Un point de vue épistémique sur la fusion de croyances : l'identification du monde réel

Patricia Everaere* patricia.everaere@univ-lille1.fr
Sébastien Konieczny° konieczny@cril.fr
Pierre Marquis° marquis@cril.fr

*Laboratoire d'Informatique Fondamentale de Lille
Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq Cedex – France

°Centre de Recherche en Informatique de Lens
Université d'Artois
62307 Lens Cedex – France

Résumé :
La fusion de croyances est souvent décrite comme un processus permettant de définir une base de croyances qui représente au mieux les croyances d'un groupe d'agents (un profil de bases de croyances). La base résultant de ce processus peut être considérée comme une synthèse du profil donné. Dans cet article, nous proposons un autre point de vue, de type épistémique, sur la fusion de croyances. Sous ce point de vue, le but de la fusion de croyances est d'approcher au mieux le monde réel. Nous étudions des généralisations du Théorème du Jury de Condorcet du point de vue de la fusion de croyances. Nous montrons que si l'on fusionne les croyances d'un nombre suffisamment grand d'agents fiables, il est possible d'assurer (en probabilité) l'identification du monde réel. Nous montrons que quelques (mais pas tous les) opérateurs de fusion de la littérature ont un bon comportement pour ce problème de suppression de l'incertitude.

Mots-clés : Fusion de croyances, théorème du jury de Condorcet

Abstract:
Belief merging is often described as the process of defining a belief base which best represents the beliefs of a group of agents (a profile of belief bases). The resulting belief base can be viewed as a synthesis of the input profile. In this paper another view, called the epistemic view, of what belief merging aims at is considered. Under this view the purpose of belief merging is to best approximate what is the true state of the world. We study generalizations of Condorcet's Jury Theorem from the belief merging perspective. Roughly we show that if we merge the beliefs of sufficiently many reliable agents then we can ensure to identify the true state of the world. We show that some (but not all) merging operators from the literature are suited to the truth tracking issue.

Keywords: Belief Merging, Condorcet's Jury Theorem

1 Introduction

De nombreux problèmes issus de domaines variés de l'informatique, comme par exemple les bases de données distribuées et les systèmes multi-agents, nécessitent de synthétiser des informations cohérentes provenant de plusieurs sources d'information. Réaliser une telle synthèse est en général difficile car, en particulier, les informations provenant de différentes sources peuvent se contredire les unes les autres. Lorsque les informations disponibles sont des croyances représentées par des formules de la logique propositionnelle, ce problème est appelé fusion de croyances (propositionnelles). Dans cet article, nous nous intéressons au cas purement logique, i.e., nous supposons que les bases de croyances sont des ensembles de formules propositionnelles (voir par exemple [2, 20, 12, 11, 9]). De nombreux opérateurs de fusion ont été introduits dans un tel cadre. Les propriétés logiques de ces opérateurs de fusion ont été largement étudiées et sont décrites dans de nombreux travaux [20, 14, 12]. Ainsi, [12] présente un ensemble de postulats permettant de caractériser la famille des opérateurs de fusion contrainte. D'autres opérateurs ont été introduits dans des cadres plus généraux, comme par exemple en logiques pondérées (logique possibiliste ou basée sur des fonctions ordinales conditionnelles) [3, 16, 4] ; ils se révèlent très utiles lorsque des informations supplémentaires sont disponibles (en particulier, lorsque toutes les croyances ne sont pas également sûres). Dans ces cadres plus généraux, le problème de la fusion devient proche de l'agrégation de préordres (préférences), très étudiée en théorie du choix social [1, 19].

Jusqu'à présent, les opérateurs de fusion de croyances existants sont aussi considérés comme adaptés à la fusion de buts. Cela a du sens, puisque dans les deux cas la fusion cherche à synthétiser les informations représentées dans le profil initial de bases propositionnelles. 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 [13] 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.

Dans cet article, nous introduisons un nouveau point de vue sur la fusion de croyances, qui est différent du point de vue synthétique habituel et qui permet de justifier une telle distinction entre fusion de croyances et fusion de buts : le point de vue épistémique.

Vue synthétique : Sous le point de vue synthétique, comme évoqué auparavant, 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.

Dans le cas général, aucun agent n'a une connaissance parfaite du monde réel, les croyances sont entachées d'incertitude :

1. Typiquement, un agent ne sait pas quel modèle de sa base correspond au monde réel,
2. Il n'est même pas certain que le monde réel se trouve réellement parmi les modèles de sa base.[1]

La fusion de croyances sous le point de vue épistémique peut être considérée comme un moyen de lever l'incertitude sur le monde réel au niveau du groupe.

Il est remarquable que la recherche du monde réel soit un moyen de différencier la fusion de croyances de la fusion de buts. En effet, alors que la recherche du monde réel peut être attendue lors de la fusion de croyances dans de nombreuses circonstances, 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.

L'intuition derrière la vue épistémique de la fusion de croyances est que si les agents sont indépendants et dignes de confiance, alors on peut espérer d'un opérateur de fusion qu'il soit capable d'identifier en probabilité le monde réel, en s'appuyant sur un nombre suffisamment grand d'agents.

Le problème de l'identification du monde réel (*Truth Tracking*) a été étudié depuis des siècles en choix social et en sciences politiques, afin de justifier les fondations des élections démocratiques ou les décisions prises par des jurys. Le principal résultat théorique est le théorème du jury de Condorcet [7]. Ce théorème établit que si un jury est composé d'individus indépendants et dignes de confiance, et qu'ils ont à trouver la véritable réponse oui/non à une question, alors la probabilité que la décision prise par le jury à la majorité soit la bonne tend vers 1 quand la taille du jury tend vers l'infini.

Dans ce travail, nous formalisons le problème de l'identification du monde réel dans une perspective de fusion de croyances. Nous montrons que certains opérateurs de fusion de croyances peuvent être utilisés pour identifier l'état du monde en considérant suffisamment d'agents indépendants et fiables. Plus précisément, la contribution de ce papier est principalement comme suit : nous présentons une généralisation du théorème du jury de Condorcet au cas incertain (i.e., lorsque chaque base peut contenir plusieurs modèles). Nous introduisons un postulat d'identifiabilité à la limite du monde réel, noté **TT**. Nous présentons certains opérateurs de fusion contrainte satisfaisant ce postulat et montrons que d'autres opérateurs ne le satisfont pas. En conséquence, nous concluons que certains, mais pas tous les opérateurs de fusion de la littérature sont adaptés à l'identification du monde réel. Nous présentons également des résultats expérimentaux sur la vitesse de convergence vers le monde réel lorsque l'opérateur de fusion considéré est $\Delta^{d_D,\Sigma}$. Dans le plupart des cas, le nombre d'agents nécessaires pour garantir que la base fusionnée identifie le monde réel avec une probabilité élevée n'est pas très grand. Cela montre la faisabilité pratique de la recherche du monde réel en utilisant $\Delta^{d_D,\Sigma}$.

Le reste de l'article est organisé ainsi : nous donnons quelques préliminaires formels en section 2. Nous rappelons le théorème du jury de Condorcet, et certaines de ses généralisations en section 3. En section 4, nous présentons de nouvelles généralisations du théorème du jury de Condorcet en présence d'incertitude, utiles pour la fusion de croyances. En section 5, nous montrons que certains opérateurs de fusion contrainte de la littérature (mais pas tous) satis-

[1] Si l'on suppose que les agents savent que le monde réel est un modèle de leur base, on peut parler de "connaissances"– cette hypothèse est la seule différence entre croyances et connaissances. Or, la *fusion de connaissances* n'est pas intéressante, puisque la seule méthode raisonnable pour fusionner des connaissances est évidemment la conjonction.

font **TT**. En Section 6, nous présentons les résultats empiriques que nous avons obtenus et les analysons. Finalement nous concluons en section 7.

2 Préliminaires

Nous considérons un langage propositionnel $\mathcal{L}$ défini à partir d'un ensemble fini de variables propositionnelles $\mathcal{P}$ et des connecteurs usuels.

Pour tout sous-ensemble c de $\mathcal{P}$, $|c|$ dénote le nombre d'éléments de c. Une interprétation (ou état du monde) est une application de $\mathcal{P}$ dans $\{0, 1\}$. L'ensemble de toutes les interprétations est notée $\mathcal{W}$. Le véritable état du monde est notée $\omega^\star$. Une interprétation ω est un modèle d'une formule $\phi \in \mathcal{L}$ si et seulement si elle la rend vraie au sens usuel. $[\phi]$ représente l'ensemble de modèles de la formule ϕ, i.e., $[\phi] = \{\omega \in \mathcal{W} \mid \omega \models \phi\}$.

Une *base* K dénote l'ensemble des croyances d'un agent, c'est un ensemble fini de formules propositionnelles, interprété conjonctivement. Nous identifions K avec la conjonction de ses éléments. Toute base K représente un ensemble $[K]$ d'états du monde.

Un *profil* E représente les croyances d'un groupe de n agents impliqués dans le processus de fusion. Dans cet article, les agents expriment parfois un seul monde possible. Dans ce cas un profil est un vecteur de bases complètes. Pour éviter des notations trop lourdes, nous identifions chaque base complète avec son modèle et écrivons ces profils $E_c = \langle \omega_1, \ldots, \omega_n \rangle$. Partout ailleurs les agents expriment des ensembles de mondes possibles, ainsi E est représenté comme un vecteur de bases $E = \langle K_1, \ldots, K_n \rangle$, comme c'est souvent l'usage en fusion propositionnelle.

3 Theorème du jury de Condorcet et extensions

Nous considérons un profil de n agents où chaque agent i vote pour une alternative, par exemple un état du monde ω_i. Parmi tous les mondes possibles se trouve le véritable monde $\omega^\star$. Les hypothèses du théorème du jury de Condorcet sont que les agents sont à la fois indépendants et fiables. Les agents sont *indépendants* si connaître le monde réel et le monde donné par n'importe quel agent ne donne aucune information supplémentaire sur le monde donné par tout autre agent (cela signifie que les choix des agents sont indépendants conditionnellement au monde réel, au sens Bayésien standard [17]). Comme de nombreuses notions de fiabilité vont être considérées dans la suite, nous appelons la première *R1-fiabilité* :

– La *R1-fiabilité* p_i d'un agent i est la probabilité que i donne le monde réel, i.e., $p_i = P(\omega_i = \omega^\star)$. **(R1)**
– Un agent i est *R1-fiable* si sa R1-fiabilité est strictement plus élevée que 0.5.

La règle majoritaire donne simplement comme résultat l'interprétation qui reçoit une majorité stricte de votes. Formellement, nous définissons une notion de score d'un monde par rapport à un profil de bases complètes :

$$s(\omega) = |\{\omega_i \in E_c \mid \omega_i = \omega\}|.$$

Définition 1 (Majorité) *Etant donné un profil E_c de n bases complètes, la règle de la* majorité m *est définie par :*

$$m(E_c) = \omega \text{ si } s(\omega) > n/2.$$

Rappelons à présent le théorème du jury de Condorcet. Dans ce théorème, deux alternatives sont considérées seulement et chaque agent vote pour une des deux (pas d'abstention possible) :

Théorème 1 ([7]) *Considérons deux mondes possibles $\mathcal{W} = \{\omega, \omega^\star\}$ et un profil E_c de bases complètes issues d'un ensemble de n agents indépendants et R1-fiables où tous les agents partagent la même R1-fiabilité p. La probabilité que la règle de la majorité sur ce profil retourne le monde réel $\omega^\star$ tend vers 1 lorsque n tend vers l'infini, i.e. :*

$$P(m(E_c) = \omega^\star) \xrightarrow[n\to\infty]{} 1.$$

Ce théorème est une conséquence de la loi faible des grands nombres. Schématiquement, il établit que si les individus d'un jury sont suffisamment fiables (i.e., que leurs décisions sont meilleures que celles obtenues par tirage au sort à pile ou face) et indépendants, alors la fiabilité du jury augmente avec sa taille et converge vers 1.

Il est assez clair que les deux hypothèses de ce théorème sont assez fortes. D'abord, il est inhabituel que les agents d'un jury soient totalement indépendants : ils ont souvent des connaissances partagées, une culture similaire, sont attentifs

aux mêmes leaders d'opinion, etc. De plus, en général, tous les agents n'ont pas exactement la même fiabilité : il y a souvent des agents plus compétents que d'autres. Des extensions du theorème du jury de Condorcet montrent qu'il est possible de relâcher certaines des hypothèses du théorème sans changer la conclusion. Ainsi, la conclusion du théorème est toujours vraie lorsque les opinions des individus ne sont pas totalement indépendantes [8]. Concernant la fiabilité, il est suffisant de supposer que la fiabilité moyenne des individus est supérieure à 0.5 [10].

Une autre limitation importante du théorème du jury de Condorcet est qu'il ne considère que deux alternatives seulement. Un résultat récent [15] permet d'étendre ce théorème à un nombre fini quelconque d'options. Afin de présenter ce résultat, nous devons d'abord rappeler la définition de la règle de la pluralité :

Définition 2 (Pluralité) *Soit un profil E_c de bases complètes, la règle de la* pluralité *pl est définie par :*

$$pl(E_c) = \{\omega \mid \forall \omega' \in \mathcal{W} \;\; s(\omega) \geq s(\omega')\}.$$

L'hypothèse de fiabilité **(R1)** doit être étendue si l'on considère plus de deux alternatives. List et Goodin [15] ont défini la notion suivante de fiabilité. Soit un ensemble $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ de k mondes possibles :

– Un agent est *R2-fiable* si la probabilité qu'il vote pour $\omega^\star$ est strictement supérieure à la probabilité qu'il vote pour n'importe quel autre monde.[2] **(R2)**

List et Goodin ont montré que :

Théorème 2 ([15]) *Soit $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ un ensemble de k mondes possibles et soit un profil E_c de bases complètes issues d'un ensemble de n agents indépendants et R2-fiables. La probabilité que la règle de la pluralité sur ce profil retourne le monde réel $\omega^\star$ tend vers 1 lorsque n tend vers l'infini, i.e.,*

$$P(pl(E_c) = \{\omega^\star\}) \xrightarrow[n\to\infty]{} 1.$$

Ce théorème est une généralisation du théorème du jury de Condorcet. On peut noter que la règle de la pluralité est utilisée (et non la règle de la majorité) et que l'hypothèse de fiabilité demande seulement que la probabilité de voter pour le monde réel soit strictement supérieure à la probabilité de voter pour un autre monde, et donc la probabilité de voter pour le monde réel peut être inférieure à 0.5.[3] Voir [15] pour une discussion de ce résultat et ses conséquences philosophiques, et pour plus de détails sur le théorème du jury de Condorcet.

4 Un théorème du jury avec incertitude

Dans tous les travaux précédents autour du théorème du jury de Condorcet, les agents choisissent toujours exactement une alternative unique. Cette hypothèse n'est pas adaptée au cadre de la fusion de croyances, dans lequel chaque agent présente typiquement une base de croyances qui peut contenir de nombreux modèles (et imposer aux agents de donner des bases de croyances complètes serait très restrictif, puisque cela nierait la nature incertaine des croyances des agents). Ainsi, à partir de maintenant, nous supposons que chaque agent i a une base de croyances K_i qui peut avoir plusieurs modèles parmi un ensemble fini $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$.

Le théorème du jury de Condorcet peut être étendu au cas où chaque agent peut voter pour plusieurs alternatives. Tout d'abord, nous avons à définir une notion de fiabilité adaptée à cette situation :

– La *R3-fiabilité* p_i d'un agent i est la probabilité que le monde réel $\omega^\star$ fasse partie des modèles de sa base de croyances K_i, i.e., $p_i = P(\omega^\star \models K_i)$. **(R3)**
– Un agent i est *R3-fiable* si $p_i > 0.5$.

De même, la notion de score d'une interprétation doit être étendue de la façon suivante :

$$s_a(\omega) = |\{K_i \in E \mid \omega \models K_i\}|.$$

Nous avons établi le résultat suivant :

Proposition 1 *Soit $p^\star \in [0, 1[$. On considère un profil E issu d'un ensemble de n agents indépendants ayant tous la même R3-fiabilité $p >$*

[2] Formellement, soit $p_1, \ldots, p_{k-1}, p^\star$ (respectivement) la probabilité qu'un agent vote pour $\omega_1, \ldots, \omega_{k-1}, \omega^\star$ (respectivement). Alors $\forall i \in 1 \ldots k-1 \; p^\star > p_i$.

[3] Clairement, si l'on considère seulement deux états du monde possibles, les hypothèses du théorème de List-Goodin sont équivalentes à celles du théorème du jury de Condorcet, et les deux théorèmes sont identiques dans ce cas.

$p^\star$. *La probabilité que le score du monde réel* $\omega^\star$ *soit supériuer à* $p^\star n$ *tend vers* 1 *lorsque* n *tend vers l'infini, i.e,*

$$P(s_a(\omega^\star) > p^\star n) \xrightarrow[n\to\infty]{} 1.$$

Preuve: Pour un agent i, la probabilité que $\omega^\star \models K_i$ est p ; donc la probabilité que $\omega^\star \not\models K_i$ est $q = 1 - p$. En supposant que les agents sont indépendants, la probabilité que $\omega^\star$ soit un modèle de k parmi n bases vaut $\binom{n}{k} p^k q^{n-k}$. Dans la suite de la preuve, nous utilisons le théorème de Bienaimé-Tchebitchev : *soit* v *une variable aléatoire de moyenne* m *et de variance* σ^2. *Etant donné* t, *un réel positif, la probabilité que la variable* v *s'écarte de* m *d'au moins* $t\sigma$ *est inférieure à* $\frac{1}{t^2}$ *:*

$$P(\mid v - m \mid \geq t\sigma] < \frac{1}{t^2} \qquad (1)$$

Soit v_i la variable aléatoire égale à 1 si $\omega^\star \models K_i$, et à 0 sinon. Comme chaque v_i suit une distribution de Bernoulli avec une probabilité de succès p, la moyenne et la variance de v_i sont égales à p et pq, respectivement. Alors le nombre de bases du profil E parmi les n ayant $\omega^\star$ comme modèle est égale à $v = \Sigma_{i=1}^{n} v_i$. Comme v suit une distribution binomiale de paramètres n, p, la moyenne de v est égale à $m = np$, et sa variance est égale à $\sigma^2 = npq$.

La probabilité que le nombre de bases ayant $\omega^\star$ comme modèle soit inférieur à $np^\star$ est donnée par $P(v \leq p^\star n)$.

Or, $\mid v - m \mid \geq m - p^\star n$ si et seulement si $v - m \geq m - p^\star n \geq 0$ ou (exclusif) $v - m \leq p^\star n - m < 0$. Donc $P(\mid v - m \mid \geq m - p^\star n) = P(v - m \geq m - p^\star n \geq 0) + P(v - m \leq p^\star n - m < 0)$.

Donc on a $P(\mid v - m \mid \geq m - p^\star n) \geq P(v - m \leq p^\star n - m < 0)$. Comme enfin $v - m \leq p^\star n - m < 0$ si et seulement si $v \leq p^\star n < m$, on obtient :

$$P(v \leq p^\star n < m) \leq P(\mid v - m \mid \geq m - p^\star n)$$

En prenant $t = \frac{m - p^\star n}{\sigma}$ ($t > 0$ si $p > p^\star$), et en utilisant l'inéquation (1), on obtient :

$$P(\mid v - m \mid \geq m - p^\star n) < \frac{\sigma^2}{(m - p^\star n)^2}.$$

En remplaçant m et σ par leur valeurs, comme on a toujours $np^\star < np$ lorsque $p^\star < p$, par transitivité, il s'ensuit que :

$$P(v \leq np^\star) < \frac{pq}{n(p - p^\star)^2}.$$

Ce qui conduit à :

$$P(v \leq p^\star n) \xrightarrow[n\to\infty]{} 0.$$

Autrement dit :

$$P(s_a(\omega^\star) > p^\star n) \xrightarrow[n\to\infty]{} 1.$$

□

Ce résultat donne (à la limite et en probabilité) un minorant du score obtenu par le monde réel à partir du moment où les agents ont la même R3-fiabilité. Il assure (toujours à la limite et en probabilité) pour certaines règles de vote que le monde réel fait partie de l'ensemble des mondes retournés par la règle. Considérons par exemple les règles de vote suivantes :

Définition 3 (M **et** Q_p**)** *Soit* E *un profil issu d'un ensemble de* n *agents.*
- *La règle de la* majorité M *est définie par :* $M(E) = \{\omega \mid s_a(\omega) > n/2\}$.
- *Plus généralement, étant donné* $k \in]0, 1[$, *la règle du* k-quota Q_k *est définie par :* $Q_k(E) = \{\omega \mid s_a(\omega) > kn\}$.

La règle de la majorité M étend au cas où les agents expriment des croyances incertaines la règle de majorité précédente m.

On obtient le corollaire suivant à la proposition précédente :

Corollaire 1 *Soit* E *un profil issu d'un ensemble de* n *agents indépendants.*
- *Si tous les agents sont R3-fiables et ont la même R3-fiabilité* p, *alors le monde réel est capturé à la limite par la règle de la majorité, i.e.,*

$$P(\omega^\star \in M(E)) \xrightarrow[n\to\infty]{} 1.$$

- *Si tous les agents ont la même R3-fiabilité* $p > k$, *alors le monde réel est capturé à la limite par la règle du* k*-quota, i.e.,*

$$P(\omega^\star \in Q_k(E)) \xrightarrow[n\to\infty]{} 1.$$

On peut souligner que cette proposition ne porte que sur la présence du monde réel dans le résultat du processus de vote. En aucun cas elle n'exclut qu'un autre monde soit présent dans ce résultat. Bien sûr, cela pose un problème pour la détermination du monde réel. En particulier, si chacun des agents i donne tous les mondes possibles ($[K_i] = \mathcal{W}$), alors toute fusion raisonnable du profil E correspondant (en particulier toute fusion réalisée via un opérateur vérifiant le postulat **(IC2)**, [12]) donnera comme résultat tous les mondes possibles (par exemple $Q_k(E) = \mathcal{W}$ quel que soit k), ce qui ne donne aucune information sur le monde réel.

Le problème est dû à la notion de R3-fiabilité, qui n'est pas assez forte pour l'identification du monde réel. Intuitivement, demander aux agents de choisir le monde réel avec une probabilité élevée est nécessaire mais pas suffisant puisque cela n'empêche pas les agents de donner (comme modèles de leurs bases) d'autres mondes possibles. Par exemple, un agent i dont la base est toujours une tautologie ($[K_i] = \mathcal{W}$), qui ne donne aucune information, est considéré complètement R3-fiable (i.e., sa R3-fiabilité p_i est égale à 1), ce qui est inadapté. Ainsi, une notion de fiabilité plus forte est nécessaire. La notion de R4-fiabilité qui suit répond à cette demande :

- L'*incompétence* q_i d'un agent i est la probabilité maximale qu'un monde différent de $\omega^\star$ appartienne à l'ensemble de modèles de sa base, i.e., $q_i = max_{\omega_j \neq \omega^\star}(\{P(\omega_j \models K_i)\})$. La *compétence* d'un agent est $c_i = 1 - q_i$.
- Un agent est *compétent* si $c_i > 0.5$.
- Un agent i est *R4-fiable* si sa R3-fiabilité p_i est strictement supérieure à son incompétence q_i : $p_i > q_i$. **(R4)**

Intuitivement, alors que la R3-fiabilité exprime la faculté d'un agent à avoir des croyances compatibles avec le monde réel, la notion de compétence traite de la quantité d'incertitude sur ses croyances. Nous pensons que la R3-fiabilité et la compétence sont des notions à la fois naturelles et importantes pour caractériser l'idée intuitive d' "agent fiable" dans le cadre de la fusion de croyances. Dans le cas spécifique où $\mathcal{W}$ est réduit à deux alternatives seulement, un agent est compétent si et seulement s'il est R3-fiable ; à l'opposé, dans le cas général, ces notions sont bien distinctes. Il est facile de montrer que la notion de R4-fiabilité étend les notions précédentes de fiabilité :

Proposition 2 – *Quand on considère seulement des profils E_c de bases complètes, les fiabilités de type R4 et R2 sont équivalentes.*
– *Quand on considère seulement des profils E_c de bases complètes et un ensemble $\mathcal{W}$ d'interprétations constitué de deux éléments $\{\omega, \omega^\star\}$, les fiabilités de type R4, R3, R2 et R1 sont équivalentes.*

Les hypothèses supplémentaires suivantes sur l'ensemble d'agents considérés vont permettre de dériver des résultats intéressants :

- Un ensemble d'agents est *homogène* si pour chaque monde possible ω_j, la probabilité $P(\omega_j \models K_i)$ que ce monde soit un modèle de la base K_i fournie par chacun des agents de l'ensemble est la même pour tous les agents i de l'ensemble. En particulier le monde réel $\omega^\star$ a la même probabilité d'apparaître comme modèle chez chacun des agents.
- Un ensemble d'agents homogène est *uniforme* si pour chaque monde possible ω_j sauf le monde réel $\omega^\star$, les probabilités $P(\omega_j \models K_i)$ coincident.

Proposition 3 *Soit E un profil issu d'un ensemble uniforme de n agents indépendants. Alors la base K la plus probable (à l'équivalence logique près) dans E est $[K] = \{\omega^\star\}$.*

Preuve: Soit p la probabilité pour un agent de choisir une base K telle que le monde réel $\omega^\star$ vérifie $\omega^\star \models K$, et soit q la probabilité pour un agent de choisir une base K telle qu'un monde quelconque ω différent de $\omega^\star$ vérifie $\omega \models K$. Soit $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ l'ensemble de tous les mondes possibles ($\mathcal{W}$ contient k éléments). Grâce à l'hypothèse d'indépendance, la probabilité pour un agent de choisir une base K telle que $[K] = \{\omega^\star\}$ est égale à $p(1-q)^{k-1}$. Plus généralement, la probabilité pour un agent de choisir une base K est égale à $p^a(1-p)^b q^c (1-q)^d$ avec a, b, c, d des entiers positifs tels que $a + b = 1$ et $a + b + c + d = k$. Sous l'hypothèse de R4-fiabilité de l'ensemble d'agents, on a $p > 0.5$ et $q < 0.5$, donc $p > (1-p)$ et $q < (1-q)$. En conséquence, $p^a(1-p)^b q^c (1-q)^d$ est maximal pour $a = 1, b = 0, c = 0, d = k-1$, seulement. Ainsi, la base K la plus probable (à l'équivalence logique près) dans E est $[K] = \{\omega^\star\}$. □

Ce résultat n'est néanmoins pas suffisant pour obtenir directement le théorème du jury sous incertitude qui suit :

Théorème 3 *Soit $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ un ensemble de mondes possibles et soit E un profil issu d'un ensemble homogène de n agent indépendants et R4-fiables. Alors $\forall i \in \{1, \ldots, k-1\}$,*

$$P(s_a(\omega^\star) > s_a(\omega_i)) \xrightarrow[n\to\infty]{} 1.$$

Preuve: Soit $(s_a(\omega_1), \ldots, s_a(\omega_{k-1}), s_a(\omega^*))$ un vecteur de variables aléatoires où $s_a(\omega_i) = j$ ($i \in \{1, \ldots, k-1\}$) (resp. $s_a(\omega^*) = j$) signifie que le score $s_a(\omega_i)$ (resp. $s_a(\omega^*)$) est égal à j ($j \in \{0, \ldots, n\}$). Chacune de ces variables $s_a(\omega_i)$ ($i \in \{1, \ldots, k-1\}$) (resp. $s_a(\omega^*)$) suit une distribution binomiale de paramètres q_i et n (resp. p et n). Ainsi, on a $\forall j \in \{0, \ldots, n\}$:

$$P(s_a(\omega_i) = j) = \binom{n}{j} q_i^j (1-q_i)^{n-j}$$

et

$$P(s_a(\omega^*) = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

La moyenne de chaque $s_a(\omega_i)$ ($i \in \{1, \ldots, k-1\}$) est nq_i, sa variance est $nq_i(1-q_i)$, la moyenne de $s_a(\omega^*)$ est np et sa variance est $np(1-p)$.

La loi faible des grands nombres appliquée à $s_a(\omega_i)$ ($i \in \{1, \ldots, k-1\}$) et $s_a(\omega^*)$ donne que $\forall \epsilon > 0$:

$$P(\mid \frac{s_a(\omega_i)}{n} - q_i \mid \geq \epsilon) \longrightarrow_{n\to\infty} 0, \qquad (2)$$

$$P(\mid \frac{s_a(\omega^*)}{n} - p \mid \geq \epsilon) \longrightarrow_{n\to\infty} 0. \qquad (3)$$

Soit $q = max_{i\in\{1,\ldots,k-1\}} q_i$ et $\epsilon_1 = \frac{p-q}{2}$. Comme chaque agent est $R4$-fiable, on a forcément $q_i < p$ pour chaque $i \in \{1, \ldots, k-1\}$, donc $q < p$. En conséquence, on a $\epsilon_1 > 0$. Grâce aux assertions (2) et (3), on obtient pour tout $i \in \{1, \ldots, k-1\}$,

$$P(\mid \frac{s_a(\omega_i)}{n} - q_i \mid \geq \epsilon_1) \longrightarrow_{n\to\infty} 0$$

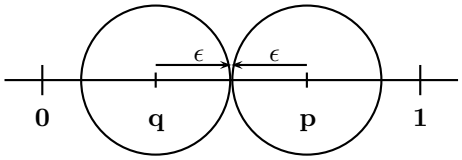
$$et\ P(\mid \frac{s_a(\omega^*)}{n} - p \mid \geq \epsilon_1) \longrightarrow_{n\to\infty} 0.$$

Ce qui permet facilement de déduire que :

$$P(\frac{s_a(\omega_i^*)}{n} > q_i + \epsilon_1) \longrightarrow_{n\to\infty} 0, \qquad (4)$$

et

$$P(\frac{s_a(\omega^*)}{n} < p - \epsilon_1) \longrightarrow_{n\to\infty} 0. \qquad (5)$$



Cette figure illustre la preuve à venir. L'idée est que lorsque la loi faible des grand nombres est utilisable, les valeurs prises se concentrent autours de la moyenne avec une probabilité importante. Schématiquement, la probabilité que toutes les valeurs soient dans une sphère centrée en la moyenne et de rayon ϵ_1 tend vers 1. Ainsi, comme $p > q$, la probabilité que les deux sphères ont une intersection non vide tend vers 0 à l'infini.

Supposons à présent que $\frac{s_a(\omega_i)}{n} \leq q_i + \epsilon_1$ et que $\frac{s_a(\omega^*)}{n} \geq p - \epsilon_1$. Alors, comme $\forall i \in 1 \ldots k-1, q_i + \epsilon_1 \leq p - \epsilon_1$, on obtient :

$$\frac{s_a(\omega_i)}{n} \leq q_i + \epsilon_1 \leq p - \epsilon_1 \leq \frac{s_a(\omega^*)}{n}.$$

Ainsi l'événement $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}$ peut se produire uniquement si $\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$, ou $\frac{s_a(\omega^*)}{n} < p - \epsilon_1$. Et donc on a :

$P(\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) = P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ et $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) + P(\frac{s_a(\omega^*)}{n} < p - \epsilon_1$ et $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) - P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ et $\frac{s_a(\omega^*)}{n} < p - \epsilon_1$ et $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n})$

Comme $P(\frac{s_a(\omega^*)}{n} < p - \epsilon_1$ et $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) \leq P(\frac{s_a(\omega^*)}{n} < p - \epsilon_1)$ et $P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ et $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) \leq P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1)$, on obtient :

$P(\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^*)}{n}) \leq$

$$P(\frac{s_a(\omega^*)}{n} < p - \epsilon_1) + P(\frac{s_a(\omega_i^*)}{n} > q_i + \epsilon_1).$$

Finalement, grâce aux assertions (4) et (5), on obtient :

$$P(\frac{s_a(\omega_i)}{n} \geq \frac{s_a(\omega^*)}{n}) \xrightarrow[n\to\infty]{} 0,$$

ou, de façon équivalente

$$P(s_a(\omega_i) \geq s_a(\omega^*)) \xrightarrow[n\to\infty]{} 0.$$

□

Ce théorème montre immédiatement l'aptitude de la méthode de vote dite par approbation [6], à identifier le monde réel (à la limite et en probabilité). Avec cette méthode, les agents sont autorisés à voter pour un nombre quelconque de mondes, et ensuite on choisit comme résultat les mondes avec les scores les plus élevés :

Définition 4 (Approbation) *Etant donné un profil de bases E, la règle de vote* par approbation av *est définie par :*

$$av(E) = \{\omega \mid \forall \omega' \in \mathcal{W} \;\; s_a(\omega) \geq s_a(\omega')\}.$$

5 Identifier le monde réel par fusion de croyances

L'aptitude d'un opérateur de fusion à identifier le monde réel peut être modélisée par un nouveau postulat, appelé postulat **T***ruth* **T***racking* :

TT Soit E un profil issu d'un ensemble homogène de n agents indépendants et R4-fiables. Soit $\omega^\star$ le monde réel.

$$P([\Delta(E)] = \{\omega^\star\}) \xrightarrow[n\to\infty]{} 1.$$

Ce postulat est vérifié par un opérateur de fusion si l'opérateur est capable d'identifier le monde réel (à la limite et en probabilité) en s'appuyant sur un ensemble homogène d'agents indépendants qui sont plus R3-fiables que incompétents. Clairement, ses conditions d'application sont fortes puisqu'il n'est pas faciles de pouvoir garantir en pratique que l'on pourra disposer des croyances d'un ensemble homogène et suffisamment arbitrairement grand d'agents indépendants et R4-fiables. Il en est d'autant plus intéressant puisqu'il assure que les opérateurs ne le vérifiant pas ne sont pas de bons candidats pour identifier le monde réel.

On peut à présent étudier le comportement par rapport à ce postulat de certains opérateurs de fusion de croyances bien connus. Nous rappelons d'abord la définition des opérateurs de fusion basés sur une distance [12].

Définition 5 (fusion à base de distance) *Soit d une pseudo-distance entre interprétations et soit f une fonction d'agrégation. L'opérateur de fusion $\Delta^{d,f}(E)$ est défini par :* $[\Delta^{d,f}_\mu(E)] = \min([\mu], \leq_E)$
où le pré-ordre $\leq_E$ sur $\mathcal{W}$ induit par E est défini par :
– $\omega \leq_E \omega'$ *si et seulement si* $d(\omega, E) \leq d(\omega', E)$, *avec*
– $d(\omega, E) = f_{K\in E}(d(\omega, K))$, *où*
– $d(\omega, K) = min_{\omega'\models K} d(\omega, \omega')$.

On notera d_D la distance drastique ($d_D(\omega, \omega') = 0$ si $\omega = \omega'$ et 1 sinon), et d_H la distance de Hamming [11].

Le résultat suivant est une conséquence directe du théorème 3 :

Proposition 4 $\Delta^{d_D,\Sigma}$ *satisfait* **TT**.

Puisque $\Delta^{d_D,\Sigma}$ est un opérateur de fusion contrainte [12], cette proposition montre qu'il n'y a pas de conflit entre **TT** et les postulats **(IC0-IC8)** qui caractérisent les opérateurs de fusion contrainte. Néanmoins, tous les opérateurs de fusion contrainte ne satisfont pas **TT**. Par exemple :

Proposition 5 $\Delta^{d_H,Gmax}$ *ne satisfait pas* **TT**.

Preuve: On considère quatre mondes possibles $\mathcal{W} = \{00, 01, 10, 11\}$, et on suppose que le monde réel est $\omega^\star = 11$. On considère la probabilité suivante p sur les mondes possibles, partagées par tous les agents (hypothèse d'homogénéité) : $p(11 \models K) = a > 0.5$, $p(00 \models K) = b$, $p(01 \models K) = c$, $p(10 \models K) = 0$, avec $a + b + c = 1$, $b > 0$, et $c > 0$. Par construction, pour chaque monde ω de $\mathcal{W}$ sauf 01, la probabilité qu'un profil E de n agents indépendants contienne une base K telle que $d_H(\omega, K) = 2$ tend vers 1 lorsque n tend vers l'infini ; par ailleurs, la probabilité qu'un profil E de n agents indépendants contienne une base K telle que $d_H(01, K) = 2$ est toujours égale à 0 (en effet on a toujours $d_H(01, K) < 2$). En conséquence, $P([\Delta^{d_H,Gmax}(E)] = \{01\}) \xrightarrow[n\to\infty]{} 1$, ce qui contredit **TT**. □

6 Résultats expérimentaux

Les résultats obtenus dans les sections précédentes au sujet de l'identification du monde réel $\omega^\star$ sont des résultats *à la limite*. Aucun d'entre eux ne donne d'information sur l'identification du monde réel d'un point de vue pratique, comme par exemple un majorant sur le nombre de bases à partir de laquelle l'identification est atteinte avec une probabilité élevée. Or, clairement, si le nombre de bases nécessaires à cette identification est très élevé, le problème de l'identification du monde réel, lorsqu'il est envisageable a priori (e.g. lorsque les agents considérés sont des experts indépendants) reste sans résolution pratique possible.

Afin d'approfondir cette question, nous avons conduit des expérimentations en utilisant $\Delta^{d_D,\Sigma}$, un opérateur de fusion contrainte qui satisfait **TT** et qui est simple à implémenter. Le protocol expérimental utilisé a été le suivant. Nous avons étudié la vitesse de convergence vers le monde réel en utilisant $\Delta^{d_D,\Sigma}$, en fonction de la R3-fiabilité des agents p et de leur incompétence q (supposée identique pour tous les agents). Nous avons considéré des ensembles d'interprétations de tailles variables (jusqu'à 2^{15}), nous avons fixé le monde réel $\omega^\star$ comme l'interprétation associant chaque variable propositionnelle à 0, et nous avons engendré des profils E à partir de n agents indépendants, avec une R3-fiabilité p et une incompétence q, pour différentes valeurs de n. Pour chaque valeur n, nous avons construit 1000 profils E. Pour chaque E, nous avons calculé $\Delta^{d_D,\Sigma}(E)$ et testé si $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$. La proportion des 1000 profils pour lesquels $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$ donne une estimation de la probabilité de succès de l'identification du monde réel.

La figure 1 donne la probabilité que $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$ étant donné le nombre n d'agents, lorsque $p = 0.3$, et $|\mathcal{W}| = 2^7$ mondes, pour différentes valeurs q, La figure 2 présente des expérimentations similaires, mais pour $p = 0.9$.

Un point intéressant est à observer sur la figure 1 : même si $p = 0.3$ est relativement faible (ici les agents considérés ne sont pas R3-fiables), la vitesse de convergence est élevée : pour obtenir $[\Delta^{d_D,\Sigma}(E)] = \{\omega^\star\}$ avec une probabilité supérieure à 90%, 1250 agents sont nécessaires pour une incompétence $q = 0.25$, 300 agents sont nécessaires pour $q = 0.2$ et seulement 60 agents

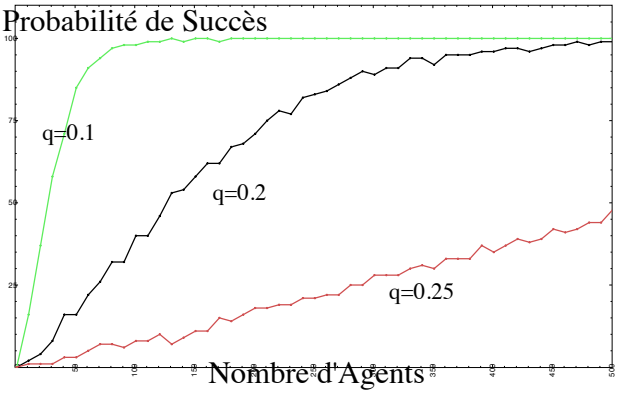


FIG. 1 – Vitesse de convergence (7 variables, p=0.3)

sont nécessaires pour $q = 0.1$.

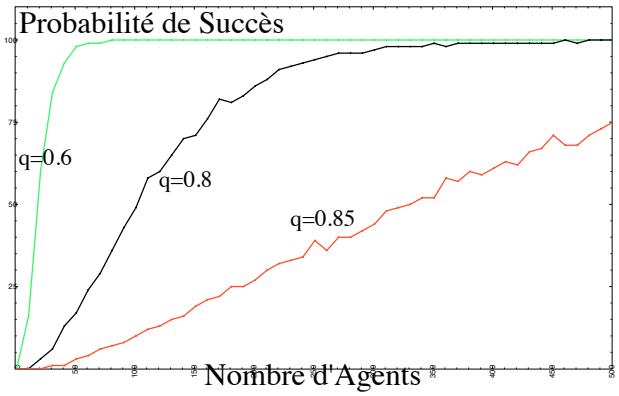


FIG. 2 – Vitesse de convergence (7 variables, p=0.9)

Ces résultats sont proches de ceux donnés à la figure 2 pour $p = 0.9$. Les valeurs correspondantes sont 800 pour $q = 0.85$, 230 pour $q = 0.8$, et 40 pour $q = 0.6$.

Empiriquement, il apparaît que le "niveau de R4-fiabilité" des agents, i.e., l'écart $p - q$ a plus d'impact sur la vitesse de convergence de l'identification du monde réel en utilisant $\Delta^{d_D,\Sigma}$ que le fait que les valeurs p et q sont plutôt proches de 1 ou de 0.

La figure 3 donne la probabilité de succès de l'identification du monde réel étant donné le nombre de variables propositionnelles (et donc le nombre d'interprétations) avec $p = 0.70$ et $q = 0.40$.

Il n'est pas étonnant que la complexité de d'identifier le monde réel augmente avec le nombre de mondes possibles. Cependant, le

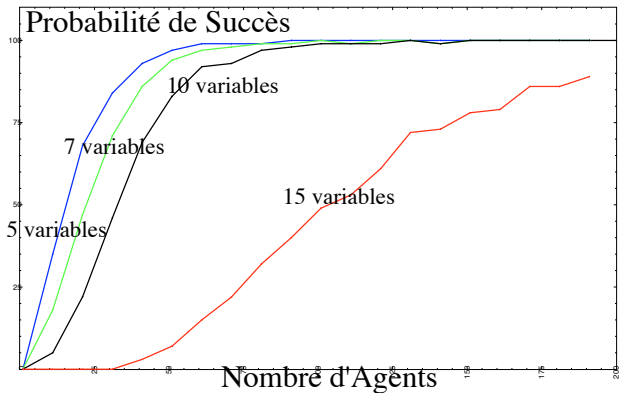


FIG. 3 – Vitesse de convergence (p=0.7, q =0.4)

nombre d'agents qui doivent être considérés pour atteindre l'objectif avec une probabilité élevée n'est pas si grand, si on le met en balance avec le nombre d'interprétations possibles. Par exemple, on peut observer sur la figure 3 que, lorsque 10 variables sont considérées, moins de 50 agents sont suffisants pour assurer que $[\Delta^{d_D,\Sigma}(E)] = \{\omega^\star\}$ avec une probabilité supérieure à 90%, en dépit du fait qu'un unique monde doit être discriminé parmi 1024 et que la R3-fiabilité et la compétence des agents ne sont pas très élevées. Tous ces résultats montrent la faisabilité pratique de la recherche de la vérité en utilisant $\Delta^{d_D,\Sigma}$.

7 Conclusion

Dans ce travail, nous avons considéré la fusion de croyances selon un point de vue original que nous avons appelé point de vue épistémique. Cette approche permet d'évaluer la capacité des opérateurs de fusion de croyances à trouver le monde réel. Nous avons démontré une généralisation du théorème du jury de Condorcet lorsque les croyances des agents sont incertaines. Nous avons également défini un postulat correspondant à l'identifiabilité à la limite du monde réel pour la fusion de croyances et montré que certains opérateurs de fusion de croyances satisfont ce postulat. Finalement, nous avons étudié la vitesse de convergence offerte par $\Delta^{d_D,\Sigma}$.

Le problème de l'identification du monde réel a aussi été étudié dans le cadre plus spécifique de l'agrégation de jugement [5]. Dans [18] les auteurs étudient les performances d'un opérateur $\Delta^{d_H,\Sigma}$ dans cet optique, et montrent que ses performances sont bonnes (typiquement meilleures que celles d'autres procédures d'agrégation de jugement) pour des agents assez fiables.

Dans nos travaux futurs, nous pensons étudier le problème de l'identification du monde réel pour d'autres opérateurs de fusion de croyances, comme ceux basés sur la distance de Hamming. Nos premières expériences suggèrent que $\Delta^{d_H,\Sigma}$ satisfait également **TT**.

Remerciements

Les auteurs remercient les relecteurs anonymes pour la qualité de leur travail. Ils ont bénéficier du support du projet ANR PHAC (ANR-05-BLAN-0384).

Références

[1] K. J. Arrow. *Social Choice and Individual Values*. Wiley, New York, second edition, 1963.

[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] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Possibilistic merging and distance-based fusion of propositional information. *Annals of Mathematics and Artificial Intelligence*, 34(1–3) :217–252, 2002.

[4] S. Benferhat, S. Lagrue, and J. Rossit. An egalitarist fusion of incommensurable ranked belief bases under constraints. In *Proceedings of AAAI'07*, pages 367–372, 2007.

[5] L. Bovens and W. Rabinowicz. Democratic answers to complex questions. *Synthese*, 150 :131–153, 2006.

[6] S. J. Brams and P. C. Fishburn. *Approval voting*. Springer Verlag, 1983.

[7] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris, 1785.

[8] D. Estlund. Opinion leaders, independence and condorcet's jury theorem. *Theory and Decision*, 36(2) :131–162, January 1994.

[9] P. Everaere, S. Konieczny, and P. Marquis. Quota and gmin merging operators. In *Proceedings of IJCAI'05*, pages 424–429, 2005.

[10] O. Guillermo, B. Grofman, and S. Feld. Proving a distribution-free generalization of the condorcet jury theorem. *Mathematical Social Sciences*, 17(1) :1–16, February 1989.

[11] S. Konieczny, J. Lang, and P. Marquis. DA^2 merging operators. *Artificial Intelligence*, 157 :49–79, 2004.

[12] S. Konieczny and R. Pino Pérez. Merging information under constraints : a logical framework. *Journal of Logic and Computation*, 12(5) :773–808, 2002.

[13] S. Konieczny and R. Pino Pérez. On the frontier between arbitration and majority. In *Proceedings of 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] C. List and R. Goodin. Epistemic democracy : Generalizing the condorcet jury theorem. *Journal of Political Philosophy*, 9(3) :277–306, 2001.

[16] T. Meyer. On the semantics of combination operations. *Journal of Applied Non-Classical Logics*, 11(1-2) :59–84, 2001.

[17] Judea Pearl. *Causality : models, reasoning, and inference*. Cambridge University Press, 2000.

[18] G. Pigozzi and S. Hartmann. Judgment aggregation and the problem of truth-tracking. In *Proceedings of TARK'07*, pages 248–252, 2007.

[19] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Aggregating partially ordered preferences : impossibility and possibility results. In *Proceedings of TARK'05*, pages 193–206, 2005.

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