity constraints if consistent, and μ otherwise.

– $\langle g^{mc2}, \blacktriangledown_\top \rangle$: In this case, the belief base result of the merging is the conjunction of the belief bases that belong to the biggest maxcons for cardinality and the integrity constraints if consistent, and μ otherwise.

These operators are not logically independent, some of them are logically stronger than others, as stated in the following proposition.

THEOREM 15. — *In figure 1 an arrow between an operator A and an operator B ($A \longrightarrow B$) means that operator A is logically stronger*[4] *(or less cautious) than operator B. Results obtained by transitivity are not represented.*

$$\begin{tikzcd}
\langle g^{mc2}, \blacktriangledown_\top \rangle \arrow[r] & \langle g^{d_D,\Sigma}, \blacktriangledown_\top \rangle \arrow[dr] & \\
& \langle g_{id}, \delta \rangle \arrow[r] & \langle g_{id}, \blacktriangledown_\top \rangle \\
\langle g^{mc1}, \blacktriangledown_\top \rangle \arrow[r] & \langle g^{d_D,\max}, \blacktriangledown_\top \rangle \arrow[ur] &
\end{tikzcd}$$

Figure 1. *Cautiousness*

5.2. *An example*

We will see on an example [REV 97], what is the behaviour or some BGM operators, namely the operators $\langle g^{d_H,h^\Sigma}, \blacktriangledown_\delta \rangle$, $\langle g^{d_H,h^{\max}}, \blacktriangledown_\delta \rangle$, $\langle g^{mc1}, \blacktriangledown_\delta \rangle$ and $\langle g^{mc2}, \blacktriangledown_\delta \rangle$. Here is the example : There are three sources $\Psi = \{\varphi_1, \varphi_2, \varphi_3\}$ with the following belief bases $Mod(\varphi_1) = \{(1,0,0),(0,0,1),(1,0,1)\}$, $Mod(\varphi_2) = \{(0,1,0),(0,0,1)\}$, $Mod(\varphi_3) = \{(1,1,1)\}$. There are no constraints on the result, so $\mu = \top$.

4. An operator A is logically stronger than an operator B iff for all profile Ψ, $A(\Psi) \vdash B(\Psi)$, where $A(\Psi)$ denotes the belief base result of the BGM A on the profile Ψ.

– $\langle g^{d_H,h^\Sigma}, \blacktriangledown_\delta\rangle$: As Ψ is not consistent, let us do the first round. $d(\varphi_1,\varphi_2) = 0$, $d(\varphi_1,\varphi_3) = 1$, $d(\varphi_2,\varphi_3) = 2$. So $h^\Sigma_\Psi(\varphi_1) = 1$, $h^\Sigma_\Psi(\varphi_2) = 2$, $h^\Sigma_\Psi(\varphi_3) = 3$. That gives $g^{d_H,h^\Sigma}(\Psi) = \{\varphi_3\}$. So φ_3 is replaced[5] by $\blacktriangledown_\delta(\varphi_3) = form(\{(1,\ 1,1),(1,1,0),(1,0,1),(0,1,1)\})$. We have not yet reached a consistent Ψ, so let us do a further round. Let us first compute the new distances. $d(\varphi_1,\varphi_2) = 0$, $d(\varphi_1,\varphi_3) = 0$, $d(\varphi_2,\varphi_3) = 1$. So $h^\Sigma_\Psi(\varphi_1) = 0$, $h^\Sigma_\Psi(\varphi_2) = 1$, $h^\Sigma_\Psi(\varphi_3) = 1$. That gives $g^{d_H,h^\Sigma}(\Psi) = \{\varphi_2,\varphi_3\}$. So φ_2 is replaced by $\blacktriangledown_\delta(\varphi_2) = form(\{(0,1,0),(0,0,1),(1,1,0),(0,0,0),(0,1,1),(1,0,1)\})$, and φ_3 is replaced by $\blacktriangledown_\delta(\varphi_3) = form(\{(1,1,1),(1,1,0),(1,0,1),(0,1,1),(0,1,0),(1,0,0),(0,0,1)\})$. We have reached a consistent belief profile, so the result is $Mod(\Psi_{\langle g^{d_H,h^\Sigma},\blacktriangledown_\delta\rangle}) = \{(0,0,1),(1,0,1)\}$.

– $\langle g^{d_H,h^{\max}}, \blacktriangledown_\delta\rangle$: As Ψ is not consistent, let us do the first round. $d(\varphi_1,\varphi_2) = 0$, $d(\varphi_1,\varphi_3) = 1$, $d(\varphi_2,\varphi_3) = 2$. So $h^{\max}_\Psi(\varphi_1) = 1$, $h^{\max}_\Psi(\varphi_2) = 2$, $h^{\max}_\Psi(\varphi_3) = 2$. That gives $g^{d_H,h^{\max}}(\Psi) = \{\varphi_2,\varphi_3\}$. So φ_2 is replaced by $\blacktriangledown_\delta(\varphi_2) = form(\{(0,1,0),(0,0,1),(1,1,0),(0,0,0),(0,1,1),\ (1,0,1)\})$, and φ_3 is replaced by $\blacktriangledown_\delta(\varphi_3) = form(\{(1,1,1),(1,1,0),(1,0,1),(0,1,1)\})$. The obtained profile is consistent, so the result is $Mod(\Psi_{\langle g^{d_H,h^{\max}},\blacktriangledown_\delta\rangle}) = \{(1,0,1)\}$.

– $\langle g^{mc1}, \blacktriangledown_\delta\rangle$: Ψ is not consistent, and $\textsc{maxcons}(\Psi) = \{\{\varphi_1,\varphi_2\},\{\varphi_3\}\}$. So $h^{mc1}_\Psi(\varphi_1) = h^{mc1}_\Psi(\varphi_2) = h^{mc1}_\Psi(\varphi_3) = 1$, and $g^{mc1}(\Psi) = \Psi$. So we weaken the three bases, which gives respectively $\blacktriangledown_\delta(\varphi_1) = form(\{(1,0,0),(0,0,1),(1,0,1),(0,0,\ 0),(1,1,0),(0,1,1),(1,1,1)\})$, $\blacktriangledown_\delta(\varphi_2) = form(\{(0,1,0),(0,0,1),(1,1,0),\ \ (0,0,0),(0,1,1),(1,0,1)\})$, and $\blacktriangledown_\delta(\varphi_3) = form(\ \{(1,1,1),(1,1,0),(1,0,1),\ (0,1,1)\})$. This belief profile is consistent, and the resulting base is $Mod(\Psi_{\langle g^{mc1},\blacktriangledown_\delta\rangle}) = \{(1,0,1),(1,1,0),(0,1,1)\}$.

– $\langle g^{mc2}, \blacktriangledown_\delta\rangle$: Ψ is not consistent, and we have $\textsc{maxcons}(\Psi) = \{\{\varphi_1,\varphi_2\},\{\varphi_3\}\}$. So $h^{mc2}_\Psi(\varphi_1) = h^{mc2}_\Psi(\varphi_2) = 2$ and $h^{mc2}_\Psi(\varphi_3) = 1$, and $g^{mc2}(\Psi) = \{\varphi_3\}$. So φ_3 is replaced by $\blacktriangledown_\delta(\varphi_3) = form(\{(1,1,1),(1,1,0),(1,0,1),(0,1,1)\})$. The belief profile is still not consistent, so we need one more round. Now we have $\textsc{maxcons}(\Psi) = \{\{\varphi_1,\varphi_2\},\{\varphi_1,\varphi_3\}\}$. So $h^{mc2}_\Psi(\varphi_1) = h^{mc2}_\Psi(\varphi_2) = h^{mc2}_\Psi(\varphi_3) = 2$, and $g^{mc2}(\Psi) = \Psi$. So we weaken the three bases, which gives respectively $\blacktriangledown_\delta(\varphi_1) = form(\{(1,0,0),(0,0,1),(1,0,1),(0,0,0),(1,1,0),(0,1,1),(1,1,1)\})$, $\blacktriangledown_\delta(\varphi_2) = form(\{(0,1,0),(0,0,1),(1,1,0),(0,0,0),(0,1,1),(1,0,1)\})$, and $\blacktriangledown_\delta(\varphi_3) = form(\ \{(1,1,1),(1,1,0),(1,0,1),(0,1,1)\})$. The belief profile is consistent, and the resulting base is $Mod(\Psi_{\langle g^{mc2},\blacktriangledown_\delta\rangle}) = \{(0,0,1),(1,0,1),(1,1,0),(0,1,1)\}$.

As we can note, on this example the four operators give different (non trivial) results. As all these operators take dilation as weakening functions, we sometimes have the interpretation $(1,1,0)$ as model of the base result of the merging, whereas

5. In order to avoid unnecessary notations, we do not use subscripts to denote the different weakening steps of the bases, we simply replace the belief bases by their weakened counterparts. Hopefully, it can not lead to confusions.

it is a model of none of the initial belief bases. This means that, conversely to usual formula-based merging operators [BAR 92, KON 00, KON 04], the result of the BGM does not (always) imply the disjunction of the belief bases of the profile.

6. Logical properties

Some work in belief merging aims at finding sets of axiomatic properties operators may exhibit in order to ensure the expected behaviour [REV 93, REV 97, LIB 98, KON 98, KON 99, KON 02b]. We focus here on the characterization of Integrity Constraints (IC) merging operators [KON 99, KON 02a].

DEFINITION 16 (IC MERGING OPERATORS). — $\triangle$ *is an* IC merging operator *if and only if it satisfies the following properties:*

(IC0) $\triangle_\mu(\Psi) \models \mu$

(IC1) *If* μ *is consistent, then* $\triangle_\mu(\Psi)$ *is consistent*

(IC2) *If* $\bigwedge \Psi$ *is consistent with* μ*, then* $\triangle_\mu(\Psi) \equiv \bigwedge \Psi \wedge \mu$

(IC3) *If* $\Psi_1 \equiv \Psi_2$ *and* $\mu_1 \equiv \mu_2$*, then* $\triangle_{\mu_1}(\Psi_1) \equiv \triangle_{\mu_2}(\Psi_2)$

(IC4) *If* $\varphi_1 \models \mu$ *and* $\varphi_2 \models \mu$*, then* $\triangle_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_1$ *is consistent if and only if* $\triangle_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_2$ *is consistent*

(IC5) $\triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2) \models \triangle_\mu(\Psi_1 \sqcup \Psi_2)$

(IC6) *If* $\triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2)$ *is consistent, then* $\triangle_\mu(\Psi_1 \sqcup \Psi_2) \models \triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2)$

(IC7) $\triangle_{\mu_1}(\Psi) \wedge \mu_2 \models \triangle_{\mu_1 \wedge \mu_2}(\Psi)$

(IC8) *If* $\triangle_{\mu_1}(\Psi) \wedge \mu_2$ *is consistent, then* $\triangle_{\mu_1 \wedge \mu_2}(\Psi) \models \triangle_{\mu_1}(\Psi)$

The intuitive meaning of the properties is the following: *(IC0)* ensures that the result of merging satisfies the integrity constraints. *(IC1)* states that, if the integrity constraints are consistent, then the result of merging will be consistent. *(IC2)* states that if possible, the result of merging is simply the conjunction of the belief bases with the integrity constraints. *(IC3)* is the principle of irrelevance of syntax : the result of merging has to depend only on the expressed opinions and not on their syntactical presentation. *(IC4)* is a fairness postulate meaning that the result of merging of *two* belief bases should not give preference to one of them (if it is consistent with one of both, it has to be consistent with the other one.) It is a symmetry condition, that aims to rule out operators that can give priority to one of the bases. *(IC5)* expresses the following idea: if belief profiles are viewed as expressing the beliefs of the members of a group, then if Ψ_1 (corresponding to a first group) compromises on a set of alternatives which A belongs to, and Ψ_2 (corresponding to a second group) compromises on another set of alternatives which contains A too, then A has to be in the chosen alternatives if we join the two groups. *(IC5)* and *(IC6)* together state that if one could find two subgroups which agree on at least one alternative, then the result of the global merging will be exactly those alternatives the two groups agree on. *(IC7)* and *(IC8)* state that the notion of closeness is well-behaved, i.e. that an alternative that is preferred among the possible alternatives (μ_1), will remain preferred if one restricts the possible choices ($\mu_1 \wedge \mu_2$). For more explanations on those properties see [KON 02a].

So, let us see now what are the properties of BGM operators.

THEOREM 17. — *BGM operators*[6] *satisfy properties (IC0), (IC1), (IC2), (IC3), (IC7), (IC8). They do not necessarily satisfy properties (IC4), (IC5), (IC6).*

PROOF. —

(IC0) is satisfied by definition (see definition 4), since the result of the BGM process is the base $\bigwedge \Psi_{\mathcal{N}}^{\mu} \wedge \mu$, so it implies μ.

(IC1) is satisfied by definition (see theorem 5).

(IC2) by definition of the BGM (see definition 4), if $\bigwedge \Psi$ is consistent with μ, then the halting condition is satisfied before any weakening round, so $\Psi_{\mathcal{N}}^{\mu} = \Psi$, and $\bigwedge \Psi_{\mathcal{N}}^{\mu} \wedge \mu \equiv \bigwedge \Psi \wedge \mu$

(IC3) is a straighforward consequence of the third condition of the weakening functions, and of the last condition of the choice functions.

(IC7) and *(IC8)* are satisfied. First note that if $\triangle_{\mu_1}(\Psi) \wedge \mu_2$ is not consistent, then $\triangle_{\mu_1}(\Psi) \wedge \mu_2 \models \triangle_{\mu_1 \wedge \mu_2}(\Psi)$, so (IC7) is trivially satisfied. Now, in the case $\triangle_{\mu_1}(\Psi) \wedge \mu_2$ consistent, let us prove (IC7) and (IC8), that is $\triangle_{\mu_1 \wedge \mu_2}(\Psi) \equiv \triangle_{\mu_1}(\Psi) \wedge \mu_2$. As $\triangle_{\mu_1}(\Psi) = \Psi_{\mathcal{N}}^{\mu_1} \wedge \mu_1$, and by definition 4, $\Psi_{\mathcal{N}}^{\mu_1}$ is the first profile Ψ_i such that $\bigwedge \Psi_i \wedge \mu_1$ is consistent. As we also know that, by hypothesis, $\bigwedge \Psi_i \wedge \mu_1 \wedge \mu_2$ is consistent, Ψ_i is also the first profile such that $\bigwedge \Psi_i \wedge \mu_1 \wedge \mu_2$ is consistent. By definition it means that $\Psi_{\mathcal{N}}^{\mu_1 \wedge \mu_2} = \Psi_i$. So we get that $\Psi_{\mathcal{N}}^{\mu_1} = \Psi_{\mathcal{N}}^{\mu_1 \wedge \mu_2}$. From that we get $(\Psi_{\mathcal{N}}^{\mu_1} \wedge \mu_1) \wedge \mu_2 = \Psi_{\mathcal{N}}^{\mu_1 \wedge \mu_2} \wedge \mu_1 \wedge \mu_2$, which is exactly $\triangle_{\mu_1}(\Psi) \wedge \mu_2 = \triangle_{\mu_1 \wedge \mu_2}(\Psi)$.

■

So, as stated in the previous proposition, BGM operators do not fit all properties of IC merging operators. On the other hand, we know for example that the operator $\langle g_{id}, \blacktriangledown_{\delta} \rangle = \triangle^{d_H, \max}$ satisfies also *(IC4), (IC5)* [KON 02a]. So the question is to know if we can ensure more logical properties by making some restrictions on the weakening and/or the choice functions.

A first remark is that *(IC4)* cannot be proved to hold for any BGM operator, but it is satisfied for all the particular operators we have defined in this paper.

THEOREM 18. — *If the weakening function is dilation or drastic weakening, and if the choice function is a symmetric model-based choice function or the formula-based choice function g^{mc1} or g^{mc2}, then the BGM operator satisfies (IC4).*

6. Defined from any choice function and any weakening function.

PROOF. — First note that for all those operators if the two bases are consistent with the constraints (i.e. $\varphi_1 \wedge \varphi_2 \wedge \mu$), then the result of the merging is (from (IC2)) $\triangle_\mu(\{\varphi_1, \varphi_2\}) = \varphi_1 \wedge \varphi_2 \wedge \mu$, so (IC4) is trivially satisfied. The interesting part of (IC4) is when $\varphi_1 \wedge \varphi_2 \wedge \mu$ is not consistent.

If the choice function is a symmetric model-based choice function, then simply note that the distance used to define the model-based functions are symmetric by definition, so $d(\varphi, \varphi') = d(\varphi', \varphi)$. Let note also that by definition of the distance $d(\varphi, \varphi) = d(\varphi', \varphi') = 0$. So, now, using the symmetry condition of the aggregation function, it is easy to show that $h(d(\varphi, \varphi), d(\varphi, \varphi')) = h(d(\varphi', \varphi), d(\varphi', \varphi'))$, that means that φ and φ' are always both maximal, and that they always both have to be weakened. If we use drastic weakening as the weakening function, we have that φ and φ' are replaced by $\top$ and $\top$. So the result of the merging is consistent with φ and φ'. So (IC4) is satisfied. If the weakening function is dilation, then when we weaken φ (resp. φ') for the n^{th} time, we get as the result the base φ^n (resp. φ'^n) that is the formula (up to logical equivalence) whose models are the ones that are at a (Dalal) distance less or equal to n to models of φ (resp. φ'). So, let note the final profile of the BGM $\Psi_i = \{\varphi^i, \varphi'^i\}$, if $\bigwedge \Psi_i \wedge \varphi$ is consistent, it means that $d_H(\varphi'^i, \varphi) = 0$, but by symmetry it implies that $d_H(\varphi^i, \varphi') = 0$, so (recall also that, by definition of dilation $d_H(\varphi^i, \varphi) = 0$ and $d_H(\varphi'^i, \varphi') = 0$) that $\bigwedge \Psi_i \wedge \varphi'$ is consistent. So (IC4) is satisfied.

If the choice function is g^{mc1} or g^{mc2}, let us note that the maximal consistent sets are $\{\varphi\}$ and $\{\varphi'\}$, so the φ and φ' are both minimal, and that they always both have to be weakened. If we use drastic weakening as the weakening function, we have that φ and φ' are replaced by $\top$ and $\top$. So the result of the merging is consistent with φ and φ'. So (IC4) is satisfied. If the weakening function is dilation, we use the same symmetry argument as for symmetric model-based choice functions. ■

The property *(IC5)* can also be recovered for some BGM operators, but *(IC6)* seems hardly recoverable. Those two properties (*(IC5)* and *(IC6)*) are important for classical merging operators, so we can wonder if the BGM operators missing those properties can still be called "merging" operators. One answer to this is that the BGM operators aim at focusing on the interaction between the beliefs of the sources, so it seems natural to loose property *(IC6).* Indeed, whereas classical merging operators aim at giving the result of the merging process in an *ideal* framework, BGM operators seem more adequately reflect the behaviour of a *real* multi-source merging process.

Another important logical link to be underlined is the relationship between BGM operators and AGM belief revision operators [ALC 85, GÄR 88, KAT 91, GÄR 92]. Belief revision aims to make the minimum change in a belief base in order to take into account new information that is more reliable than the current belief base (and that usually contradicts the current belief base). Technically those operators can be described as follows : until the belief base is consistent with the new item of information (seen as an integrity constraint) then weaken the belief base[7]. Stated this way, we can immediately see the parallel with BGM operators since they are described as follows : until the belief profile is consistent with the constraint then weaken some belief bases. The following result shows more formally that, as explained above, BGM operators can be seen as a direct generalization of AGM belief revision operators.

THEOREM 19. — *Let $\mathcal{N} = \langle g, \blacktriangledown \rangle$ be a BGM operator. Let φ and μ be two belief bases. The operator $\circ$ defined as $\varphi \circ \mu = \mathcal{N}_\mu(\{\varphi\})$ is an AGM belief revision operator (i.e. it satisfies properties (R1-R6) of [KAT 91]).*

PROOF. — It is easy to see that if we restrict belief profiles Ψ to singletons $\{\varphi\}$, then postulates (IC0), (IC1), (IC2), (IC3), (IC7), (IC8) directly translate to (R1-R6). After that the proof follows from theorem 17. ■

In particular, we have that each BGM using the dilation weakening function is a generalization of Dalal's revision operator [DAL 88].

Finally let us see another cardinality restriction on the belief profile.

THEOREM 20. — *Let $\mathcal{N} = \langle g^{d,h}, \blacktriangledown_\delta \rangle$ be a BGM operator defined from a symmetric model-based choice function and dilation weakening function. Let φ_1, φ_2 and μ be three belief bases, then the operator $\mathcal{N}_\mu(\{\varphi_1, \varphi_2\})$ is the model-based merging operator $\triangle_\mu^{d_H,\max}(\{\varphi_1, \varphi_2\})$ [KON 02a].*

Note that the previous result holds only when we merge two belief bases.

7. Comparison between BGM and BNM

In this section we will mainly compare our proposal with Booth's Belief Negotiation Model (BNM) [BOO 02]. Let us first briefly recall Booth's proposal.

Belief profiles in this framework are no longer multi-sets but vectors of belief bases, noted $\vec{\Psi}$. Let us note $\vec{\mathcal{E}}$ the set of belief profiles, and let us note $\vec{\Sigma}$ the set of all sequences (vectors) of belief profiles, and $\vec{\sigma}$ one element of this set. When $\vec{X}$ is a vector, we will note $\vec{X}^i$ the ith element of the vector and $\vec{X}^m$ the last element of the vector.

So a sequence of belief profiles $\vec{\sigma}$ is of the form $\vec{\sigma} = \{\vec{\Psi}_1, \ldots, \vec{\Psi}_n\}$, with each $\vec{\Psi}_i$ being a vector of belief bases, i.e. $\vec{\Psi}_i = (\varphi_{i,1}, \ldots, \varphi_{i,n_i})$. And, for example, the notation $\vec{\sigma}^3$ stands for $\vec{\Psi}_3$, and the notation $\vec{\sigma}^m$ stands for $\vec{\Psi}_n$.

Then a BNM negotiation (choice) function is defined as :

DEFINITION 21. — *A* BNM negotiation function *is a function $g^{\text{BNM}} : \vec{\Sigma} \rightarrow \vec{\mathcal{E}}$ such that:*

7. It is the intuitive meaning behind Katsuno and Mendelzon representation theorem in terms of faithful assignments [KAT 91].

– $g^{\text{BNM}}(\vec{\sigma}) \sqsubseteq \vec{\sigma}^m$

– $g^{\text{BNM}}(\vec{\sigma}) \neq \emptyset$

– *If* $\varphi_i \in g^{\text{BNM}}(\vec{\sigma})$, *then* $\varphi_i \not\equiv \top$

And a BNM weakening function is defined as :

DEFINITION 22. — *A* BNM weakening function *is a function* $\blacktriangledown^{\text{BNM}} : \vec{\Sigma} \rightarrow \vec{\mathcal{E}}$ *such that:*

– $(\vec{\sigma}^m)^i \vdash \blacktriangledown^{\text{BNM}}(\vec{\sigma})^i$

– *If* $(\vec{\sigma}^m)^i \equiv \blacktriangledown^{\text{BNM}}(\vec{\sigma})^i$, *then* $(\vec{\sigma}^m)^i \equiv \top$

Finally the solution to a BNM is defined as :

DEFINITION 23. — *The solution to a belief profile* $\vec{\Psi}$ *for a Belief Negotiation Model* $\mathcal{N}^{\text{BNM}} = \langle g^{\text{BNM}}, \blacktriangledown^{\text{BNM}} \rangle$ *under the integrity constraints* μ, *noted* $\mathcal{N}_\mu^{\text{BNM}}(\Psi)$, *is given by the function* $f^{\mathcal{N}} : \vec{\mathcal{E}} \rightarrow \vec{\Sigma}$ *defined as:*

– $f^{\mathcal{N}}(\vec{\Psi}) = \vec{\sigma} = (\vec{\Psi}_0, \ldots, \vec{\Psi}_k)$

with $\vec{\Psi}_0 = \vec{\Psi}$, k *is the smallest integer such that* $\bigwedge \vec{\Psi}_k \wedge \mu$ *is consistent, and for each* $0 \leq j < k$ *we have (* $\vec{\sigma_j}$ *denotes* $(\vec{\Psi}_0, \ldots, \vec{\Psi}_j)$*):*

$$(\vec{\Psi}_{j+1})^i = \begin{cases} \blacktriangledown^{\text{BNM}}(\vec{\sigma_j})^i & \textit{if } (\vec{\Psi}_j)^i \in g^{\text{BNM}}(\vec{\sigma_j}) \\ (\vec{\Psi}_j)^i & \textit{otherwise} \end{cases}$$

Finally, the belief base result of the BNM is $\bigwedge \vec{\Psi}_k \wedge \mu$.

We change some notation, in order to show the closeness with our present work. For the original presentation and detailed explanations on the definitions see [BOO 02, BOO 01].

The main differences between BNM and BGM are :

– BNM's definition of belief profile as vectors allows us to speak about sources separately. So when there are two identical belief bases in the belief profile, it is possible to weaken only one of these bases. This is not possible in the BGM framework.

– The BNM negotiation function takes as input the whole negotiation history from the initial belief profile. So it is possible to implement a negotiation process such that each source weakens after the previous one (for example, source 1, then source $2, \ldots$), or such that we prevent a source to weaken two times successively. The BGM choice functions are more Markovian, taking only into account the current belief profile.

– Similarly, the BNM weakening function also takes as input the whole negotiation history. It allows us to weaken differently two identical belief bases obtained at different rounds or to weaken differently two identical belief bases of the same belief profile.

– According to the previous items ideas, the irrelevance of syntax condition of the BGM weakening function, and the anonymity condition of the BGM choice function are not required in the BNM framework.

The main difference between Booth's proposal and our is that Booth's takes the sources as candidates to weakening, whereas we restrict ourselves to *"points of view"* (logical content of the bases). That means that in Booth's if one source has to weaken, it can be the case that another source with exactly the same beliefs do not have to weaken too. Our proposal adds more anonymity by saying that only beliefs decide who has to weaken, not the identity of one source. Similarly, the choice functions are more "Markovian" in our framework than in Booth's one. We think that those hypothesis are more realistic (and necessary) on a belief merging point of view. Whereas Booth's framework allows us to model more generalized negotiation frameworks, where one can decide for example that each source has to weaken one after the other. So, despite the closeness of the models, and the objective fact that our proposal is a particular case of Booth's one (i.e. each of our operators can be defined in Booth's framework), the intended applications of those two frameworks are quite different. And the particular properties achieved by adding those restrictions shows that this framework forms a consistent family of merging operators. It explains why it is worth focusing on the model we have defined.

A last difference is that, in this paper, we are interested on the result of the process (as a belief base), whereas BNM framework aims at studying the resulting profile, in connection with a notion of *"social contraction"*. See [BOO 02] for a study of logical properties for social contraction.

An additional contribution of this work is to give examples of purely propositional logic BNM operators. In [BOO 02], Booth propose two examples of BNM, both working on ordinal conditional functions (OCF) [SPO 88], but none on a propositional belief base. So this work can be seen as an investigation of what kind of operators this definition can give on propositional belief bases (through adding additional requirements).

8. Conclusion

We have proposed in this paper a new family of belief merging operators, that we call Belief Game Model (BGM) operators. The hypothesis for those operators is that all the sources are *a priori* reliable, or that we know that some sources are less reliable than the others, but without knowing which ones. This hypothesis leads to choosing a majority approach, justified by Condorcet's Jury Theorem. The idea behind the Belief Game Model is simple : Until a coherent set of sources is reached, at each round a contest is organized to find out the weakest sources, then those sources have to concede (weaken their point of view). This idea leads to numerous new interesting operators and opens new perspectives for belief merging. Some existing operators are also recovered as particular cases.

Not surprisingly, the operators defined do not satisfy all logical properties proposed for IC merging operators. The reason is that those logical properties aim at giving constraints on the result of the merging in an *ideal* framework, whereas BGM operators aim at describing more accurately what can happen in a *real* multi-source environment. So usual IC merging operators can be seen as a *normative* approach to merging. They show the way to a purely logical result. Conversely, BGM operators adopt a *descriptive* approach to merging, taking into account the interaction between the sources. They try to simulate more adequately what can happen in a group-decision process. So they are maybe more realistic.

This paper mainly aims to introduce BGM operators, but it provides several open questions that are left for further research.

The first one is about the definition of BGM operators and the computation of the result. We give an iterative definition of BGM operators, that leads to an iterative computation of the result. The question is to know if we can find a non-iterative equivalent definition. We know that some simple operators can be defined non-iteratively. But the question is to know if all operators or a non-trivial subclass of them are also definable non-iteratively.

Another open question is about the logical characterization of this family. In this paper we study the logical properties of this family with respect to the general definition of IC merging operators. The question is to know if we can find a set of logical properties that characterizes BGM operators.

Finally, we have recently studied the strategy-proofness of the usual propositional merging operators, showing that most of them are not strategy-proof [EVE 04]. And we have exhibited several restrictions on which strategy-proofness can be achieved. So an interesting question is to compare the strategy-proofness of BGM operators with the one of the classical merging operators.

Acknowledgements

The author wish to thank Richard Booth for hepful discussions and for its comments on a previous version of this paper, the anonymous referees for their interesting comments and Anthony Hunter for its help. The author has been supported by the IUT de Lens, the Université d'Artois, the Région Nord/Pas-de-Calais under the TACT-TIC project, and by the European Community FEDER Program.

9. References

[ALC 85] ALCHOURRÓN C. E., GÄRDENFORS P., MAKINSON D., "On the logic of theory change: Partial meet contraction and revision functions", *Journal of Symbolic Logic*, vol. 50, 1985, p. 510-530.

[BAR 92] BARAL C., KRAUS S., MINKER J., SUBRAHMANIAN V. S., "Combining knowledge bases consisting of first-order theories", *Computational Intelligence*, vol. 8, num. 1, 1992, p. 45-71.

[BEN 97] BENFERHAT S., DUBOIS D., PRADE H., “Some syntactic approaches to the handling of inconsistent knowledge bases: a comparative study, part 1: the flat case”, *Studia Logica*, vol. 58, 1997, p. 17-45.

[BEN 98a] BENFERHAT S., DUBOIS D., LANG J., PRADE H., SAFFIOTI A., SMETS P., “A general approach for inconsistency handling and merging information in prioritized knowledge bases”, *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR’98)*, 1998, p. 466-477.

[BEN 98b] BENFERHAT S., DUBOIS D., PRADE H., “Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study. Part II: The prioritized case”, ORLOWSKA E., Ed., *Logic at work*, vol. 24 of *Physica-Verlag*, Heidelberg, 1998, p. 473-511.

[BEN 99] BENFERHAT S., DUBOIS D., PRADE H., “An overview of inconsistency-tolerant inferences in prioritized knowledge bases”, *Fuzzy Sets, Logic and Reasoning about Knowledge*, vol. 15 of *Applied Logic Series*, p. 395-417, Kluwer, 1999.

[BLO 00] BLOCH I., LANG J., “Towards mathematical morpho-logics”, *Proceedings of the Eigth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU’00)*, 2000, p. 1405-1412.

[BLO 04] BLOCH I., PINO PÉREZ R., UZCÁTEGUI C., “A Unified Treatment for Knowledge Dynamics”, *Proceedings of the Ninth Conference on Principles of Knowledge Representation and Reasoning (KR’04)*, 2004, p. 329-337.

[BOO 01] BOOTH R., “A negociation-style framework for non-prioritised revision”, *Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK’01)*, 2001, p. 137-150.

[BOO 02] BOOTH R., “Social contraction and belief negociation”, *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR’02)*, 2002, p. 374-384.

[CHO 93] CHOLVY L., “Proving theorems in a multi-sources environment”, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI’93)*, 1993, p. 66-71.

[CHO 95] CHOLVY L., “Automated reasoning with merged contradictory information whose reliability depends on topics”, *Proceedings of the Third European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU’95)*, 1995.

[CHO 98] CHOLVY L., “Reasoning about data provided by federated deductive databases”, *Journal of Intelligent Information Systems*, vol. 10, num. 1, 1998.

[DAL 88] DALAL M., “Investigations into a theory of knowledge base revision: preliminary report”, *Proceedings of the Seventh American National Conference on Artificial Intelligence (AAAI’88)*, 1988, p. 475-479.

[EVE 04] EVERAERE P., KONIECZNY S., MARQUIS P., “On merging strategy-proofness”, *Proceedings of the Ninth Conference on Principles of Knowledge Representation and Reasoning (KR’04)*, 2004, p. 357-358.

[GÄR 88] GÄRDENFORS P., *Knowledge in flux*, MIT Press, 1988.

[GÄR 92] GÄRDENFORS P., Ed., *Belief Revision*, Cambridge University Press, 1992.

[KAT 91] KATSUNO H., MENDELZON A. O., “Propositional knowledge base revision and minimal change”, *Artificial Intelligence*, vol. 52, 1991, p. 263-294.

[KON 98] KONIECZNY S., PINO PÉREZ R., "On the logic of merging", *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, 1998, p. 488-498.

[KON 99] KONIECZNY S., PINO PÉREZ R., "Merging with Integrity Constraints", *Proceedings of the Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99)*, LNAI 1638, 1999, p. 233-244.

[KON 00] KONIECZNY S., "On the difference between merging knowledge bases and combining them", *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR'00)*, 2000, p. 135-144.

[KON 02a] KONIECZNY S., PINO PÉREZ R., "Merging information under constraints: a qualitative framework", *Journal of Logic and Computation*, vol. 12, num. 5, 2002, p. 773-808.

[KON 02b] KONIECZNY S., PINO PÉREZ R., "On the Frontier between Arbitration and Majority", *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, 2002, p. 109-118.

[KON 04] KONIECZNY S., LANG J., MARQUIS P., "DA^2 Merging Operators", *Artificial Intelligence*, vol. 157, num. 1-2, 2004, p. 49-79.

[LIB 98] LIBERATORE P., SCHAERF M., "Arbitration (or how to merge knowledge bases)", *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, num. 1, 1998, p. 76-90.

[LIN 99] LIN J., MENDELZON A. O., "Knowledge base merging by majority", *Dynamic Worlds: From the Frame Problem to Knowledge Management*, Kluwer, 1999.

[REV 93] REVESZ P. Z., "On the semantics of theory change: arbitration between old and new information", *Proceedings of the Twelfth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Databases*, 1993, p. 71-92.

[REV 97] REVESZ P. Z., "On the semantics of arbitration", *International Journal of Algebra and Computation*, vol. 7, num. 2, 1997, p. 133-160.

[SPO 88] SPOHN W., "Ordinal conditional functions: a dynamic theory of epistemic states", HARPER W. L., SKYRMS B., Eds., *Causation in Decision, Belief Change, and Statistics*, vol. 2, p. 105-134, Kluwer, 1988.

CONFLUENCE OPERATORS

Sébastien Konieczny, Ramón Pino Pérez.
Eleventh European Conference on Logics in Artificial Intelligence (JELIA'08).
pages 272-284.
2008.

Confluence Operators

Sébastien Konieczny[1] and Ramón Pino Pérez[2]

[1] CRIL - CNRS
Faculté des Sciences, Université d'Artois, Lens, France
e-mail: konieczny@cril.fr
[2] Departamento de Matemáticas
Facultad de ciencias, Universidad de Los Andes, Mérida, Venezuela
e-mail: pino@ula.ve

Abstract. In the logic based framework of knowledge representation and reasoning many operators have been defined in order to capture different kinds of change: revision, update, merging and many others. There are close links between revision, update, and merging. Merging operators can be considered as extensions of revision operators to multiple belief bases. And update operators can be considered as pointwise revision, looking at each model of the base, instead of taking the base as a whole. Thus, a natural question is the following one: Are there natural operators that are pointwise merging, just as update are pointwise revision? The goal of this work is to give a positive answer to this question. In order to do that, we introduce a new class of operators: the confluence operators. These new operators can be useful in modelling negotiation processes.

1 Introduction

Belief change theory has produced a lot of different operators that models the different ways the beliefs of one (or some) agent(s) evolve over time. Among these operators, one can quote revision [1, 5, 10, 6], update [9, 8], merging [19, 14], abduction [16], extrapolation [4], etc.

In this paper we will focus on revision, update and merging. Let us first briefly describe these operators informally:

Revision Belief revision is the process of accomodating a new piece of evidence that is more reliable than the current beliefs of the agent. In belief revision the world is static, it is the beliefs of the agents that evolve.

Update In belief update the new piece of evidence denotes a change in the world. The world is dynamic, and these (observed) changes modify the beliefs of the agent.

Merging Belief merging is the process of defining the beliefs of a group of agents. So the question is: Given a set of agents that have their own beliefs, what can be considered as the beliefs of the group?

Apart from these intuitive differences between these operators, there are also close links between them. This is particularly clear when looking at the technical definitions. There are close relationship between revision [1, 5, 10] and KM update operators [9]. The first ones looking at the beliefs of the agents globally, the second ones looking at them locally (this sentence will be made formally clear later in the paper)[3]. There is

[3] See [8, 4, 15] for more discussions on update and its links with revision.

also a close connection between revision and merging operators. In fact revision operators can be seen as particular cases of merging operators. From these two facts a very natural question arises: What is the family of operators that are a generalization of update operators in the same way merging operators generalize revision operators? Or, equivalently, what are the operators that can be considered as pointwise merging, just as KM update operators can be considered as pointwise belief revision. This can be outlined in the figure below. The aim of this paper is to introduce and study the operators corresponding to the question mark. We will call these new operators confluence operators.

$$\begin{tikzcd}
\text{Revision} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{Update} \arrow[d, leftrightarrow] \\
\text{Merging} \arrow[r, leftrightarrow] & \textbf{?}
\end{tikzcd}$$

Fig. 1. Revision - Update - Merging - Confluence

these new operators are more cautious than merging operators. This suggest that they can be used to define negotiation operators (see [2, 20, 18, 17, 12]), or as a first step of a negotiation process, in order to find all the possible negotiation results.

In order to illustrate the need for these new operators and also the difference of behaviour between merging and confluence we present the following small example.

Example 1. Mary and Peter are planning to buy a car. Mary does not like a German car nor an expensive car. She likes small cars. Peter hesitates between a German, expensive but small car or a car which is not German, nor expensive and is a big car. Taking three propositional variables `German_car`, `Expensive_car` and `Small_car` in this order, Mary's desires are represented by $mod(A) = \{001\}$ and Peter's desires by $mod(B) = \{111, 000\}$. Most of the merging operators[4] give as solution (in semantical terms) the set $\{001, 000\}$. That is the same solution obtained when we suppose that Peter's desires are only a car which is not German nor expensive but a big car ($mod(B') = \{000\}$). The confluence operators will take into account the disjunctive nature of Peter's desires in a better manner and they will incorporate also the interpretations that are a trade-off between 001 and 111. For instance, the worlds 011 and 101 will be also in the solution if one use the confluence operator $\diamond^{d_H, Gmax}$ (defined in Section 7).

This kind of operators is particularly adequate when the base describes a situation that is not perfectly known, or that can evolve in the future. For instance Peter's desires can either be imperfectly known (he wants one of the two situations but we do not know which one), or can evolve in the future (he will choose later between the two situations). In these situations the solutions proposed by confluence operators will be

[4] Such as $\triangle^{d_H, \Sigma}$ and $\triangle^{d_H, Gmax}$ [14].

more adequate than the one proposed by merging operators. The solutions proposed by the confluence operators can be seen as all possible agreements in a negotiation process.

In the next section we will give the required definitions and notations. In Section 3 we will recall the postulates and representation theorems for revision, update, and merging, and state the links between these operators. In Section 4 we define confluence operators. We provide a representation theorem for these operators in Section 5. In Section 6 we study the links between confluence operators and update and merging. In Section 7 we give examples of confluence operators. And we conclude in Section 8.

2 Preliminaries

We consider a propositional language $\mathcal{L}$ defined from a finite set of propositional variables $\mathcal{P}$ and the standard connectives, including $\top$ and $\bot$.

An interpretation ω is a total function from $\mathcal{P}$ to $\{0, 1\}$. The set of all interpretations is denoted by $\mathcal{W}$. An interpretation ω is a model of a formula $\phi \in \mathcal{L}$ if and only if it makes it true in the usual truth functional way. $mod(\varphi)$ denotes the set of models of the formula φ, i.e., $mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$. When M is a set of models we denote by φ_M a formula such that $mod(\varphi_M) = M$.

A *base* K is a finite set of propositional formulae. In order to simplify the notations, in this work we will identify the base K with the formula φ which is the conjunction of the formulae of K[5].

A *profile* Ψ is a non-empty multi-set (bag) of bases $\Psi = \{\varphi_1, \ldots, \varphi_n\}$ (hence different agents are allowed to exhibit identical bases), and represents a group of n agents.

We denote by $\bigwedge \Psi$ the conjunction of bases of $\Psi = \{\varphi_1, \ldots, \varphi_n\}$, i.e., $\bigwedge \Psi = \varphi_1 \wedge \ldots \wedge \varphi_n$. A profile Ψ is said to be consistent if and only if $\bigwedge \Psi$ is consistent. The multi-set union is denoted by $\sqcup$.

A formula φ is complete if it has only one model. A profile Ψ is complete if all the bases of Ψ are complete formulae.

If $\leq$ denotes a pre-order on $\mathcal{W}$ (i.e., a reflexive and transitive relation), then $<$ denotes the associated strict order defined by $\omega < \omega'$ if and only if $\omega \leq \omega'$ and $\omega' \not\leq \omega$, and $\simeq$ denotes the associated equivalence relation defined by $\omega \simeq \omega'$ if and only if $\omega \leq \omega'$ and $\omega' \leq \omega$. A pre-order is *total* if $\forall \omega, \omega' \in \mathcal{W}$, $\omega \leq \omega'$ or $\omega' \leq \omega$. A pre-order that is not total is called *partial*. Let $\leq$ be a pre-order on A, and $B \subseteq A$, then $\min(B, \leq) = \{b \in B \mid \nexists a \in B\ a < b\}$.

3 Revision, Update and Merging

Let us now recall in this section some background on revision, update and merging, and their representation theorems in terms of pre-orders on interpretations. This will allow us to give the relationships between these operators.

[5] Some approaches are sensitive to syntactical representation. In that case it is important to distinguish between K and the conjonction of its formulae (see e.g. [13]). But operators of this work are all syntax independant.

3.1 Revision

Definition 1 (Katsuno-Mendelzon [10]). *An operator $\circ$ is an AGM belief revision operator if it satisfies the following properties:*

(R1) $\varphi \circ \mu \vdash \mu$
(R2) *If* $\varphi \wedge \mu \nvdash \bot$ *then* $\varphi \circ \mu \equiv \varphi \wedge \mu$
(R3) *If* $\mu \nvdash \bot$ *then* $\varphi \circ \mu \nvdash \bot$
(R4) *If* $\varphi_1 \equiv \varphi_2$ *and* $\mu_1 \equiv \mu_2$ *then* $\varphi_1 \circ \mu_1 \equiv \varphi_2 \circ \mu_2$
(R5) $(\varphi \circ \mu) \wedge \phi \vdash \varphi \circ (\mu \wedge \phi)$
(R6) *If* $(\varphi \circ \mu) \wedge \phi \nvdash \bot$ *then* $\varphi \circ (\mu \wedge \phi) \vdash (\varphi \circ \mu) \wedge \phi$

When one works with a finite propositional language the previous postulates, proposed by Katsuno and Mendelzon, are equivalent to AGM ones [1, 5]. In [10] Katsuno and Mendelzon give also a representation theorem for revision operators, showing that each revision operator corresponds to a faithful assignment, that associates to each base a plausibility preorder on interpretations (this idea can be traced back to Grove systems of spheres [7]).

Definition 2. *A faithful assignment is a function mapping each base φ to a pre-order $\leq_\varphi$ over interpretations such that:*

1. *If $\omega \models \varphi$ and $\omega' \models \varphi$, then $\omega \simeq_\varphi \omega'$*
2. *If $\omega \models \varphi$ and $\omega' \not\models \varphi$, then $\omega <_\varphi \omega'$*
3. *If $\varphi \equiv \varphi'$, then $\leq_\varphi = \leq_{\varphi'}$*

Theorem 1 (Katsuno-Mendelzon [10]). *An operator $\circ$ is a revision operator (ie. it satisfies (R1)-(R6)) if and only if there exists a faithful assignment that maps each base φ to a total pre-order $\leq_\varphi$ such that*

$$mod(\varphi \circ \mu) = \min(mod(\mu), \leq_\varphi).$$

This representation theorem is important because it provides a way to easily define revision operators by defining faithful assignments. But also because their are similar such theorems for update and merging (we will also show a similar result for confluence), and that these representations in term of assignments allow to more easily find links between these operators.

3.2 Update

Definition 3 (Katsuno-Mendelzon [9, 11]). *An operator $\diamond$ is a (partial) update operator if it satisfies the properties* **(U1)-(U8)**. *It is a total update operator if it satisfies the properties* **(U1)-(U5), (U8), (U9)**.

(U1) $\varphi \diamond \mu \vdash \mu$
(U2) *If* $\varphi \vdash \mu$, *then* $\varphi \diamond \mu \equiv \varphi$
(U3) *If* $\varphi \nvdash \bot$ *and* $\mu \nvdash \bot$ *then* $\varphi \diamond \mu \nvdash \bot$
(U4) *If* $\varphi_1 \equiv \varphi_2$ *and* $\mu_1 \equiv \mu_2$ *then* $\varphi_1 \diamond \mu_1 \equiv \varphi_2 \diamond \mu_2$

(U5) $(\varphi \diamond \mu) \wedge \phi \vdash \varphi \diamond (\mu \wedge \phi)$
(U6) *If* $\varphi \diamond \mu_1 \vdash \mu_2$ *and* $\varphi \diamond \mu_2 \vdash \mu_1$*, then* $\varphi \diamond \mu_1 \equiv \varphi \diamond \mu_2$
(U7) *If* φ *is a complete formula, then* $(\varphi \diamond \mu_1) \wedge (\varphi \diamond \mu_2) \vdash \varphi \diamond (\mu_1 \vee \mu_2)$
(U8) $(\varphi_1 \vee \varphi_2) \diamond \mu \equiv (\varphi_1 \diamond \mu) \vee (\varphi_2 \diamond \mu)$
(U9) *If* φ *is a complete formula and* $(\varphi \diamond \mu) \wedge \phi \nvdash \bot$*, then* $\varphi \diamond (\mu \wedge \phi) \vdash (\varphi \diamond \mu) \wedge \phi$

As for revision, there is a representation theorem in terms of faithful assignment.

Definition 4. *A faithful assignment is a function mapping each interpretation* ω *to a pre-order* $\leq_\omega$ *over interpretations such that if* $\omega \neq \omega'$*, then* $\omega <_\omega \omega'$*.*

One can easily check that this faithful assignment on interpretations is just a special case of the faithful assignment on bases defined in the previous section on the complete base corresponding to the interpretation.

Katsuno and Mendelzon give two representation theorems for update operators. The first representation theorem corresponds to partial pre-orders.

Theorem 2 (Katsuno-Mendelzon [9, 11]). *An update operator* $\diamond$ *satisfies (U1)-(U8) if and only if there exists a faithful assignment that maps each interpretation* ω *to a partial pre-order* $\leq_\omega$ *such that*

$$mod(\varphi \diamond \mu) = \bigcup_{\omega \models \varphi} \min(mod(\mu), \leq_{\varphi_{\{\omega\}}})$$

And the second one corresponds to total pre-orders.

Theorem 3 (Katsuno-Mendelzon [9, 11]). *An update operator* $\diamond$ *satisfies (U1)-(U5), (U8) and (U9) if and only if there exists a faithful assignment that maps each interpretation* ω *to a total pre-order* $\leq_\omega$ *such that*

$$mod(\varphi \diamond \mu) = \bigcup_{\omega \models \varphi} \min(mod(\mu), \leq_{\varphi_{\{\omega\}}})$$

3.3 Merging

Definition 5 (Konieczny-Pino Pérez [14]). *An operator* $\triangle$ *mapping a pair* Ψ, μ *(profile, formula) into a formula denoted* $\triangle_\mu(\Psi)$ *is an IC merging operator if it satisfies the following properties:*

(IC0) $\triangle_\mu(\Psi) \vdash \mu$
(IC1) *If* μ *is consistent, then* $\triangle_\mu(\Psi)$ *is consistent*
(IC2) *If* $\bigwedge \Psi$ *is consistent with* μ*, then* $\triangle_\mu(\Psi) \equiv \bigwedge \Psi \wedge \mu$
(IC3) *If* $\Psi_1 \equiv \Psi_2$ *and* $\mu_1 \equiv \mu_2$*, then* $\triangle_{\mu_1}(\Psi_1) \equiv \triangle_{\mu_2}(\Psi_2)$
(IC4) *If* $\varphi_1 \vdash \mu$ *and* $\varphi_2 \vdash \mu$*, then* $\triangle_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_1$ *is consistent if and only if* $\triangle_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_2$ *is consistent*
(IC5) $\triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2) \vdash \triangle_\mu(\Psi_1 \sqcup \Psi_2)$
(IC6) *If* $\triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2)$ *is consistent, then* $\triangle_\mu(\Psi_1 \sqcup \Psi_2) \vdash \triangle_\mu(\Psi_1) \wedge \triangle_\mu(\Psi_2)$
(IC7) $\triangle_{\mu_1}(\Psi) \wedge \mu_2 \vdash \triangle_{\mu_1 \wedge \mu_2}(\Psi)$

(IC8) *If $\triangle_{\mu_1}(\Psi) \wedge \mu_2$ is consistent, then $\triangle_{\mu_1 \wedge \mu_2}(\Psi) \vdash \triangle_{\mu_1}(\Psi)$*

There is also a representation theorem for merging operators in terms of pre-orders on interpretations [14].

Definition 6. *A* syncretic assignment *is a function mapping each profile Ψ to a total pre-order $\leq_\Psi$ over interpretations such that:*

1. *If $\omega \models \Psi$ and $\omega' \models \Psi$, then $\omega \simeq_\Psi \omega'$*
2. *If $\omega \models \Psi$ and $\omega' \not\models \Psi$, then $\omega <_\Psi \omega'$*
3. *If $\Psi_1 \equiv \Psi_2$, then $\leq_{\Psi_1} = \leq_{\Psi_2}$*
4. *$\forall \omega \models \varphi\ \exists \omega' \models \varphi'\ \omega' \leq_{\{\varphi\} \sqcup \{\varphi\}'} \omega$*
5. *If $\omega \leq_{\Psi_1} \omega'$ and $\omega \leq_{\Psi_2} \omega'$, then $\omega \leq_{\Psi_1 \sqcup \Psi_2} \omega'$*
6. *If $\omega <_{\Psi_1} \omega'$ and $\omega \leq_{\Psi_2} \omega'$, then $\omega <_{\Psi_1 \sqcup \Psi_2} \omega'$*

Theorem 4 (Konieczny-Pino Pérez [14]). *An operator $\triangle$ is an IC merging operator if and only if there exists a syncretic assignment that maps each profile Ψ to a total pre-order $\leq_\Psi$ such that*

$$mod(\triangle_\mu(\Psi)) = \min(mod(\mu), \leq_\Psi)$$

3.4 Revision vs Update

Intuitively revision operators bring a minimal change to the base by selecting the most plausible models among the models of the new information. Whereas update operators bring a minimal change to each possible world (model) of the base in order to take into account the change described by the new infomation whatever the possible world. So, if we look closely to the two representation theorems (propositions 1, 2 and 3), we easily find the following result:

Theorem 5. *If $\circ$ is a revision operator (i.e. it satisfies (R1)-(R6)), then the operator $\diamond$ defined by:*

$$\varphi \ \diamond \ \mu = \bigvee_{\omega \models \varphi} \varphi_{\{\omega\}} \ \circ \ \mu$$

is an update operator that satisfies (U1)-(U9).

Moreover, for each update operator $\diamond$, there exists a revision operator $\circ$ such that the previous equation holds.

As explained above this proposition states that update can be viewed as a kind of pointwise revision.

3.5 Revision vs Merging

Intuitively revision operators select in a formula (the new evidence) the closest information to a ground information (the old base). And, identically, IC merging operators select in a formula (the integrity constraints) the closest information to a ground information (a profile of bases).

So following this idea it is easy to make a correspondence between IC merging operators and belief revision operators [14]:

Theorem 6 (Konieczny-Pino Pérez [14]). *If $\triangle$ is an IC merging operator (it satisfies (IC0-IC8)), then the operator $\circ$, defined as $\varphi \circ \mu = \triangle_\mu(\varphi)$, is an AGM revision operator (it satisfies (R1-R6)).*

See [14] for more links between belief revision and merging.

4 Confluence operators

So now that we have made clear the connections sketched in figure 1 between revision, update and merging, let us turn now to the definition of confluence operators, that aim to be a pointwise merging, similarly as update is a pointwise revision, as explained in Section 3.4. Let us first define p-consistency for profiles.

Definition 7. *A profile $\Psi = \{\varphi_1, \dots \varphi_n\}$ is* p-consistent *if all its bases are consistent, i.e $\forall \varphi_i \in \Psi$, φ_i is consistent.*

Note that p-consistency is much weaker than consistency, the former just asks that all the bases of the profile are consistent, while the later asks that the conjunction of all the bases is consistent.

Definition 8. *An operator $\diamond$ is a confluence operator if it satisfies the following properties:*

(UC0) $\diamond_\mu(\Psi) \vdash \mu$
(UC1) *If μ is consistent and Ψ is p-consistent, then $\diamond_\mu(\Psi)$ is consistent*
(UC2) *If Ψ is complete, Ψ is consistent and $\bigwedge \Psi \vdash \mu$, then $\diamond_\mu(\Psi) \equiv \bigwedge \Psi$*
(UC3) *If $\Psi_1 \equiv \Psi_2$ and $\mu_1 \equiv \mu_2$, then $\diamond_{\mu_1}(\Psi_1) \equiv \diamond_{\mu_2}(\Psi_2)$*
(UC4) *If φ_1 and φ_2 are complete formulae and $\varphi_1 \vdash \mu$, $\varphi_2 \vdash \mu$,*
then $\diamond_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_1$ is consistent if and only $\diamond_\mu(\{\varphi_1, \varphi_2\}) \wedge \varphi_2$ is consistent
(UC5) $\diamond_\mu(\Psi_1) \wedge \diamond_\mu(\Psi_2) \vdash \diamond_\mu(\Psi_1 \sqcup \Psi_2)$
(UC6) *If Ψ_1 and Ψ_2 are complete profiles and $\diamond_\mu(\Psi_1) \wedge \diamond_\mu(\Psi_2)$ is consistent,*
then $\diamond_\mu(\Psi_1 \sqcup \Psi_2) \vdash \diamond_\mu(\Psi_1) \wedge \diamond_\mu(\Psi_2)$
(UC7) $\diamond_{\mu_1}(\Psi) \wedge \mu_2 \vdash \diamond_{\mu_1 \wedge \mu_2}(\Psi)$
(UC8) *If Ψ is a complete profile and if $\diamond_{\mu_1}(\Psi) \wedge \mu_2$ is consistent*
then $\diamond_{\mu_1 \wedge \mu_2}(\Psi) \vdash \diamond_{\mu_1}(\Psi) \wedge \mu_2$
(UC9) $\diamond_\mu(\Psi \sqcup \{\varphi \vee \varphi'\}) \equiv \diamond_\mu(\Psi \sqcup \{\varphi\}) \vee \diamond_\mu(\Psi \sqcup \{\varphi'\})$

Some of the (UC) postulates are exactly the same as (IC) ones, just like some (U) postulates for update are exactly the same as (R) ones for revision.

In fact, (UC0), (UC3), (UC5) and (UC7) are exactly the same as the corresponding (IC) postulates. So the specificity of confluence operators lies in postulates (UC1), (UC2), (UC6), (UC8) and (UC9). (UC2), (UC4), (UC6) and (UC8) are close to the corresponding (IC) postulates, but hold for complete profiles only. The present formulation of (UC2) is quite similar to formulation of (U2) for update. Note that in the case of a complete profile the hypothesis of (UC2) is equivalent to ask coherence with the constraints, *i.e.* the hypothesis of (IC2). Postulates (UC8) and (UC9) are the main difference with merging postulates, and correspond also to the main difference between revision and KM update operators. (UC9) is the most important postulate, that defines confluence operators as pointwise agregation, just like (U8) defines update operators as pointwise revision. This will be expressed more formally in the next Section (Lemma 1).

5 Representation theorem for confluence operators

In order to state the representation theorem for confluence operators, we first have to be able to "localize" the problem. For update this is done by looking to each model of the base, instead of looking at the base (set of models) as a whole. So for "localizing" the aggregation process, we have to find what is the local view of a profile. That is what we call a state.

Definition 9. *A multi-set of interpretations will be called a* state*. We use the letter e, possibly with subscripts, for denoting states. If $\Psi = \{\varphi_1, \dots, \varphi_n\}$ is a profile and $e = \{\omega_1, \dots, \omega_n\}$ is a state such that $\omega_i \models \varphi_i$ for each i, we say that e is a state of the profile Ψ, or that the state e models the profile Ψ, that will be denoted by $e \models \Psi$. If $e = \{\omega_1, \dots, \omega_n\}$ is a state, we define the profile Ψ_e by putting $\Psi_e = \{\varphi_{\{\omega_1\}}, \dots, \varphi_{\{\omega_n\}}\}$.*

State is an interesting notion. If we consider each base as the current point of view (goals) of the corresponding agent (that can be possibly strengthened in the future) then states are all possible negotiation starting points.

States are the points of interest for confluence operators (like interpretations are for update), as stated in the following Lemma:

Lemma 1. *If $\diamond$ satisfies (UC3) and (UC9) then $\diamond$ satisfies the following*

$$\diamond_\mu(\Psi) \equiv \bigvee_{e \models \Psi} \diamond_\mu(\Psi_e)$$

Defining profile entailment by putting $\Psi \vdash \Psi'$ iff every state of Ψ is a state of Ψ', the previous Lemma has as a corollary the following:

Corollary 1. *If $\diamond$ is a confluence operator then it is monotonic in the profiles, that means that if $\Psi \vdash \Psi'$ then $\diamond_\mu(\Psi) \vdash \diamond_\mu(\Psi')$*

This monotony property, that is not true in the case of merging operators, shows one of the big differences between merging and confluence operators. Remark that there is a corresponding monotony property for update.

Like revision's faithful assignments that have to be "localized" to interpretations for update, merging's syncretic assignments have to be localized to states for confluence.

Definition 10. *A* distributed assignment *is a function mapping each state e to a total pre-order $\leq_e$ over interpretations such that:*

1. $\omega <_{\{\omega,\dots,\omega\}} \omega'$ *if* $\omega' \neq \omega$
2. $\omega \simeq_{\{\omega,\omega'\}} \omega'$
3. *If* $\omega \leq_{e_1} \omega'$ *and* $\omega \leq_{e_2} \omega'$, *then* $\omega \leq_{e_1 \sqcup e_2} \omega'$

4. *If* $\omega <_{e_1} \omega'$ *and* $\omega \leq_{e_2} \omega'$, *then* $\omega <_{e_1 \sqcup e_2} \omega'$

Now we can state the main result of this paper, that is the representation theorem for confluence operators.

Theorem 7. *An operator* $\diamondsuit$ *is a confluence operator if and only if there exists a distributed assignment that maps each state* e *to a total pre-order* $\leq_e$ *such that*

$$mod(\diamondsuit_\mu(\Psi)) = \bigcup_{e \models \Psi} \min(mod(\mu), \leq_e) \tag{1}$$

Unfortunately, we have to omit the proof for space reasons. Nevertheless, we indicate the most important ideas therein. As it is usual, the *if* condition is done by checking each property without any major difficulty. In order to verify the *only if* condition we have to define a distributed assigment. This is done in the following way: for each state e we define a total pre-order $\leq_e$ by putting $\forall \omega, \omega' \in \mathcal{W}\ \omega \leq_e \omega'$ if and only if $\omega \models \diamondsuit_{\varphi_{\{\omega,\omega'\}}}(\Psi_e)$. Then, the main difficulties are to prove that this is indeed a distributed assigment and that the equation (1) holds. In particular, Lema 1 is very helpful for proving this last equation.

Note that this theorem is still true if we remove respectively the postulate (UC4) from the required postulates for confluence operators and the condition 2 from distributed assignments.

6 Confluence vs Update and Merging

So now we are able to state the proposition that shows that update is a special case of confluence, just as revision is a special case of merging.

Theorem 8. *If* $\diamondsuit$ *is a confluence operator (i.e. it satisfies (UC0-UC9)), then the operator* $\diamond$*, defined as* $\varphi \diamond \mu = \diamondsuit_\mu(\varphi)$*, is an update operator (i.e. it satisfies (U1-U9)).*

Concerning merging operators, one can see easily that the restriction of a syncretic assignment to a complete profile is a distributed assignment. From that we obtain the following result (the one corresponding to Theorem 5):

Theorem 9. *If* $\triangle$ *is an IC merging operator (i.e. it satisfies (IC0-IC8)) then the operator* $\diamondsuit$ *defined by*

$$\diamondsuit_\mu(\Psi) = \bigvee_{e \models \Psi} \triangle_\mu(\Psi_e)$$

is a confluence operator (i.e. it satisfies (UC0-UC9)).

Moreover, for each confluence operator $\diamondsuit$*, there exists a merging operator* $\triangle$ *such that the previous equation holds.*

It is interesting to note that this theorem shows that every merging operator can be used to define a confluence operator, and explains why we can consider confluence as a pointwise merging.

Unlike Theorem 5, the second part of the previous theorem doesn't follow straightforwardly from the representation theorems. We need to build a syncretic assignment extending the distributed assignment representing the confluence operator. In order to do that we can use the following construction: Each pre-order $\leq_e$ defines naturally a rank function r_e on natural numbers. Then we put

$$\omega \leq_{\Psi} \omega' \quad \text{if and only if} \quad \sum_{e \models \Psi} r_e(\omega) \leq \sum_{e \models \Psi} r_e(\omega')$$

As a corollary of the representation theorem we obtain the following

Corollary 2. *If $\diamond$ is a confluence operator then the following property holds:*

If $\bigwedge \Psi \vdash \mu$ and Ψ is consistent then $\bigwedge \Psi \wedge \mu \vdash \diamond_{\mu}(\Psi)$

But unlike merging operators, we don't have generally $\diamond_{\mu}(\Psi) \vdash \bigwedge \Psi \wedge \mu$.

Note that this "half of (IC2)" property is similar to the "half of (R2)" satisfied by update operators.

This corollary is interesting since it underlines an important difference between merging and confluence operators. If all the bases agree (i.e. if their conjunction is consistent), then a merging operator gives as result exactly the conjunction, whereas a confluence operator will give this conjunction plus additional results. This is useful if the bases do not represent interpretations that are considered equivalent by the agent, but uncertain information about the agent's current or future state of mind.

7 Example

In this section we will illustrate the behaviour of confluence operators on an example. We can define confluence operators very similarly to merging operators, by using a distance and an aggregation function.

Definition 11. *A* pseudo-distance *between interpretations is a total function $d : \mathcal{W} \times \mathcal{W} \mapsto \mathbb{R}^+$ s.t. for any ω, $\omega' \in \mathcal{W}$: $d(\omega, \omega') = d(\omega', \omega)$, and $d(\omega, \omega') = 0$ if and only if $\omega = \omega'$.*

A widely used distance between interpretations is the Dalal distance [3], denoted d_H, that is the Hamming distance between interpretations (the number of propositional atoms on which the two interpretations differ).

Definition 12. *An* aggregation function *f is a total function associating a nonnegative real number to every finite tuple of nonnegative real numbers s.t. for any $x_1, \ldots, x_n, x$, $y \in \mathbb{R}^+$:*

- *if $x \leq y$, then $f(x_1, \ldots, x, \ldots, x_n) \leq f(x_1, \ldots, y, \ldots, x_n)$* (non-decreasingness)
- *$f(x_1, \ldots, x_n) = 0$ if and only if $x_1 = \ldots = x_n = 0$* *(minimality)*
- *$f(x) = x$* *(identity)*

Sensible aggregation functions are for instance $\max$, sum, or leximax ($Gmax$) [6] [14].

Definition 13 (distance-based confluence operators). *Let d be a pseudo-distance between interpretations and f be an aggregation function. The result $\diamond_{\mu}^{d,f}(\Psi)$ of the confluence of Ψ given the integrity constraints μ is defined by: $mod(\diamond_{\mu}^{d,f}(\Psi)) = \bigcup_{e \models \Psi} \min(mod(\mu), \leq_e)$, where the pre-order $\leq_e$ on $\mathcal{W}$ induced by e is defined by:*

- *$\omega \leq_e \omega'$ if and only if $d(\omega, e) \leq d(\omega', e)$, where*
- *$d(\omega, e) = f(d(\omega, \omega_1) \ldots, d(\omega, \omega_n))$ with $e = \{\omega_1, \ldots, \omega_n\}$.*

It is easy to check that by using usual aggregation functions we obtain confluence operators.

Proposition 1. *Let d be any distance, $\diamond_{\mu}^{d,\Sigma}(\Psi)$ and $\diamond_{\mu}^{d,Gmax}(\Psi)$ are confluence operators (i.e. they satisfy (UC0)-(UC9)).*

Example 2. Let us consider a profile $\Psi = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$ and an integrity constraint μ defined on a propositional language built over four symbols, as follows: $mod(\mu) = \mathcal{W} \setminus \{0110, 1010, 1100, 1110\}$, $mod(\varphi_1) = mod(\varphi_2) = \{1111, 1110\}$, $mod(\varphi_3) = \{0000\}$, and $mod(\varphi_4) = \{1110, 0110\}$.

$\mathcal{W}$	1111	1110	0000	0110	e_1		e_2		e_3		e_4		e_5		e_6		$\diamond_{\mu}^{d,\Sigma}$	$\diamond_{\mu}^{d,\mathsf{Gmax}}$
					Σ	Gmax	Σ	Gmax	Σ	Gmax	Σ	Gmax	Σ	Gmax	Σ	Gmax		
0000	4	3	0	2	11	4430	10	4420	10	4330	9	4320	9	3330	8	3320		
0001	3	4	1	3	11	4331	10	3331	12	4431	11	4331	13	4441	12	4431		
0010	3	2	1	1	9	3321	8	3311	8	3221	7	**3211**	**7**	**2221**	**6**	**2211**	$\times$	$\times$
0011	2	3	2	2	9	3222	8	**2222**	10	3322	9	3222	11	3332	10	3322		$\times$
0100	3	2	1	1	9	3321	8	3311	8	**3221**	**7**	**3211**	**7**	**2221**	**6**	**2211**	$\times$	$\times$
0101	2	3	2	2	9	3222	8	**2222**	10	3322	9	3222	11	3332	10	3322		$\times$
0110	2	1	2	0	7	2221	6	2220	6	2211	5	2210	5	2111	4	2110		
0111	1	2	3	1	7	**3211**	**6**	3111	8	**3221**	**7**	**3211**	9	3222	8	3221	$\times$	$\times$
1000	3	2	1	3	9	3321	10	3331	8	**3221**	9	3321	**7**	**2221**	8	3221	$\times$	$\times$
1001	2	3	2	4	9	3222	10	4222	10	3322	11	4322	11	3332	12	4332		
1010	2	1	2	2	7	2221	8	2222	6	2211	7	2221	5	2111	6	2211		
1011	1	2	3	3	7	**3211**	8	3311	8	**3221**	9	3321	9	3222	10	3322		$\times$
1100	2	1	2	2	7	2221	8	2222	6	2211	7	2221	5	2111	6	2211		
1101	1	2	3	3	7	**3211**	8	3311	8	**3221**	9	3321	9	3222	10	3322		$\times$
1110	1	0	3	1	5	3110	6	3111	4	3100	5	3110	3	3000	4	3100		
1111	0	1	4	2	**5**	4100	**6**	4200	**6**	4110	**7**	4210	**7**	4111	8	4211	$\times$	

Table 1.

The computations are reported in Table 1. The shadowed lines correspond to the interpretations rejected by the integrity constraints. Thus the result has to be taken among

[6] leximax ($Gmax$) is usually defined using lexicographic sequences, but it can be easily represented by reals to fit the above definition (see e.g. [13]).

the interpretations that are not shadowed. The states that model the profile are the following ones:
$e_1 = \{1111, 1111, 0000, 1110\}, e_2 = \{1111, 1111, 0000, 0110\}$,
$e_3 = \{1111, 1110, 0000, 1110\}, e_4 = \{1110, 1111, 0000, 0110\}$,
$e_5 = \{1110, 1110, 0000, 1110\}, e_6 = \{1110, 1110, 0000, 0110\}$.

For each state, the Table gives the distance between the interpretation and this state for the Σ aggregation function, and for the $Gmax$ one. So one can then look at the best interpretations for each state.

So for instance for $\diamond_\mu^{d,\Sigma}(\Psi)$, e_1 selects the interpretation 1111, e_2 selects 0111 and 1111, etc. So, taking the union of the interpretations selected by each state, gives $mod(\diamond_\mu^{d,\Sigma}(\Psi)) = \{0010, 0100, 0111, 1000, 1111\}$.

Similarly we obtain $mod(\diamond_\mu^{d,Gmax}(\Psi)) = \{0100, 0011, 0010, 0101, 0111, 1000, 1011, 1101\}$.

8 Conclusion

We have proposed in this paper a new family of change operators. Confluence operators are pointwise merging, just as update can be seen as a pointwise revision. We provide an axiomatic definition of this family, a representation theorem in terms of pre-orders on interpretations, and provide examples of these operators.

In this paper we define confluence operators as generalization to multiple bases of total update operators (i.e. which semantical counterpart are total pre-orders). A perspective of this work is to try to extend the result to partial update operators.

As Example 1 suggests, these operator can prove meaningful to aggregate the goals of a group of agents. They seem to be less adequate for aggregating beliefs, where the global minimization done by merging operators is more appropriate for finding the most plausible worlds. This distinction between goal and belief aggregation is a very interesting perspective, since, as far as we know, no such axiomatic distinction as been ever discussed.

Acknowledgements

The idea of this paper comes from discussions in the 2005 and 2007 Dagstuhl seminars ($\#5321$ and $\#7351$) *"Belief Change in Rational Agents"* . The authors would like to thank the Schloss Dagstuhl institution, and the participants of the seminars, especially Andreas Herzig for the initial question *"If merging can be seen as a generalization of revision, what is the generalization of update ?"*. Here is an answer !

The authors would like to thank the reviewers of the paper for their helpful comments.

The second author was partially supported by a research grant of the Mairie de Paris and by the project CDCHT-ULA N° C-1451-07-05-A. Part of this work was done when the second author was a visiting professor at CRIL (CNRS UMR 8188) from September to December 2007 and a visiting researcher at TSI Department of Telecom ParisTech (CNRS UMR 5141 LTCI) from January to April 2008. The second author thanks to CRIL and TSI Department for the excellent working conditions.

References

1. 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.
2. R. Booth. Social contraction and belief negotiation. In *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 374–384, 2002.
3. M. Dalal. Investigations into a theory of knowledge base revision: preliminary report. In *Proceedings of the American National Conference on Artificial Intelligence (AAAI'88)*, pages 475–479, 1988.
4. F. Dupin de Saint-Cyr and J. Lang. Belief extrapolation (or how to reason about observations and unpredicted change). *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 497–508, 2002.
5. P. Gärdenfors. *Knowledge in flux*. MIT Press, 1988.
6. P. Gärdenfors, editor. *Belief Revision*. Cambridge University Press, 1992.
7. A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17(157-180), 1988.
8. A. Herzig and O. Rifi. Update operations: a review. In *Proceedings of the Thirteenth European Conference on Artificial Intelligence (ECAI'98)*, pages 13–17, 1998.
9. H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 387–394, 1991.
10. H. Katsuno and A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.
11. H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In P. Gärdenfors, editor, *Belief Revision*. Cambridge University Press, 1992.
12. S. Konieczny. Belief base merging as a game. *Journal of Applied Non-Classical Logics*, 14(3):275–294, 2004.
13. S. Konieczny, J. Lang, and P. Marquis. DA^2 merging operators. *Artificial Intelligence*, 157(1-2):49–79, 2004.
14. S. Konieczny and R. Pino Pérez. Merging information under constraints: a logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.
15. J. Lang. Belief update revisited. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'07)*, pages 2517–2522, 2007.
16. J. Lobo and C. Uzcátegui. Abductive change operators. *Fundamenta Informaticae*, 27(4):385–411, 1996.
17. T. Meyer, N. Foo, D. Zhang, and R. Kwok. Logical foundations of negotiation: Outcome, concession and adaptation. In *Proceedings of the American National Conference on Artificial Intelligence (AAAI'04)*, pages 293–298, 2004.
18. T. Meyer, N. Foo, D. Zhang, and R. Kwok. Logical foundations of negotiation: Strategies and preferences. In *Proceedings of the Ninth Conference on Principles of Knowledge Representation and Reasoning (KR'04)*, pages 311–318, 2004.
19. P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2):133–160, 1997.
20. D. Zhang, N. Foo, T. Meyer, and R. Kwok. Negotiation as mutual belief revision. In *Proceedings of the American National Conference on Artificial Intelligence (AAAI'04)*, pages 317–322, 2004.

BIBLIOGRAPHIE

[AA98] O. Arieli et A. Avron. The value of four values. *Artificial Intelligence*, 102 :97–141, 1998. 23

[AGM85] C. E. Alchourrón, P. Gärdenfors et D. Makinson. On the logic of theory change : Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50 :510–530, 1985. 9, 14, 31, 33, 37, 38, 45, 47

[ALP⁺00] J. J. Alferes, J. A. Leite, L. M. Pereira, H. Przymusinska et T. C. Przymusinski. Dynamic updates of non-monotonic knowledge bases. *Journal of Logic Programming*, 45(1-3) :43–70, 2000. 65

[ALS98] M. H. F. Alves, D. Laurent et N. Spyratos. Update rules in datalog programs. *Journal of Logic and Computation*, 8(6) :745–775, 1998. 65

[AM82] C. E. Alchourrón et D. Makinson. The logic of theory change : Contraction functions and their associated revision functions. *Theoria*, 48 :14–37, 1982. 37

[AM85] C. E. Alchourrón et D. Makinson. On the logic of theory change : Safe contraction. *Studia Logica*, 44 :405–422, 1985. 36

[Ari08] O. Arieli. Distance-based paraconsistent logics. *International Journal of Approximate Reasoning*, 48(3) :766–783, 2008. 20

[Arr63] K. J. Arrow. *Social choice and individual values*. Wiley, New York, second édition, 1963. 14, 47, 65

[ASS02] K. Arrow, A. K. Sen et K. Suzumura, éditeurs. *Handbook of Social Choice and Welfare*, tome 1. North-Holland, 2002. 65

[BBN04] S. Benferhat, J.-F. Bonnefon et R. D. S. Neves. An experimental analysis of possibilistic default reasoning. Dans D. Dubois, C. A. Welty et M.-A. Williams, éditeurs, *Principles of the Ninth International Conference on Knowledge Representation and Reasoning (KR'04)*, pages 130–140. AAAI Press, 2004. 12

[BBN05] S. Benferhat, J. Bonnefon et R. Neves. An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese*, 146(1) :53–70, 2005. 12

[BDKP02] S. Benferhat, D. Dubois, S. Kaci et H. Prade. Possibilistic merging and distance-based fusion of propositional information. *Annals of Mathematics and Artificial Intelligence*, 34(1-3) :217–252, 2002. 64

[BDL+98] S. Benferhat, D. Dubois, J. Lang, H. Prade, A. Saffioti et P. Smets. A general approach for inconsistency handling and merging information in prioritized knowledge bases. Dans *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 466–477. 1998. 52

[BDP92] S. Benferhat, D. Dubois et H. Prade. Representing default rules in possibilistic logic. Dans *Proceedings of the Third International Conference on the Principles of Knowledge Representation and Reasoning (KR'92)*, pages 673–684. 1992. 19

[BDP93] S. Benferhat, D. Dubois et H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. Dans D. Heckerman et E. H. Mamdani, éditeurs, *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI '93)*, pages 411–419. Morgan Kaufmann, 1993. 19

[BDP97] S. Benferhat, D. Dubois et H. Prade. Some syntactic approaches to the handling of inconsistent knowledge bases : a comparative study, part 1 : the flat case. *Studia Logica*, 58 :17–45, 1997. 19, 20

[BDP98] S. Benferhat, D. Dubois et H. Prade. *Logic at work (Ewa Orlowska, (eds))*, tome 24, chapitre Some syntactic approaches to the handling of inconsistent knowledge bases : A comparative study. Part 2 : the prioritized case, pages 473–511. Physica-Verlag, Heidelberg, 1998. 19

[Bel77a] N. D. Belnap. *Contemporary aspects of philosophy*, chapitre How a computer should think, pages 30–56. Oriel Press, Stocksfield, GB, 1977. 23, 27

[Bel77b] N. D. Belnap. *Modern uses of multiple-valued logic*, chapitre A useful four-valued logic, pages 8–37. Reidel, Dordrecht, NL, 1977. 23, 27

[BF83] S. J. Brams et P. C. Fishburn. *Approval voting*. Springer Verlag, 1983. 60

[BH95] P. Besnard et A. Hunter. Quasi-classical logic : Non-trivializable classical reasoning from inconsistent information. Dans *Proceedings of the 3rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'95)*, tome LNAI 946, pages 44–51. Springer Verlag, 1995. 21, 27

[BK03] S. Benferhat et S. Kaci. Fusion of possibilistic knowledge bases from a postulate point of view. *International Journal of Approximate Reasoning*, 33(3) :255–285, 2003. 64

[BKM91] C. Baral, S. Kraus et J. Minker. Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 3(2) :208–220, 1991. 53

[BKMS92] C. Baral, S. Kraus, J. Minker et V. S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1) :45–71, 1992. 19, 52, 53

[BL00] I. Bloch et J. Lang. Towards mathematical morpho-logics. Dans *Proceedings of the Eigth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'00)*, pages 1405–1412. 2000. 65

[BLP05] S. Benferhat, S. Lagrue et O. Papini. Revision of partially ordered information : Axiomatization, semantics and iteration. Dans L. P. Kaelbling et A. Saffiotti, éditeurs, *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*, pages 376–381. 2005. 43

[BLR07] S. Benferhat, S. Lagrue et J. Rossit. An egalitarist fusion of incommensurable ranked belief bases under constraints. Dans *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI'07)*, pages 367–372. AAAI Press, 2007. 65

[BLR09] S. Benferhat, S. Lagrue et J. Rossit. An analysis of sum-based incommensurable belief base merging. Dans L. Godo et A. Pugliese, éditeurs, *Proceedings of the third International Conference on Scalable Uncertainty Management (SUM'08)*, tome 5785 de *Lecture Notes in Computer Science*, pages 55–67. Springer, 2009. 65

[BM06] R. Booth et T. Meyer. Admissible and restrained revision. *Journal of Artificial Intelligence Research*, 26 :127–151, 2006. 44, 45

[BNDP08] J.-F. Bonnefon, R. D. S. Neves, D. Dubois et H. Prade. Predicting causality ascriptions from background knowledge : model and experimental validation. *International Journal of Approximate Reasoning*, 48(3) :752–765, 2008. 12

[Boo02] R. Booth. Social contraction and belief negotiation. Dans *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 374–384. 2002. 69

[Boo06] R. Booth. Social contraction and belief negotiation. *Information Fusion*, 7(1) :19–34, 2006. 69

[Bor81] J. C. Borda. Mémoire sur les élections au srutin. Histoire de l'Académie Royale des Sciences, 1781. 14

[Bor85] A. Borgida. Language features for flexible handling of exceptions in information systems. *ACM Transactions on Database Systems*, 10 :563–603, 1985. 14, 56

[Bou93] C. Boutilier. Revision sequences and nested conditionals. Dans *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 519–525. 1993. 31, 43

[Bou96] C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3) :262–305, 1996. 43

[Bre89] G. Brewka. Preferred subtheories : an extended logical framework for default reasoning. Dans *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI'89)*, pages 1043–1048. 1989. 19

[BS96] P. Besnard et T. Schaub. A simple signed system for paraconsistent reasoning. Dans J. J. Alferes, L. M. Pereira et E. Orlowska, éditeurs, *Proccedings of the European Workshop on Logics in Artificial Intelligence (JELIA'96)*, tome 1126 de *Lecture Notes in Computer Science*, pages 404–416. Springer, 1996. 21

[BSZ91] S. Barberà, H. Sonnenschein et L. Zhou. Voting by committees. *Econometrica*, 59(3) :595–609, 1991. 55

[Béz00] J. Béziau. *Frontiers of Paraconsistent Logic (D. Batens et al. eds)*, chapitre What is Paraconsistent Logic ?, pages 95–111. Research Studies Press, Baldock, 2000. 21

[CKMS09a] J.-F. Condotta, S. Kaci, P. Marquis et N. Schwind. Merging qualitative constraint networks in a piecewise fashion. Dans *Proceedings of the twenty first International Conference on Tools with Artificial Intelligence (ICTAI'09)*, pages 605–608. 2009. 66

[CKMS09b] J.-F. Condotta, S. Kaci, P. Marquis et N. Schwind. Merging qualitative constraints networks using propositional logic. Dans *10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, pages 347–358. 2009. 66

[CM01] W. Carnielli et J. Marcos. A taxonomy of C-systems. *Paraconsistency-the Logical Way to the Inconsistent, Lecture Notes in Pure and Applied Mathematics*, 228 :1–94, 2001. 21

[Con85] M. de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris, 1785. 14

[CP06] J. L. Chacón et R. P. Pérez. Merging operators : Beyond the finite case. *Information Fusion*, 7(1) :41–60, 2006. 50

[Dal88] M. Dalal. Investigations into a theory of knowledge base revision : preliminary report. Dans *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'88)*, pages 475–479. 1988. 14, 56, 65

[DDL06] J. P. Delgrande, D. Dubois et J. Lang. Iterated revision as prioritized merging. Dans P. Doherty, J. Mylopoulos et C. A. Welty, éditeurs, *Proceedings of the tenth International Conference on Knowledge Representation and Reasoning (KR'06)*, pages 210–220. AAAI Press, 2006. 63

[DLP94] D. Dubois, J. Lang et H. Prade. *Handbook of Logic in Artificial Intelligence and Logic Programming*, tome 3 : Nonmonotonic Reasoning and Uncertain Reasoning, chapitre Possibilistic Logic. Oxford Science Publications, 1994. 64

[DLST07] J. P. Delgrande, D. H. Liu, T. Schaub et S. Thiele. COBA 2.0 : A consistency-based belief change system. Dans K. Mellouli, éditeur, *Proceedings of the ninth International Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'07)*, tome 4724 de *Lecture Notes in Computer Science*, pages 78–90. Springer, 2007. 58

[DP91] D. Dubois et H. Prade. Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50 :223–239, 1991. 41

[DP94] A. Darwiche et J. Pearl. On the logic of iterated belief revision. Dans M. Kaufmann, éditeur, *Proceedings of the International Conference on Theoretical Aspects of Reasoning about Knowledge : (TARK'94)*, pages 5–23. 1994. 42

[DP97] A. Darwiche et J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89 :1–29, 1997. 31, 42, 43, 45

[DS07] J. P. Delgrande et T. Schaub. A consistency-based framework for merging knowledge bases. *Journal of Applied Logic*, 5(3) :459–477, 2007. 57

[DSTW08] J. P. Delgrande, T. Schaub, H. Tompits et S. Woltran. Belief revision of logic programs under answer set semantics. Dans G. Brewka et J. Lang, éditeurs, *Principles of the eleventh International Conference on Knowledge Representation and Reasoning (KR'08)*, pages 411–421. AAAI Press, 2008. 65

[DSTW09] J. P. Delgrande, T. Schaub, H. Tompits et S. Woltran. Merging logic programs under answer set semantics. Dans P. M. Hill et D. S. Warren, éditeurs, *Proceedings of the twenty fifth International Conference on Logic Programming (ICLP'09)*, tome 5649 de *Lecture Notes in Computer Science*, pages 160–174. Springer, 2009. 66

[Dub86] D. Dubois. Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence*, 5(5) :403–416, 1986. 41

[Dun95] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77 :321–357, 1995. 66

[FH96a] L. Fariñas del Cerro et A. Herzig. Belief change and dependence. Dans *Proceedings of the Sixth conference on Theoretical aspects of rationality and knowledge (TARK '96)*, pages 147–161. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996. 36

[FH96b] N. Friedman et J. Halpern. Belief revision : a critique. Dans *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*, pages 421–431. 1996. 63

[FL94] M. Freund et D. Lehmann. Belief revision and rational inference. Rapport technique TR-94-16, Institute of Computer Science, The Hebrew University of Jerusalem, 1994. 41, 42

[FUV83] R. Fagin, J. D. Ullman et M. Y. Vardi. On the semantics of updates in databases. Dans *Proceedings of the 2nd ACM SIGACT-SIGMOD Symposium on the Principles of Database Systems*, pages 352–365. 1983. 14

[Gab85] D. M. Gabbay. Theoretical foundations for nonmonotonic reasoning in experts systems. Dans K. Apt, éditeur, *Logic and Models of Concurrent Systems*. Springer Verlag, 1985. 13

[Gär88] P. Gärdenfors. *Knowledge in flux*. MIT Press, 1988. 9, 14, 33, 35, 36, 38, 45, 47

[Gär90] P. Gärdenfors. Belief revision and nonmonotonic logic : two sides of the same coin ? Dans *Proceeding of the ninth European Conference on Artificial Intelligence (ECAI'90)*, pages 768–773. 1990. 40

[GH06] J. Grant et A. Hunter. Measuring inconsistency in knowledgebases. *Journal of Intelligent Information Systems*, 27 :159–184, 2006. 26

[GH08] N. Gorogiannis et A. Hunter. Merging first-order knowledge using dilation operators. Dans S. Hartmann et G. Kern-Isberner, éditeurs, *Proceedings of the fifth International Symposium on Foundations of Information and Knowledge Systems (FoIKS'08)*, tome 4932 de *Lecture Notes in Computer Science*, pages 132–150. Springer, 2008. 65

[Gib73] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41 :587–602, 1973. 60

[Gin88] M. L. Ginsberg. Multivalued logics : A uniform approach to reasoning in AI. *Computer Intelligence*, 4 :256–316, 1988. 23

[GMP06] É. Grégoire, B. Mazure et C. Piette. Extracting muses. Dans G. Brewka, S. Coradeschi, A. Perini et P. Traverso, éditeurs, *Proceedings of the seventeenth European Conference on Artificial Intelligence (ECAI'06)*, pages 387–391. IOS Press, 2006. 26

[GMP09] É. Grégoire, B. Mazure et C. Piette. Does this set of clauses overlap with at least one mus ? Dans R. A. Schmidt, éditeur, *Proceedings of the twenty second International Conference on Automated Deduction (CADE'09)*, tome 5663 de *Lecture Notes in Computer Science*, pages 100–115. Springer, 2009. 26

[Gra78] J. Grant. Classifications for inconsistent theories. *Notre Dame Journal of Formal Logic*, 19 :435–444, 1978. 26

[Gro88] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17(157-180), 1988. 38, 39

[Han98] S. O. Hansson. What's new isn't always best. *Theoria*, pages 1–13, 1998. Special issue on non-prioritized belief revision. 45

[HPW09] J. Hué, O. Papini et E. Würbel. Merging belief bases represented by logic programs. Dans C. Sossai et G. Chemello, éditeurs, *Proceedings of the tenth International Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, tome 5590 de *Lecture Notes in Computer Science*, pages 371–382. Springer, 2009. 65

[HR99] A. Herzig et O. Rifi. Propositional belief base update and minimal change. *Artificial Intelligence*, 115(1) :107–138, 1999. 71

[Hun96] A. Hunter. Paraconsistent logics. Dans *Handbook of Defeasible Reasoning and Uncertain Information*, pages 11–36. Kluwer, 1996. 21

[Hun02] A. Hunter. Measuring inconsistency in knowledge via quasi-classical models. Dans *Proceedings of the American National Conference on Artificial Intelligence (AAAI'02)*, pages 68–73. 2002. 24, 26, 27

[JT07] Y. Jin et M. Thielscher. Iterated belief revision, revised. *Artificial Intelligence*, 171 :1–18, 2007. 44, 45

[KBDP00] S. Kaci, S. Benferhat, D. Dubois et H. Prade. A principled analysis of merging operations in possibilistic logic. Dans C. Boutilier et M. Goldszmidt, éditeurs, *Proceedings of the sixteenth Conference in Uncertainty in Artificial Intelligence (UAI'00)*, pages 24–31. Morgan Kaufmann, 2000. 64

[KLM90] S. Kraus, D. Lehmann et M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44 :167–207, 1990. 13, 40

[KM91a] H. Katsuno et A. O. Mendelzon. On the difference between updating a knowledge base and revising it. Dans *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 387–394. 1991. 71

[KM91b] H. Katsuno et A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52 :263–294, 1991. 9, 35, 39, 42, 50

[Kni01] K. M. Knight. Measuring inconsistency. *Journal of Philosophical Logic*, 31 :77–98, 2001. 25, 26

[Kni02] K. M. Knight. *A theory of inconsistency*. Thèse de doctorat, The university of Manchester, 2002. 24, 25, 27

[Kni03] K. M. Knight. Two information measures for inconsistent sets. *Journal of Logic, Language and Information*, 12 :227–248, 2003. 24, 25, 27

[KW85] A. M. Keller et M. Winslett. On the use of an extended relational model to handle changing incomplete information. *IEEE Transaction on Software Engineering*, SE-11(7) :620–633, 1985. 14

[Leh95] D. Lehmann. Belief revision, revised. Dans *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, pages 1534–1540. 1995. 31, 42, 45, 63

[Lis10] C. List. The theory of judgment aggregation : An introductory review. *Synthese (to appear)*, 2010. 75

[LM92] D. Lehmann et M. Magidor. What does a conditional knowledge base entail ? *Artificial Intelligence*, 55 :1–60, 1992. 13, 40

[LM99] J. Lin et A. O. Mendelzon. Knowledge base merging by majority. Dans *Dynamic Worlds : From the Frame Problem to Knowledge Management*. Kluwer, 1999. 51

[LM02] J. Lang et P. Marquis. Resolving inconsistencies by variable forgetting. Dans *Proceedings of the eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 239–250. Morgan Kaufmann Publishers, Inc., Toulouse, France, 2002. 20

[Loz94] E. Lozinskii. Information and evidence in logic systems. *Journal of Experimental and Theoretical Artificial Intelligence*, 6 :163–193, 1994. 24, 25, 27

[LP04] C. List et P. Pettit. Aggregating Sets of Judgments : Two Impossibility Results Compared. *Synthese*, 140(1) :207–235, 2004. 75

[Mak94] D. Makinson. *Handbook of Logic in Artificial Intelligence and Logic Programming*, tome III, chapitre General Pattern in nonmonotonic reasoning, pages 35–110. Clarendon Press, Oxford, 1994. 6, 13, 40

[Mey01] T. A. Meyer. On the semantics of combination operations. *Journal of Applied Non-Classical Logics*, 11(1-2) :59–84, 2001. 64

[MFZK04] T. Meyer, N. Foo, D. Zhang et R. Kwok. Logical foundations of negotiation : Outcome, concession and adaptation. Dans *Proceedings of the American National Conference on Artificial Intelligence (AAAI'04)*, pages 293–298. 2004. 69

[MG89] D. Makinson et P. Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. Dans *The Logic of Theory Change, Workshop, Konstanz, FRG*, tome 465 de *LNAI*, pages 185–205. 1989. 41

[MPP05] T. A. Meyer, P. P. Parra et L. Perrussel. Mediation using m-states. Dans L. Godo, éditeur, *Proceedings of the eighth International Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, tome 3571 de *Lecture Notes in Computer Science*, pages 489–500. Springer, 2005. 69

[NJ50] J. Nash Jr. The bargaining problem. *Econometrica*, 18(2) :155–162, 1950. 69, 72

[Par94] J. B. Paris. *The uncertain reasoner's companion : a mathematical perspective*. Numéro 39 dans Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1994. 25

[Pig06] G. Pigozzi. Belief merging and the discursive dilemma : an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2) :285–298, 2006. 75

[Pri91] G. Priest. Minimally inconsistent LP. *Studia Logica*, 50 :321–331, 1991. 22

[QLB06] G. Qi, W. Liu et D. A. Bell. Merging stratified knowledge bases under constraints. Dans *Proceedings of the twenty-first AAAI Conference on Artificial Intelligence (AAAI'06)*. AAAI Press, 2006. 64

[Rev97] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2) :133–160, 1997. 48, 51

[RM70] N. Rescher et R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1 :179–219, 1970. 19

[Rot09] H. Rott. Shifting priorities : Simple representations for twenty-seven iterated theory change operators. Dans D. Makinson, J. Malinowski et H. Wansing, éditeurs, *Towards Mathematical Philosophy*, pages 269–296. Springer, 2009. 40, 44

[Sat75] M. Satterthwaite. Strategy-proofness and Arrow's conditions. *Journal of Economic Theory*, 10 :187–217, 1975. 60

[Sav54] L. Savage. *The Foundations of Statistics*. J. Wiley, New York, 1954. Second revised edition, 1972. 47

[Sha53] L. S. Shapley. A value for n-person games. Dans H. W. Kuhn et A. W. Tucker, éditeurs, *Contributions to the Theory of Games II (Annals of Mathematics Studies 28)*, pages 307–317. Princeton University Press, 1953. 72

[Sho87] Y. Shoham. A semantical approach to nonmonotonic logics. Dans *Proc. of the* 2^{nd} *IEEE Symposium on Logic in Computer Science (LICS'87)*, pages 275–279. Washington (DC), 1987. 13, 18

[SP09] S. Schockaert et H. Prade. Merging conflicting propositional knowledge by similarity. Dans *Proceedings of the twenty first International Conference on Tools with Artificial Intelligence (ICTAI'09)*, pages 224–228. IEEE Computer Society, 2009. 58

[Spo87] W. Spohn. Ordinal conditional functions : a dynamic theory of epistemic states. Dans W. L. Harper et B. Skyrms, éditeurs, *Causation in Decision, Belief Change, and Statistics*, tome 2, pages 105–134. 1987. 64

[Tar36] A. Tarski. Über den begriff der logischen folgerung. *Actes du Congrès International de Philosophie Scientifique*, 7 :1–11, 1936. 17

[TCW83] A. Tarski, J. Corcoran et J. Woodger. *Logic, semantics, metamathematics : papers from 1923 to 1938*. Hackett publishing company, 1983. 17

[Wil94] M. A. Williams. Transmutations of knowledge systems. Dans *Proceedings of the Fourth International Conference on the Principles of Knowledge Representation and Reasoning (KR'94)*, pages 619–629. 1994. 31

[Win88] M. Winslett. Reasoning about action using a possible model approach. Dans *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI'88)*, pages 89–93. 1988. 14

[Win90] M. Winslett. *Updating logical databases*. Cambridge University Press, 1990. 19

[ZF98] Y. Zhang et N. Y. Foo. Updating logic programs. Dans *Proceedings of the European Conference on Artificial Intelligence (ECAI'98)*, pages 403–407. 1998. 65

[Zha05] D. Zhang. A logical model for Nash bargaining solution. Dans *Proceedings of the nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 983–988. 2005. 69

[Zha07] D. Zhang. Reasoning about bargaining situations. Dans *Proceedings of the National Conference on Artificial Intelligence (AAAI'99)*, tome 22, page 154. 2007. 69

[ZZ08] D. Zhang et Y. Zhang. An ordinal bargaining solution with fixed-point property. *Journal of Artificial Intelligence Research*, 33(1) :433–464, 2008. 69

Publications

Revues Internationales

1. Sébastien Konieczny, Ramón Pino Pérez. A framework for iterated revision. *Journal of Applied Non-Classical Logics*. 10(3-4). 339-367. 2000.
2. Sébastien Konieczny, Ramón Pino Pérez. Merging information under constraints : a logical framework. *Journal of Logic and Computation*. 12(5). 773-808. 2002.
3. Sébastien Konieczny, Didier Dubois, Henri Prade. Quasi-possibilistic logic and its measures of information and conflict. *Fundamenta Informaticae*. 57(2-4). 101-125. 2003.
4. Sébastien Konieczny. Belief base merging as a game. *Journal of Applied Non-Classical Logics*. 14(3). 275-294. 2004.
5. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. DA^2 merging operators. *Artificial Intelligence*. 157(1-2). 49-79. 2004.
6. Sébastien Konieczny, Ramón Pino Pérez. Propositional belief base merging or how to merge beliefs/goals coming from several sources and some links with social choice theory. *European Journal of Operational Research*. 160(3). 785-802. 2005.
7. Olivier Gauwin, Sébastien Konieczny, Pierre Marquis. Conciliation through iterated belief merging. *Journal of Logic and Computation*. 17(5). 909-937. 2007.
8. Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, Pierre Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence*. 171. 740-753. 2007.
9. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*. 28. 49-105. 2007.
10. Philippe Besnard, Sébastien Konieczny, Pierre Marquis. Bipolarity in bilattice logics. *International Journal of Intelligent Systems*. 23(9). 1046-1061. 2008.

Edition de Livres

11. Eric Grégoire, Sébastien Konieczny Ed. *Information Fusion*. 7(1). *Special issue on Logic-based approaches to information fusion*. 2006.

Chapitres de Livre

12. Hassan Bezzazi, Stéphane Janot, Sébastien Konieczny, Ramón Pino Pérez. Analysing rational properties of change operators based on forward chaining. *Transaction and Change in Logic Databases*. Springer. LNCS 1472. 317-339. 1998.
13. Sébastien Konieczny, Ramón Pino Pérez. Sur la représentation des états épistémiques et la révision itérée. *Révision des Croyances*. Hermes. 181-202. 2002.
14. Sébastien Konieczny, Ramón Pino Pérez. De la révision à la fusion de croyances. *Révision des Croyances*. Hermes. 203-226. 2002.
15. Anthony Hunter, Sébastien Konieczny. Approaches to measuring inconsistent information. *Inconsistency tolerance*. Springer. LNCS 3300. 189-234. 2005.

Revues Nationales

16. Sébastien Konieczny, Ramón Pino Pérez. Sur la modélisation de la concertation entre agents : fusion de bases de croyances. *Revue Information-Interaction-Intelligence*. Cepadues. Hors Série.91-113. 2002.

Conférences Internationales

17. Sébastien Konieczny, Ramón Pino Pérez. On the logic of merging. *Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*. 488-498. 1998.
18. Sébastien Konieczny, Ramón Pino Pérez. Merging with integrity constraints. *Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99)*. 233-244. 1999.
19. Salem Benferhat, Sébastien Konieczny, Odile Papini, Ramón Pino Pérez. Iterated revision by epistemic states : axioms, semantics and syntax. *Fourteenth European Conference on Artificial Intelligence (ECAI'00)*. 13-17. 2000.
20. Sébastien Konieczny. On the difference between merging knowledge bases and combining them. *Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR'00)*. 135-144. 2000.
21. Sébastien Konieczny, Ramón Pino Pérez. Some operators for iterated revision. *Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01)*. 498-509. 2001.
22. Sébastien Konieczny, Pierre Marquis. Three-valued logics for inconsistency handling. *Eighth European Conference on Logic in Artifical Intelligence (JELIA'02)*. 332-344. 2002.
23. Sébastien Konieczny, Ramón Pino Pérez. On the frontier between arbitration and majority. *Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*. 109-118. 2002.
24. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Distance-based merging : a general framework and some complexity results. *Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*. 97-108. 2002.
25. Andreas Herzig, Sébastien Konieczny, Laurent Perussel. On iterated revision in the AGM framework. *Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03)*. 477-488. 2003.
26. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Quantifying information and contradiction in propositional logic through test actions. *Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*. 106-111. 2003.
27. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. On merging strategy-proofness. *Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR'04)*. 357-368. 2004.
28. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Reasoning under inconsistency : the forgotten connective. *Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*. 484-489. 2005.

29. Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, Pierre Marquis. Merging Argumentation Systems. *Twentieth American National Conference on Artificial Intelligence (AAAI'05).* 614-619. 2005.

30. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Quota and Gmin merging operators. *Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05).* 424-429. 2005.

31. Olivier Gauwin, Sébastien Konieczny, Pierre Marquis. Conciliation and consensus in iterated belief merging. *Eigth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05).* 514-526. 2005.

32. Anthony Hunter, Sébastien Konieczny. Shapley Inconsistency Values. *Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR'06).* 249-259. 2006.

33. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. Extending classical planning to the multi-agent case : a game-theoretic approach. *Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'07).* 731-742. 2007.

34. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. A model for multiple outcomes games. *Twentieth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'08).* 27-34. 2008.

35. Sébastien Konieczny, Ramón Pino Pérez. Improvement operators. *Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08).* 177-186. 2008.

36. Anthony Hunter, Sébastien Konieczny. Measuring inconsistency through minimal inconsistent sets. *Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08).* 358-366. 2008.

37. Sébastien Konieczny, Ramón Pino Pérez. Confluence operators. *Eleventh European Conference on Logics in Artificial Intelligence (JELIA'08).* 272-284. 2008.

38. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Propositional merging operators based on set-theoretic closeness (short paper). *Eighteenth European Conference on Artificial Intelligence (ECAI'08).* 737-738. 2008.

39. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Conflict-based merging operators. *Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08).* 348-357. 2008.

40. Sébastien Konieczny. Using transfinite ordinal conditional functions. *Tenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09).* 396-407. 2009.

Workshops Internationaux

41. Hassan Bezzazi, Stéphane Janot, Sébastien Konieczny, Ramón Pino Pérez. Forward chaining and change operators. *Workshop on (Trans)Actions and Change in Logic Programming and Deductive Databases (DYNAMICS'97), in conjunction with International Logic Programming Symposium (ILPS'97).* 135-146. 1997.

42. Sébastien Konieczny, Ramón Pino Pérez. On the difference between arbitration and majority merging. *Fourth Workshop on Logic, Language, Information and Computation (WoLLIC'97)*. 1997.
43. Sébastien Konieczny, Ramón Pino Pérez. An analysis of merging from a logical point of view. *Eleventh Latin American Symposium on Mathematical Logic (SLALM'98)*. 1998.
44. Sébastien Konieczny, Ramón Pino Pérez. Dynamical revision operators with memory. *Ninth International Workshop on Non-Monotonic Reasoning (NMR'02)*. 171-179. 2002.
45. Didier Dubois, Sébastien Konieczny, Henri Prade. Quasi-possibilistic logic and measures of information and conflict. *First International Workshop on Knowkedge Representation and Approximate reasoning (KR&AR'03)*. 2003.
46. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Measures of contradiction and information in propositional bases. *Third World Congress on Paraconsistency (WCP'03)*. 2003.
47. Sébastien Konieczny. Belief as a game on beliefs. *Uncertainty, Incompleteness, Imprecision and Conflict in Multiple Data Sources (Affiliated workshop to ECSQARU'03)*. 2003.
48. Sébastien Konieczny. Propositional belief merging and belief negotiation model. *Tenth International Workshop on Non-Monotonic Reasoning (NMR'04)*. 249-257. 2004.
49. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. On the logic of merging : Quota and Gmin merging operators. *Seventh International Symposium on Logical Formalizations of Commonsense Reasoning (CommonSense'05)*. 2005.
50. Olivier Gauwin, Sébastien Konieczny, Pierre Marquis. Iterated belief merging as conciliation operators. *Seventh International Symposium on Logical Formalizations of Commonsense Reasoning (CommonSense'05)*. 2005.
51. Sébastien Konieczny. More games in belief game model : a preliminary report. *Formal Approaches to Multi-Agent Systems (FAMAS'06) - Affiliated workshop to ECAI'06*. 2006.
52. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Propositional merging operators based on set-theoretic closeness. *Twelfth International Workshop on Non-Monotonic Reasoning (NMR'08)*. 2008.
53. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. A Characterization of an optimality criterion for decision making under complete ignorance. *Twelfth International Workshop on Non Monotonic Reasoning (NMR'08)*. 2008.
54. Sébastien Konieczny, Ramón Pino Pérez. Confluence operators. *Second International Workshop on Computational Social Choice (COMSOC'08)*. 2008.

Workshops Internationaux (sans comité de selection)

55. Sébastien Konieczny. Conciliation, negotiation and merging. *Dagsthul Seminar #53-21 : Belief Change in Rational Agents : Perspectives from Artificial Intelligence, Philosophy, and Economics*. 2005.

56. Sébastien Konieczny. Inconsistency measures and their application to belief change. *Dagsthul Seminar #7351 : Formal Models of Belief Change in Rational Agents.* 2007.

57. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. Qualitative multiple outcomes games with consensus. *Third World Congress of the Game Theory Society (GAME'08).* 2008.

58. Sébastien Konieczny, Ramón Pino Pérez. Improvement operators. *Dagsthul Seminar #9351 : Information processing, rational belief change and social interaction.* 2009

Conférences Nationales

59. Sébastien Konieczny. Vers un modèle de la coopération. *6eme colloque de l'Association pour la Recherche Cognitive (ARC'96).* 227-231. 1996.

60. Sébastien Konieczny, Ramón Pino Pérez. Fusion avec contraintes d'intégrité. *Journées Nationales sur les Modèles de Raisonnement (JNMR'99).* 1999.

61. Salem Benferhat, Sébastien Konieczny, Odile Papini, Ramón Pino Pérez. Révision itérée basée sur la primauté forte des observations. *Journées Nationales sur les Modèles de Raisonnement (JNMR'99).* 1999.

62. Sébastien Konieczny. Sur la logique de la concertation entre agents. *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'01).* 289-302. 2001.

63. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Fusion de bases de croyances à partir de distances : modèle général et complexité algorithmique. *Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'02).* 695-704. 2002.

64. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. De la manipulabilité des opérateurs de fusion de croyances. *Journées Nationales sur les Modèles de Raisonnement (JNMR'03).* 107-121. 2003.

65. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Opérateurs de fusion à quota. *Journées Nationales sur les Modèles de Raisonnement (JNMR'03).* 91-105. 2003.

66. Sébastien Konieczny, Jérôme Lang, Pierre Marquis. Raisonnement en présence d'incohérence : le connecteur oublié. *Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'04).* 1091-1099. 2004.

67. Didier Dubois, Sébastien Konieczny, Henri Prade. Logique quasi-possibiliste et mesures de conflit. *Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'04).* 1081-1090. 2004.

68. Sébastien Konieczny. SBMG : conciliation et mesures de conflit. *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'07).* 189-200. 2007.

69. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. Planification multi-agent et diagnostique stratégique. *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'07).* 25-36. 2007.

70. Ramzi Ben Larbi, Sébastien Konieczny, Pierre Marquis. Jeux qualitatifs à résultats multiples. *Journées Francophones des Systèmes Multi-Agents (JFSMA'08)*. 65-74. 2008.

71. Patricia Everaere, Sébastien Konieczny, Pierre Marquis. Un point de vue épistémique sur la fusion de croyances : l'identification du monde réel. *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'09)*. 135-145. 2009.

72. Sébastien Konieczny, Ramón Pino Pérez. Opérateurs de confluence. *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'09)*. 193-202. 2009.

73. Sébastien Konieczny, Ramón Pino Pérez. Opérateurs d'amélioration. *Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'10)*. 50-57. 2010.

74. Sébastien Konieczny. Sur les fonctions ordinales conditionnelles transfinies. *Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'10)*. 302-309. 2010.

Conférences Invitées

75. Sébastien Konieczny. Around propositional base merging. 8th Augustus De Morgan Workshop : Belief revision, belief merging and social choice (ADMW'06). King's College, London, UK. 8-10 Novembre 2006.

76. Sébastien Konieczny. Iterated belief revision and improvement operators. Preference Change Workshop, London School of Economics (LSE). London, UK. 28-30 mai 2009.

Théses et Mémoires

77. Sébastien Konieczny. Sur la logique du changement : révision et fusion de bases de connaissance. *Thèse de Doctorat*. Laboratoire d'Informatique Fondamentale de Lille. 1999.

INDEX

RÉSUMÉ

Ce travail se situe dans le domaine des logiques pour l'intelligence artificielle, et concerne plus précisément la modélisation et la représentation des connaissances et des raisonnements (KR).

Nous nous intéressons au problème du raisonnement en présence d'incohérence. Selon les informations supplémentaires dont on dispose, cela conduit à des cadres différents, nécessitant des méthodes de raisonnement spécifiques :

- Incohérence : on ne dispose d'aucune information supplémentaire. Il est donc nécessaire d'obtenir des conclusions cohérentes (raisonnables) à partir d'un ensemble d'information incohérent.
- Révision : une des formules est plus importante que les autres. Il faut donc garder cette formule, tout en éliminant les incohérences.
- Fusion : les formules proviennent de sources différentes. Il faut alors définir une base cohérente à partir de ces informations, en prenant en compte la localisation de ces informations (avec des arguments majoritaires par exemple).
- Négociation : les formules proviennent de sources différentes, comme pour la fusion, mais les sources gardent la maîtrise des modifications de leurs formules. Il faut donc tenir compte de possibles interactions (coalitions, etc.).

Notre démarche est principalement axiomatique, c'est-à-dire que nous tentons de trouver des caratérisations logiques pour les différentes méthodes de raisonnement. Nous présentons nos travaux concernant ces quatre cadres, en particulier les caractérisations obtenues pour les mesures d'incohérence, la révision itérée, et la fusion ; ainsi que les premiers résultats concernant la négociation.

Raisonnement & Incohérence

Habilitation à diriger des recherches

Sébastien Konieczny

Composition du jury

Gerhard Brewka (rapporteur)
Professeur à l’Université de Leipzig
Marie-Christine Rousset (rapporteur)
Professeur à l’Université de Grenoble
Torsten Schaub (rapporteur)
Professeur à l’Université de Potsdam
Salem Benferhat
Professeur à l’Université d’Artois
Andreas Herzig
Directeur de Recherche au CNRS - IRIT
David Makinson
Professeur à la London School of Economics
Pierre Marquis
Professeur à l’Université d’Artois