De la manipulabilité des opérateurs de fusion de croyances

Patricia Everaere[1] Sébastien Konieczny[2] Pierre Marquis[1]

[1] Centre de Recherche en Informatique de Lens
Université d'Artois - 62300 Lens
{everaere,marquis}@cril.univ-artois.fr

[2] Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier - 31062 Toulouse
konieczny@irit.fr

Résumé *Les opérateurs de fusion de croyances ont pour but de déterminer les croyances d'un groupe d'agents à partir des croyances individuelles des agents. Dans de nombreuses situations, certains agents ont des préférences sur le résultat de la fusion. Il peut être alors tentant pour eux d'essayer de manipuler l'opération de fusion en mentant sur leurs véritables croyances afin de mieux convaincre les autres agents participant au processus. Cette possibilité de manipulation est nuisible lorsqu'un comportement coopératif des agents est attendu. La résistance à la manipulation est de fait une propriété importante pour les opérateurs de fusion. Nous montrons dans cet article que la grande majorité des opérateurs de fusion de croyances propositionnelles existants sont manipulables et présentons plusieurs restrictions sous lesquelles la manipulation n'est pas possible.*

1 Introduction

Les opérateurs de fusion de croyances ont pour but de déterminer les croyances d'un groupe d'agents à partir des croyances individuelles. La fusion de croyances propositionnelles met en jeu trois notions clés : contrainte d'intégrité, base de croyances et ensemble de croyances. Une contrainte d'intégrité représente une information qui contraint le résultat de la fusion (i.e. un ensemble de contraintes physiques, de normes, etc). Une base de croyances représente les croyances d'un agent et un ensemble de croyances regroupe les bases de croyances du groupe d'agents considéré. Formellement, ces notions sont représentées comme suit dans notre cadre :

– une *contrainte d'intégrité* μ est une formule propositionnelle cohérente ;

- une *base de croyances* K est un ensemble fini et cohérent de formules propositionnelles, interprété conjonctivement ;
- soient $K_1, \ldots, K_n$, n bases de croyances (non nécessairement distinctes), on appelle *ensemble de croyances* le multi-ensemble E constitué de ces n bases de croyances : $E = \{K_1, \ldots, K_n\}$.

Le résultat de la fusion de E sous la contrainte μ, noté $\triangle_\mu(E)$, est représenté par une base de croyances, ou indifféremment par une formule ou un ensemble d'interprétations lorsque cela est possible sans perte de généralité.

Quoique croyance et but soient des notions dissemblables, les opérateurs de fusion de croyances propositionnelles peuvent également être utilisés pour fusionner des buts, sans aucun changement du point de vue technique. Dans tous les cas, les agents fournissent une information interprétée comme un ensemble de mondes propositionnels et le résultat de la fusion est aussi de ce type. Dans le cas de la fusion de croyances, les mondes d'un agent (ou du groupe) sont ceux que l'agent considère comme possibles (i.e. l'agent suppose que le "monde réel" est parmi eux), alors que dans le cas de la fusion de buts, les mondes d'un agent (ou du groupe) représentent ses désirs.

Qu'il s'agisse de croyances ou de buts, nombreuses sont les situations où certains agents ont des préférences sur le résultat de la fusion. Dans le cas de la fusion de buts, cela est immédiat, puisqu'un agent est certainement satisfait si ses buts individuels font partie des buts du groupe. Dans le cas de fusion de croyances, un agent peut avoir intérêt à imposer ses croyances au groupe, i.e. convaincre les membres du groupe. Ainsi, dès qu'un agent impliqué dans une opération de fusion de croyances[1] a des préférences sur l'issue de la fusion, se pose le *problème de la manipulabilité* : il peut être tentant pour l'agent d'améliorer de son point de vue le résultat de la fusion en mentant sur ses croyances ou ses buts véritables.

Prenons, pour illustrer notre propos, l'exemple suivant (de type fusion de buts).

Exemple 1 *Marie, Alain et Pierre sont collocataires et ont décidé de passer la soirée ensemble. Marie ne veut pas aller au restaurant. Pierre veut aller au restaurant, mais pas au cinéma, et Alain ne veut pas rester à la maison. En utilisant un opérateur de fusion de buts usuel[2], on obtiendra comme résultat de la fusion que les trois amis doivent aller au restaurant et pas au cinéma, car cette situation maximise la satisfaction générale. Toutefois, si Marie affirmait qu'elle veut aller au cinéma et pas au restaurant, le même opérateur de fusion donnerait un résultat différent. En effet, dans ce cas, le résultat de la fusion indiquerait que les collocataires doivent sortir, sans préciser où, et donc Marie aurait encore toutes les chances de ne pas aller au restaurant. On comprend bien alors l'intérêt évident qu'elle a à mentir sur ses préférences pour passer une bonne soirée.*

[1]Dans la suite de l'article et afin d'alléger le style, on écrira *fusion de croyances* pour dénoter indifféremment fusion de croyances proprement dite ou fusion de buts.

[2]Formellement, un opérateur à sélection de modèles basé sur la distance de Dalal et utilisant la fonction d'agrégation Σ, défini paragraphe 3.1.

Dans le cadre de la théorie du choix social [1], qui s'intéresse à la définition d'une préférence sociale à partir de préférences individuelles, le problème se pose de manière similaire et a reçu une large attention (la manipulation des méthodes de vote étant bien sûr le cas particulier le plus connu). Un des plus célèbres résultats de la théorie du choix social est que toute méthode de vote est manipulable. Ainsi, lorsque trois candidats au moins sont en lice, toute méthode de vote - associant un candidat à un profil (i.e. un classement des candidats pour chaque votant) - qui est surjective (i.e. permet a priori l'élection de tout candidat) et non dictatoriale (i.e. ne conduit pas à élire systématiquement le candidat préféré d'un votant fixé) est manipulable : en mentant sur ses préférences véritables et en connaissant les préférences des autres agents, le votant manipulateur peut conduire à faire élire un candidat meilleur pour lui (i.e. mieux classé selon ses préférences). Ce résultat est connu sous le nom de théorème de Gibbard-Satterthwaite [7, 19, 16].

L'objet de notre étude est de déterminer si les opérateurs d'agrégation que sont les opérateurs de fusion de croyances propositionnelles sont manipulables ou pas. Les résultats que nous présentons dans cet article sont principalement des résultats négatifs : la quasi-totalité des opérateurs de fusion de croyances proposés dans la littérature sont manipulables dans le cas général. Notre objectif est de déterminer la frontière entre manipulabilité et non-manipulabilité en imposant des conditions supplémentaires sur la fusion ou la possibilité de manipulation.

La suite de cet article s'organise comme suit. Au paragraphe 2, nous présentons quelques définitions de base. Au paragraphe 3, nous rappelons les définitions des principaux opérateurs de fusion de croyances existants dans la littérature. Après avoir défini ce que nous entendons par manipulabilité (paragraphe 4), nous présentons au paragraphe 5 les résultats de manipulabilité/ non-manipulabilité obtenus. Nous concluons au pararaphe 6 en présentant une synthèse des résultats obtenus. Par manque d'espace, les preuves sont omises dans la suite. Le lecteur intéressé peut se référer à [6].

2 Définitions de base

On considère un langage propositionnel $\mathcal{L}$ sur un alphabet fini $\mathcal{P}$ de variables propositionnelles. Une interprétation est une application de $\mathcal{P}$ vers $\{0, 1\}$, représentée par un vecteur de bits une fois un ordre total sur $\mathcal{P}$ fixé. L'ensemble de toutes les interprétations est noté $\mathcal{W}$. Une interprétation ω est un modèle d'une formule $\phi \in \mathcal{L}$ si et seulement si elle la rend vraie au sens usuel. $[\phi]$ dénote l'ensemble des modèles de la formule ϕ, i.e. $[\phi] = \{\omega \in \mathcal{W} \mid \omega \models \phi\}$. ϕ est cohérente si et seulement si elle possède au moins un modèle. Deux formules sont logiquement équivalentes ($\equiv$) si et seulement si elles possèdent les mêmes modèles. $\top$ désigne la constante booléenne toujours vraie et $\bot$ la constante booléenne toujours fausse.

On note $\bigwedge E$ la conjonction des bases de croyances de E, c'est-à-dire $\bigwedge E = K_1 \wedge \ldots \wedge K_n$. On dit que l'ensemble de croyances E est cohérent, si $\bigwedge E$

est cohérent. L'union sur les multi-ensembles est notée $\sqcup$. Le cardinal d'un ensemble ou d'un multi-ensemble E est noté $\#(E)$.

On dit qu'une base de croyances est complète si son ensemble de modèles est un singleton. On note K_ω la base de croyances complète (à l'équivalence logique près) dont le seul modèle est ω.

Un pré-ordre $\leq$ sur un ensemble A est une relation binaire sur A qui est réflexive et transitive. Etant donné un pré-ordre $\leq$, on notera $<$ le pré-ordre strict associé défini par : $\omega < \omega'$ si et seulement si $\omega \leq \omega'$ et $\omega' \not\leq \omega$. Etant donné un ensemble A muni d'un pré-ordre $\leq$, on peut définir l'ensemble des éléments minimaux A par $min(A, \leq) = \{a \in A \mid \nexists b, b < a\}$.

3 Opérateurs de fusion

Nous rappelons dans cette partie les définitions des deux principales familles d'opérateurs de fusion de la littérature. La première famille comprend des opérateurs définis sémantiquement, par sélection de certaines interprétations, usuellement à l'aide d'une notion de distance. La seconde famille est définie syntaxiquement, par sélection de certaines formules présentes dans les croyances des différents agents. Pour plus de détails sur ces opérateurs, voir par exemple [9, 10].

3.1 Opérateurs à sélection de modèles

Cette première famille d'opérateurs est basée sur la sélection de certaines interprétations, considérées comme les plus "proches" de l'ensemble de bases de croyances. Cette notion de proximité est usuellement capturée par une notion de distance. Ces travaux sont reliés aux caractérisations axiomatiques des opérateurs de fusion [18, 11, 12, 15, 14].

Définition 1 *Soit $\leq_E$ une relation*[3] *sur $\mathcal{W}$ induite à partir d'un ensemble de croyances E. Le résultat $\triangle_\mu(E)$de la fusion de E par sélection de modèles étant donnée la contrainte d'intégrité μ est défini par :*

$$[\triangle_\mu(E)] = \min([\mu], \leq_E).$$

Cette définition indique que l'on choisit comme modèles de la fusion les modèles de μ minimaux pour $\leq_E$. La relation de proximité $\leq_E$ est usuellement induite à partir d'une notion de distance :

$$\omega \leq_E \omega' \text{ si et seulement si } d(\omega, E) \leq d(\omega', E).$$

Pour définir la distance entre une interprétation et un ensemble de croyances, on commence par définir une distance entre interprétations. Intuitivement, une telle distance $d(\omega, \omega')$ indique à quel point un monde possible (une interprétation) ω' est considéré crédible lorsque l'on se trouve dans le monde ω.

[3] Intuitivement, $\omega \leq_E \omega'$ signifie que ω est au moins aussi proche de l'ensemble E que ω'.

Définition 2 *Une* pseudo-distance *entre interprétations est une application* $d : \mathcal{W} \times \mathcal{W} \mapsto \mathbb{R}^+$ *telle que pour tout* $\omega,\ \omega' \in \mathcal{W}$, *on a :*

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

Une distance entre interprétations est une pseudo-distance qui vérifie l'inégalité triangulaire : pour tout $\omega,\ \omega',\ \omega'' \in \mathcal{W}$, *on a :*

- $d(\omega, \omega') \leq d(\omega, \omega'') + d(\omega'', \omega')$.

Deux distances entre interprétations couramment utilisées sont la distance de Dalal [5], notée d_H, qui est la distance de Hamming entre les interprétations, soit le nombre de variables propositionnelles sur lesquelles les deux interprétations considérées diffèrent ; et la distance drastique, notée d_D, qui est la pseudo-distance la plus simple qu'on peut définir entre deux interprétations. Elle vaut 0 si les deux interprétations sont égales, et 1 sinon.

Toute pseudo-distance entre interprétations induit une notion de distance entre une interprétation et une base de croyances K :

$$d(\omega, K) = min_{\omega' \models K} d(\omega, \omega').$$

Enfin, pour définir la distance d'une interprétation ω à un ensemble de croyances $E = \{K_1, \ldots, K_n\}$, on utilise une fonction d'agrégation f, qui nous permet de combiner les distances de ω à chaque K_i :

$$d(\omega, E) = f_{\{K \in E\}}(d(\omega, K)).$$

La fonction f doit satisfaire certaines propriétés pour être considérée comme une fonction d'agrégation acceptable et garantir de bonnes propriétés à l'opérateur ainsi défini [10]. Les fonctions couramment utilisées sont le max [18, 13], la somme Σ [18, 15, 12], ou le leximax $GMax$ [12, 13].

Dans la suite, on note $\Delta_\mu^{d,f}$ les opérateurs de fusion obtenus à partir d'une distance ou d'une pseudo-distance d entre interprétations et d'une fonction d'agrégation f, et Δ_μ^f si la pseudo-distance n'est pas précisée (i.e. elle peut être quelconque).

Considérons, pour illustrer le fonctionnement des opérateurs de fusion à sélection de modèles, l'exemple suivant, qui reprend l'exemple donné en introduction.

Exemple 2 *On considère un alphabet* $\mathcal{P}$ *à deux symboles c(inéma) et r(estaurant), pris dans cet ordre. Les croyances des trois agents considérés sont alors représentées par les trois bases :* K_1 *d'ensemble de modèles* $\{00, 10\}$ *(les choix de Marie),* K_2 *d'ensemble de modèles* $\{01, 10, 11\}$ *(les choix d'Alain) et* K_3 *d'ensemble de modèles* $\{01\}$ *(les choix de Pierre). Il n'y a pas de contraintes* $(\mu = \top)$*.*

Le tableau 1 donne les résultats obtenus en utilisant la distance de Hamming et la fonction d'agrégation Σ *(les modèles de la fusion sont en gras). Selon le résultat de la fusion, les trois amis doivent aller au cinéma et pas au restaurant.*

ω	$d_H(\omega, K_1)$	$d_H(\omega, K_2)$	$d_H(\omega, K_3)$	$\Delta_\mu^{d_H,\Sigma}(\{K_1, K_2, K_3\})$
00	0	1	1	2
01	**1**	**0**	**0**	**1**
10	0	0	2	2
11	1	0	1	2

TAB. 1: Fusion avec $\Delta_\mu^{d_H,\Sigma}$.

3.2 Opérateurs à sélection de formules

L'autre famille très représentative d'opérateurs de fusion de la littérature est composé de ce que l'on nomme souvent les "opérateurs syntaxiques", car les formules utilisées pour représenter les croyances influent sur le résultat. Avec les opérateurs du paragraphe précédent, on peut considérer les bases K conjonctivement : remplacer chaque $K = \{\varphi_1, \ldots, \varphi_n\}$ par $K = \{\varphi_1 \wedge \ldots \wedge \varphi_n\}$ ne change pas le résultat de la fusion. En revanche, avec ces opérateurs à sélection de formules, des bases de croyances logiquement équivalentes peuvent conduire à des résultats différents une fois fusionnées. Ces opérateurs sont basés sur la sélection de sous-ensembles cohérents de formules parmi l'union des bases de croyances de l'ensemble E. Le critère de sélection retenu diffère selon les opérateurs qui s'appuient sur le principe implicatif sceptique : seules les conséquences logiques de *tous* les sous-ensembles sélectionnés sont considérées comme des conséquences de la fusion. Voir par exemple [4, 17, 8] pour plus de détails.

Définition 3 *$MAXCONS(K, \mu)$ est l'ensemble de tous les sous-ensembles maximaux (au sens de l'inclusion ensembliste) cohérents de $K \cup \{\mu\}$ qui contiennent μ, i.e. $MAXCONS(K, \mu)$ est l'ensemble de tous les M qui vérifient :*
- *$M \subseteq K \cup \{\mu\}$,*
- *$\mu \in M$,*
- *Si $M \subset M' \subseteq K \cup \{\mu\}$, alors M' n'est pas cohérent.*

Lorsque la maximalité est définie au sens de la cardinalité, on utilise la notation $\textsc{maxcons}_{card}(K, \mu)$.

Soit $MAXCONS(E, \mu) = MAXCONS(\bigcup_{K_i \in E} K_i, \mu)$. Les opérateurs suivants ont été définis [2, 3, 8] :

Définition 4 *Soient un ensemble de croyances E et une contrainte d'intégrité μ :*
$\triangle^{C1}{}_\mu(E) = \bigvee\{M \in \textsc{maxcons}(E, \mu)\}$.
$\triangle^{C3}{}_\mu(E) = \bigvee\{M \mid M \in \textsc{maxcons}(E, \top) \text{ et } M \cup \{\mu\} \text{ cohérent}\}$.
$\triangle^{C4}{}_\mu(E) = \bigvee\{M \in \textsc{maxcons}_{card}(E, \mu)\}$.
$\triangle^{C5}{}_\mu(E) = \bigvee\{M \cup \{\mu\} \mid M \in \textsc{maxcons}(E, \top) \text{ et } M \cup \{\mu\} \text{ cohérent}\}$
si cet ensemble est non vide et μ sinon.

Reprenons l'exemple des collocataires.

Exemple 3 *Les choix de Marie, Alain et Pierre peuvent être représentés respectivement par les bases* $\{\neg r\}$, $\{c \vee r\}$ *et* $\{\neg c \wedge r\}$. *Les sous-ensembles maximaux (pour l'inclusion ensembliste) cohérents de l'union de ces bases sont les suivants (pour une contrainte* $\mu = \top$*) :* $\{\neg r, c \vee r, \top\}$ *et* $\{c \vee r, \neg c \wedge r, \top\}$. *Pour cet exemple, on obtient* ${\triangle^{C1}}_{\mu}(E) = {\triangle^{C3}}_{\mu}(E) = {\triangle^{C4}}_{\mu}(E) = {\triangle^{C5}}_{\mu}(E) \equiv (c \wedge \neg r) \vee (\neg c \wedge r)$ *(ce qui conduit à conclure que les trois compères décident d'aller au cinéma ou au restaurant, mais pas au deux).*

Dans la suite, nous noterons $\triangle^{C}$ lorsque nous souhaiterons présenter des propriétés communes à ces quatre opérateurs. Clairement, l'union multi-ensembliste pourrait être utilisée au lieu de l'union ensembliste des bases ; les opérateurs de fusion fondés sur l'inclusion (multi-ensembliste) que l'on pourrait ainsi obtenir coïncideraient avec $\triangle^{C1}$, $\triangle^{C3}$ et $\triangle^{C5}$ mais celui défini en s'appuyant sur la maximalité au sens de la cardinalité différerait de $\triangle^{C4}$. D'autres opérateurs à sélection de formules existent, voir par exemple [8] pour plus de détails.

Enfin, notons que des variantes des opérateurs $\triangle^{C}$ peuvent être obtenues en remplaçant d'abord chaque base de croyances par le singleton contenant la conjonction de ses formules avant d'effectuer la fusion. Comme la virgule - en tant que connecteur non vérifonctionnel - n'équivaut en général pas à la conjonction logique dans le cadre des “opérateurs syntaxiques”, ces variantes fournissent souvent des résultats différents, comparés aux opérateurs primitifs $\triangle^{C}$ dont ils sont issus. Par exemple, avec $\mu = \top, K_1 = \{a \wedge b\}, K_2 = \{\neg(a \wedge b)\}$ et $K'_1 = \{a, b\}$, le fait que $K'_1 \equiv K_1$ n'entraîne pas que ${\triangle^{C1}}_{\mu}(\{K_1, K_2\}) \equiv {\triangle^{C1}}_{\mu}(\{K'_1, K_2\})$, puisque ${\triangle^{C1}}_{\mu}(\{K_1, K_2\}) \equiv \top$ et ${\triangle^{C1}}_{\mu}(\{K'_1, K_2\}) \equiv a \vee b$.

4 Manipulabilité

Intuitivement, un opérateur de fusion est manipulable si l'on peut répondre positivement à la question suivante : est-il possible qu'un agent, en supposant qu'il connaît l'opérateur de fusion utilisé et les croyances $K_1, K_2, ..., K_n$ des autres agents, modifie ce qu'il déclare être ses croyances K de manière à améliorer le résultat de son point de vue ? Ainsi, on peut dire qu'il y a manipulabilité d'un opérateur de fusion si on peut trouver un ensemble de croyances $E = \{K_1, K_2, ..., K_n\}$, une contrainte d'intégrité μ, deux bases de croyances K et K' tels que le résultat de la fusion de E et K' est meilleur pour l'agent dont les croyances sont K que le résultat de la fusion de E avec K.

Que veut dire “meilleur” ici ? Pour le définir formellement, nous utilisons une notion d'indice de satisfaction.

Définition 5 (indice de satisfaction)
Un indice de satisfaction i *est une fonction calculable de* $\mathcal{L} \times \mathcal{L}$ *dans* $\mathbb{R}$.

Nous pouvons à présent définir formellement la notion d'opérateur manipulable.

Définition 6 (opérateur manipulable)
Un opérateur de fusion contrainte Δ *est* manipulable pour un indice de satisfaction i *si et seulement si il existe une contrainte d'intégrité* μ*, un ensemble de croyances* $E = \{K_1, K_2, ...K_n\}$*, une base de croyances* K *et une base de croyances* K' *tels que*

$$i(K, \Delta_\mu(E \sqcup \{K'\})) > i(K, \Delta_\mu(E \sqcup \{K\})).$$

On peut également considérer des restrictions de cette notion générale. Deux cas particuliers de manipulabilité se présentent lorsque l'on autorise l'agent soit à simplement "oublier" certains de ses modèles, c'est-à-dire à remplacer K par une base K' logiquement plus forte : $K' \models K$ (on parle de *manipulation par érosion*), soit à ajouter des interprétations à l'ensemble de ses modèles, c'est-à-dire remplacer K par une base K' logiquement plus faible : $K \models K'$ (on parle de *manipulation par dilatation*).

Par abus de langage, on utilisera aussi la notion de *"base manipulable"* : on dira qu'une base de croyances K est manipulable pour i étant donnés Δ, μ et E si et seulement si il existe une base K' telle que

$$i(K, \Delta_\mu(E \sqcup \{K'\})) > i(K, \Delta_\mu(E \sqcup \{K\})).$$

Clairement, on peut définir de nombreuses façons différentes la satisfaction d'un agent étant donné une formule représentant le résultat d'une fusion. Dans le cadre d'une étude de la manipulabilité, il est nécessaire d'utiliser les indices les plus "objectifs" possibles, c'est-à-dire ceux nécessitant peu d'hypothèses sur les préférences des agents. Il n'y a d'après nous que trois indices de satisfaction satisfaisant cette condition : les deux premiers sont purement logiques, le dernier est probabiliste.

Définition 7 (indice drastique faible)

$$i_{d_f}(K, K_\Delta) = \begin{cases} 1 & \textit{si } K \wedge K_\Delta \textit{ est cohérent,} \\ 0 & \textit{sinon.} \end{cases}$$

Cet indice vaut 1 si le résultat de la fusion (noté K_Δ) est cohérent avec les croyances de l'agent (K), et 0 sinon.

Le second est l'indice drastique fort i_{d_F} :

Définition 8 (indice drastique fort)

$$i_{d_F}(K, K_\Delta) = \begin{cases} 1 & \textit{si } K_\Delta \models K, \\ 0 & \textit{sinon.} \end{cases}$$

Cet indice vaut 1 si la base de croyances de l'agent est une conséquence du résultat de la fusion, et 0 sinon. Dans ce cas (i.e. si un opérateur est manipulable au sens de l'indice drastique fort), cela signifie que l'agent peut faire accepter l'intégralité de ses croyances par le groupe.

Le dernier indice est probabiliste. Il repose sur l'idée que le résultat de la fusion sera utilisé ensuite pour prendre des décisions, et que plus le résultat de la fusion comprend de modèles de l'agent, plus il est probable que les décisions ultérieures soient conformes aux vœux de celui-ci.

Définition 9 (indice probabiliste)
Soient K et K_Δ deux bases de croyances. On définit l'indice de satisfaction probabiliste $i_p(K, K_\Delta)$ comme la probabilité d'obtenir un modèle de K en faisant un tirage uniforme d'un modèle de K_Δ. On a donc :

$$i_p(K, K_\Delta) = \frac{\#([K] \cap [K_\Delta])}{\#([K_\Delta])}.$$

Dans le cas où $\#([K_\Delta]) = 0$, on pose $i_p(K, K_\Delta) = 0$.

Cet indice prend une valeur minimale quand aucun modèle de K n'appartient à la fusion K_Δ. Il est maximal si tous les modèles de K sont dans la fusion et si ce sont les seuls modèles de la fusion. Les trois indices définis ci-dessus ne sont pas indépendants. En particulier, on peut considérer l'indice probabiliste comme une généralisation des indices drastiques. Le résultat suivant illustre également le lien entre ces définitions.

Proposition 1 *Si un opérateur de fusion est manipulable pour i_{d_f} ou i_{d_F}, alors il est manipulable pour i_p.*

En revanche, la manipulabilité pour i_{d_F} et celle pour i_{d_f} sont logiquement indépendantes.

5 Résultats de manipulabilité

En toute généralité, i.e. sans hypothèse supplémentaire, les opérateurs de fusion à sélection de modèles sont manipulables pour i_{d_f}, donc pour i_p. Toutefois, le choix de certaines pseudo-distances et de certaines fonctions d'agrégation peut conduire à des situations de non-manipulabilité. Nous nous sommes efforcés de préciser les frontières entre manipulabilité et non-manipulabilité, en fontion des restrictions imposées à la pseudo-distance, la fonction d'agrégation ou le processus même de manipulation.

Considérons à nouveau l'exemple introductif :

Exemple 4 *On considère les trois bases K_1 d'ensemble de modèles $\{00, 10\}$ (choix de Marie), K_2 d'ensemble de modèles $\{01, 10, 11\}$ (choix d'Alain) et K_3 d'ensemble de modèles $\{01\}$ (choix de Pierre) et une contrainte $\mu = \top$. Alors $\Delta_\mu^{d_H,\Sigma}(\{K_1, K_2, K_3\})$ a pour ensemble de modèles $\{01\}$ et $i_{d_f}(K_1, \Delta_\mu^{d_H,\Sigma}(\{K_1, K_2, K_3\})) = 0$, ce qui indique que Marie n'est pas du tout satisfaite. Si, en revanche, Marie déclare K_1' d'ensemble de modèles $\{10\}$ à la place de K_1, alors $\Delta_\mu^\Sigma(E)$ a pour ensemble de modèles $\{01, 10, 11\}$ et $i_{d_f}(K_1, \Delta_\mu^{d_H,\Sigma}(\{K_1', K_2, K_3\})) = 1$, ce qui est plus satisfaisant pour elle. Le tableau 2 détaille les calculs effectués sur cet exemple.*

ω	K_1	K'_1	K_2	K_3	$\Delta_{\mu}^{d_H,\Sigma}(\{K_1, K_2, K_3\})$	$\Delta_{\mu}^{d_H,\Sigma}(\{K'_1, K_2, K_3\})$
00	0	1	1	1	2	3
01	**1**	**2**	**0**	**0**	**1**	**2**
10	0	**0**	**0**	**2**	2	**2**
11	1	**1**	**0**	**1**	2	**2**

TAB. 2: Manipulabilité de $\Delta^{d_H,\Sigma}$.

La famille des opérateurs de fusion contrainte étendant trivialement celle des opérateurs de fusion (sans contrainte), la première remarque que l'on peut faire est que tous les exemples de manipulation obtenus pour des opérateurs de fusion sans contrainte d'intégrité (i.e., avec $\mu = \top$) restent des exemples de manipulation des opérateurs de fusion contrainte.

Notons que l'influence des contraintes d'intégrité sur la manipulabilité est bien réelle. D'une part, la présence de contraintes non triviales ($\mu \neq \top$) peut conduire à rendre manipulable un opérateur qui ne le serait pas sinon. D'autre part, elle peut aussi conduire à rendre la manipulation impossible (il suffit en effet de choisir une contrainte μ incohérente avec la base manipulable K). Cependant, ce choix ne peut se faire qu' *a posteriori*.

Le premier résultat montre la non-manipulabilité par dilatation des opérateurs à sélection de modèles définis à partir d'une pseudo-distance :

Proposition 2 *La manipulabilité par dilatation est impossible pour tout opérateur de fusion contrainte $\Delta_{\mu}^{d,f}$ (où d est une pseudo-distance quelconque et f une fonction d'agrégation quelconque), pour les indices de satisfaction i_p, i_{d_f} et i_{d_F}.*

En revanche, si on ne force pas l'utilisation de la dilatation, ces opérateurs sont typiquement manipulables. Il existe même des situations où la manipulation est possible par érosion seule et en l'absence de contrainte (comme l'exemple introductif l'illustre). Donnons un premier résultat utile pour la suite.

Proposition 3 *Pour toute pseudo-distance d et toute fonction d'agrégation f, si l'opérateur de fusion contrainte $\Delta_{\mu}^{d,f}$ est manipulable pour l'indice de satisfaction drastique faible i_{d_f} ou l'indice de satisfaction drastique fort i_{d_F}, alors cet opérateur est manipulable avec K' — les croyances que K déclare à la place de ses croyances réelles, afin d'augmenter son indice de satisfaction — complète.*

Cette propriété permet de simplifier le test de non-manipulabilité, via le corollaire suivant :

Corollaire 4 *K n'est pas manipulable pour i_{d_f} ou i_{d_F} étant donnés $\Delta_{\mu}^{d,f}$ et E si et seulement si $\forall K_\omega$, $i(K, \Delta_{\mu}^{d,f}(E \sqcup \{K_\omega\})) \leq i(K, \Delta_{\mu}^{d,f}(E \sqcup \{K\}))$, où i est l'indice de satisfaction utilisé (i_{d_f} ou i_{d_F}).*

Des opérateurs à sélection de modèles définis à partir d'une pseudo-distance qui sont non manipulables dans le cas général sont les opérateurs définis à partir de la distance drastique :

Proposition 5 *Les opérateurs de fusion contrainte $\Delta_\mu^{d_D,f}$ définis à partir de la distance drastique ne sont pas manipulables pour les indices de satisfaction i_p, i_{d_f} et i_{d_F}, quelle que soit la fonction d'agrégation f utilisée.*

Si deux agents seulement sont pris en compte, on obtient des résultats différents selon que l'on utilise ou pas une contrainte d'intégrité. En effet, la manipulation de $\Delta_\mu^{d_H,\Sigma}$ est possible, alors qu'on peut montrer la non-manipulabilité de $\Delta_\top^{d,\Sigma}$, pour toute distance d :

Proposition 6 *L'opérateur de fusion contrainte $\Delta_\mu^{d_H,\Sigma}$ est manipulable pour les indices de satisfaction i_{d_f}, i_{d_F} et i_p à partir de deux bases.*

En revanche, dans le cas où on considère l'opérateur de fusion $\Delta^{d,\Sigma}$ où d est une distance quelconque, sans introduction de contrainte d'intégrité (i.e. avec une contrainte égale à $\top$), on obtient un résultat différent pour deux bases, puisqu'on a non-manipulabilité pour i_{d_f} et i_{d_F} :

Proposition 7 *La fusion de deux bases avec $\Delta_\top^{d,\Sigma}$ n'est pas manipulable pour i_{d_f} et i_{d_F}, quelle que soit la distance d utilisée.*

Ce résultat n'est vrai que pour deux agents, puisque l'on a :

Proposition 8 *$\Delta_\mu^{d_H,\Sigma}$ est manipulable pour i_{d_f} et i_{d_F} si le nombre de bases de croyances est au moins égal à trois, même si $\mu = \top$.*

On a vu que la manipulation par dilatation est impossible pour tout opérateur de fusion contrainte à sélection de modèles. Pour la manipulation par érosion, le résultat analogue ne tient pas. Par ailleurs, pour certains de ces opérateurs comme $\Delta_\mu^{d_H,G_{max}}$, il est facile de trouver des exemples où la manipulation est possible, sans qu'une manipulation par érosion le soit. En revanche, la situation est différente pour l'opérateur $\Delta_\mu^{d,\Sigma}$: la propriété suivante montre que si la manipulation est possible pour cet opérateur pour l'indice drastique faible, la manipulation par érosion l'est aussi.

Proposition 9 *Si l'opérateur de fusion contrainte $\Delta_\mu^{d,\Sigma}$ est manipulable pour i_{d_f}, alors il est manipulable par érosion, quelle que soit la distance d considérée.*

Ce résultat montre que tester si une base K est manipulable pour i_{d_f} peut se faire en prenant en compte seulement les bases complètes correspondant aux modèles de K.

Corollaire 10 *K n'est pas manipulable pour i_{d_f} étant donnés $\Delta_\mu^{d,\Sigma}$ (où d est une distance quelconque) et E si et seulement si $K \wedge \Delta_\mu^{d,\Sigma}(E \sqcup \{K\})$ est cohérent ou $\forall \omega \models K, K \wedge \Delta_\mu^{d,\Sigma}(E \sqcup \{\omega\})$ est incohérent.*

Le second résultat que l'on peut déduire de la propriété 9 est un résultat de non-manipulabilité :

Corollaire 11 *Si la base de croyances initiale K est complète, alors K n'est pas manipulable pour l'indice drastique faible i_{d_f} et l'opérateur de fusion contrainte $\Delta_\mu^{d,\Sigma}$ (quelle que soit la distance d utilisée).*

En ce qui concerne l'opérateur d'agrégation $GMax$, les résultats sont encore "pires" :

Proposition 12 *$\Delta_\mu^{d_H,GMax}$ est manipulable pour les trois indices de satisfaction i_{d_f}, i_{d_F} et i_p, même si la base de croyances initiale K de l'agent manipulateur est complète, même si on ne considère que deux agents, et même si la contrainte $\mu = \top$.*

Les résultats de manipulabilité en ce qui concerne les opérateurs à sélection de modèles ne sont donc pas très encourageants. Voyons à présent ce que donnent les opérateurs à sélection de formules.

Nous nous posons la question de la manipulabilité de ces opérateurs dans le cas général et dans un cas particulier, lorsque toutes les bases de croyances sont des singletons, c'est-à-dire composées d'une seule formule (ou encore en considérant comme base le singleton formé par la conjonction des formules de la base initiale).

Pour l'indice probabiliste, les opérateurs $\triangle^C$ sont manipulables :

Proposition 13 *Les opérateurs $\triangle_\mu^C$ sont manipulables pour i_p, même si l'on ne considère que deux agents.*

En revanche, en ce qui concerne l'indice drastique faible :

Proposition 14

- *L'opérateur $\triangle_\mu^{C1}$ n'est pas manipulable pour i_{d_f}.*
- *Les opérateurs $\triangle_\mu^{C3}$ et $\triangle_\mu^{C5}$ sont manipulables pour i_{d_f}, mais ne sont pas manipulables lorsque $IC = \top$.*
- *L'opérateur $\triangle_\mu^{C4}$ est manipulable pour i_{d_f}, mais n'est pas manipulable lorsque les K_i sont des singletons.*

Pour l'indice drastique fort, on a un résulat similaire au précédent :

Proposition 15

- *L'opérateur $\triangle_\mu^{C1}$ n'est pas manipulable pour i_{d_F}.*
- *Les opérateurs $\triangle_\mu^{C3}$ et $\triangle_\mu^{C5}$ sont manipulables pour i_{d_F}, mais ne sont pas manipulables lorsque $IC = \top$.*
- *L'opérateur $\triangle_\mu^{C4}$ est manipulable pour i_{d_F} dans le cas général, mais n'est pas manipulable lorsque les K_i sont des singletons.*

6 Conclusion

Nous nous sommes intéressés à la manipulabilité des opérateurs de fusion de croyances usuels. Nous avons obtenu des résultats généraux sur les opérateurs de fusion contrainte à sélection de modèles. En particulier, nous avons montré que la manipulation par dilatation est impossible pour ces opérateurs. Nous avons également prouvé que si l'opérateur de fusion contrainte est basé sur la distance drastique, il est non manipulable quelle que soit la fonction d'agrégation considérée.

	$\triangle_\mu^{d,f}$	$\triangle_\mu^{d_D,f}$	$\triangle^{d,\Sigma}$	$\triangle_\mu^{d_H,\Sigma}$	$\triangle_\mu^{d_H,GMax}$	$\triangle^{C1}$	$\triangle^{C3}$	$\triangle^{C4}$	$\triangle^{C5}$
i_p	$\overline{sp}$	**sp**	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$
i_{d_f}	$\overline{sp}$	**sp**	**c**	$\overline{sp}$	$\overline{sp}$	**sp**	**uc**	**s**	**uc**
i_{d_F}	$\overline{sp}$	**sp**	$\overline{sp}$	$\overline{sp}$	$\overline{sp}$	**sp**	**uc**	**s**	**uc**

TAB. 3: Synthèse des résultats obtenus.

Le tableau 3 résume certains des résultats de ce papier. **sp** (pour *strategy-proof*) signifie que l'opérateur est non manipulable pour l'indice correspondant, **c** signifie que la non-manipulabilité est obtenue lorsque les bases considérées sont complètes, **s** signifie que la non-manipulabilité est obtenue lorsque les bases sont des singletons, et **uc** (pour *unconstrained*) signifie que la non-manipulabilité est obtenue lorsqu'il n'y a pas de contraintes d'intégrité (i.e. $IC = \top$). Finalement, les cas de manipulabilité sont notés $\overline{sp}$.

Remerciements

Un grand merci aux relecteurs pour leurs commentaires et leurs suggestions. Patricia Everaere et Pierre Marquis remercient l'IUT de Lens, la Région Nord/Pas-de-Calais et les Communautés Européennes pour leur support.

Références

[1] K.J. Arrow, A. K. Sen, and K. Suzumura, editors. *Handbook of social choice and Welfare*, volume 1. North-Holland, 2002.

[2] C. Baral, S. Kraus, and J. Minker. Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 3(2) :208–220, 1991.

[3] C. Baral, S. Kraus, J. Minker, and V. S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1) :45–71, 1992.

[4] J. Minker C. Baral, S. Kraus and V.S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8 :45–71, 1992.

[5] 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)*, pages 475–479, 1988.

[6] P. Everaere. Manipulabilité des opérateurs de fusion de croyances. Mémoire de DEA. Centre de Recherche en Informatique de Lens - Université d'Artois. `http ://www.irit.fr/recherches/RPDMP/persos/Konieczny/manipulation.html`, 2003.

[7] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41 :587–602, 1973.

[8] 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)*, pages 135–144, 2000.

[9] S. Konieczny, J. Lang, and P. Marquis. Distance-based merging : a general framework and some complexity results. In *Proceedings of the eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 97–108, 2002.

[10] S. Konieczny, J. Lang, and P. Marquis. DA^2 merging operators. Artificial Intelligence. To appear, 2004.

[11] S. Konieczny and R. Pino Pérez. On the logic of merging. In *Proceedings of the sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 488–498, 1998.

[12] S. Konieczny and 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, pages 233–244, 1999.

[13] S. Konieczny and 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)*, pages 109–118, 2002.

[14] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering*, 10(1) :76–90, 1998.

[15] J. Lin and A. O. Mendelzon. Knowledge base merging by majority. In *Dynamic Worlds : From the Frame Problem to Knowledge Management*. Kluwer, 1999.

[16] H. Moulin. *Axioms of cooperative decision making*, chapter 9. Econometric society monographs. Cambridge University Press, 1988.

[17] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1 :179–219, 1970.

[18] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2) :133–160, 1997.

[19] M.A. Satterthwaite. Strategy-proofness and arrow's conditions. *Journal of Economic Theory*, 10 :187–217, 1975.