THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par **l'Institut Supérieur de l'Aéronautique et de l'Espace**
Spécialité : STIC Intelligence artificielle

Présentée et soutenue par **Stéphanie ROUSSEL**
le 18 octobre 2010

Apports de la logique mathématique
pour la modélisation de l'information échangée
dans des systèmes multiagents interactifs

JURY

Mme Claudette Cayrol, présidente
M. Éloi Bossé, rapporteur
Mme Laurence Cholvy, directrice de thèse
M. Frédéric Cuppens, rapporteur
Mme Camilla Schwind
M. Pierre Siegel, rapporteur

École doctorale : **Mathématiques,, informatique et télécommunications de Toulouse**

Unité de recherche : **Équipe d'accueil ISAE-ONERA MOIS**

Directrice de thèse : **Mme Laurence Cholvy**

Déroulement de la thèse

La thèse s'est déroulée à l'ONERA sur le site de Toulouse, sous la direction de Laurence Cholvy.

Elle a été financée par la DGA (Direction Générale de l'Armement). Mon correspondant était Mr Blanc-Talon. N'ayant pu se déplacer pour la soutenance, Mme Crück l'a représenté.

L'école doctorale dont je dépendais était l'EDMITT (Ecole Doctorale Mathématiques Informatique et Télécommunications de Toulouse) et j'étais inscrite à l'ISAE (Institut Supérieur de l'Aéronautique et de l'Espace - Supaéro).

Remerciements

Dans le cadre de l'échange d'information entre différents agents humains, en l'occurence moi et les autres, il est un besoin dont je ne parlerai que dans cette partie du manuscrit : *le besoin de remercier*. Les informations utiles pour un tel besoin sont les informations i du type : « je tiens à remercier a pour φ » telles que les personnes désignées par a, après avoir pris connaissance de cette information i, sachent que leur contribution φ, quelle qu'elle soit, a été importante pour moi durant ces trois années de thèse.

Je vous propose donc ici une liste que j'espère exhaustive de toutes les informations utiles pour ce besoin particulier. Evidemment, les différents éléments de cette liste ne sauraient être ordonnés ...

Je tiens à remercier les différents membres du jury pour leurs remarques et leurs questions (même les difficiles) qui ont, j'espère, contribué à améliorer ce manuscrit et qui m'ont ouvert de nouvelles perspectives de recherche. Je les remercie également de s'être déplacés pour assister à ma soutenance sans savoir s'il leur serait possible de rentrer dans leurs régions et contrées respectives.

Un grand merci à Laurence pour ces trois années et quelques mois. Grâce à elle, j'ai découvert un cadre de recherche, et plus généralement un cadre de travail, stimulant et enthousiasmant. Merci de m'avoir guidée et écoutée dans les hauts et surtout dans les bas. Je suis sûre que cette thèse n'est que le début d'une collaboration qui sera fructueuse, d'un point de vue professionnel mais aussi personnellement.

Je tiens également à remercier Christophe pour m'avoir supportée en cours pendant deux ans, pour avoir accepté de travailler avec moi sur une partie de ma thèse, pour les mangas et les séries et tout simplement pour avoir toujours été présent pour me conseiller.

Evidemment, merci à toute les membres de l'équipe DTIM pour leur accueil chaleureux, leurs conseils et leur bonne humeur. Merci à tous les thésards passés et présents. Une pensée particulièrement affectueuse à Maria qui a partagé mon bureau pendant cette dernière année, et à Pierre et Cédric pour l'organisation de tous ces évènements réussis !

Merci à tous les copains, Toulousains d'origine, d'adoption ou ex-Toulousains pour toutes les soirées, les week-end, les apéros en terrasse, les matchs de rugby et les coups de téléphone interminables pour certains...

Ces trois années ne se seraient pas aussi bien passées sans la présence de mes deux collocs Ben et Mathieu. Malgré les petites chamailleries et quelques histoires de rideaux, cette collocation est et restera un souvenir très cher à mes yeux.

Merci à ma famille pour son soutien, même si celui-ci était souvent à distance. Je ne serai pas allée si loin sans eux.

Finalement, un dernier merci à Julien pour tout ce qu'il m'apporte en étant à mes côtés....

Table des matières

Introduction

Les *systèmes multiagents interactifs* sont des systèmes dans lesquels différentes entités (informatiques ou physiques) que l'on appelle des *agents* intéragissent. Selon [CDJM01], les systèmes multiagents possèdent les caractéristiques principales suivantes :

- chaque agent a un point de vue limité (car il a des informations ou des capacités de résolution de problèmes incomplètes) ;
- il n'y a pas forcément d'entité centrale qui assure le contrôle global ;
- les données sont distribuées ;

Ainsi, les systèmes dont les entités sont réparties géographiquement (par exemple dans le domaine du trafic aérien), ou les systèmes dont les entités sont des organisations qui ont chacune leur propre mode de fonctionnement et leurs règles (par exemple des systèmes de coordination entre différents pays), ou plus généralement les systèmes de systèmes ([LR08]) sont des exemples de systèmes multiagents interactifs.

Cette thèse s'intéresse plus particulièrement aux systèmes multiagents interactifs dans lesquels les agents doivent interagir pour réaliser une tâche commune.

Au sein d'un même système, les agents ont différents moyens d'interagir. En effet, ils peuvent communiquer verbalement, diffuser des messages à tous les agents, etc. Nous nous concentrons dans ce mémoire sur un type d'interaction : la *communication interpersonnelle*, plus précisément sur *l'échange d'information* entre un agent et un autre agent.

Dans [CdD02], les auteurs dressent une liste des différents éléments qui déterminent les actions des agents, notamment leurs actions communicationnelles. Parmi ces éléments, on trouve, entre autres, l'*attitude sociale* des agents, les *obligations, permissions et interdictions* que les agents doivent respecter, les *lois de l'environnement*, les *buts* ou *intentions* antérieures des agents, leurs *préférences*, etc. En ce qui concerne l'attitude sociale, les auteurs mettent en avant trois types d'attitude sociale :

- *obéissant*. Un agent est obéissant lorsqu'il cherche à respecter les différentes obligations, permissions et interdictions auxquels il est soumis.
- *négociateur*. Un agent est négociateur s'il cherche à obtenir des contreparties lorsque d'autres agents lui font des demandes.
- *coopératif* ou *altruiste*. Un agent est coopératif lorsque, spontanément, il va effectuer des actions qui peuvent contribuer à la résolution des intentions des autres agents.

Les attitudes sociales des agents sont en général déterminantes pour le bon fonctionnement d'un système multiagent. Par exemple, un système peut très bien être soumis à une réglementation censée garantir son fonctionnement et être complètement chaotique si les agents ne respectent pas cette réglementation, c'est-à-dire s'ils ne sont pas obéissants. Il est donc important de s'assurer que les différents agents du système ont des attitudes sociales satisfaisantes pour que le système puisse réaliser ce pour quoi il a été conçu.

Dans ce travail, nous nous intéressons tout particulièrement à deux des attitudes sociales précédentes, la **coopération** et l'**obéissance**, et en proposons des modèles formels. En effet, dans l'objectif de s'assurer qu'un système multiagent répond aux spécifications qui ont précédé sa constitution, il est nécessaire d'utiliser un modèle formel et les méthodes formelles qui vont avec, afin de décrire et analyser les systèmes, les interactions entre les agents, ... Parmi les formalismes candidats, la logique mathématique est un des formalismes les plus utilisés et certainement un des plus adaptés : « *La modélisation logique est particulièrement bien adaptée pour aborder les problèmes du futur qui concerneront de plus en plus l'organisation de la coopération entre des agents existants, et la prise en compte d'agents humains et/ou artificiels, dans une vision* *globale.* »([DdC00]).

Dans une première partie, nous nous concentrons sur la notion de coopération. Plus précisément, nous proposons une définition en logique modale du caractère coopératif d'un agent vis-à-vis d'un autre.

Le caractère obéissant d'un agent dans le cadre d'un échange d'information présuppose l'existence d'une politique d'échange d'informations au sein du système dans lequel se trouve l'agent. Pour pouvoir définir l'obéissance, la politique d'échange doit satisfaire certaines propriétés. Par exemple, elle doit être cohérente (que signifie sinon d'obéir à une politique qui stipule qu'une action et son contraire sont toutes deux obligatoires ?). Dans une deuxième partie, nous nous intéressons donc aux politiques d'échange d'informations et à deux de leurs propriétés : la cohérence et la complétude.

Nous proposons dans une troisième partie d'étudier dans quelle mesure il est possible d'être à la fois coopératif et obéissant.

Finalement, la dernière partie présente un résumé des travaux ainsi que les perspectives pour les travaux futurs.

Toutes les preuves des propriétés et théorèmes présentés dans ce mémoire sont données en annexe.

Finalement, précisons que le but des modèles que nous proposons dans ce manuscrit est de clarifier des concepts qui nous paraissent être non triviaux. Notre intention n'est pas de proposer des algorithmes qui permettent par exemple de calculer dans un temps raisonnable si un agent est coopératif ou non mais de comprendre quels sont les différents paramètres qui interviennent dans la notion de coopération. Nous préférons utiliser des logiques multimodales dont l'application algorithmique est certes peu évidente, mais qui offre une expressivité que nous jugeons indispensable à ce stade, tout en restant dans un cadre formel.

Première partie

Information utile et agent coopératif

Introduction

Nous nous intéressons dans cette première partie à la modélisation formelle du caractère *coopératif* des agents d'un système.

Selon le dictionnaire Larousse, coopérer signifie « prendre part, concourir à une œuvre commune ; contribuer, participer ». Les différents agents d'un système ayant un certain nombre de buts ou d'intentions à réaliser, être coopératif, dans un tel système, signifie donc prendre part à la réalisation du but ou de l'intention de l'autre en effectuant les actions qui contribuent à cette réalisation. Nous nous concentrons ici sur une action particulière qui est l'échange d'information. Ainsi, dans ce cadre, coopérer signifie échanger les informations qui contribuent à la réalisation du but ou de l'intention de celui à qui on les envoie. Autrement dit, coopérer signifie échanger les informations qui sont *utiles* pour ce but ou cette intention.

La notion d'utilité d'une information est donc un concept sous-jacent à la notion de coopération. Ainsi, dans cette partie, avant de proposer une définition formelle d'un agent coopératif, nous travaillons d'abord sur la notion d'utilité d'une information.

Plus précisément, cette partie est structurée de la façon suivante :

Tout d'abord, nous réalisons un état de l'art qui traite à la fois des concepts d'utilité et de coopération (chapitre 1). Cet état de l'art nous permet d'établir un certain nombre de propriétés que doivent vérifier ces deux concepts. Dans le chapitre 2, nous définissons le cadre formel, une logique multimodale propositionnelle, sur lequel nous basons nos différents modèles. Le chapitre 3 est consacré à la formalisation de la notion d'utilité d'une information, l'étude des propriétés des informations utiles ainsi que quelques extensions de cette notion. Nous proposons ensuite plusieurs définitions pour la notion d'agent coopératif et nous comparons ces définitions entre elles ainsi qu'avec des travaux de la littérature (chapitre 4). Finalement, le chapitre 5 conclut sur le travail réalisé dans cette partie, analyse ses limites et en propose des perspectives.

Chapitre 1

État de l'art

Les concepts d'*utilité d'une information* et de *coopération* sont au cœur de nombreux domaines de recherche. Notre état de l'art ne se cantonne donc pas au domaine de l'Intelligence Artificielle mais survole également des domaines tels que la pragmatique ou la philosophie. Nous ne prétendons pas faire pas un état de l'art détaillé de chacun des domaines de recherche que nous abordons mais voulons donner au lecteur une idée des différentes façons d'appréhender ces concepts.

Le problème que l'on se pose ici est de caractériser formellement, dans un système où les agents échangent des informations, quelles sont les informations utiles à échanger et quels sont les agents coopératifs dans leurs échanges d'information. Ce n'est généralement pas le cadre d'étude des différents domaines de recherche que nous présentons. Ainsi, pour chacun des domaines de recherche, nous mettons en avant les caractéristiques de l'utilité ou de la coopération utilisées qui peuvent s'appliquer à notre problématique. Plus précisément, pour chacun des domaines que nous parcourons, nous commençons par faire une présentation des principaux travaux puis nous analysons ces travaux en terme d'application à notre cadre de travail.

Nous commençons par étudier le Principe de Coopération et la Théorie de la Pertinence développés dans le domaine de la pragmatique inférentielle. Puis, nous présentons dans la deuxième section une formalisation en logique du Principe de Coopération. La troisième section est consacrée à une définition de la pertinence épistémique en philosophie. La quatrième section adresse la caractérisation des réponses coopératives dans le domaine des bases de données. Puis, en Intelligence Artificielle, nous analysons les relations entre pertinence et inférence. La sixième section porte sur les définitions et caractérisations de la pertinence dans le domaine de la Recherche d'Information. Nous clôturons cet état de l'art par une conclusion en section 1.7.

Finalement, il convient de remarquer que, dans la littérature, le terme « pertinence » est en général préféré au terme « utilité ». Cependant, l'utilisation de « pertinence » peut conduire à différents contresens, notamment dans sa version anglaise « relevance ». Ainsi, dans cet état de l'art et dans chacun des domaines que nous parcourons, nous utiliserons les termes du domaine étudié et notamment le terme « pertinence » mais dans les chapitres suivants, nous préférerons le terme « utilité ».

1.1 Pragmatique inférentielle : Principe de Coopération et Théorie de la Pertinence

1.1.1 Présentation des travaux

Selon la définition donnée par Moeschler ([Moe07], [MR06]), la pragmatique est la discipline des sciences du langage qui étudie et analyse les processus de compréhension intervenant pendant les actes de communication ou d'interaction.

Plus particulièrement, la pragmatique inférentielle étudie les inférences qui peuvent être déduites de façon non logique du discours (également appelées *implicatures*). Par exemple, si Alice pose la question « Viens-tu déjeuner avec nous ? » et que Bernard répond « Je ne suis pas prêt pour mon cours à 13h00 », Alice peut déduire que Bernard ne viendra pas déjeuner, même si la réponse à la question n'a pas été explicitement donnée.

En général, le calcul des implicatures repose sur des *hypothèses* sur le monde et sur les locuteurs. Dans cet état de l'art, on ne s'intéresse pas au calcul même de ces implicatures mais à une hypothèse utilisée pour ces calculs, hypothèse que l'on peut résumer ainsi : *la communication humaine est guidée par un ensemble de principes généraux qui traitent de la rationalité, de la coopération ou de la cognition.* Ainsi, les pragmaticiens ont étudié les règles et principes qui permettent de juger si un discours ou plus généralement un échange est coopératif (ou rationnel, ou cognitif, ...). Bien que ces règles soient le plus souvent informelles, elles permettent de définir une base des différents éléments qui doivent apparaître dans une formalisation de la coopération entre agents. Ce sont ces éléments que nous mettons en évidence.

La nature des principes qui régissent la communication diffère suivant les courants pragmatiques. On s'intéresse plus précisément aux trois courants de la pragmatique inférentielle (comme présenté dans [Moe07]) : tout d'abord Grice dont la théorie est une des théories fondatrices de la pragmatique inférentielle, puis les néo-Gricéens considérés comme héritiers de Grice et finalement, les post-Gricéens qui sont à l'origine de la théorie de la Pertinence.

Le principe de coopération de Grice

Grice ([Gri75]) suppose que la communication humaine est guidée par le *Principe de Coopération.* Les participants à un discours, s'ils sont coopératifs, doivent respecter ce principe ou alors l'exploiter. Ce principe peut s'énoncer de la façon suivante : *Fais en sorte que ta contribution soit, au moment de l'échange, conforme à la direction et au but exigés par cet échange.*
Grice décompose ce principe en quatre maximes.

La première maxime est la maxime de **Quantité** :
- *Fais en sorte que ta contribution soit aussi informative que nécessaire (pour les besoins de l'échange).*
- *Fais en sorte que ta contribution ne soit pas plus informative que nécessaire.*

Vient ensuite la maxime de **Qualité** : *Fais en sorte que ta contribution soit vraie.* Elle se décompose en deux sous-maximes :
- *Ne dis pas ce que tu sais être faux* ;
- *Ne dis pas ce pour quoi tu n'as pas de preuves.*

La troisième maxime est la maxime de **Relation** : *Sois pertinent.*

Finalement, la dernière maxime est la maxime de **Manière** : *Fais en sorte que ta contribution soit clairement exprimée* (ou encore *facile à comprendre*).
- *Évite d'être obscur* ;
- *Évite l'ambiguïté* ;
- *Sois bref* ;
- *Sois ordonné* ;

Pour être coopératif, il faut donc être aussi informatif que possible pour le besoin de l'échange mais pas trop informatif (quantité), il faut dire des informations vraies (qualité) et claires (manière) et il faut être pertinent (relation).

Sur cette dernière maxime (relation), Grice reste relativement imprécis. Dans [Gri75], il fait l'analogie entre un échange verbal et une transaction. Dans le cas de la transaction, être pertinent signifie par exemple que si Alice a besoin de sel alors lui donner le sel est pertinent alors que lui donner du sucre ne l'est pas, même si Alice aura besoin de sucre plus tard. Dans ce cas, être pertinent pourrait signifier que l'échange doit être fait au moment où il est requis. Dans [RM98], Reboul et Moeschler estiment que « pertinent » dans cette maxime doit être pris dans son sens usuel, c'est-à-dire « qui est approprié à son objet, justifié » (Larousse). Selon eux, il est difficile de discriminer cette maxime des autres maximes. Dans ce travail, nous préférons donc nous focaliser sur les autres maximes dont l'interprétation porte moins à ambiguïté.

Prenons l'exemple d'Alice qui cherche à savoir quel plat Bernard a prévu pour le dîner. Supposons que Bernard lui réponde une des trois réponses :

(1) De la viande sera servie.

(2) Du poulet sera servi.

(3) Du poulet sera servi ou alors $(7^2 - 3)$ n'est pas égal à 46.

Ici, le besoin d'Alice est de savoir quel plat est servi. Bernard peut supposer qu'elle cherche à savoir plus que « viande » ou « poisson ». Si tel est le cas, alors (2) est, selon Bernard, plus informatif que (1) pour le besoin d'Alice. (2) et (3) apportent la même quantité d'information sur le plat mais (3) est moins claire que (2) : il faut faire le calcul pour se rendre compte que (3) et (2) sont équivalents. La réponse (2) est donc celle qui répond au mieux aux maximes de quantité et de manière. Si Bernard respecte uniquement ces maximes, ce sera donc cette réponse qu'il fera à Alice.

Supposons que Bernard croit que du poulet sera servi et qu'il a des preuves de cela (par exemple, il a lui-même acheté le poulet). En accord avec les différentes maximes, il peut donc répondre (2). S'il ne croit pas que du poulet sera servi (ou qu'il n'a pas de preuve de cela) mais qu'il croit que le plat sera tout de même composé de viande, alors Bernard ne peut pas faire mieux que de répondre (1) pour être en accord avec la maxime de qualité.

Néo-Gricéens

Les maximes de Grice ont donné lieu à de nombreuses discussions. En particulier, les néo-Gricéens en proposent des simplifications. Par exemple, Horn ([Hor84]) réduit les maximes à deux principes :
- le *principe Q* (centré sur celui qui écoute) : Fais en sorte que ta contribution soit suffisante ; dis autant que possible (modulo le principe R et la maxime de Qualité).
- le *principe R* (centré sur celui qui parle) : Fais en sorte que ta contribution soit nécessaire ; ne dis pas plus que ce qui est requis (modulo le principe Q).

Bien qu'ayant des débouchés importants pour le calcul des implicatures (voir [Car98] pour l'exemple des implicatures scalaires), les simplifications des maximes de Grice n'apportent pas beaucoup quant aux principes qui régissent la communication. En effet, il y a correspondance entre les maximes de Grice et les deux principes énoncés ici. La deuxième simplification majeure est présentée dans [Lev00] et on peut également établir une correspondance entre les « nouveaux » principes et les « anciens ». Ainsi, les différents éléments proposés dans les maximes gricéennes se retrouvent chez les néo-gricéens. Nous ne nous attardons donc pas sur la théorie néo-gricéenne.

Théorie de la pertinence

La théorie de la pertinence est une théorie développée par Sperber et Wilson ([SW86] à la base, [SW04] pour une présentation plus récente). Cette théorie s'inscrit dans le courant post-Gricéen, dans lequel se reconnaissent aujourd'hui beaucoup de pragmaticiens ([Moe07]). Elle se démarque des théories gricéennes et néo-gricéennes par un certain nombre de points, dont principalement :

- le principe de coopération et les maximes associées sont regroupées dans un principe : le principe de pertinence.
- la démarche utilisée est une démarche cognitiviste (et non pas juste linguistique ou pragmatique).

Pour Sperber et Wilson, la pertinence se définit informellement de la façon suivante : une contribution est pertinente pour un individu si et seulement si elle produit des effets cognitifs positifs, c'est à dire si elle entraîne une modification valable de l'état mental de l'individu qui la reçoit. Sperber et Wilson notent à ce sujet que les contributions qui entraînent des conclusions fausses ne peuvent pas être considérées comme étant valables. Cela signifie qu'une information fausse n'est pas pertinente.

La notion de pertinence est une notion qualitative qui vérifie les deux principes suivants :

- Toutes choses étant égales par ailleurs, plus une contribution apporte d'effets cognitifs positifs à un individu, plus elle est pertinente pour lui.
- Toutes choses étant égales par ailleurs, plus une contribution demande d'efforts d'interprétation pour un individu, moins elle est pertinente pour lui.

Reprenons l'exemple donné dans [SW04] et que nous avons déjà donné précédemment. Alice veut savoir ce qui est au menu pour le dîner. Bernard lui répond :

(1) De la viande sera servie.
(2) Du poulet sera servi.
(3) Du poulet sera servi ou bien $(7^2 - 3)$ n'est pas égal à 46.

Supposons que ces énoncés soient vrais. Selon la définition de la pertinence, les trois énoncés sont pertinents puisque chacun d'eux a un effet cognitif positif sur Alice : sa connaissance du menu du dîner est plus précise.

On peut qualitativement déterminer quel énoncé est le plus pertinent. En effet, l'effet cognitif de (2) est plus important que celui de (1) car il est plus précis ; (2) est donc plus pertinent que (1).

L'effet cognitif de (2) et (3) est le même. Par contre, (3) demande plus d'effort d'interprétation que (2) ; (2) est donc plus pertinent que (3). On en conclue que (2) est l'énoncé le plus pertinent pour Alice.

Le principe de pertinence stipule que *chaque énoncé comporte la présomption de sa pertinence optimale*. Cela signifie que l'interprétation d'un énoncé par celui qui le reçoit doit être celle qui rend l'énoncé le plus pertinent pour lui.

Reprenons l'exemple précédent. Si Bernard a la possibilité de communiquer les trois énoncés, il transmettra celui qu'il pense être le plus pertinent, c'est-à-dire le (2). S'il transmet l'énoncé (1), Alice pourra certes déduire que de la viande sera servie au dîner mais elle pourra également déduire que Bernard ne sait pas plus précisément quelle viande sera servie ou ne veut pas le dire, sinon il l'aurait communiqué.

1.1.2 Analyse

Le Principe de Coopération, bien qu'énoncé pour caractériser un discours coopératif, se transpose facilement à l'échange d'informations dans un système multiagents. Ainsi, un agent a est coopératif dans son acte d'échange d'information vers un agent b si les conditions suivantes sont vérifiées :

– selon a, l'information transmise est suffisamment informative pour le besoin de b mais pas trop informative ;
– selon a, l'information transmise est vraie ;
– selon a, l'information transmise est claire et ordonnée pour b.

Tout comme le Principe de Coopération, la Théorie de la Pertinence s'applique facilement au cadre des agents qui échangent des informations dans un système.

Même si la démarche est différente, on trouve des éléments communs entre les différentes théories :

– les énoncés communiqués sont vrais ou du moins considérés comme vrai par celui qui les émet (qualité pour Grice et effet positif pour Sperber et Wilson)
– il faut transmettre suffisamment d'information (quantité pour Grice et effet positif pour Sperber et Wilson)
– il ne faut pas transmettre trop d'information (quantité pour Grice et effort d'interprétation pour Sperber et Wilson)
– il faut être clair et non-ambigu (manière pour Grice, effort d'interprétation pour Sperber et Wilson)

Pour Grice, le jugement de celui qui transmet l'information fait partie de l'aspect coopératif. En effet, les maximes commencent toutes par « Fais en sorte que... ». Cet aspect est moins net dans la Théorie de la Pertinence. En effet, cette théorie est plus centrée sur celui qui reçoit et non celui qui informe.

Ces théories sont considérées comme des fondements des théories de la communication et de la coopération. Les différents éléments listés ci-dessus peuvent donc être considérés comme des caractéristiques essentielles de la coopération et doivent se retrouver dans toute définition de ce concept.

La limite de ces théories est que les définitions proposées sont informelles. Il est vrai que de nombreuses formalisations des travaux de Grice existent mais ces travaux portent plus sur une formalisation des implicatures que du Principe de Coopération lui-même. Or, c'est précisément le Principe de Coopération qui nous intéresse dans le cadre de modélisation des échanges entre les agents. Nous présentons dans la section suivante un des travaux portant sur la modélisation du Principe de Coopération : la logique de l'interrogation de Groenendijk.

1.2 Pragmatique et logique : formalisation du Principe de Coopération

Cette section est dédiée à une formalisation en logique du Principe de Coopération de Grice. Nous nous intéressons notamment à la logique de l'interrogation développée par Groenendijk ([Gro99]). Le but de cette logique est de pouvoir juger de la qualité d'une formule, non pas en terme de validité, mais en terme d'accord ou non avec le Principe de Coopération de Grice.

1.2.1 Présentation des travaux

Groenendijk met en place un jeu d'interrogation. Ce jeu est composé de deux joueurs : un interrogateur et un témoin. Les règles du jeu sont les suivantes :

1. L'interrogateur soulève des problèmes en posant des questions non-superflues au témoin.
2. Le témoin fait des énoncés crédibles non redondants qui répondent exclusivement aux problèmes soulevés par l'interrogateur.

Ce jeu est censé être une représentation idéaliste d'un échange coopératif selon la définition donnée par Grice. La maxime de Quantité est respectée car les questions de l'interrogateur sont non superflues et les énoncés du témoin non-redondantes. La maxime de Qualité est respectée car le témoin fait des énoncés crédibles. Finalement, la maxime de Relation est respectée car les énoncés du témoin traitent exclusivement les questions posées par l'interrogateur. Le travail de Groenendijk consiste alors en la définition formelle ce que signifie respecter ces différentes maximes.

Groenendijk commence par définir le langage avec lequel l'interrogateur et le témoin vont pouvoir communiquer. Ce langage est basé sur une logique des prédicats à laquelle est rajoutée une couche pour pouvoir poser des questions. Les formules de la logique des prédicats sont appelées des *indicatives* et sont utilisées pour les énoncés du témoin. Les questions, appelées *interrogatives* sont de type $?\vec{x}Px$ où $\vec{x}$ est un ensemble de n variables ($n \geq 0$). Si $n = 0$, alors la question est de type oui/non. Par exemple $?\exists xPx$: y a t-il un objet qui a la propriété P ? Pour $n = 1$, un exemple est $?xPx$ (quels sont les objets qui vérifient la propriété P ?). Pour $n = 2$, un exemple est $?xyRxy$ (quelles paires vérifient la relation R ?). Les interrogatives peuvent uniquement être posées par l'interrogateur.

La sémantique de cette logique [1] n'est pas basée sur la notion d'interprétation classique mais sur le potentiel de changement de contexte. Développée par Groenendijk, Stokhof et Veltman ([GSV96]), cette sémantique permet l'interprétation des formules en terme d'opération sur le contexte dans lequel la formule est énoncée et est qualifiée de sémantique dynamique (contrairement à l'interprétation qui est considérée comme étant statique). La dynamique vient du fait que le contexte évolue au fur et à mesure que des informations sont énoncées.

Ici, le contexte dépend de deux éléments : les données et les problèmes. Le contexte relatif aux données est représenté par un ensemble de mondes : ceux qui sont compatibles avec les données déjà énoncées. Une indicative φ a pour effet de réduire le contexte à l'ensemble des mondes dans lesquels φ est vrai.

Par exemple, considérons les cinq mondes suivants (nous supposons que $P(a)$, $P(b)$ et $P(c)$ sont des formules du langage) : $w_1 = \{P(a), P(b), P(c)\}$, $w_2 = \{P(a), \neg P(b), P(c)\}$, $w_3 = \{\neg P(a), P(b), P(c)\}$, $w_4 = \{\neg P(a), \neg P(b), P(c)\}$ et $w_5 = \{P(a), P(b), \neg P(c)\}$. Supposons que la

1. Dans [Gro99], Groenendijk définit uniquement une sémantique pour sa logique. Une axiomatique est définie plus tard dans [tCS07].

seule donnée énoncée soit $P(c)$. Dans ce cas, les mondes compatibles avec cette donnée sont w_1, w_2, w_3 et w_4 et cet ensemble de quatre monde représente donc le contexte relatif aux données. Après avoir énoncé l'indicative $!P(a)$, le contexte est réduit aux mondes w_1 et w_2 c'est à dire aux mondes qui sont en accord avec les données, ici $P(c)$ et $P(a)$.

Le contexte relatif aux problèmes est représenté par une relation d'équivalence sur l'ensemble des mondes possibles : deux mondes sont dans la même classe d'équivalence si les réponses à tous les problèmes déjà posés ne les distinguent pas. L'effet d'une interrogative est donc de déconnecter des mondes pour lesquels la réponse n'est pas la même.

Par exemple, considérons de nouveau les cinq mondes $w_1, \ldots, w_5$. L'interrogative $?P(b)$ divise l'ensemble des mondes en deux classes : ceux dans lesquels $P(b)$ est vrai à savoir w_1, w_3 et w_5 et ceux dans lesquels $P(b)$ est faux à savoir w_2 et w_4.

Pour regrouper les deux parties du contexte dans un seul élément, le contexte est défini comme étant une relation symétrique et transitive sur un sous-ensemble des mondes possibles. Dans notre exemple, lorsque la seule donnée énoncée est $P(c)$ et que l'interrogative $?P(b)$ est demandée, le contexte est représenté par les deux classes de mondes $\{w_1, w_3\}$ et $\{w_2, w_4\}$: chacun des mondes de ces classes est en accord avec les données (c'est pour cela que w_5 n'apparaît dans aucun sous-ensemble) et les deux classes représentent les deux réponses possibles à la question $?P(b)$. Si l'indicative $!P(a)$ est énoncée, le contexte est alors réduit aux deux classes $\{w_1\}$ et $\{w_2\}$ dans lesquels les mondes sont en accord avec la nouvelle donnée $P(a)$.

Groenendijk caractérise alors les différentes maximes composant le Principe de Coopération de Grice dans sa logique.

- **Qualité** : le témoin respecte la maxime de qualité lorsqu'il ne se contredit pas. L'interrogateur ne peut pas violer cette maxime. En terme de contexte, cela signifie qu'après avoir énoncé son témoignage, l'ensemble des mondes ne se réduit pas à néant.
- **Quantité** : le témoin ne fait pas d'énoncé redondant et l'interrogateur ne pose pas de question superflue. Cela signifie que les témoignages ou les questions ne doivent pas déjà être induits par le contexte avant d'être énoncés ou posés.
- **Relation** : le témoin adresse uniquement les problèmes soulevés par l'interrogateur. En terme de contexte, cela signifie que si un monde est éliminé du contexte de données, alors tous les mondes qui lui étaient liés sont également éliminés. Autrement dit, un énoncé est informatif s'il élimine des mondes mais il répond complètement et uniquement à une question posée s'il élimine complètement au moins une alternative. Dans ce cas, Groenendijk dit que l'énoncé est agréé par le contexte. Sinon, cela signifie que l'énoncé ne répond que partiellement au problème soulevé.

Groenendijk définit ensuite la notion de **pertinence** [2]. Être pertinent signifie respecter les règles du jeu, c'est-à-dire respecter les trois maximes de Grice. Un énoncé est donc pertinent s'il est cohérent avec le contexte, non déjà induit et agréé par celui-ci.

Finalement, Groenendijk s'intéresse à la notion de réponse pertinente. Un énoncé est une réponse si elle est agréée par la question (c'est-à-dire si elle répond complètement à celle-ci). Une réponse est pertinente si elle est cohérente, non déjà induite et agréée après que la question ait été posée.

Les réponses pertinentes ont des propriétés intéressantes :

- Une information pertinente pour une question ssi [3] sa négation est pertinente pour la même question.

2. On peut noter que le terme anglais utilisé par Groenendijk est « pertinence » et non pas « relevance ». Cette notion est donc à distinguer de la pertinence (« relevance ») utilisée dans la maxime de relation de Grice.

3. Dans toute la suite du manuscrit, l'abréviation « ssi » sera utilisée à la place de « si et seulement si ».

– Si φ et ψ sont deux réponses à une même question, alors la conjonction $\varphi \wedge \psi$ est une réponse à cette question.
– Soient φ et ψ deux formules pertinentes pour une même question. φ est plus informative que ψ ssi après avoir dit φ, le contexte permet d'induire ψ et la réciproque (après avoir dit ψ, le contexte permet d'induire φ) est fausse.

1.2.2 Analyse

Ce travail est certainement la formalisation des travaux de Grice qui s'approche le plus de ce que nous souhaitons définir dans le cadre d'échange d'informations dans un système multi-agents. En effet, le jeu d'interrogation défini par Groenendijk permet de caractériser le fait qu'une réponse soit pertinente et à fortiori qu'un agent (le témoin) soit coopératif dans ses réponses.

Néanmoins, certains aspects ne sont pas considérés :
– le témoin peut énoncer des faits vrais ou faux. En effet, les réponses pertinentes ne sont pas forcément des réponses vraies. Par exemple, si une formule est pertinente alors sa négation l'est aussi. Cela vient du fait que dans ce travail, le respect de la maxime de Qualité consiste à ce que les énoncés du témoin soient cohérents entre eux. Ainsi, le témoin peut délibérément mentir à l'interrogateur, dès lors qu'il ne se contredit pas dans ses mensonges. Il y a, dans ce cas, violation de la maxime de Qualité telle qu'elle a été énoncée dans le Principe de Coopération.
– le témoin ne peut pas anticiper sur les questions de l'interrogateur, il ne peut que répondre aux questions qui sont posées.
– les réponses que le témoin donne à l'interrogateur sont des réponses directes. Le témoin ne peut pas faire des énoncés qui permettent à l'interrogateur de répondre à la question. Or, il se peut que le témoin ne soit pas capable de donner la réponse à la question mais qu'il possède des informations à partir desquelles l'interrogateur pourrait déduire la réponse à la question posée.
– Groenendijk caractérise les informations pertinentes. Cette caractérisation définit de nombreuses informations pertinentes (si deux informations sont une réponse alors leur conjonction l'est également). Une hiérarchie entre ces informations est introduite par Groenendijk dans le sens où il caractérise quelles informations sont les plus informatives. Néanmoins, le témoin n'est en aucun cas tenu de transmettre l'information la plus informative. Or, cette hiérarchie sur les informations devrait se répercuter sur la caractérisation de l'aspect coopératif du témoin : il est plus coopératif s'il transmet les informations les plus pertinentes.

1.3 Philosophie : une caractérisation de la pertinence épistémique

1.3.1 Présentation des travaux

Dans le domaine de la philosophie, Floridi ([Flo07]) propose une caractérisation de la pertinence épistémique. Par pertinence épistémique, il entend une pertinence qui prend en compte celui à qui l'information est destinée.

La démarche du travail de Floridi est particulièrement intéressante car il part d'une définition relativement simple et analyse ses limites. Il complique alors cette définition et réitère le processus. La définition qu'il obtient est relativement complexe. Selon lui, le degré de pertinence d'une information i pour un agent a est une fonction de l'adéquation avec laquelle i peut être considérée comme réponse à une question q, étant donnée la probabilité que cette question q soit posée par l'agent a.

Sans rentrer dans les détails, la définition finale proposée par Floridi met en avant les éléments suivants :

– *une information i est pertinente pour un agent a, en référence à un domaine d, dans un contexte c et à un certain niveau d'abstraction l.* Cela signifie que tous ces éléments interviennent dans la caractérisation de la pertinence d'une information.
– *une information pertinente i répond à une question q.* C'est la problématique du besoin. Une information est pertinente si elle répond à un besoin ou à une question qui importe à l'agent a.
– *le degré de pertinence d'une information i dépend de l'adéquation de celle-ci comme réponse à la question q.* Floridi met ici en avant le fait que certaines informations peuvent ne pas répondre complètement à des questions. C'est pourquoi il introduit des degrés de pertinence pour représenter l'adéquation de la réponse.
– *la probabilité qu'un agent a pose une question q dépend des informations qu'il possède déjà et des informations qu'il pense être disponibles.* Ceci souligne le fait que les agents ne sont pas forcément conscients de tous leurs besoins et que les questions qu'ils se posent peuvent évoluer en fonction de leurs connaissances.
– *la désinformation (information bien formée, qui a du sens, mais qui est fausse) ne peut pas être pertinente.* Floridi argumente que la désinformation est délétère et ne peut pas être considérée comme étant un apport positif pour celui qui la reçoit.

1.3.2 Analyse

Ce travail est très intéressant et répond à de nombreux problèmes qui ont été soulevés au sujet de la pertinence [4].

Parmi les différents éléments mis en avant par Floridi, on retrouve certains éléments qui relèvent de la communication en général et qui ont notamment été mis en avant en pragmatique, comme présenté en sous-section 1.1 (par exemple, la vérité de l'information qui est échangée). Mais on retrouve aussi des éléments plus particuliers à l'échange d'information entre les agents, notamment tout ce qui touche le besoin en information de celui qui reçoit l'information.

Même si le travail de Floridi se veut être beaucoup plus formel que ce qui est habituellement présenté en philosophie, il reste relativement imprécis sur la façon dont les différentes probabilités qu'il utilise doivent être définies. En effet, celles-ci paraissent délicates à évaluer (par exemple, quelle est la probabilité qu'un agent pose une question donnée après avoir appris une nouvelle information ?) et aucun élément n'est apporté à ce sujet.

Une autre limite de ce travail pour l'application à des agents qui échangent des informations est que Floridi ne prend en compte qu'un agent. En effet, le seul agent qui intervient dans la définition est l'agent qui reçoit l'information (cela peut d'ailleurs être mis en rapport avec la logique « d'être informé » que Floridi a développée). Or, l'échange d'information doit prendre en compte deux agents : celui qui reçoit mais également celui qui transmet l'information. C'est d'ailleurs ce deuxième agent qui juge de la pertinence de l'information pour celui qui va la recevoir et ce jugement n'est pas considéré dans le travail de Floridi.

4. L'article [Flo07] de Floridi lui permet notamment de répondre à plusieurs critiques ou remarques qui ont été faites sur sa définition.

1.4 Bases de données : caractérisation des réponses coopératives

1.4.1 Présentation des travaux

Dans le domaine des bases de données, de nombreux travaux ont été réalisés, notamment dans les années 1990, sur la caractérisation des « réponses coopératives » ([CD86, Min98, GM87, CD91, CD89]). Dans ce cadre, un utilisateur soumet une requête à un système coopératif (ou à une base de données) et ce dernier doit répondre à cette requête. La réponse coopérative peut être vue comme un type de réponse particulière qui va aider l'utilisateur, au delà de ce qu'il avait exprimé directement dans sa requête.

Dans [GGM92], les auteurs présentent un aperçu des différents travaux sur les réponses coopératives. On trouve également un tel état de l'art dans [Bre95].

À la base de ces travaux, on trouve le Principe de Coopération de Grice. Ce principe permet de caractériser les réponses appropriées à une requête : ce sont les réponses correctes, non-trompeuses et utiles pour une requête. Les réponses appropriées peuvent être considérées comme des réponses directes. Par exemple, à la question « Où est le bureau du Professeur Tournesol ? », une réponse appropriée serait « Au fond du couloir à droite. ». Les réponses coopératives, bien qu'étant également correctes, non-trompeuses et utiles, sont des réponses indirectes car elles sont jugées plus utiles à l'utilisateur que des réponses directes. Ces réponses passent par la reconnaissance d'un besoin de l'utilisateur qui n'était pas clairement exprimé dans sa requête. Par exemple, à la même question « Où est le bureau du Professeur Tournesol ? », une réponse coopérative pourrait être « Au fond du couloir à droite, mais le Professeur est absent aujourd'hui. ». Dans cet exemple, celui qui répond suppose que l'autre cherche le bureau du Professeur afin de le voir. Ainsi, dans sa réponse, il répondra non seulement à la question posée mais il apportera également une réponse au besoin présupposé.

Les réponses coopératives peuvent être classées en quatre grandes catégories ([Bre95, Sad96, Sad04]) [5] :

- *les réponses correctives* : la plupart des requêtes sous-entendent un certain nombre de présupposés. Par exemple, la question « Comment s'appelle le mari de Jeanne » sous-entend que Jeanne est mariée. Le rôle des réponses correctives est d'informer l'interlocuteur des présupposés de la requête qui son faux et qui empêchent une réponse directe à celle-ci. Une réponse corrective à la question « Comment s'appelle le mari de Jeanne ? » serait donc « Jeanne n'est pas mariée ».
- *les réponses complétives* ou *réponses sur-informatives* : ces réponses fournissent une information supplémentaire, non explicitement demandée par l'interlocuteur. Par exemple, une réponse complétive à la question « À quelle heure part le prochain train pour Toulouse ? » est « Prochain train pour Toulouse : 17h15, quai n°3 ».
- *les réponses suggestives* : ces réponses sont en général utilisées lorsqu'il n'y a pas de réponse directe à la requête. Elles représentent pour l'interlocuteur une ouverture ou une négociation possible. Par exemple, une réponse suggestive à la requête « Pourrais-je avoir l'heure de départ du prochain TGV à destination de Toulouse ? » est « Il n'y a plus de TGV à destination de Toulouse ce soir mais un train Corail part dans 30 minutes. »
- *les réponses conditionnelles* : ce type de réponse est utilisé lorsque la réponse directe à la requête est soumise à une ou des conditions. Par exemple à la question « Est-ce qu'il y a des vols pour Toulouse le matin avant 7 heures ? » on répond « Il y a un vol à 6h10 en semaine. »

5. Las catégories peuvent légèrement différer en fonction des sources. Nous donnons ici les catégories telles que définies par Sadek dans [Sad04].

– *les réponses intensionnelles* : les réponses intensionnelles (par opposition aux réponses extensionnelles) ne sont pas données sous la forme d'une liste d'éléments qui vérifient la requête mais sous la forme d'une propriété vérifiée par les éléments qui répondent à la requête. Par exemple, une réponse intensionnelle à la requête « Quels sont les employés qui gagnent plus de 4500 € par mois ? » est « Tous les cadres du service commercial. ». Une réponse extensionnelle serait du type « J.Dupont, M. Dupond et P. Tournesol gagnent plus de 4500 € par mois ». Cette propriété n'est pas obtenue en factorisant une propriété commune de tous les éléments qui vérifient la requête mais par un raisonnement sur la partie « intelligente » de la base de données dans laquelle sont décrites les propriétés des faits de la base.

Ces différents types de réponses peuvent être donnés à partir du moment où les besoins de celui qui est à l'origine de la requête ont été reconnus. Cette reconnaissance de l'intention est une problématique au centre de la coopération. Elle est présente dans la définition informelle de la coopération proposée par Sadek ([Sad94]) : *Un agent i est qualifié de coopératif à l'égard d'un agent j si (1) i tente d'aider j à réaliser ses buts tant que cela ne contredit pas ses propres buts et si (2) i ne se contente pas dans son aide à j, des buts de j exactement tels que celui-ci les explicite, mais va au-delà en essayant de découvrir ses intentions « ultimes » afin de l'aider à les réaliser.*

Sadek et Brétier utilisent une logique modale dont les modalités sont la croyance et l'intention pour formaliser cette reconnaissance d'intention. Ils utilisent alors la théorie des actes de langage développée par Austin ([Aus62]) pour reconnaître les différentes intentions qui interviennent dans la communication. Ils la mettent d'ailleurs en œuvre dans le projet ARTIMIS ([SBP97]).

Bien que se plaçant plus du côté Intelligence Artificielle, Herzig et Longin ont également développé une logique de l'intention basée sur les actes de langage [HL02]. Ils associent à leur logique des principes dits coopératifs car le but de la logique est de modéliser la reconnaissance d'intention.

De nombreuses techniques ont été développées pour les systèmes coopératifs. Dans [GGM92], les auteurs proposent de classifier ces techniques dans cinq catégories. Par exemple, la première catégorie est la considération des attentes de l'utilisateur. Cette catégorie rejoint le travail de Sadek et Brétier sur la reconnaissance d'intention. Une autre approche est celle développée par Cuppens et Demolombe ([CD89, CD91]) et consiste en la réécriture des requêtes posées par l'utilisateur pour inclure plus d'informations que n'aurait apporté une réponse à la requête de base. Les autres catégories sont : l'évaluation des présuppositions de la requête ; la détection et la correction des inconsistances dans une requête ; la formulation de réponses intensionnelles ; la généralisation des requêtes et des réponses.

1.4.2 Analyse

La problématique des systèmes coopératifs et des bases de données est relativement proche de la question des échanges d'informations dans un système multi-agent. En effet, l'utilisateur qui pose une requête peut être vu comme un agent qui a un besoin, la requête étant alors une expression de ce besoin. Le système coopératif peut également être vu comme un agent qui va apporter de l'information au premier, par la réponse qu'il va faire à la requête.

De nombreuses techniques et de formalismes ont été définis et développés pour le calcul des réponses coopératives. Le caractère coopératif du système a été majoritairement abordé d'un point de vue opérationnel, débouchant sur des algorithmes qui permettent de calculer les réponses. L'aspect sémantique n'a été, quant à lui, que très peu abordé.

Bien qu'une partie des techniques développées aient pour but de reconnaître les intentions de l'utilisateur, les connaissances ou les croyances de celui-ci ne sont pas prises en compte. Au mieux, le formalisme passe par la construction d'un profil d'utilisateur dans lequel celui-ci est défini en terme de sujet d'intérêt, de niveau d'expertise. Seuls Brétier et Sadek utilisent une logique modale dans laquelle ils peuvent exprimer les croyances des agents mais ils ne se sont pas intéressés à la formalisation des informations utiles et à l'étude de leurs propriétés ou du caractère coopératif des agents.

1.5 Intelligence Artificielle : pertinence et inférence

Dans le domaine de l'intelligence artificielle, le concept de pertinence a été étudié pour caractériser une relation qui existe entre deux éléments. L'exemple cité dans [DP98] illustre bien cette relation :
– Les oiseaux volent (d'habitude).
– Les oiseaux verts volent.
– Les oiseaux qui ont une aile cassée ne volent pas.
– Les oiseaux verts avec une aile cassée ne volent pas.

Intuitivement, on voit que le fait qu'un oiseau soit vert n'est pas pertinent pour savoir s'il vole ou pas. Par contre, avoir une aile cassée semble l'être.

De nombreuses approches pour formaliser cette relation de pertinence existent. La première approche qui peut venir à l'esprit est celle prise dans les logiques qui portent le nom de logiques de la pertinence, que nous commençons par présenter. Dans un deuxième temps, nous présentons les différents travaux développés autour de la pertinence pour améliorer l'inférence.

1.5.1 Les logiques de la pertinence

Présentation des travaux

L'implication matérielle est la solution classique proposée en logique pour modéliser la conditionnelle *si ... alors ...* que l'on utilise dans le langage naturel. Lorsque l'on utilise cette conditionnelle, on s'attend à ce qu'il existe une relation de cause à effet ou plus généralement une relation de pertinence entre la prémisse et la conclusion. Or l'implication matérielle pêche pour la représentation de cette pertinence. En effet, on a par exemple la propriété $p \to (q \to p)$. Si *l'eau est bouillante à 100˚C* est vrai, alors il est vrai que si *les oiseaux volent* alors *l'eau est bouillante à 100˚C*. Ce résultat est contre-intuitif car le fait que *les oiseaux volent* n'est pas pertinent pour conclure que *l'eau est bouillante à 100˚C*.

Pour résoudre ces problèmes de pertinence, les logiciens (dont Lewis comme présenté dans [Pri01]) ont introduit l'implication stricte. p implique (strictement) q est noté $\Box(p \to q)$ et signifie que dans tous les mondes où p est vrai alors q l'est aussi [6]. Cependant, cette implication pose aussi des problèmes de pertinence, comme $\Box(p \to (q \vee \neg q))$. Cela signifie que dans tous les mondes où p est vrai alors q ou $\neg q$ est vrai. Une illustration est si *la Terre est ronde* alors *soit il pleut à Paris, soit il ne pleut pas*. Ici encore, le fait que la Terre soit ronde n'est pas pertinent pour déduire le fait qu'il pleuve ou ne pleuve pas à Paris.

C'est dans cette problématique que les logiques de la pertinence ont été introduites ([AB75]). Ces logiques s'appuient sur un principe de partage de variable dans le cas de calcul propositionnel (il faut que le prémisse et la conclusion aient au moins une variable commune). Ce principe de partage est nécessaire pour assurer la pertinence mais non-suffisant dans le cas des logiques plus

6. Ce sont les prémisses de la logique modale (voir [HC96] pour plus de détails).

compliquées. De nombreux systèmes ont été développés (voir [DR02] pour une synthèse complète des travaux faits en logiques de la pertinence depuis Anderson et Belnap).

Analyse

Bien qu'approchant une définition réaliste de la pertinence, ces logiques sont en général assez compliquées à mettre en œuvre. En effet, ces logiques sont parfois vues comme des logiques modales car elles sont valides pour une classe de modèles défini sur une structure de Kripke. La différence avec la logique modale classique est que la relation d'accessibilité entre les mondes est un relation ternaire au lieu de l'usuelle relation binaire.

1.5.2 La pertinence pour améliorer l'inférence

Présentation des travaux

Avec l'explosion d'internet et de l'accessibilité à l'information en général, faire une requête dans une base de données est devenu un problème central. Dans ce cadre, la notion de pertinence a un rôle à jouer pour améliorer le raisonnement dans ces bases. Par exemple, si une requête porte sur « Titi » , qui est un oiseau, on va pouvoir accélérer le raisonnement en ne cherchant pas tout ce qui ne traite pas des oiseaux dans la base. En ne cherchant pas dans une partie de la base, on rend la recherche plus efficace. Dans ce cas, on *oublie* tout ce qui n'est pas pertinent avec le fait que Titi soit un oiseau.

Dans ce cadre, c'est plutôt la notion de non-pertinence (« irrelevance ») qui est caractérisée : on veut oublier ou ne pas chercher les informations qui ne sont pas pertinentes. Les informations pertinentes sont dans ce cas les informations qui ne sont pas non-pertinentes. [7]

Selon [LLM03], la pertinence utilisée dans ce cadre est fortement liée à deux notions qui ont eu de nombreuses dénominations dans la littérature et sont appliquées pour la résolution de nombreux problèmes :

- *la FV-indépendance*. Cette notion est également connue sous le nom de pertinence pour un sujet d'intérêt (Lakemeyer [Lak97]). Une formule est indépendante d'une variable si elle peut être réécrite de façon équivalente dans une forme dans laquelle la variable en question n'apparaît pas.
- *l'oubli d'un ensemble de variables dans une formule*. Cette notion se ramène à la précédente : oublier une variable dans une formule, c'est considérer la conséquence maximale de cette formule qui est indépendante de la variable.

Dans [LLM03], les auteurs proposent une étude systématique de ces notions et présentent les différents travaux dans lesquels elles sont utilisées. Nous reviendrons sur quelques définitions qu'ils proposent au cours du manuscrit (section 3.3), notamment pour comparer formellement la définition que nous donnons de l'utilité avec ces notions de pertinence.

Analyse

Cette approche de la pertinence est assez éloignée du problème que nous nous posons. En effet, les agents sont complètement éludés. Au mieux, les formules et les variables traitées sont issues de bases de connaissances que l'on pourrait supposer être des bases de connaissances d'agents

7. Dans [SG87], les auteurs considèrent que la pertinence et la non-pertinence ne peuvent pas être considérées comme des notions duales mais plutôt comme des notions complémentaires : l'une exprime la dépendance entre deux faits et l'autre exprime le manque de dépendance.

(par exemple dans [Lak97]). Un certain nombre de concepts comme le besoin en information, indispensables pour la représentation de ce qu'est une information utile ou pertinente pour un agent, n'apparaissent pas non plus. Il faut donc utiliser un cadre plus large que ceux proposés par cette approche.

1.6 Recherche d'information : caractérisation des documents pertinents

1.6.1 Présentation des travaux

« Le but d'un système de recherche d'information est de trouver les documents pertinents, ainsi la pertinence est un (si ce n'est pas le) concept clé de la recherche d'information » (Mizzaro [Miz98]). Dans le cadre de la recherche d'information (RI), un utilisateur a besoin d'information pour réaliser une tâche. Il dispose (généralement via un système de RI) d'un corpus de documents (de tout type et de tout format). Quand l'utilisateur propose au système une requête décrivant son besoin en information, on souhaite que les documents du corpus contenant des informations répondant à sa requête (et donc son besoin) soient retournés et classés selon leur pertinence par rapport à la requête. L'objectif de la RI est donc de sélectionner dans une collection les *documents pertinents* qui répondent à des *besoins utilisateurs*.

Classification de la pertinence

Les chercheurs en recherche d'information ont cherché à comprendre et clarifier le concept de pertinence.

Dans [SEN90], les auteurs caractérisent la nature et le rôle de la notion de pertinence :

- La pertinence est un *concept cognitif multidimensionnel* dont la définition dépend des *perceptions de l'utilisateur* sur ses besoins en information.
- La pertinence est un *concept dynamique* qui dépend du jugement de l'utilisateur sur la qualité de la relation entre une information qu'on lui propose et l'information dont il a réellement besoin à un instant donné.
- La pertinence est un concept *complexe mais systématique et mesurable* s'il est approché de façon conceptuelle et opérationnelle à partir du point de vue de l'utilisateur.

Borlund, dans [Bor03], fait une synthèse des différents travaux menés sur la pertinence ces dernières années. Elle distingue tout d'abord deux classes de pertinence : la pertinence orientée système et la pertinence orientée agent. La pertinence orientée système est calculée alors que la pertinence orientée agent est jugée par l'utilisateur.

Dans la classe de la pertinence orientée système, on trouve la pertinence algorithmique (ou pertinence thématique orientée système). C'est cette pertinence qui est utilisée dans la plupart des système de recherche d'information. Elle est en général calculée en fonction du nombre de caractéristiques communes entre la requête posée par l'utilisateur et les documents auxquels le système a accès.

Dans la classe de la pertinence orientée agent, on trouve plusieurs types de pertinence. Ceux ci ont été mis en évidence par Saracevic ([Sar96]) [8] :

8. Saracevic propose un dernier type de pertinence que nous ne présentons pas ici, la pertinence affective ou émotionnelle. Un document est pertinent émotionnellement s'il aide l'utilisateur à atteindre son but. Pour Borlund, ce dernier type n'est pas indépendant des autres et devrait être une caractéristique inhérente à toutes les pertinences orientées agent déjà citées.

- *la pertinence thématique orientée agent* : l'utilisateur juge du fait que le sujet du document correspond au sujet décrit par la requête. Contrairement à la pertinence algorithmique, cette pertinence n'est pas calculée par le système.
- *la pertinence cognitive* : le document est jugé pertinent si son sujet est connecté au besoin en information de l'utilisateur.
- *la pertinence situationnelle* : le document est jugé pertinent s'il contient des informations que l'utilisateur considère comme contribuant à la réalisation de sa tâche.

De nombreux autres types de pertinence existent. Par exemple, Mizzaro ([Miz98]) et Reid ([Rei99]) définissent la pertinence pour une tâche mais selon Borlund, cette pertinence est équivalente à la pertinence situationnelle. De même, les autres types de pertinence peuvent plus ou moins être ramenés à un des quatre types présentés ici.

Parmi ces différents types de pertinence, Borlund préconise l'utilisation de la pertinence situationelle car elle permet l'évolution de la perception de l'utilisateur par rapport à son besoin.

Une fois la pertinence déterminée, il faut choisir des critères qui permettent de dire si un document est pertinent. De nombreux critères ont été proposés ([Par93, Sch94, Bar94]). Ces critères portent en général sur les caractéristiques du document potentiellement pertinent (notamment les sources du document) et sur l'utilisateur, en particulier son expérience passée, sa motivation, son niveau d'expertise dans le domaine de la requête, etc.

Formalismes pour la recherche d'information

De nombreux formalismes existent pour représenter ces différentes pertinences. Chevallet s'est particulièrement intéressé aux formalismes basés sur la logique. De nombreux travaux ont été menés, et *il s'est avéré qu'un modèle théorique de RI devenait nécessaire. La modélisation par la logique a alors été choisie et plus particulièrement la logique modale* (Chevallet [Che04]). Dans ce même article, Chevallet fait une synthèse des différentes approches logiques utilisées en RI. Dans cet état de l'art, nous en donnons un bref aperçu.

Chevallet fait l'hypothèse suivante : *Pour qu'un document soit pertinent pour une requête, il doit exister une chaîne de déductions logiques incertaines*[9] *commençant au document et finissant à la requête. Le calcul de la pertinence se résume donc à prouver que* $D \rightarrow Q$, $\rightarrow$ étant une implication dite de recherche d'information qui doit être définie.

Chevallet propose alors des définitions de l'implication de recherche d'information pour plusieurs logiques.

- Logique des propositions

En logique des propositions, un document est représenté par une interprétation logique des termes atomiques et une requête est une formule appartenant au langage de la logique. Ainsi, un document D est pertinent pour un requête Q ssi $D \vDash Q$.

Pour introduire une notion de degré de pertinence, il est possible d'utiliser une logique trivaluée. Dans ce cas, les interprétations sont partielles et un document (représenté par une interprétation) est pertinent pour une requête si il peut être étendu en une interprétation qui est un modèle pour la requête.

9. Les déductions sont dites incertaines car le document peut être transformé. En particulier, Lalmas ([Lal96]) propose la transformation suivante : *Pour une représentation d d'un document, une représentation q de la requête et un ensemble de connaissances K, la mesure de pertinence, notée $d \rightarrow q$ relative à K, est déterminée par la transformation minimale, appliquée à d pour obtenir un d' tel que d' contienne q, noté $d' \Rightarrow q$, $\Rightarrow$ étant cette fois l'implication logique.*

- Logique modale

En logique modale, les documents sont des mondes. Les mondes sont connectés par une relation d'accessibilité qui représente les transformations possibles : deux documents sont connectés si l'un peut être transformé en l'autre et réciproquement. Plusieurs conséquences logiques (et donc plusieurs pertinences) peuvent être définies dans ce type de logique et chacune peut être interprétée en RI. Par exemple, si $D \models Q$ alors D est une réponse possible à Q ; si $D \models \Diamond Q$ alors D est une bonne réponse pour Q, après une possible transformation de D.

La logique modale peut être étendue en un modèle modal flou. Le principe est alors le même mais les différents éléments sont pondérés par des valeurs comprises entre 0 et 1. Cela permet de représenter les probabilités que des termes soient vrais ou faux ou de représenter la proportion de sources qui disent qu'un terme est vrai, etc. De même, la relation entre les mondes est pondérée par une valeur comprise entre 0 et 1, ceci pour modéliser le fait qu'un document peut « plus ou moins » être transformé en un autre. On obtient alors une mesure de la pertinence d'un document pour une requête donnée. Cette notion de modèle modal probabiliste a été étudiée notamment par Nie ([Nie92]).

Chevallet donne d'autres pistes de modélisations logiques (logique abductive, logique non monotone, ...) qui ont été utilisées pour modéliser la pertinence. Nous ne les présentons pas dans cet état de l'art.

1.6.2 Analyse

Parmi l'ensemble des pertinences qui sont définies pour le domaine de la RI, ce sont celles qui prennent en compte les agents qui nous intéressent. En effet, l'agent qui va recevoir l'information joue un rôle central dans la définition de ce que va être une information utile pour lui. Plus précisément, le concept d'utilité que nous cherchons à modéliser est la pertinence situationelle, c'est-à-dire la pertinence qui prend en compte l'apport de l'information pour l'agent par rapport à son besoin.

Nous retrouvons certains éléments que nous avons déjà rencontrés :
- l'utilité d'une information pour un utilisateur dépend du besoin en information de celui-ci et de la perception qu'il en a ;
- l'information utile est connectée au besoin en information de l'utilisateur ;
- l'utilité d'une information dépend des croyances et des connaissances de l'utilisateur.

Même si nous n'en avons présenté qu'un nombre restreint, de nombreux modèles de pertinence existent en RI. Néanmoins, ces modèles posent quelques problèmes pour s'appliquer à l'échange d'informations entre systèmes :
- la grande majorité des modèles formalisent la pertinence système et non pas la pertinence orientée agent. L'agent (ses connaissances ou son expérience, son besoin, etc.) n'est donc pas pris en compte.
- le domaine de la recherche d'information ne définit pas quelles sont les *informations pertinentes* pour une requête mais quels sont les *documents pertinents* pour une requête. Le document, support de l'information, introduit une couche de modélisation supplémentaire, le plus souvent très problématique. Cet aspect de la modélisation ne nous concerne pas puisque nous considérons que les agents échangent les informations directement.

1.7 Conclusion

Floridi affirme dans [Flo07] qu'il manque une définition formelle de la pertinence d'une information pour un agent. Ce n'est pas complètement vrai au vu de l'état de l'art qui vient d'être dressé. Pourtant, parmi tous les travaux présentés ici, il est difficile d'en trouver qui s'applique exactement au cas des agents qui échangent des informations dans un système. En effet, l'agent pour qui l'information est pertinente n'est pas considéré dans l'ensemble des définitions dites orientées système. Pour les définitions orientées agent, l'agent qui possède l'information pertinente, celui qui va donc informer l'autre de l'information pertinente n'est, lui, que rarement pris en considération. Or, c'est lui qui doit juger de la pertinence ou non de l'information pour celui à qui il va l'envoyer. Les seuls travaux dans lesquels cet agent qui possède l'information est pris en compte sont les travaux de Grice et de Groenendijk. Dans le premier cas, la définition de l'aspect coopératif de l'agent est informel. Dans le second, cet agent est relativement contraint car il ne peut pas anticiper sur les questions ou les besoins que l'interrogateur pourrait avoir. De plus, dès lors qu'il est cohérent, il peut mentir sur les informations qu'il énonce.

Néanmoins, ces différentes approches permettent de déterminer les caractéristiques principales des informations utiles et des agents coopératifs. Ces caractéristiques devront donc se retrouver dans le formalisme proposé.

- une information utile pour un agent répond à un besoin de celui-ci et doit donc être (réellement) connectée à celui-ci ;
- plus une information répond à un besoin, plus elle est utile pour ce besoin ;
- une information utile pour un agent est une information vraie ;
- plus une information est facile à comprendre (au sens complexité de l'information), plus elle est utile ;
- une information utile pour un agent dépend des croyances de celui-ci et l'utilité de l'information peut changer en fonction de l'évolution des croyances de l'agent ;
- un agent est coopératif si l'information qu'il transmet répond à un besoin de celui qui la reçoit ou tout du moins si elle répond à ce qu'il juge être un besoin de celui qui la reçoit ;
- un agent est coopératif s'il transmet des informations qui sont vraies ou tout du moins qu'il juge vraies ;
- un agent est coopératif vis à vis d'un autre s'il anticipe sur les besoins de celui-ci ;
- un agent est d'autant plus coopératif qu'il transmet des informations faciles à interpréter ;

On remarque que les informations que l'agent doit transmettre afin d'être coopératif ont beaucoup de caractéristiques communes avec les informations utiles pour celui qui les reçoit. Cela rejoint l'intuition qu'être coopératif revient à échanger les informations utiles ou tout moins jugées utiles par celui qui les envoie pour celui qui les reçoit.

Pour Sperber et Wilson, l'information pertinente a un effet cognitif positif et change la perception du monde de l'agent pour qui elle est pertinente. Le changement du monde correspond à un besoin de l'agent et peut être l'apprentissage d'une nouvelle information, la confirmation d'une information, la correction d'une information, etc., Il est possible de classer les informations utiles selon le besoin auxquelles elles répondent. Réciproquement, pour un besoin donné, les informations utiles vont avoir certaines caractéristiques. Par exemple, si un agent a un besoin en information (qui peut également être vu comme un manque d'information ([Tri04]), alors l'agent n'a a priori aucune croyance sur la vérité ou pas de l'information. De même, si un agent a besoin de vérifier ou de confirmer une information, alors c'est qu'il a déjà une croyance sur la vérité ou non de l'information en question. Si l'agent a besoin de compléter ces croyances sur un domaine donné, alors sont utiles les informations nouvelles de ce domaine ou les informations qui viennent corriger sa connaissance du domaine. Il serait possible de trouver d'autres besoins que ceux énoncés ci-dessus.

Pour chacun de ces besoins, correspond donc une définition différente de ce qu'est l'utilité.

Dans ce travail, nous nous concentrons sur un besoin particulier : **le besoin en information**. Ainsi, nous caractérisons les informations utiles pour un besoin en information. La coopération que nous définirons par la suite sera également une coopération pour un besoin en information.

Chapitre 2

Modélisation en logique modale

2.1 Quel formalisme choisir ?

L'état de l'art du chapitre précédent montre que de nombreux formalismes peuvent être utilisés pour travailler sur les concepts d'utilité d'une information et de coopération. La question de savoir lequel choisir se pose donc.

Tout d'abord, le formalisme doit permettre de travailler avec des agents. Plus précisément, le formalisme doit permettre de travailler avec les besoins des agents, leurs croyances, leurs capacités déductives, etc. Ces différents éléments peuvent être vus comme des états mentaux des agents. Ainsi, très classiquement ([RG91, Woo00] par exemple), nous choisissons d'utiliser un modèle de type BDI (Belief Desire Intention) pour représenter les différents états mentaux et pouvoir raisonner avec ceux-ci. Plus particulièrement, nous choisissons d'utiliser comme formalisme une logique modale dans laquelle les modalités représentent différents états mentaux.

Il faut ensuite choisir ces différentes modalités ainsi que leurs axiomatiques. Les possibilités sont nombreuses : croyance, connaissance, intention, désir, pro-attitude etc.

2.1.1 Connaissance ou croyance ?

Nous voulons pouvoir raisonner sur ce que les agents savent. Les deux modalités généralement utilisées pour représenter cela sont la connaissance et la croyance. La différence entre les deux tient dans le fait qu'une connaissance est toujours vraie alors qu'une croyance peut être fausse. Dans notre cas, les agents échangent des informations mais nous ne voulons pas contraindre les informations échangées à être vraies. De plus, les agents vont devoir juger de la pertinence d'une information pour les autres. Nous voulons pouvoir modéliser le fait qu'un agent se trompe dans son jugement de la pertinence d'une information pour un autre ou qu'il se trompe dans le fait de croire qu'une information est vraie. Nous préférons donc ici raisonner avec les croyances des agents plutôt qu'avec leurs connaissances.

2.1.2 Désir, intention et pro-attitude ...

Nous voulons pouvoir raisonner sur les besoins en information des agents. Parmi les différentes modalités généralement utilisées en logique modale, la notion de besoin peut se rapprocher de la notion de désir, mais également de la notion d'intention, de la notion de pro-attitude, de la notion de but,... En effet, avoir besoin de quelque chose peut être vu comme désirer être dans l'état où je possède ce quelque chose, ou avoir l'intention d'avoir ce quelque chose, ou avoir ce quelque chose est bon pour moi, ...

La modalité du désir est la modalité la plus « faible » parmi celles citées ci-dessus. En effet, les chercheurs considèrent que peu de contraintes s'appliquent aux désirs. Par exemple, il est possible d'avoir des désirs inconsistants entre eux. Dans la littérature, le désir est souvent mis en relation avec la croyance (par exemple [Woo00]) : a t-on conscience ou non de ses propres désirs ?

La modalité qui représente l'intention a reçu beaucoup d'attention dans la littérature ([CL90, Bra87]). Cohen et Levesque ([CL90]) font une étude approfondie de l'intention. Pour eux, l'intention peut être vue comme un désir pour lequel on s'engage. Les intentions sont plus contraintes que les désirs : il n'est en général pas possible d'avoir des intentions incohérentes. À cause de la notion d'engagement sous-jacente au concept d'intention, cette modalité a souvent été mise en relation avec une logique des actions. En effet, l'agent essaie de mettre en place les actions qu'il juge nécessaires à la réalisation de son intention. La notion de but est en général très proche de la notion d'intention. En effet, avoir l'intention de quelque chose sous-entend que ce quelque chose est un but à atteindre.

Comme son nom l'indique, la modalité de pro-attitude est une modalité qui représente le fait qu'un agent ait une attitude positive envers une formule. Les notions de désir, d'intention ou de but peuvent alors être vues comme des pro-attitudes particulières. La notion de pro-attitude développée dans [SSLR07] est relativement contrainte : les agents ont conscience de leurs pro-attitudes, ils ne peuvent avoir de pro-attitude envers les tautologies et les contradictions, les effets collatéraux de la pro-attitude doivent être limités ...

Pour nous, la notion de besoin n'implique pas de notion d'engagement. Ainsi, la modalité d'intention, telle qu'elle est généralement utilisée dans la littérature n'est pas appropriée. Il en est de même pour la notion de but.

La notion de pro-attitude est trop générale pour représenter le besoin. Avoir un besoin de quelque chose est plus fort qu'une simple attitude positive envers ce quelque chose.

De même, le désir tel qu'il est utilisé habituellement n'est pas assez contraint (par des axiomes) pour représenter le besoin.

Aucune des modalités décrites ci-dessus ne convient exactement pour l'expression du besoin. Ainsi, nous choisissons d'utiliser la modalité du désir mais avec des contraintes plus fortes que celles qu'on lui attribue généralement. L'ensemble des axiomes qui décrivent le comportement du désir est donné dans la section suivante.

2.2 Cadre Logique

Le cadre formel avec lequel nous travaillons est une logique multimodale propositionnelle que nous notons $\mathcal{L}$. Les modalités qui nous intéressent ici et qui nous permettent de donner des définitions formelles de l'utilité et de la coopération sont la croyance et le désir.

Nous présentons tout d'abord le langage, puis nous proposons une axiomatique et une sémantique. Nous établissons finalement la validité et la complétude du système de preuves pour une classe de modèles que nous caractérisons.

2.2.1 Langage

L'alphabet du langage de $\mathcal{L}$ est composé de :

– un ensemble d'agents $\mathcal{A} = \{a, b, c, \dots\}$

- un ensemble de variables propositionnelles $\mathcal{V} = \{p, q, r, \ldots\}$
- des connecteurs : $\neg$, $\vee$
- des parenthèses (et)
- pour chaque agent a de $\mathcal{A}$, on définit deux modalités B_a et D_a

Définition I.1 (Formules de $\mathcal{L}$)**.** *Les formules de $\mathcal{L}$ sont définies récursivement comme suit :*
- *si $p \in \mathcal{V}$ alors p est une formule de $\mathcal{L}$*
- *si φ est une formule de $\mathcal{L}$ alors $\neg\varphi$ est également une formule de $\mathcal{L}$*
- *si φ est une formule de $\mathcal{L}$ et a un agent de $\mathcal{A}$, alors $B_a\varphi$ et $D_a\varphi$ sont des formules de $\mathcal{L}$. $B_a\varphi$ se lit l'agent a croit que φ et $D_a\varphi$ se lit l'agent a désire φ.*
- *si φ_1 et φ_2 sont des formules de $\mathcal{L}$ alors $\varphi_1 \vee \varphi_2$ est également une formule de $\mathcal{L}$.*

Nous définissons également quelques abréviations et notations.

Notation I.1. Soient φ_1 et φ_2 des formules et a un agent de $\mathcal{A}$, on note alors :
- $\varphi_1 \wedge \varphi_2 \equiv \neg(\neg\varphi_1 \vee \neg\varphi_2)$
- $\varphi_1 \rightarrow \varphi_2 \equiv \neg\varphi_1 \vee \varphi_2$
- $\varphi_1 \leftrightarrow \varphi_2 \equiv (\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_2 \rightarrow \varphi_1)$
- $\varphi_1 \otimes \varphi_2 \equiv (\varphi_1 \wedge \neg\varphi_2) \vee (\varphi_2 \wedge \neg\varphi_1)$
- $Bif_a\varphi_1 \equiv B_a\varphi_1 \vee B_a\neg\varphi_1$. $Bif_a\varphi$ se lit *l'agent a sait si φ* [10]

Soit n un entier naturel tel que $n \geq 1$. Soient φ_1, ... φ_n n formules du langage.
- $\bigwedge_{i=1}^n \varphi_i \equiv \varphi_1 \wedge \ldots \wedge \varphi_n$
- $\bigvee_{i=1}^n \varphi_i \equiv \varphi_1 \vee \ldots \vee \varphi_n$
- $\bigotimes_{i=1}^n \varphi_i \equiv \bigvee_{i=1}^n(\varphi_i \wedge \bigwedge_{j=1, j\neq i}^n \neg\varphi_j)$
- Pour $n = 1$, on a $\bigwedge_{i=1}^n \varphi_i \equiv \varphi_1$, $\bigvee_{i=1}^n \varphi_i \equiv \varphi_1$ et $\bigotimes_{i=1}^n \varphi_i \equiv \varphi_1$

De façon à alléger l'écriture, lorsqu'il n'y a pas d'ambiguïté, nous écrirons $\bigwedge_i$ (resp. $\bigvee_i$ et $\bigotimes_i$) à la place de $\bigwedge_{i=1}^n$ (resp. $\bigvee_{i=1}^n$ et $\bigotimes_{i=1}^n$).

Nous désignerons par $\top$ n'importe quelle tautologie et par $\bot$ n'importe quelle contradiction.

Finalement, on dira qu'une formule est *objective* lorsqu'elle ne contient pas d'opérateur modal.

2.2.2 Axiomatique

Soient a un agent de $\mathcal{A}$, φ et ψ des formules du langage. Les règles d'inférence et les schémas d'axiomes sont présentés sur les figures 2.1 et 2.2. Les théorèmes de la logique (notés $\vdash \varphi$) sont toutes les formules déductibles des axiomes et des règles d'inférence.

Nous reprenons tout d'abord des bases de la logique propositionnelle. (MP) est la règle de Modus Ponens. Nous considérons que toutes les formules valides du calcul propositionnel sont des axiomes de notre logique. Pour plus de détail sur le calcul propositionnel et les différentes règles de celui-ci, le lecteur pourra se référer à [End72].

Très classiquement, nous considérons que l'opérateur de croyance B_a est un opérateur normal régi par le schéma d'axiomes KD45.
- (BK) représente la capacité déductive de l'agent : si celui-ci croit que $\varphi \rightarrow \psi$ et qu'il croit que φ est vrai alors il croit également que ψ est vrai.
- (BD) signifie que que les agents ont des croyances cohérentes.

10. Notons ici que l'utilisation du verbe savoir n'a aucun lien avec la modalité épistémique généralement notée K mais relève de l'utilisation, en langue française, de “savoir si” au lieu de “croire si”.

(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$
(BNec)	$\dfrac{\varphi}{B_a\varphi}$
(DUE)	$\dfrac{\varphi}{\neg D_a(\varphi)}$

FIGURE 2.1 – Règles d'inférence

(Taut)	Les tautologies de la logique propositionnelle
(BK)	$B_a(\varphi \to \psi) \land B_a\varphi \to B_a\psi$
(BD)	$B_a\varphi \to \neg B_a\neg\varphi$
(B4)	$B_a\varphi \to B_aB_a\varphi$
(B5)	$\neg B_a\varphi \to B_a\neg B_a\varphi$
(D4)	$D_a\varphi \to B_aD_a\varphi$
(D5)	$\neg D_a\varphi \to B_a\neg D_a\varphi$
(DRE)	$B_a(\varphi \leftrightarrow \psi) \to (D_a\varphi \leftrightarrow D_a\psi)$

FIGURE 2.2 – Schéma d'axiomes

– (B4) représente l'introspection positive des croyances (les agents sont conscients de leurs croyances).
– (B5) représente l'introspection négative des croyances (les agents sont conscients de ce qu'ils croient être faux).
– (BNec) est la règle d'inférence nécessitation. Elle signifie que les agents sont conscients des théorèmes de la logique.

L'opérateur de désir est un opérateur non-normal. Cela signifie notamment que les agents ne désirent pas forcément les conséquences de ce qu'ils désirent (problème de "side effect" ou effet collatéral mis en évidence par Bratman [Bra87]).

– La règle d'inférence (DUE) signifie que les agents n'ont pas le désir de ce qui est déjà tautologiquement vrai. Le désir correspond donc à un manque. Nous verrons que cela a un impact sur les relations entre les désirs des agents et leurs croyances.
– (D4) et (D5) représentent respectivement les introspections positive et négative des désirs. Autrement dit, les agents ont conscience de leurs désirs et également de leurs "non-désirs".
– (DRE) signifie que si deux formules sont équivalentes pour l'agent, alors désirer la première équivaut à désirer la seconde.

À ce stade, il est intéressant de regarder un des théorèmes qui découle des différents axiomes et qui illustre particulièrement bien la relation qui existe entre la croyance et le désir.

Théorème I.1 (Relation désir-croyance)**.** *Soit a un agent de $\mathcal{A}$ et φ une formule.*
Alors $\vdash D_a\varphi \to \neg B_a\varphi$ *(ou* $\vdash B_a\varphi \to \neg D_a\varphi$*)*

Ce théorème signifie que si un agent a un désir φ, alors il ne croit pas que φ soit vrai. Autrement dit, les agents ne peuvent pas avoir le désir des choses qu'ils croient déjà vraies, ils ont des désirs de choses nouvelles. Par exemple, si un agent désire que le courrier soit trié, alors il croit que le courrier n'est pas trié ou s'il croit que le courrier est trié alors il ne désire pas qu'il le soit.

Par contre (et c'est une des différences avec la pro-attitude de [SSLR07]), nous considérons que les agents peuvent avoir le désir de choses qu'ils croient fausses, c'est-à-dire que $\nvDash B_a\varphi \to D_a\neg\varphi$. Par exemple, si un agent croit que le courrier n'est pas trié, alors il peut avoir le désir qu'il soit trié.

2.2.3 Sémantique

Dans [SSLR07], les auteurs proposent une sémantique pour la notion de pro-attitude et prouvent la validité-complétude de leur système de preuve pour la classe de modèles qu'ils spécifient. Comme nous l'avons vu en section 2.1.2, la notion de pro-attitude est relativement contrainte et nous ne désirons pas autant contraindre la notion de désir. Ainsi, la sémantique de notre logique est une généralisation de la sémantique de [SSLR07].

Notre opérateur de désir est un opérateur non-normal. En conséquence, nous ne pouvons pas utiliser une structure de Kripke classique mais nous devons utiliser une sémantique du voisinage. Cette sémantique est une généralisation de la sémantique relationnelle classique pour les logiques modales normales ([Che80]).

Afin de ne pas perdre la relation d'accessibilité classique que l'on a pour la croyance, les cadres que nous considérons ici sont des hybrides de structures de Kripke et de cadres de voisinage.

Définition I.2. *Soit W un ensemble non vide de mondes. Pour chaque agent a de $\mathcal{A}$, on définit une relation d'accessibilité R_a sur W^2 et une fonction de voisinage N_a qui va de W vers 2^{2^W}. On note R l'ensemble des relations d'accessibilité R_a et N l'ensemble des fonctions de voisinage N_a. Le tuple $\langle W, R, N\rangle$ est alors appelé un cadre de voisinage hybride.*

Dans un cadre $\langle W, R, N\rangle$, R_a (élément de R) sera la relation d'accessibilité associée à la croyance B_a et N_a (élément de N) la fonction de voisinage associée au désir D_a.

Notation I.2. Pour tout w de W, on note $R_a[w]$ l'ensemble $\{w' \in W | wR_aw'\}$ qui représente l'ensemble des mondes en relation R_a avec w.

Définition I.3. *Un modèle de voisinage hybride est un tuple $\mathbb{M} = \langle W, R, N, V\rangle$ où $\mathbb{F} = \langle W, R, N\rangle$ est un cadre de voisinage hybride et $V : \mathcal{V} \to 2^W$ est une fonction de valuation qui à chaque variable propositionnelle p de $\mathcal{V}$ associe un ensemble de mondes..*

On utilisera le terme cadre pour désigner les cadres de voisinage hybrides et le terme modèle pour désigner les modèles de voisinage hybrides.

Définition I.4 (Satisfaction d'une formule)**.** *Étant donné un modèle $\mathbb{M} = \langle W, R, N, V\rangle$ et $w \in W$, la satisfaction d'une formule φ est définie récursivement, pour chaque agent a de $\mathcal{A}$ de la façon suivante :*

- $\mathbb{M}, w \vDash p$ *ssi* $w \in V(p)$. $(p \in \mathcal{V})$
- $\mathbb{M}, w \vDash \neg\varphi$ *ssi* $\mathbb{M}, w \nvDash \varphi$.
- $\mathbb{M}, w \vDash \varphi \vee \psi$ *ssi* $\mathbb{M}, w \vDash \varphi$ *ou* $\mathbb{M}, w \vDash \psi$.
- $\mathbb{M}, w \vDash B_a\varphi$ *ssi* $\forall\ w' \in R_a[w]$, $\mathbb{M}, w' \vDash \varphi$.
- $\mathbb{M}, w \vDash D_a\varphi$ *ssi* $\{w | \mathbb{M}, w \vDash \varphi\} \in N_a(w)$

Notation I.3. Étant donné un modèle $\mathbb{M}$ et une formule φ, $(\varphi)^{\mathbb{M}}$ est l'ensemble $\{w|\mathbb{M}, w \vDash \varphi\}$, c'est-à-dire l'ensemble des mondes de $\mathbb{M}$ dans lesquels φ est satisfaite.

Définition I.5 (Satisfaisabilité - validité d'une formule)**.** *Soit φ une formule.*
- *φ est* ***satisfaite dans le monde*** *w* ***du modèle*** *$\mathbb{M}$ ssi $\mathbb{M}, w \vDash \varphi$*
- *φ est* ***satisfaite dans le modèle*** *$\mathbb{M}$ (noté $\mathbb{M} \vDash \varphi$) ssi elle est satisfaite dans tous les mondes de ce modèle.*
- *φ est* ***valide dans la classe de modèles*** *$\mathbb{C}$ (noté $\mathbb{C} \vDash \varphi$) ssi elle est satisfaite dans tous les modèles de la classe $\mathbb{C}$. Lorsqu'il n'y a pas d'ambiguïté, nous dirons que φ est valide et nous le noterons $\vDash \varphi$.*

Nous utilisons une classe de modèles qui modélisent le comportement des opérateurs B_a et D_a en contraignant la relation d'accessibilité R_a et la fonction de voisinage N_a.

Définition I.6. *Soit $\mathbb{M} = \langle W, R, N, V\rangle$ un modèle et a un agent de $\mathcal{A}$.*
- *La relation R_a est* ***sérielle*** *ssi $\forall w \in W, R_a[w] \neq \{\}$. Cela signifie que tout monde a au moins un successeur.*
- *La relation R_a est* ***transitive*** *ssi $\forall w_1, w_2, w_3 \in W$, si $w_1 R_a w_2$ et $w_2 R_a w_3$ alors $w_1 R_a w_3$.*
- *La relation R_a est* ***euclidienne*** *ssi $\forall w_1, w_2, w_3 \in W$, si $w_1 R_a w_2$ et $w_1 R_a w_3$ alors $w_2 R_a w_3$.*

Un modèle $\mathbb{M}$ est sériel (resp. transitif et euclidien) si et seulement si toutes les relations d'accessibilité R_a de R sont sérielles (resp. transitives et euclidiennes).

Définition I.7. *Soit $\mathbb{M} = \langle W, R, N, V\rangle$ un modèle et a un agent de $\mathcal{A}$. On dit que la fonction de voisinage N_a est* ***unit-exclusive*** *si et seulement si $\forall w \in W, W \notin N_a(w)$. Le modèle $\mathbb{M}$ est unit-exclusif si et seulement si toutes les fonctions de voisinage de l'ensemble N sont unit-exclusives.*

Définition I.8. *Soit $\mathbb{M} = \langle W, R, N, V\rangle$ un modèle. On dit que $\mathbb{M}$ est* introspectif *si, pour tout agent a de $\mathcal{A}$*

(intro-1) $\forall w, w' \in W$ tels que $wR_a w'$, on a $N_a(w) = N_a(w')$;

(intro-2) $\forall w \in W$ et $\forall X, X' \subseteq W$ tels que $X \cap R_a[w] = X' \cap R_a[w']$, on a $X \in N_a(w)$ ssi $X' \in N_a(w)$.

Définition I.9. *On appelle* **C** *la classe des modèles qui sont sériels, transitifs, euclidiens, unit-exclusifs et introspectifs.*

Exemple I.1. Un exemple de modèle de la classe **C** est présenté sur la figure 2.3. Dans cet exemple, a est un agent de $\mathcal{A}$ et p une variable.

Sur ce schéma, sont présentées la relation d'accessibilité R_a et la fonction de voisinage N_a. Plus précisément on a :
- $R[w_0] = R[w_1] = R[w_2] = \{w_1, w_2\}$
- $N_a(w_0) = N_a(w_1) = N_a(w_2) = \{\{w_1\}, \{w_0, w_1\}\}$

2.2.4 Validité - Complétude

Théorème I.2. *Soit φ une formule. φ est un théorème de notre logique ssi φ est valide dans la classe* **C***.*

La preuve de la validité est relativement directe. Pour la complétude, elle consiste à construire un modèle canonique. Cette preuve se trouve en annexe de ce manuscrit.

Corollaire I.1. *Le système de preuves étant valide pour la classe* **C** *qui est non-vide (voir exemple I.1), on peut en déduire que le système est consistant.*

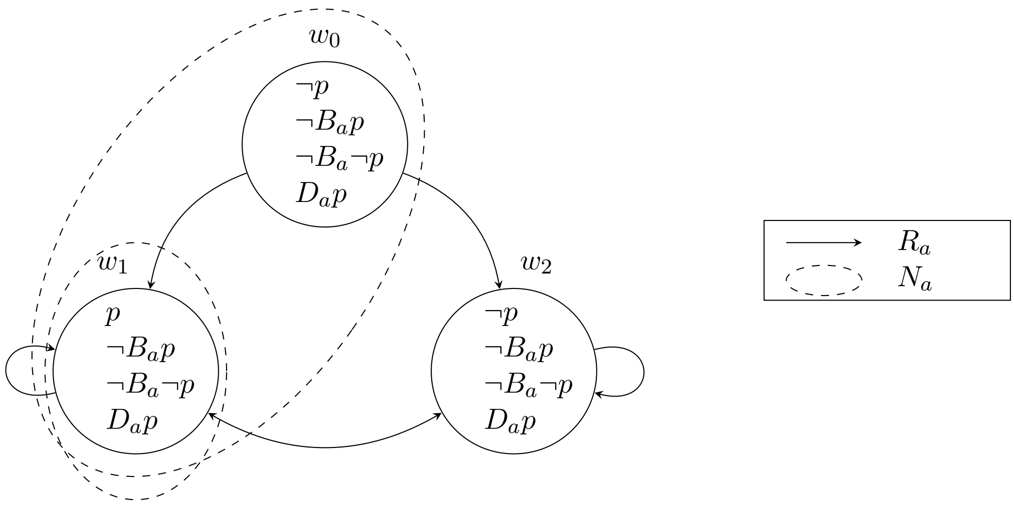

FIGURE 2.3 – Exemple de modèle

Chapitre 3

Information utile

Dans ce chapitre, nous proposons tout d'abord une définition formelle de la notion d'utilité d'une information. Dans un deuxième temps, nous étudions les propriétés des informations utiles. Nous définissons ensuite quelles sont les informations les plus utiles. L'avant-dernière partie est consacrée à la comparaison avec une notion de pertinence, celle proposée par Lakemeyer dans le domaine de l'Intelligence Artificielle. Finalement, nous développons quelques extensions de la notion d'utilité d'une information.

3.1 Définition

L'utilité d'une information est définie pour un agent et pour une formule objective (c'est-à-dire sans modalité) qui est la requête à partir de laquelle on déduit le besoin en information de l'agent.

Notation I.4. Soit a un agent de $\mathcal{A}$, soit φ une formule et soit Q une formule objective. On note $U_a^Q\varphi$ la formule

$$D_a Bif_a Q \wedge (B_a(\varphi \rightarrow Q) \otimes B_a(\varphi \rightarrow \neg Q)) \wedge \varphi$$

$U_a^Q\varphi$ se lit **l'information φ est utile pour l'agent a par rapport à sa requête Q**.

Définition I.10 (Utilité de l'information dans un monde)**.** *Soit a un agent de $\mathcal{A}$, soit φ une formule et soit Q une formule objective. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w un monde de ce modèle. L'information φ est* ***utile*** *pour l'agent a par rapport à sa requête Q dans le monde w si et seulement si $U_a^Q\varphi$ est satisfaite dans w, c'est à dire :*

$$\mathbb{M}, w \vDash U_a^Q\varphi$$

De façon à alléger l'écriture, lorsqu'il n'y a pas d'ambiguïté, nous ne précisons ni le modèle $\mathbb{M}$ ni le monde w dans laquelle l'utilité est vraie.

Cette définition est composée de trois éléments.

1. Le besoin en information de l'agent
L'information utile doit permettre à un agent de répondre à un besoin de celui-ci. Ce besoin

apparaît donc dans la définition de l'utilité. Le besoin qui nous intéresse ici est le besoin en information. Nous supposons dans un premier temps que ce besoin est relativement simple et peut s'exprimer de la façon suivante : « L'agent a veut savoir si une information Q est vraie ou non, Q étant une formule objective ». Dans le cadre logique que nous nous sommes donné, le besoin en information se représente par la formule $\mathbf{D_a Bif_a Q}$, c'est-à-dire par « l'agent a désire être dans un état où il a une croyance sur Q ». L'information Q à l'origine du besoin en information est appelée une *requête*.

2. Les croyances de l'agent
L'information utile doit être connectée, dans la base de croyances, à ce que pour quoi elle est utile, c'est-à-dire le besoin de l'agent. Plus précisément, l'information utile doit répondre au besoin de l'agent. Ici, le besoin de l'agent est de savoir si la formule Q est vraie ou non. L'information utile doit donc être connectée, pour l'agent, soit à Q, soit à $\neg Q$. Nous choisissons d'utiliser l'implication matérielle pour représenter cette connexion. Ainsi, une information est utile si elle implique Q ou si elle implique $\neg Q$ dans la base de croyances de l'agent.
Si une information permet à la fois de déduire Q et $\neg Q$, nous considérons qu'elle ne permet pas de répondre au besoin de l'agent. C'est pourquoi nous utilisons l'opérateur $\otimes$ (ou exclusif).
Formellement, la représentation de l'utilisation des connaissances de l'agent dans la définition de l'utilité est la formule $\mathbf{B_a}(\varphi \to \mathbf{Q}) \otimes \mathbf{B_a}(\varphi \to \neg\mathbf{Q})$.

3. La vérité de l'information
Seules les informations vraies sont utiles pour l'agent. Une information fausse, même si elle répond au besoin de l'agent, ne peut pas être considérée comme étant utile à celui-ci. Ce point a été particulièrement abordé lors de l'état de l'art et nous ne le redéveloppons pas ici.

De la définition de l'utilité d'une information, nous pouvons déduire la définition de l'inutilité d'une information pour un besoin.

Définition I.11 (Inutilité d'une information)**.** *Soit a un agent, φ une formule et Q une formule objective. Soit $\mathbb{M}$ un modèle et w un monde de ce modèle. On dit que l'information φ est* ***inutile*** *pour l'agent a par rapport à une requête Q dans le monde w si et seulement si la formule $\neg U_a^Q \varphi$ est satisfaite dans le monde w. Cela signifie qu'au moins une des trois formules qui suit est satisfaite dans w :*
- $\neg D_a Bif_a Q$ *: l'agent n'a pas besoin de savoir si Q*
- $\neg B_a(\varphi \to Q) \wedge \neg B_a(\varphi \to \neg Q)$ *: l'agent ne peut rien déduire sur Q à partir de φ*
- $B_a(\varphi \to Q) \wedge B_a(\varphi \to \neg Q)$ *: l'information φ permet à la fois de déduire Q et $\neg Q$* [11]
- $\neg\varphi$ *: l'information φ est fausse dans w*

Illustrons cette définition de l'utilité sur un exemple que nous reprendrons tout au long de cette partie.

Exemple I.2. Soient a et b deux agents qui doivent prendre le train. Ces agents aimeraient savoir si leur train est en retard. Supposons qu'un incident sur la ligne ait eu lieu. « Le train est en retard » est modélisé par *retard* et « Il y a un incident » est modélisé par *inc*. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w_0 un monde de $\mathbb{M}$ qui va représenter cet état du monde réel.

11. On remarquera que ce cas est le cas où l'agent croit $\neg\varphi$. Nous reviendrons sur cette propriété dans la section 3.2 Propriétés.

L'agent a a besoin de savoir si son train est en retard. Formellement, cela s'exprime par $D_a Bif_a(retard)$. Supposons qu'il croit que s'il y a un incident alors son train est en retard [12]. Cette croyance est modélisée par la formule $B_a(inc \to retard)$. On a donc :

- $\mathbb{M}, w_0 \models D_a Bif_a(retard)$
- $\mathbb{M}, w_0 \models B_a(inc \to retard)$
- $\mathbb{M}, w_0 \models inc$

On peut donc déduire que $\mathbb{M}, w_0 \models U_a^{retard} inc$, c'est-à-dire que l'information inc est utile à l'agent a pour la requête $retard$ dans le monde w_0.

L'agent b a également besoin de savoir si son train est en retard. Formellement, on a donc $D_b Bif_b(retard)$. Supposons que les croyances de l'agent b soient différentes de celles de l'agent a. Plus précisément, b croit que s'il n'y a pas d'incident alors le train ne sera pas en retard [13]. Cette croyance est modélisée par $B_b(\neg inc \to \neg retard)$. On a donc :

- $\mathbb{M}, w_0 \models D_b Bif_b(retard)$
- $\mathbb{M}, w_0 \models B_b(\neg inc \to \neg retard)$
- $\mathbb{M}, w_0 \models inc$

Cette fois, nous ne pouvons déduire que ni l'information $\neg inc$, ni l'information inc ne sont utiles pour le besoin de l'agent b dans le monde w_0. En effet, l'information inc ne permet pas à l'agent de déduire quoi que ce soit à propos du retard du train et lui est donc inutile dans w_0. L'information $\neg inc$ est fausse dans w_0. Elle est donc également inutile dans ce monde. On remarque que la vérité de l'information utile est indispensable pour ne pas induire l'agent en erreur.

3.2 Propriétés

Cette section est dédiée à l'étude formelle des propriétés des informations utiles. Cette étape est importante car elle permet la « validation » de la définition que nous donnons à l'utilité. En effet, il faut que les différentes propriétés obtenues correspondent à ce que l'on attend d'une caractérisation de l'utilité (voir état de l'art).

Dans l'ensemble des propriétés que nous énonçons, a est un agent de $\mathcal{A}$, φ, φ_1 et φ_2 sont des formules et Q, Q_1 et Q_2 sont des formules objectives. Les propositions qui suivent sont théorèmes de notre logique.

Proposition I.1.

$$\vdash D_a Bif_a Q \to \neg Bif_a Q$$

Si un agent a le besoin de savoir si une requête Q est vraie, alors il ne sait pas si celle-ci est vraie ou fausse, c'est-à-dire qu'il ne croit ni Q ni $\neg Q$. Cette propriété correspond au fait que le besoin en information est un manque d'information.

Proposition I.2.

$$\vdash U_a^Q \varphi \to \neg B_a \varphi$$

12. Nous supposons que ce sont les seules croyances que l'agent possède sur l'incident et le retard. En particulier, $\neg B_a(inc \to \neg retard)$, $\neg Bif_a(retard)$, $\neg Bif_a(inc)$ sont satisfaites dans w_0.

13. Comme pour l'agent a, nous supposons que b n'a pas d'autre croyance. En particulier, nous supposons que $\neg B_b(\neg inc \to retard)$, $\neg Bif_b(retard)$, $\neg Bif_b(inc)$ sont satisfaites dans w_0.

Si une information est utile pour un agent alors elle est nouvelle pour lui. Cela vient du fait que si l'agent connaît déjà l'information utile, alors il est capable de répondre à son besoin en information (il peut déduire Q ou $\neg Q$). Or, nous avons vu dans la proposition précédente que le besoin en information est un manque d'information. Il y a donc contradiction.

Proposition I.3.

$$\vdash U_a^Q \varphi \to \neg B_a \neg \varphi$$

Si une information est utile pour un agent, alors celle-ci n'est pas en contradiction avec ce que croit l'agent. Cette propriété vient des paradoxes intuitifs liés à l'utilisation de l'implication matérielle. En effet, à partir d'une information que l'agent croit fausse, il peut déduire n'importe quoi dans sa base de croyances, dont Q et $\neg Q$. Or, si une information permet de déduire à la fois Q et $\neg Q$, elle ne peut pas être considérée comme utile. Nous reviendrons sur ce point dans la partie Limites 5.2.

Corollaire I.2. *Si une information φ est utile pour un agent a par rapport à une requête Q, alors φ n'est ni une tautologie, ni une contradiction.*

Cela signifie que les contradictions et les tautologies ne sont pas considérées comme étant utiles.

Proposition I.4. *Soit $*$ un opérateur de révision de croyances satisfaisant les postulats AGM (postulats 1 à 4) [AGM85]. Bel_a représente l'ensemble des croyances de l'agent a dans un monde w et $Bel_a * \varphi$ l'ensemble des croyances de l'agent après avoir été révisé par φ en utilisant l'opérateur $*$. Si $U_a^Q \varphi$ est satisfaite dans w alors soit $Q \in Bel_a * \varphi$ soit $\neg Q \in Bel_a * \varphi$.*

Cette propriété montre que si l'agent décide d'intégrer l'information utile à sa base de croyances, alors il aura, dans sa base de croyances révisée, la réponse à son besoin en information. L'information utile comble donc le manque d'information.

Proposition I.5.

$$\vdash D_a Bif_a Q \to U_a^Q Q \otimes U_a^Q \neg Q$$

Si un agent a un besoin de savoir si l'information Q est vraie ou non, alors soit l'information Q est utile pour lui, soit l'information $\neg Q$ l'est. Cela signifie que les réponses directes au besoin en information sont des informations utiles.

Proposition I.6.

$$\vdash B_a(Q_1 \leftrightarrow Q_2) \to (U_a^{Q_1} \varphi \leftrightarrow U_a^{Q_2} \varphi)$$

Si deux informations Q_1 et Q_2 sont équivalentes pour un agent, alors toute information utile par rapport à Q_1 est utile par rapport à Q_2.

Proposition I.7.

$$\vdash U_a^Q \varphi \leftrightarrow U_a^{\neg Q} \varphi$$

Toute information utile par rapport à une requête Q est également utile par rapport à la requête $\neg Q$. En effet, savoir si Q est vrai est équivalent à savoir si $\neg Q$ est vrai.

Proposition I.8.

$$\vdash \neg(\varphi_1 \wedge \varphi_2) \rightarrow \neg(U_a^Q \varphi_1 \wedge U_a^Q \varphi_2)$$

Deux informations contradictoires ne peuvent pas être toutes deux utiles. Cette propriété est souhaitable car elle implique les informations utiles sont des informations qui sont cohérentes entre elles.

Proposition I.9.

$$\vdash \neg B_a U_a^Q \varphi$$

Les agents ne sont pas conscients des informations qui sont utiles pour eux. Cela vient du fait que les informations utiles sont des informations vraies. Supposons qu'un agent croit qu'une information est utile par rapport à une requête. Il croit donc que cette information est vraie. Or, les informations utiles sont des informations nouvelles (proposition I.2).

Cette propriété de non-conscience des informations utiles peut paraître au premier abord contre-intuitive pour le lecteur. Néanmoins, l'intuition que peut avoir le lecteur sur la conscience des informations utiles ne porte pas en fait sur les informations utiles mais sur les informations potentiellement utiles. Les informations potentiellement utiles pour un agent sont les informations qui, selon cet agent, peuvent l'aider pour répondre à son besoin en information mais pour lesquelles il ne connaît pas la valeur de vérité. Nous verrons dans la section 3.5 que les agents ont conscience des informations qui sont potentiellement utiles pour eux.

Proposition I.10.

$$\vdash (\varphi_1 \leftrightarrow \varphi_2) \wedge B_a(\varphi_1 \leftrightarrow \varphi_2) \rightarrow (U_a^Q \varphi_1 \leftrightarrow U_a^Q \varphi_2)$$

Si deux informations sont équivalentes et également équivalentes pour un agent, alors elles sont utiles pour cet agent de façon équivalente. Cette propriété montre que l'aspect syntaxique n'intervient pas dans la caractérisation de l'utilité. Autrement dit, c'est le contenu de l'information qui est jugé utile, indépendamment de la forme de celle-ci. Par exemple, si l'information *inc* est utile, alors l'information $(inc \wedge retard) \vee (inc \wedge \neg retard)$ l'est également. Nous reviendrons sur cette propriété lors de la comparaison de nos travaux avec ceux de la littérature.

Proposition I.11.

$$\vdash U_a^Q \varphi_1 \wedge \varphi_2 \wedge \neg B_a(\neg(\varphi_1 \wedge \varphi_2)) \rightarrow U_a^Q(\varphi_1 \wedge \varphi_2)$$

Si une information φ_1 est utile par rapport à une requête Q pour un agent a, qu'une information φ_2 est vraie et que l'agent a ne croit pas que les informations φ_1 et φ_2 soient contradictoires, alors la conjonction $\varphi_1 \wedge \varphi_2$ est une information utile pour l'agent a par rapport à sa requête Q.

Proposition I.12.

$$\vdash U_a^Q \varphi_1 \wedge U_a^Q \varphi_2 \wedge \neg B_a(\neg(\varphi_1 \wedge \varphi_2)) \rightarrow U_a^Q(\varphi_1 \vee \varphi_2)$$

Si deux informations φ_1 et φ_2 sont utiles pour un agent a et que celui-ci ne croit pas qu'elles soient contradictoires alors la disjonction $\varphi_1 \vee \varphi_2$ est utile pour cet agent.

Ces trois propositions montrent que beaucoup d'informations sont caractérisées comme étant utiles. Nous illustrons cela dans l'exemple qui suit.

Exemple I.3. Reprenons l'exemple d'un agent a qui veut prendre son train, train potentiellement en retard à cause d'un incident. L'agent a veut savoir si son train est en retard et on suppose que l'information *inc* est utile pour lui pour répondre à ce besoin.

Supposons que l'information « il pleut », modélisée par *pluie*, soit une information vraie, que l'agent a croit vraie. Alors, l'information $inc \wedge pluie$ est utile pour l'agent a. En effet, elle est vraie et contient l'élément *inc* qui permet à l'agent a de répondre à son besoin en information. Néanmoins, intuitivement, l'information *inc* est plus utile que $inc \wedge pluie$ car cette dernière contient *pluie* qui est un élément non-nécessaire à la résolution du besoin.

Il faut donc, parmi les informations utiles, déterminer quelles sont les informations les plus utiles. Pour cela, nous proposons de caractériser les informations les plus utiles.

3.3 Informations les plus utiles

Dans cette section, nous proposons une caractérisation des informations les plus utiles. Informellement, les informations les plus utiles sont les informations qui contiennent suffisamment d'éléments pour répondre au besoin en information mais qui ne contiennent pas ou peu d'élément non-nécessaires pour y répondre.

Toutes les informations que nous caractérisons par la définition I.10 sont utiles **de manière suffisante** car elles permettent de répondre complètement au besoin en information. Par contre, cette définition ne nous permet pas de caractériser le caractère **nécessaire** des informations utiles, c'est-à-dire les informations sans lesquelles on ne peut pas répondre au besoin en information. Les informations les plus utiles peuvent alors être vues comme les informations suffisantes et nécessaires pour la résolution du besoin en information. [14]

Notation I.5. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et w un monde de $\mathbb{M}$. On note $\mathcal{U}_a^Q$ l'ensemble des formules objectives utiles pour l'agent a par rapport à sa requête Q dans le monde w. [15]

Remarquons que l'ensemble $\mathcal{U}_a^Q$ ne contient ni tautologies ni contradictions (corollaire I.2).

Le critère que nous choisissons pour déterminer les informations les plus utiles est la **brièveté** des informations. En effet, le caractère bref d'une information peut être assimilé au caractère nécessaire que l'on souhaite définir pour une information utile : une information qui ne contient pas d'éléments non nécessaire est plus succincte qu'une information qui en contient. [16]

Nous adaptons alors la définition **d'explication minimale** proposée par Lakemeyer ([Lak97]).

Définition I.12 (Explication)**.** *Soient deux formules objectives φ et ψ. Soit a un agent de $\mathcal{A}$. Soient $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et w un monde de ce modèle.*
*ψ est une **explication** de φ pour l'agent a dans le monde w si et seulement si $B_a(\psi \to \varphi)$ est satisfaite dans le monde w et $B_a(\neg\psi)$ n'est pas satisfaite dans le monde w.*

Proposition I.13. *Soient Q et φ des formules objectives. Si φ appartient à l'ensemble $\mathcal{U}_a^Q$ (informations utiles dans w), alors φ est soit une explication de Q dans w, soit une explication de $\neg Q$ dans w.*

14. Notons qu'il serait possible d'utiliser une autre terminologie pour qualifier ces informations. Le terme d'information critique serait tout à fait acceptable.

15. Afin d'alléger les notations, le modèle $\mathbb{M}$ et le monde w n'apparaissent pas explicitement dans la notation.

16. Remarquons que des critères autres que la brièveté de l'information pourraient être pris en compte pour déterminer quelles sont les informations critiques ou les plus utiles.

Les informations utiles sont des explications de la requête (ou de sa négation).

Nous nous intéressons maintenant au caractère minimal d'une explication. Pour cela, nous reprenons des notions de [LLM03].

Définition I.13 (Forme Normale Négative)**.** *Soit φ une formule objective. φ est dans sa forme normale négative (NNF) si et seulement si l'opérateur de négation $\neg$ ne s'applique qu'à des symboles propositionnels. On note alors $Lit(\varphi)$ l'ensemble des littéraux*[17] *apparaissant dans la forme NNF de φ.*

Par exemple, la forme NNF de la formule φ égale à $\neg((\neg p \wedge q) \vee r)$ est $(p \vee \neg q) \wedge \neg r$. On a alors $Lit(\varphi) = \{p, \neg q, \neg r\}$.

Définition I.14 (Lit-dépendance / Lit-indépendance syntaxique)**.** *Soit φ une formule objective et l un littéral.*
*φ est **syntaxiquement Lit-dépendante** de l (respectivement syntaxiquement Lit-indépendante de l) si et seulement si $l \in Lit(\varphi)$ (respectivement $l \notin Lit(\varphi)$).*

Définition I.15 (Lit-dépendance / Lit-indépendance (sémantique))**.** *Soit φ une formule objective et l un littéral.*
*φ est (sémantiquement) **Lit-indépendante** de l si et seulement s'il existe une formule ψ telle que φ et ψ soient équivalentes et que ψ soit syntaxiquement Lit-indépendante de l. Dans le cas contraire, φ est (sémantiquement) Lit-dépendante de l.*
Étant donné un langage, $DepLit(\varphi)$ désigne l'ensemble des littéraux de ce langage dont φ est Lit-dépendante.

Exemple I.4. Soit φ la formule $p \wedge \neg q \wedge (p \vee q)$. On a $DepLit(\varphi) = \{p, \neg q\}$. Notons que φ est Lit-indépendante du littéral q car φ est équivalente à la formule $p \wedge \neg q$ dans laquelle q n'apparaît que sous sa forme négative.

Appliqué à une formule qui représente une information, l'ensemble *DepLit* caractérise les éléments de cette information qui sont nécessaires. Autrement dit, si on enlève un élément de *DepLit* à l'information, celle-ci n'a plus la même signification.

Définition I.16 (Explication minimale)**.** *Soient deux formules objectives φ et ψ. Soit a un agent de $\mathcal{A}$. Soient $\mathbb{M}$ un modèle et w un monde de ce modèle.*
*ψ est une **explication minimale** de φ dans le monde w si et seulement si ψ est une explication de φ dans w et qu'il n'existe pas d'explication ψ' de φ dans w telle que $DepLit(\psi') \subset DepLit(\psi)$.*

L'explication minimale est une généralisation de la subsomption pour les clauses et du dual de la subsomption pour les cubes. En effet étant données deux clauses φ et ψ qui sont toutes deux des explications d'une formule objective Q, si φ subsume ψ alors ψ ne peut pas être une explication minimale car dans ce cas $DepLit(\varphi) \subset DepLit(\psi)$. Les clauses minimales sont donc les clauses qui ne sont subsumées par aucune autre.

Définition I.17 (Informations les plus utiles)**.** *Soit $\mathcal{U}_a^Q$ l'ensemble des informations les plus utiles pour a par rapport à Q. On note $\mathcal{U}m_a^Q$ le sous-ensemble de $\mathcal{U}_a^Q$ qui contient les explications minimales de Q et les explications minimales de $\neg Q$*[18]*.*
On notera $Um_a^Q \varphi$ le fait que la formule objective φ soit une formule de l'ensemble $\mathcal{U}m_a^Q$.

17. Nous rappelons qu'un littéral est soit une variable propositionnelle, soit la négation d'une variable propositionnelle.

18. On laisse également implicite le modèle et le monde dans lesquels les informations sont les plus utiles.

Exemple I.5. Soit a un agent de $\mathcal{A}$. Considérons que l'ensemble des informations utiles dans un monde w_0 par rapport à une requête *retard* soit égal à $\{inc \wedge pluie, inc \vee greve, greve\}$. L'ensemble des informations les plus utiles dans w_0 est alors l'ensemble $\{greve, inc \wedge pluie\}$.

3.4 Comparaison avec les travaux de Lakemeyer

Lakemeyer propose dans l'article [Lak97] une caractérisation formelle de la pertinence. Plus précisément, il cherche à répondre à la question : quand peut-on dire qu'une information est pertinente pour un sujet d'intérêt ou qu'un ensemble d'informations est pertinent pour un sujet d'intérêt ? Son travail est un des plus reconnus dans le domaine de l'Intelligence Artificielle, domaine dans lequel la notion de pertinence a notamment été étudiée pour accélérer l'inférence dans les bases de connaissance (voir état de l'art 1.5.2). De plus, dans l'article [Lak97], de nombreuses comparaisons formelles avec d'autres définitions du domaine sont réalisées. Il est donc intéressant de situer formellement notre travail par rapport à celui de Lakemeyer.

Tout d'abord, rappelons quelques définitions et notations de l'article de Lakemeyer.

- une formule φ (resp. un ensemble de formules Δ) **mentionne** une variable propositionnelle p si et seulement si p ou si $\neg p$ apparaît dans φ (resp. Δ).
- Un **sujet d'intérêt** π est un ensemble de variables propositionnelles. Comme son nom l'indique, il représente un certain nombre de « centres d'intérêt ».
- On note π_φ (resp. π_Δ) l'ensemble des variables propositionnelles mentionnées dans φ (resp. Δ).
- Finalement, une formule **triviale** est une tautologie.

Pour éviter toute confusion entre la notion d'utilité avec laquelle nous travaillons et la définition de Lakemeyer, nous appelons la notion qu'il définit L-pertinence.

Pour répondre à la problématique posée, Lakemeyer utilise également une notion d'explication minimale [19]. Afin de ne pas introduire un nouveau formalisme, nous adaptons les définitions de Lakemeyer. Ainsi, l'agent pour lequel l'information est L-pertinente n'apparaît pas explicitement dans l'article d'origine (il est implicite à l'utilisation de la modalité de croyance B) ; ici, nous le faisons explicitement apparaître car notre modalité de croyance a l'agent comme index.

Définition I.18 (L-pertinence)**.** *Soit a un agent de $\mathcal{A}$ et soit φ une formule objective. Un sujet d'intérêt π est L-pertinent dans un monde w pour φ si et seulement s'il existe une explication minimale de φ pour l'agent a dans w qui ne soit pas triviale et qui mentionne une variable propositionnelle de π.*

Proposition I.14. *Soit a un agent de $\mathcal{A}$, soient φ et Q deux formules objectives. Si φ est une formule utile pour l'agent a par rapport à sa requête Q alors il existe une explication minimale de Q ou une explication minimale de $\neg Q$ qui mentionne une variable propositionnelle de φ. Autrement dit π_φ est L-pertinent pour Q ou pour $\neg Q$.*

Si une information est utile par rapport à un besoin, alors elle contient des éléments nécessaires pour répondre à la requête ou à sa négation. Cela signifie que l'utilité implique la L-pertinence.

La comparaison entre l'utilité et la L-pertinence est d'autant plus forte lorsque l'on considère

19. Dans [Lak97], l'agent n'apparaît pas aussi clairement que dans la définition que nous donnons de l'explication minimale. Cependant, Lakemeyer utilise des ensembles de formules qui peuvent tout à fait être assimilés à des ensembles de croyances d'agents.

les formules les plus utiles. En effet, alors que la L-pertinence exprime le fait qu'il existe une explication minimale, l'utilité réduite aux informations les plus utiles, caractérise directement quelles sont ces explications minimales.

Corollaire I.3. *Soit a un agent de $\mathcal{A}$, soient φ et Q deux formules objectives. Supposons que φ appartient à $\mathcal{U}m_a^Q$, c'est-à-dire que φ fait partie des informations les plus utiles pour l'agent a par rapport à sa requête Q (dans un monde w). Alors non seulement il existe une explication minimale de Q ou une explication minimale de $\neg Q$ dans w qui mentionne une variable propositionnelle de φ, mais φ est cette explication minimale de Q ou de $\neg Q$ dans w.*

Cela montre que si une information est utile de façon suffisante et nécessaire, non seulement alors elle est L-pertinente mais nous pouvons caractériser cette L-pertinence plus précisément que Lakemeyer car l'explication minimale à laquelle il fait référence est l'information que nous qualifions de plus utile.

3.5 Utilité potentielle

Dans cette partie, nous nous intéressons à une définition légèrement différente de l'utilité, que nous appelons utilité potentielle.

3.5.1 Définition formelle

L'utilité potentielle ressemble énormément à l'utilité sur laquelle nous avons travaillé jusque maintenant (définition I.10). La différence entre les deux est que l'utilité potentielle ne prend pas en compte la valeur de vérité de l'information. La présence de la valeur de vérité dans la définition de l'utilité, bien qu'étant justifiée dans la littérature ([SW04, Flo07, Gri75]), nous a souvent été reprochée. Nous maintenons que seules les informations vraies sont des informations utiles. Néanmoins, la notion d'utilité sans la valeur de vérité possède des propriétés intéressantes qui méritent d'être notées. De plus, cette notion nous servira dans une des sections portant sur la coopération (section 4.2.3).

Notation I.6. Soit a un agent de $\mathcal{A}$, soit φ une formule et soit Q une formule objective. On note $P_a^Q\varphi$ la formule

$$D_a Bif_a Q \wedge (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))$$

$P_a^Q\varphi$ se lit **l'information φ est potentiellement utile pour l'agent a par rapport à sa requête Q.**

Définition I.19 (Utilité potentielle)**.** *Soient a un agent de $\mathcal{A}$, Q une formule objective et φ une formule. Soient $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et w un monde de ce modèle. L'information φ est **potentiellement utile** pour l'agent a par rapport à la requête Q dans le monde w si et seulement si on a :*

$$\mathbb{M}, w \vDash P_a^Q\varphi$$

Cette définition de l'utilité contient deux des trois éléments qui apparaissent dans la définition de l'utilité : le besoin en information et les croyances de l'agent. Nous ne redonnons pas une description de ces éléments, elle serait identique à celle donnée en commentaire de la définition I.10.

L'utilité potentielle peut être vue comme le résultat d'un raisonnement de l'agent sur ses croyances et ses désirs. En effet, les informations potentiellement utiles sont les informations à partir desquelles un agent pense pouvoir répondre à son besoin en information. Cependant, comme il n'y a aucune condition de vérité sur les informations potentiellement utiles, l'agent ne peut pas les utiliser pour répondre à son besoin en information. En effet, il peut potentiellement tirer une fausse conclusion de ces informations.

Illustrons cette notion d'utilité potentielle sur un exemple.

Exemple I.6. Reprenons l'exemple des agents a et b qui doivent prendre leur train. Ils souhaitent tous deux savoir si leur train est en retard (modélisé par *retard*). On se place dans le monde w_0 du modèle $\mathbb{M}$.
L'agent a croit que s'il y a un incident alors son train est en retard et il ne croit rien d'autre sur le retard du train, l'incident ou la relation entre les deux informations. Nous avons donc :
- $\mathbb{M}, w_0 \vDash D_a Bif_a retard$
- $\mathbb{M}, w_0 \vDash B_a(inc \to retard)$
- $\mathbb{M}, w_0 \vDash \neg B_a(inc \to \neg retard)$

On peut en déduire que $\mathbb{M}, w_0 \vDash P_a^{retard} inc$, c'est-à-dire que l'information *inc* est une information potentiellement utile pour l'agent a par rapport à sa requête *retard* dans w_0.
L'agent b croit que s'il n'y a pas d'incident, alors son train n'est pas en retard (nous supposons qu'il ne croit rien d'autre). Nous avons donc :
- $\mathbb{M}, w_0 \vDash D_b Bif_b retard$
- $\mathbb{M}, w_0 \vDash B_b(\neg inc \to \neg retard)$
- $\mathbb{M}, w_0 \vDash \neg B_b(\neg inc \to retard)$

On peut en déduire que $\mathbb{M}, w_0 \vDash P_b^{retard} \neg inc$, c'est-à-dire que l'information *inc* est une information potentiellement utile pour l'agent b par rapport à sa requête *retard* dans w_0.
Les deux informations *inc* et $\neg inc$ sont potentiellement utiles respectivement pour les agents a et b. Les agents ne peuvent pourtant pas utiliser ces informations pour répondre à leur besoin : s'ils le faisaient, un des agents tirerait une conclusion fausse sur le retard de son train.

3.5.2 Propriétés

Dans cette partie, nous étudions quelques-unes des propriétés des informations potentiellement utiles. Nous considérons un agent a de $\mathcal{A}$, Q une formule objective et φ une formule.

Proposition I.15.

$$\vdash D_a Bif_a Q \to P_a^Q Q \wedge P_a^Q \neg Q$$

Les réponses directes au besoin en information sont toutes deux des informations potentiellement utiles pour l'agent. Pour les informations utiles, seule une des réponses directes est utile : la réponse directe qui est vraie.

Proposition I.16.

$$\vdash P_a^Q \varphi \to \neg B_a \varphi \wedge \neg B_a \neg \varphi$$

Comme pour les informations utiles, les informations potentiellement utiles sont des informations nouvelles pour l'agent qui ne sont pas en contradiction avec sa base de croyances.

Proposition I.17.

$$\vdash P_a^Q \varphi \to B_a P_a^Q \varphi$$

Les agents sont conscients des informations qui sont potentiellement utiles pour eux. Cela vient du fait que les agents sont conscients de leur croyances, de leur non-croyances et de leurs désirs. Les agents sont donc conscients des informations qui sont susceptibles de résoudre leur besoin en information. Cette proposition constitue la différence majeure avec la définition I.10.

Exemple I.7. Reprenons l'exemple précédent. L'agent a est conscient que l'information *inc* est potentiellement utile pour lui pour sa requête *retard*. De même, l'agent b est conscient que l'information $\neg inc$ est potentiellement utile pour lui pour sa requête *retard*. Seulement, aucun des deux agents ne peut utiliser ces informations car ils ne savent pas si elles sont vraies ou non.

En comparaison aux informations potentiellement utiles, les informations que nous avons qualifiées d'utiles peuvent être vues comme des informations utiles contingentes [20]. La contingence des informations utiles vient dans ce cas de la valeur de vérité de l'information.

3.5.3 Information potentiellement utile maximale

Finalement, il est également possible d'extraire de l'ensemble des informations potentiellement utiles l'ensemble des informations potentiellement utiles maximales.

Définition I.20 (Information potentiellement utile maximale)**.** *Soit a un agent de $\mathcal{A}$. Soient Q et φ deux formules objectives. On note $\mathcal{P}_a^Q$ l'ensemble des formules objectives potentiellement utiles pour l'agent a par rapport à sa requête Q. On note $\mathcal{P}m_a^Q$ le sous-ensemble de $\mathcal{P}m_a^Q$ qui contient les informations les plus potentiellement utiles (au sens de l'explication minimale). On note par $Pm_a^Q\varphi$ le fait que l'information φ fasse partie de l'ensemble $\mathcal{P}m_a^Q$.*

3.6 Généralisation du besoin en information

Cette section adresse une extension que nous pouvons apporter à la définition I.10 de l'utilité : la généralisation du besoin en information.

Jusqu'à présent, nous nous sommes intéressés à des besoins en information pouvant s'écrire sous la forme « l'agent a a besoin de savoir si Q est vrai, Q étant une formule objective ». Dans cette partie, nous nous intéressons à un besoin pouvant s'écrire sous la forme « l'agent a a besoin de savoir Q_1 ou savoir Q_2 ... ou savoir Q_n ou savoir qu'aucun des $Q_1 \ldots Q_n$ n'est vrai », Q_1, $\ldots Q_n$ étant n formules que l'agent a croit mutuellement exclusives.

3.6.1 Définition

Définition I.21 (Formules mutuellement exclusives pour un agent dans un monde)**.** *Soit a un agent de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle de la classe* **C** *et soit w un monde de ce modèle. Soit n un entier naturel supérieur ou égal à 2. Soient $Q_1, \ldots Q_n$ n formules objectives.*
Les formules $Q_1, \ldots Q_n$ sont ***mutuellement exclusives pour l'agent a dans le monde w*** *si et seulement si pour tout entier i et pour tout entier j tels que $1 \leq i, j \leq n$ et $i \neq j$, on a* $\mathbb{M}, \boldsymbol{w} \vDash \boldsymbol{B_a}\neg(\boldsymbol{Q_i} \wedge \boldsymbol{Q_j})$.

20. Nous remercions Robert Demolombe pour sa suggestion des termes « contingent » et « potentiel » pour qualifier les deux notions d'utilité.

Notation I.7. Soit a un agent de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w un monde de ce modèle. Soit n un entier supérieur ou égal à 1. Soit $\mathbf{Q} = \{Q_1, ..., Q_n\}$ un ensemble de n formules objectives. Si $n > 1$, alors nous supposons que ces formules sont mutuellement exclusives pour l'agent a dans w. Soit φ une formule objective. On note $\boldsymbol{U_a^{\mathbf{Q}}\varphi}$ la formule

$$D_a\left(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i\right) \wedge \left[\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)\right] \wedge \varphi$$

Cette formule se lit **l'information φ est utile pour l'ensemble de requêtes $\mathbf{Q}$**.

Définition I.22 (Utilité - besoin en information généralisé)**.** *Soit a un agent de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w un monde de ce modèle. Soit n un entier supérieur ou égal à 1. Soit $\mathbf{Q} = \{Q_1, ..., Q_n\}$ un ensemble de n formules objectives mutuellement exclusives pour l'agent a dans le monde w. Soit φ une formule objective. La formule φ est utile pour l'agent a par rapport à l'ensemble de requêtes $\mathbf{Q}$ dans le monde w si et seulement si on a :*

$$\mathbb{M}, w \models U_a^{\mathbf{Q}}\varphi$$

Le besoin en information étant plus général que dans la définition I.10, on retrouve les mêmes trois éléments dans cette définition de l'utilité mais leur expression est plus complexe.

1. **le besoin en information**
   Le besoin en information est du type « l'agent a veut savoir si Q_1 ou ...ou Q_n ou si aucun des Q_i n'est vrai. ». Par exemple, un agent veut savoir si son train part du quai 1 ou du quai 2, ou si celui-ci ne part d'aucun des deux (par exemple si le train est annulé). Le besoin peut donc être divisé en deux parties : désirer croire un des Q_i ou croire qu'ils sont tous faux. Ceci est modélisé par la formule $\boldsymbol{D_a\left(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i\right)}$.
2. **les croyances de l'agent**
   La réponse utile doit permettre à l'agent de répondre à son besoin en information. Le besoin en information est divisé en deux parties. Les informations utiles vont donc être de deux types :
   - celles qui permettent de déduire un des Q_i. Ce sont les formules φ telles que $\boldsymbol{\bigvee_{i=1}^{n} B_a(\varphi \to Q_i)}$.
   - celles qui permettent de déduire que tous les Q_i sont faux. Ce sont les formules φ telles que $\boldsymbol{\bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)}$ [21].
3. **la valeur de vérité de l'information**
   Comme précédemment, seules les informations vraies sont utiles.

Exemple I.8. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w_0 un monde de ce modèle. Dans ce monde w_0, a est un agent qui veut prendre son train et qui a besoin de savoir de quel quai part son train.
Il croit qu'il y a 3 quais dans la gare et désire donc savoir si le train partira du quai 1, du quai 2 ou du quai 3 ou si le train ne part d'aucun de ces trois quais. Nous modélisons par q_1 (resp. q_2 et q_3) le fait que le train parte du quai 1 (resp. quai 2 et quai 3) et nous appelons $\mathbf{Q}$ l'ensemble $\{q_1, q_2, q_3\}$. Le désir de l'agent s'exprime formellement par $D_a(\bigvee_i B_aq_i \vee \bigwedge_i B_a\neg q_i)$.
Nous supposons que les seules croyances que cet agent possède sur le train qu'il doit prendre est que c'est un TGV (modélisé par tgv). En particulier, nous supposons que l'agent n'a aucune croyance (positive ou négative) sur le quai de départ de son train, c'est-à-dire que l'on a

21. On peut noter que cette formule est équivalente à $\boldsymbol{B_a(\varphi \to \bigwedge_{i=1}^{n} \neg Q_i)}$

$\mathbb{M}, w_0 \vDash \bigwedge_i \neg Bif_a q_i$.

1. Supposons que l'information exprimant que si le train est un TGV alors il part du quai 1, modélisée par $tgv \to q_1$ soit une information vraie dans le monde w_0. On a donc :
   – $\mathbb{M}, w_0 \vDash D_a(\bigvee_i B_a q_i \vee \bigwedge_i B_a \neg q_i)$
   – $\mathbb{M}, w_0 \vDash B_a tgv$
   – $\mathbb{M}, w_0 \vDash B_a((tgv \to q_1) \to q_1)$ [22]
   – $\mathbb{M}, w_0 \vDash \neg B_a((tgv \to q_1) \to \neg q_1)$ [23]
   – $\mathbb{M}, w_0 \vDash tgv \to q_1$
   On peut donc en déduire que $\mathbb{M}, w_0 \vDash U_a^{\mathbf{Q}}(tgv \to q_1)$, c'est-à-dire que l'information $tgv \to q_1$ est une information utile pour l'agent a par rapport à son ensemble de requêtes $\mathbf{Q}$ dans le monde w_0.
2. Supposons que l'information stipulant que le train est annulé, modélisée par *annule* soit une information vraie dans w_0. Supposons également que l'agent croit que si le train est annulé, alors il ne part d'aucun des quais et que l'agent n'a aucune croyance sur le fait que le train soit annule ou non. On a donc :
   – $\mathbb{M}, w_0 \vDash D_a(\bigvee_i B_a q_i \vee \bigwedge_i B_a \neg q_i)$
   – $\mathbb{M}, w_0 \vDash annule$
   – $\mathbb{M}, w_0 \vDash \bigwedge_{i=1}^{3} B_a(annule \to \neg q_i)$
   – $\mathbb{M}, w_0 \vDash \neg B_a(annule \to q_1)$ [24]
   On peut donc déduire que $\mathbb{M}, w_0 \vDash U_a^{\mathbf{Q}} annule$, c'est-à-dire que l'information *annule* est une information utile pour l'agent a par rapport à son ensemble de requêtes $\mathbf{Q}$ dans le monde w_0.

Notation I.8 (Utilité simplifiée). Avec les hypothèses de la définition I.22, on note $Us_a^{\mathbf{Q}}$ la formule

$$D_a\left(\bigvee_{i=1}^{n} B_a Q_i\right) \wedge \bigotimes_{i=1}^{n} B_a(\varphi \to Q_i) \wedge \varphi$$

Dans le cas où l'agent a croit non seulement que les formules Q_i sont mutuellement exclusives mais aussi qu'au moins une de ces formules Q_i est vraie, la définition de l'utilité se simplifie grandement, comme le montre la proposition suivante.

Proposition I.18. *Soit a un agent de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et soit w un monde de ce modèle. Soit n un entier naturel supérieur ou égal à 1. Soient $Q_1, \ldots Q_n$ n formules mutuellement exclusives pour l'agent a dans le monde w. On note $\mathbf{Q}$ l'ensemble $\{Q_1, \ldots Q_n\}$. On a alors :*

$$\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n} Q_i \to \left[Us_a^{\mathbf{Q}}\varphi \leftrightarrow U_a^{\mathbf{Q}}\varphi\right]$$

Remarquons que dans le cas où la croyance de l'agent qu'au moins un des Q_i est vrai est une croyance erronée, alors il n'y a pas d'équivalence entre l'utilité et sa version simplifiée. Les informations utiles sont celles caractérisées par la définition I.22.

22. En effet, $(tgv \to q_1) \to q_1$ est équivalent à $tgv \vee q_1$. Or, on a $\mathbf{M}, w_0 \vDash B_a(tgv)$. Nous pouvons donc déduire que $\mathbf{M}, w_0 \vDash B_a(tgv \vee q_1)$ et donc que $\mathbf{M}, w_0 \vDash B_a((tgv \to q_1) \to q_1)$.
23. En effet, $(tgv \to q_1) \to \neg q_1$ est équivalent à $(tgv \vee \neg q_1) \wedge \neg q_1$. Par hypothèse, $\mathbb{M}, w_0 \vDash \neg B_a(\neg q_1)$. On a donc $\mathbb{M}, w_0 \vDash \neg B_a(\neg q_1 \wedge (tgv \vee \neg q_1))$.
24. En effet, si $\mathbb{M}, w_0 \vDash B_a(annule \to q_1)$, alors comme $\mathbb{M}, w_0 \vDash B_a(annule \to \neg q_1)$, on a $\mathbb{M}, w_0 \vDash B_a \neg annule$. C'est en contradiction avec l'hypothèse que l'agent ne sait pas si le train est annulé.

Corollaire I.4. *Soit a un agent de $\mathcal{A}$. Soit n un entier naturel supérieur ou égal à 1. Soit w un monde d'un modèle $\mathbb{M}$ de la classe* **C***.*
Soient Q_1, ..., Q_n n formules mutuellement exclusives pour l'agent a dans w.
On note **Q** *l'ensemble $\{Q_1, ...Q_n\}$. On a alors :*

$$\textit{Si} \vdash \bigvee_{i=1}^{n} Q_i \textit{ alors } \mathbb{M}, w \vDash Us_a^{\mathbf{Q}}\varphi \leftrightarrow U_a^{\mathbf{Q}}\varphi$$

Comme le montre le corollaire suivant, nous avons deux manières de nous ramener à l'utilité que nous avons étudiée en section 3.1.

Corollaire I.5. *Soient φ une formule et Q une formule objective. La formule φ est utile pour l'agent a par rapport à l'ensemble de requêtes $\{Q, \neg Q\}$ si et seulement si elle est utile pour l'agent a par rapport à sa requête Q ou encore si et seulement si elle est utile pour l'agent par rapport à l'ensemble de requêtes $\{Q\}$.*
Formellement, cela signifie que

$$\vdash U_a^{Q}\varphi \leftrightarrow U_a^{\{Q,\neg Q\}}\varphi \leftrightarrow U_a^{\{Q\}}\varphi$$

En effet, une première manière est l'utilisation de la définition I.22 avec $n = 1$. Nous pouvons également appliquer le corollaire précédent avec l'ensemble $\{Q, \neg Q\}$.

3.6.2 Propriétés

La définition de l'utilité pour un besoin en information généralisé possède des propriétés relativement similaires à celle définie dans la définition I.10. Nous en détaillons quelques-unes.

Pour chacune de ces propriétés, n est un entier naturel supérieur ou égal à 1, a est un agent de $\mathcal{A}$, $\mathbb{M}$ est un modèle et w un monde de ce modèle, **Q** est un ensemble de n formules objectives mutuellement exclusives pour l'agent a dans le monde w notées Q_1, ... Q_n, et φ, φ_1 et φ_2 sont des formules.

Proposition I.19.

$$\vdash D_a\left(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i\right) \to \bigwedge_{i=1}^{n} \neg B_aQ_i \wedge \bigvee_{i=1}^{n} \neg B_a\neg Q_i$$

Cette proposition montre que le besoin en information, bien que généralisé, correspond toujours à un manque d'information : si un agent désire savoir si Q_1 ou Q_2 ... ou Q_n ou si aucun des Q_i n'est vrai, c'est qu'il ne croit aucune de ces informations. Cela signifie également qu'il ne croit pas que tous les Q_i sont faux.

Proposition I.20.

$$\vdash U_a^{\mathbf{Q}}\varphi \to \neg B_a\varphi$$

Si une information est utile, alors elle est nouvelle pour l'agent. Si ce n'était pas le cas, alors l'agent serait en mesure de répondre à son besoin en information, qui n'en serait donc pas vraiment un.

Proposition I.21.

$$\vdash U_a^{\mathbf{Q}}\varphi \rightarrow \neg B_a \neg\varphi$$

Si une information est utile, alors elle n'est pas en contradiction avec ce que croit l'agent.

Proposition I.22.

$$\vdash \left(\bigotimes_{i=1}^{n} Q_i \wedge \bigwedge_{i=1}^{n} \neg B_a(\neg Q_i) \wedge D_a(\bigvee_{i=1}^{n} B_a Q_i \vee \bigwedge_{i=1}^{n} B_a \neg Q_i)\right) \rightarrow \bigotimes_{i=1}^{n} U_a^{\mathbf{Q}} Q_i$$

Cette proposition montre que si au moins un des Q_i est vrai ($\bigotimes_{i=1}^{n} Q_i$) et que l'agent ne croit pas que les Q_i sont faux, alors s'il a le besoin de savoir si Q_1 ou ... ou Q_n ou s'il sont tous faux alors un et un seul des Q_i est utile pour l'agent pour l'ensemble de requêtes $\mathbf{Q}$.

Proposition I.23.

$$\vdash \neg B_a U_a^{\mathbf{Q}}\varphi$$

L'agent n'a pas conscience des informations qui sont utiles pour lui.

Proposition I.24.

$$\mathbb{M}, w \models B_a \bigvee_{i=1}^{n+1} Q_i \rightarrow (U_a^{\{Q_1,\dots Q_{n+1}\}}\varphi \leftrightarrow U_a^{\{Q_1,\dots Q_n\}}\varphi)$$

Toute information utile pour un agent a par rapport au désir de savoir que Q_1, ou $\dots$, ou Q_n ou qu'ils sont tous faux est une information utile pour lui par rapport au désir de savoir que Q_1, ou $\dots$ ou Q_{n+1} ou qu'ils sont tous faux, sous réserve que celui-ci croit que la disjonction de Q_1, ... Q_n, Q_{n+1} est vraie. Par exemple, si l'agent a croit que q_1 ou q_2 ou q_3 ou *annule* est vraie, alors les informations utiles pour $\{q_1, q_2, q_3\}$ sont également les informations utiles pour $\{q_1, q_2, q_3, \mathit{annule}\}$.

3.7 Conclusion intermédiaire

À ce stade, il est intéressant de comparer les propriétés de la définition de l'utilité avec celles que nous nous sommes fixées en objectif en conclusion de l'état de l'art.

1. *Une information utile pour un agent répond à un besoin de celui-ci.* Le besoin est en effet partie intégrante de la définition de l'utilité.
2. *L'information utile est réellement connectée au besoin de l'agent.* Dans la caractérisation que nous proposons, l'information utile est connectée via l'implication matérielle au besoin dans la base de croyances de l'agent. Nous revenons dans la partie 5.2 sur les limites que ce choix entraîne.
3. *Plus une information répond à un besoin, plus elle est utile pour ce besoin.* Les informations utiles pour un besoin en information répondent complètement à ce besoin. Ainsi, la caractérisation des informations les plus utiles ne se définit pas en terme de réponse plus ou moins complète au besoin.
4. *Une information utile pour un agent est une information vraie.* La vérité de l'information utile est un élément qui apparaît dans la caractérisation formelle.

5. *Plus une information est facile à comprendre (au sens complexité de l'information), plus elle est utile.* La facilité de compréhension de l'information fait intervenir de nombreux critères qui sont basés sur la signification de l'information mais aussi sur la forme de celle-ci. La caractérisation des informations les plus utiles que nous proposons est uniquement basée sur la signification des informations. En effet, l'ensemble *DepLit*, qui permet de juger de la brièveté des informations, ne tient pas compte de leur aspect syntaxique : deux formules équivalentes ont le même *DepLit*. Par exemple, la formule $p \wedge (q \vee r)$ est équivalente à $p \wedge (q \vee r \vee \neg p) \wedge (q \vee r) \wedge p$. Or, la première est plus brève (syntaxiquement) que la deuxième. Nous reviendrons sur ce point lorsque nous vérifierons l'adéquation de nos travaux sur la coopération avec le principe de Grice (section 4.4). Nous prenons donc en compte certains des critères de facilité de compréhension mais pas ceux qui sont basés sur la forme de l'information.
6. *une information utile pour un agent dépend des croyances de celui-ci et l'utilité de l'information peut changer en fonction de l'évolution des croyances de l'agent.* Les croyances de l'agent et ce qu'il peut déduire sont pris en considération dans la définition de l'utilité. L'utilité de l'information à un instant pour un agent dépend des croyances de celui-ci à cet instant. Si ces dernières évoluent, alors les informations utiles ne seront plus les mêmes.

Chapitre 4

Agent coopératif

Nous revenons maintenant sur la notion de coopération dans les systèmes multi-agents. Nous avons vu en conclusion de l'état de l'art qu'un agent peut être considéré comme **coopératif** s'il transmet aux autres les informations qui sont utiles ou tout du moins qu'il pense être utiles pour eux. L'utilité que nous avons définie est une utilité pour un besoin en information. Ainsi, la coopération que nous définissons dans ce chapitre est également une coopération pour un besoin en information.

Dans une première section, nous commençons par introduire deux nouveaux opérateurs dans notre cadre logique. Nous nous appuyons alors sur la première définition de l'utilité que nous avons formulée dans la partie précédente pour proposer plusieurs définitions du caractère coopératif d'un agent vis-à-vis d'un autre agent. Dans la section 4.2, nous proposons une première définition de la coopération qui prend en compte les croyances de celui qui envoie l'information et que nous appelons « coopération subjective ». Puis, nous proposons dans la section 4.3 une deuxième définition dans laquelle ces croyances sont ignorées. Nous appelons alors cette coopération la « coopération objective ». La section 4.4 adresse alors une comparaison entre la coopération selon Grice (section 1 de l'état de l'art) et une des définitions que nous proposons.

4.1 Extensions du cadre formel

4.1.1 Un agent informe un autre agent que

Afin de pouvoir modéliser l'échange d'information entre les agents, nous introduisons un nouvel opérateur, défini par Demolombe [Dem04].

Définition I.23 (Opérateur non-normal pour le fait d'informer)**.** *À chaque couple d'agents a et b de $\mathcal{A}$, on associe l'opérateur modal $Inf_{a,b}$. Étant donnée une formule φ, $Inf_{a,b}\varphi$ est également une formule du langage et se lit « l'agent a informe l'agent b de l'information φ. ».*
La seule règle associée à cet opérateur est la règle de substitution de formules équivalentes : soient φ et ψ deux formules,

$$\text{(InfRE)}\ \frac{\varphi \leftrightarrow \psi}{Inf_{a,b}\varphi \leftrightarrow Inf_{a,b}\psi}$$

Nous voulons que l'opérateur Informer soit le plus simple possible. Ainsi, nous préférons utiliser un opérateur non-normal.

Ce choix implique qu'il n'est pas possible d'utiliser une sémantique classique basée sur des structures de Kripke. Nous utilisons donc des fonctions de voisinage pour la sémantique de cet opérateur. Plus précisément, il faut compléter la notion de modèle vue dans la partie 2.2.3 avec un ensemble de fonctions de voisinage $N^{inf}_{a,b}$ pour chaque couple d'agents (a,b) de $\mathcal{A}^2$.

Définition I.24. *On définit alors la satisfaisabilité des nouvelles formules de la façon suivante :*

$$\mathbb{M}, w \models Inf_{b,a}\varphi \textit{ ssi } (\varphi)^{\mathbb{M}} \in N^{inf}_{a,b}(w)$$

Proposition I.25. *Le système de preuves complété avec l'axiome (InfRE) est valide et complet pour la classe de modèles étendue* **C**.

4.1.2 Croyance distribuée

On reprend ici la notion de croyance distribuée décrite dans [HL09].

Définition I.25. *À chaque sous-ensemble I de $\mathcal{A}$, on associe un opérateur DB_I. Étant donnée une formule φ, $DB_I\varphi$ est également une formule et se lit : « les agents de l'ensemble I ont la croyance distribuée que φ ».*
À cette modalité, nous associons le schéma d'axiomes et règles présentés sur la figure 4.1. I est alors un sous-ensemble de $\mathcal{A}$.

(DBK)	$DB_I\varphi \wedge DB_I(\varphi \rightarrow \psi) \rightarrow DB_I\psi$
(DB4)	$DB_I\varphi \rightarrow DB_I DB_I\varphi$
(DB5)	$\neg DB_I\varphi \rightarrow DB_I\neg DB_I\varphi$
(DB-Int)	$B_a\varphi \leftrightarrow DB_{\{a\}}\varphi$
(DB-Mon)	$DB_J\varphi \rightarrow DB_I\varphi$ si $J \subseteq I$
(DB-Nec)	$\dfrac{\varphi}{DB_I\varphi}$

FIGURE 4.1 – Schéma d'axiomes et règles d'inférence de la croyance distribuée

Au niveau de la sémantique, il faut étendre la notion de modèle avec un ensemble de relations R^{DB}_I pour chaque sous-ensemble I de $\mathcal{A}$.

Définition I.26. *Pour chaque couple de mondes (w,v) de W^2, on a $wR^{DB}_I v$ ssi pour tout a de I on a $wR_a v$. On définit alors la satisfaisabilité de φ dans le monde w par*
$\mathbb{M}, w \models DB_I\varphi$ ssi pour tout $v \in W$ tel que $wR^{DB}_I v$, on a $\mathbb{M}, v \models \varphi$.

L'exemple qui suit illustre la relation entre croyances individuelles et croyances distribuées.

Exemple I.9. Soient a et b deux agents de $\mathcal{A}$. Soient $\mathbb{M}$ un modèle et w_0, w_1, w_2, w_3 et w_4 des mondes de ce modèle. Supposons que :
- $\mathcal{M}, w_0 \models B_a\varphi \wedge \neg Bif_a\psi$
- $\mathcal{M}, w_0 \models B_b\psi \wedge \neg Bif_a\varphi$

La figure 4.2 présente alors les relations qui existent entre les différents mondes.

Proposition I.26. *La théorie de la preuve complétée avec le schéma d'axiomes et la règle d'inférence est valide et complète pour la classe de modèles* **C** *étendue avec l'ensemble de relations R^{DB}_I.*

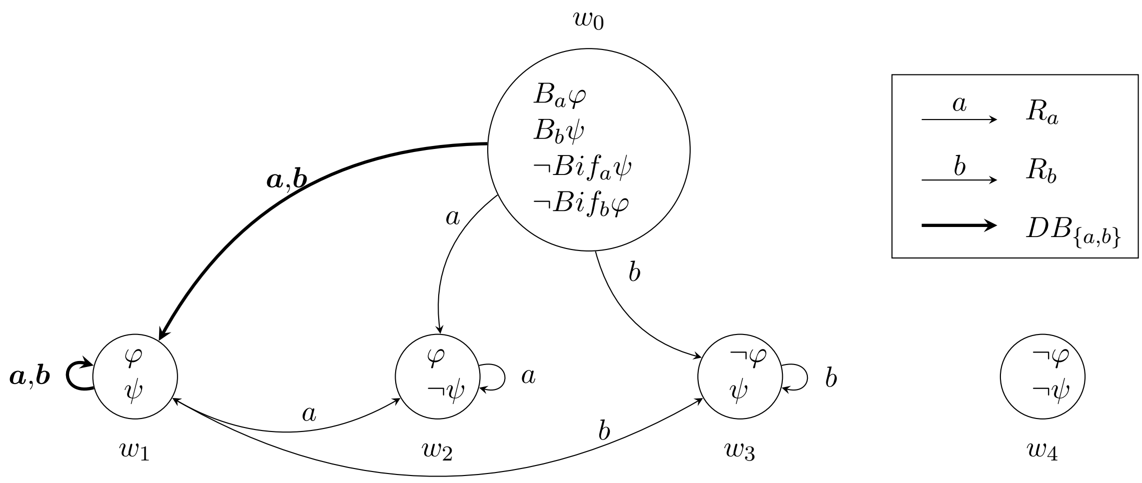

FIGURE 4.2 – Relation croyances individuelles - croyances distribuées

Notons qu'il est possible d'avoir $DB_I\bot$ pour un sous-ensemble contenant au moins deux agents. Cela correspond au cas où les agents n'ont pas de croyance distribuée, autrement dit au cas où ils ont des croyances contradictoires.

Cette notion de croyance distribuée permet de modéliser le cas où l'information est distribuée entre plusieurs agents : si l'agent a croit φ et que l'agent b croit ψ alors $\varphi \wedge \psi$ est une croyance distribuée de l'ensemble $\{a, b\}$. Remarquons que si φ et ψ sont contradictoires, alors on retombe sur $DB_{\{a,b\}}\bot$.

4.2 Coopération subjective

Dans cette partie, nous proposons une définition de la coopération qui tient compte des croyances de l'agent qui transmet l'information. Pour cette définition, nous avons besoin de formaliser le fait qu'un agent croit qu'une information fait partie des informations les plus utiles pour un autre agent. Cette notion est légèrement différente de la notion d'utilité maximale vue dans la section 3.3 car c'est un agent qui juge de l'utilité maximale de l'information pour un autre. Nous commençons donc par adapter la notion d'utilité maximale.

Nous donnons ensuite une version simple de cette coopération, puis nous en proposons des extensions.

4.2.1 Les informations les plus utiles pour l'agent a selon l'agent b

Parmi les différentes croyances d'un agent, certaines peuvent concerner les autres agents. Par exemple, Alice croit que Bernard croit qu'il pleut ou encore Alice croit que Bernard a le désir de savoir s'il fait beau, etc.

Dans ce paragraphe, nous nous intéressons aux croyances des agents qui portent sur l'utilité des informations pour d'autres agents. Plus précisément, nous caractérisons formellement les formules qu'un agent juge les plus utiles pour un autre agent.

Notation I.9. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et w un monde de ce modèle. On note $\mathcal{B}^Q_{a/b}(w)$ l'ensemble des formules objectives φ telles que $\mathbb{M}, w \vDash B_b U^Q_a \varphi$.

Pour alléger l'écriture, lorsqu'il n'y a pas d'ambiguïté sur le monde dans lequel on se place, on note cet ensemble $\mathcal{B}^{Q}_{a/b}$.

L'ensemble $\mathcal{B}^{Q}_{a/b}$ représente donc l'ensemble des formules que b juge être utiles pour l'agent a. Si l'on reprend la définition I.10 de l'utilité, les formules que l'agent b juge utile pour l'agent a sont les formules φ telles que :

- $B_b I_a Bif_a Q$: b croit que a a besoin de savoir si Q ;
- $B_b(B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))$: b croit que a peut répondre à son besoin Q à partir de l'information φ ;
- $B_b\varphi$: b croit que l'information φ est vraie.

Dans cette thèse, nous ne nous sommes pas intéressés au mécanisme par lequel les agents pouvaient obtenir de telles croyances sur les autres agents. En effet, nous ne nous intérrogeons pas sur la provenance des croyances des agents. Cependant, nous proposons quelques pistes sur ce sujet dans les perspectives de cette partie (section 5.3.1).

Définition I.27 (Réponse selon un agent). *Soient deux formules objectives φ et ψ. Soient a et b deux agents de $\mathcal{A}$. Soient $\mathbb{M}$ un modèle de la classe $\mathbf{C}$ et w un monde de ce modèle.*
*φ est une **réponse** à ψ pour l'agent a selon l'agent b dans le monde w si et seulement si $\mathbb{M}, w \models B_b(B_a(\varphi \to \psi) \vee B_a(\varphi \to \neg\psi))$ et $\mathbb{M}, w \models B_b \neg B_a(\neg\varphi)$.*

Les réponses à une formule pour a selon b sont les formules φ telles que b croit qu'à partir de φ, a pourra déduire ψ ou déduire $\neg\psi$. A priori, b n'est pas en mesure de savoir lequel de ψ ou $\neg\psi$ l'agent a pourra déduire.

Proposition I.27. *Soient Q et φ des formules objectives. Si φ appartient à l'ensemble $\mathcal{B}^{Q}_{a/b}$, alors φ est une réponse à Q pour a selon b.*

Définition I.28 (Réponse minimale selon un agent). *Soient φ et ψ deux formules objectives. Soit a un agent de $\mathcal{A}$. Soient $\mathbb{M}$ un modèle et w un monde de ce modèle.*
*φ est une **réponse minimale** de ψ pour l'agent a selon l'agent b dans le monde w si et seulement si φ est une réponse à ψ pour a selon b dans w et qu'il n'existe pas d'explication φ' de ψ pour a selon b dans w telle que $DepLit(\varphi') \subset DepLit(\varphi)$.*

Définition I.29 (Informations les plus utiles). *Soit $\mathcal{B}^{Q}_{a/b}$ l'ensemble des informations jugées utiles pour a selon b. On note $\mathcal{B}m^{Q}_{a/b}$ le sous-ensemble de $\mathcal{B}^{Q}_{a/b}$ qui contient les réponses minimales à Q pour a selon b[25]. Cet ensemble est appelé l'ensemble des informations les plus utiles pour a selon b.*
On notera $Bm^{Q}_{a/b}\varphi$ le fait que la formule objective φ soit une formule de l'ensemble $\mathcal{B}m^{Q}_{a/b}$.

La figure 4.3 représente, étant donné une requête Q, deux exemples de relation entre les informations utiles et les informations utiles pour un agent. On remarque que les informations jugées utiles par l'agent b pour l'agent a ne sont pas nécessairement les informations qui sont vraiment utiles pour a. En effet, le jugement de b est basé sur ses croyances qui peuvent être fausses. Par exemple, l'agent b peut croire que l'agent a a certaines capacités de déduction que celui-ci n'a pas en réalité. Réciproquement, l'agent a peut avoir certaines croyances ignorées de b. Il est possible (comme sur le deuxième schéma), que les croyances de b sur les informations utiles pour a aient une intersection vide avec les informations utiles pour a.

Il est également possible de définir les formules que l'agent b pense être les plus potentiellement utiles pour l'agent a.

25. On laisse également implicite le modèle et le monde dans lesquels les informations sont les plus utiles.

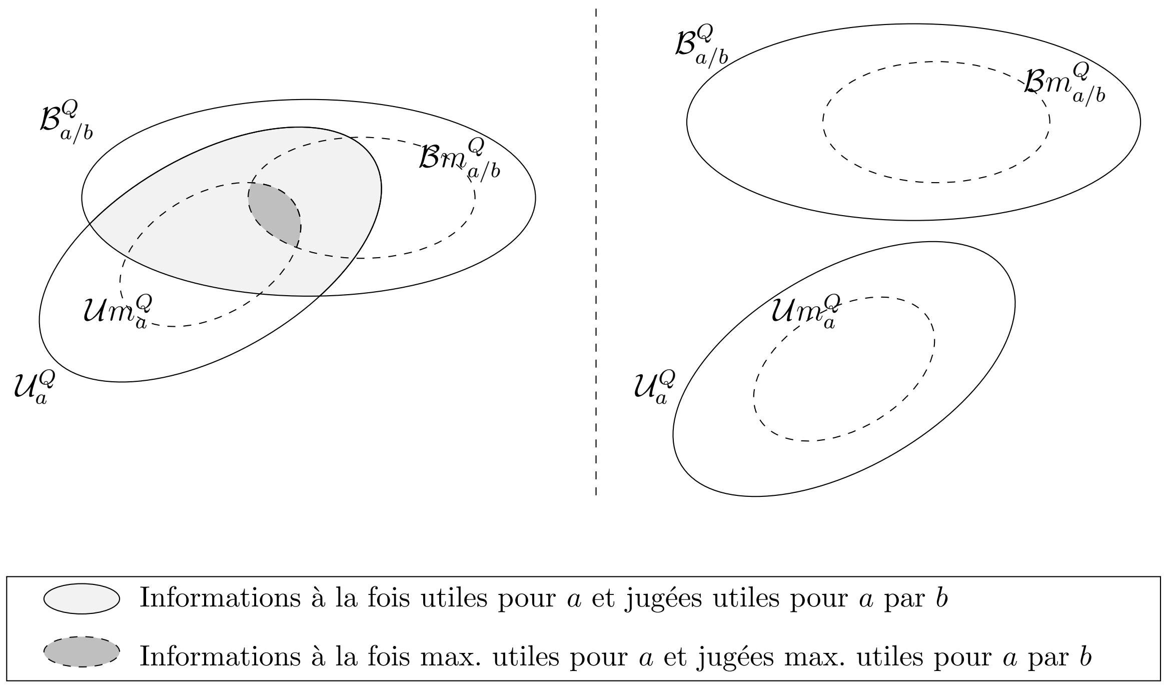

FIGURE 4.3 – Informations utiles, informations utiles selon un agent, informations les plus utiles et les plus utiles selon un agent

Définition I.30. *On note* $\mathcal{B}P^Q_{a/b}$ *l'ensemble des informations que l'agent b pense être potentiellement utiles pour l'agent a par rapport à Q.*
De cet ensemble, on extrait le sous-ensemble des informations les plus potentiellement utiles pour a selon b et on note ce sous-ensemble $\mathcal{B}Pm^Q_{a/b}$*.*
Finalement, on note $BPm^Q_{a/b}\varphi$ *le fait que la formule* φ *fasse partie de l'ensemble* $\mathcal{B}Pm^Q_{a/b}$*.*

4.2.2 Une première définition de la coopération

Dans un premier temps, nous considérons qu'un agent a est coopératif vis-à-vis d'un autre agent b si a transmet à b les informations qu'il pense être les plus utiles pour lui.

Nous commençons par définir la coopération pour deux formules : une information transmise et une requête.

Définition I.31 (Coopération subjective pour une information et une requête)**.** *Un agent b est coopératif vis-à-vis d'un agent a pour une information* φ *et une requête Q ssi b informe a de* φ *ssi l'information* φ *fait partie des informations que b juge être les plus utiles par rapport à Q pour l'agent a.*

$$Coop(a,b)_{\varphi,Q} \equiv Inf_{b,a}\varphi \leftrightarrow Bm^Q_{a/b}\varphi$$

Bien que n'étant pas précisé, cette définition se place dans un monde d'un modèle de la classe **C**. En effet, pour être tout à fait rigoureux, il faudrait écrire que l'agent b est coopératif avec l'agent a pour l'information φ et la requête Q *dans le monde* w, si et seulement si b informe a de φ *dans le monde* w ssi l'information φ fait partie des informations les plus utiles pour l'agent a par rapport à la requête Q, selon l'agent b *dans le monde* w.

Bien qu'étant formelle, remarquons que cette définition de la coopération est donnée en dehors de la logique. Cette définition est donc une notation (à cause de l'utilité maximale présente dans la définition).

Nous pouvons généraliser cette définition à l'ensemble des informations. Pour cela, nous considérons l'ensemble des formules objectives $\mathcal{O}$.

Définition I.32 (Coopération subjective)**.** *Soient a et b deux agents de $\mathcal{A}$. L'agent b est* ***(subjectivement) coopératif*** *vis-à-vis de l'agent a si et seulement si pour toute formule φ de $\mathcal{O}$, b informe a de φ si et seulement s'il existe une formule objective Q dans $\mathcal{O}$ telle que b croit que φ fait partie des informations les plus utiles pour a par rapport à Q. Ceci est représenté par la formule :*

$$Coop(b,a) \equiv \forall\varphi \in \mathcal{O},\ Inf_{b,a}\varphi \leftrightarrow \exists Q \in \mathcal{O}, Bm^{Q}_{a/b}\varphi$$

L'agent b est donc coopératif avec l'agent a si et seulement l'ensemble des informations échangées de b à a est exactement l'ensemble des informations que l'agent b pense être les plus utiles pour l'agent a par rapport à n'importe quel besoin qu'il puisse avoir.

À partir de cette définition, nous pouvons déduire la définition d'un agent non-coopératif vis-à-vis d'un autre agent. L'agent b est non-coopératif vis-à-vis de l'agent a si et seulement si :

1. l'agent b informe l'agent a d'une information φ qu'il ne croit pas être une information maximalement utile pour l'agent a (soit b ne croit pas que l'information soit utile, soit il existe une information plus utile que φ pour a selon b)
2. ou l'agent b n'informe pas l'agent a d'une information qu'il pense être une information des plus utiles pour un besoin de a.

Exemple I.10. Soient a et b deux agents de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle et w_0 un monde de ce modèle. Supposons que a désire prendre le train et que b soit un agent de gare. L'agent b croit que les agents comme a, qui doivent prendre le train, désirent savoir si leur train est en retard ou non. L'agent b croit également qu'un agent tel que a déduira du fait qu'il y ait un incident que son train est en retard. Nous supposons que l'agent b ne suppose rien d'autre sur les croyances et les intentions de a.
L'agent b croit qu'il y a un incident. Nous avons donc :

- $\mathbb{M}, w_0 \vDash B_b(D_a Bif_a retard)$
- $\mathbb{M}, w_0 \vDash B_b(inc)$
- $\mathbb{M}, w_0 \vDash B_b(B_a(inc \to retard)) \wedge B_b(\neg B_a(inc \to \neg retard))$ [26]

On a donc $B_b U_a^{retard} inc$. D'après la définition des informations les plus utiles que nous proposons, il ne peut y avoir d'information plus utiles que inc [27]. On a donc $Bm^{retard}_{a/b} inc$. Supposons que l'agent b informe a de inc ($\mathbb{M}, w \vDash Inf_{b,a} inc$). L'agent b est donc subjectivement coopératif vis-à-vis de l'agent a pour l'information φ et la requête Q, c'est-à-dire $Coop(a,b)_{inc,retard}$.

Si inc est la seule information dont l'agent b informe a, alors l'agent b est subjectivement coopératif vis-à-vis de l'agent a. En effet, l'ensemble des informations échangées de b vers a est exactement l'ensemble des informations les plus utiles pour l'agent a selon l'agent b.

26. En effet, b ne suppose rien d'autre sur les croyances de a, en particulier, $\neg B_b(B_a \neg inc)$ est satisfaite dans w_0.
27. En effet, il ne peut y avoir de formule pour laquelle l'ensemble DepLit est plus réduit qu'un ensemble contenant un seul littéral.

Nous pouvons remarquer que les agents coopératifs sont des agents sincères dans le sens où ils ne transmettent que des informations qu'ils jugent être vraies.

4.2.3 Définition étendue

Dans cette partie, nous proposons une définition un peu plus faible du caractère coopératif d'un agent. Nous introduisons cette notion sur un exemple.

Exemple I.11. Soient a, b et c trois agents de $\mathcal{A}$. L'agent a voudrait savoir si le nombre de personnes présentes à la manifestation est supérieur à 1000 personnes. L'agent c est à la manifestation et estime le nombre de personnes présentes à 1500. L'agent b croit que l'agent a désire savoir combien de personnes sont présentes. Il croit également que c croit que ce nombre est de 1500. L'agent b n'est pas sûr de l'objectivité de c quant à cette estimation et n'a donc aucune croyance sur le nombre de personnes.
Ne sachant pas si cette information est vraie ou fausse, l'agent b ne croit pas que l'information « il y a 1500 personnes » soit utile pour a. Il ne serait donc pas coopératif selon la définition précédente en transmettant cette information à a. Par contre, l'agent b croit que cette information est potentiellement utile pour l'agent a et il croit que l'agent c la croit vraie. Dans ce cas précis, nous disons que l'agent b est faiblement coopératif vis-à-vis de l'agent a s'il lui transmet l'information « il y a 1500 personnes », non pas telle que mais en précisant que cette information provient de l'agent c.

Définition I.33 (Agent faiblement coopératif pour une information, une requête et un agent)**.** *Soient a, b et c trois agents de $\mathcal{A}$. Soient φ et Q deux formules objectives. L'agent b est faiblement coopératif vis-à-vis de l'agent a pour l'information φ, la requête Q et l'agent c si et seulement si*

1. *b est coopératif vis-à-vis de a pour φ et Q*
2. *et b informe a de l'information « l'agent c croit que φ » si et seulement si b croit que φ est maximalement potentiellement utile pour a par rapport à Q et b croit que c croit φ.*

Ceci est représenté formellement par :

$$\begin{aligned} Coop_f(b,a)_{\varphi,Q,c} \equiv \quad & (Inf_{b,a}\varphi \leftrightarrow Bm^{Q}_{a/b}\varphi) \\ & \wedge(Inf_{b,a}B_c\varphi \leftrightarrow BPm^{Q}_{a/b}\varphi \wedge B_bB_c\varphi) \end{aligned}$$

Encore une fois, cette notion de coopération se place dans un monde d'un modèle de la classe **C**.

Nous pouvons généraliser cette définition à l'ensemble des formules objectives et à l'ensemble des agents.

Définition I.34 (Agent faiblement coopératif)**.** *Soient a et b deux agents de $\mathcal{A}$. L'agent b est faiblement coopératif vis-à-vis de l'agent a si et seulement si l'ensemble des informations échangées de b vers a est exactement l'union de :*

1. *l'ensemble des informations que b croit être les plus utiles pour l'agent a.*
2. *et de l'ensemble des informations du type « l'agent c croit que φ » avec c un agent de $\mathcal{A}$, et φ une information que b croit être maximalement potentiellement utile pour l'agent a et que b croit être vraie pour l'agent c.*

Ceci est représenté formellement par :

$$Coop_f(b,a) \equiv \forall\varphi \in \mathcal{O}\ \forall c \in \mathcal{A},\quad \begin{array}{l}(Inf_{b,a}\varphi \leftrightarrow \exists Q \in \mathcal{O}\ Bm^Q_{a/b}\varphi)\\ \wedge(Inf_{b,a}B_c\varphi \leftrightarrow \exists Q \in \mathcal{O}\ BPm^Q_{a/b}\varphi \wedge B_bB_c\varphi)\end{array}$$

Cette définition de la coopération peut être mise en relation avec la notion de discours rapporté. En effet, lorsque l'agent b informe l'agent a d'une information que l'agent c croit vraie, alors b rapporte à a un discours de c. L'agent b est alors coopératif car il juge que le discours qu'il rapporte est utile à celui à qui il le transmet. De même qu'il est possible d'étendre la notion de discours rapporté à plusieurs agents (a croit que b croit que c croit que ...), nous pourrions étendre cette notion de coopération à plus de 3 agents.

Exemple I.12. Formalisons l'exemple précédent. Soit w_0 un monde tel que :
- L'agent b croit que l'agent a désire savoir si le nombre de personnes est supérieur à mille personnes. On modélise par *Plus1000* l'information indiquant qu'il y a plus de mille personnes présentes.
  $\mathbb{M}, w_0 \vDash B_bD_aBif_a\mathit{Plus1000}$
- L'agent b croit que c croit qu'il y a 1500 personnes
  $\mathbb{M}, w_0 \vDash B_bB_c\mathit{1500Pers}$
- L'agent b sait qu'à partir de l'information *1500Pers*, a pourra répondre à son besoin en information.
  $\mathbb{M}, w_0 \vDash B_b(B_a(\mathit{1500Pers} \rightarrow \mathit{Plus1000}) \wedge \neg B_a(\mathit{1500Pers} \rightarrow \neg\mathit{Plus1000}))$
- L'agent b croit donc que l'information *1500Pers* est potentiellement utile pour l'agent a. De plus, cette information est une information potentielle maximale pour b. $\mathbb{M}, w_0 \vDash BPm_{a/b}\mathit{1500Pers}$
- Supposons que b informe a de $B_c\mathit{1500Pers}$, alors b est faiblement coopératif pour φ, Q et c.
- Supposons que b ne croit rien d'autre sur les besoins de l'agent a et que b ne l'informe que de $B_c\mathit{1500Pers}$, alors b est faiblement coopératif vis-à-vis de l'agent a.

4.2.4 Cas de l'information distribuée

Finalement, nous étendons la définition précédente au cas où l'information rapportée par l'agent n'est pas possédée par un seul agent mais est distribuée entre plusieurs agents.

Par exemple, considérons un agent a qui a besoin de savoir si plus de mille personnes sont présentes à la manifestation. L'information suivante est potentiellement utile pour lui : « Cinq cents personnes sont présentes dans la rue A et sept cents personnes sont présentes dans la rue B. ». Supposons que l'agent b croit que la première partie de l'information est vraie et que l'agent c croit que la deuxième partie de l'information est vraie. À eux deux, ils croient donc vraie une information qui répond au besoin de l'agent a. Dans ce cas, transmettre à l'agent a le fait que les agents b et c ont cette croyance distribuée est coopératif vis-à-vis de lui, car l'agent a pourra en déduire la réponse à son besoin en information.

Définition I.35 (Coopération faible pour une information, une requête et un ensemble d'agents)**.** *Soient a et b deux agents de $\mathcal{A}$ et I un sous-ensemble de $\mathcal{A}$. Soient φ et Q deux formules objectives. On dit que l'agent b est faiblement coopératif vis-à-vis de l'agent a pour l'information φ, la requête Q et l'ensemble d'agents I si et seulement si*

1. *b est coopératif vis-à-vis de a pour φ et Q*
2. *et b informe a de l'information « φ est une croyance distribuée de l'ensemble d'agents I » si et seulement si l'agent b croit que φ est maximalement potentiellement utile pour l'agent*

a par rapport à Q et si l'agent b croit que φ est une croyance distribuée dans l'ensemble d'agents I.

Ceci peut être plus formellement représenté par la formule suivante :

$$\begin{aligned} Coop_f(b,a)_{\varphi,Q,I} \equiv \quad & (Inf_{b,a}\varphi \leftrightarrow Bm^Q_{a/b}\varphi) \\ & \wedge Inf_{b,a}DB_I\varphi \leftrightarrow BPm^Q_{a/b}\varphi \wedge B_bDB_I\varphi) \end{aligned}$$

Dans le cas où l'ensemble d'agents I est réduit à un agent c, alors on se ramène à la définition de la coopération faible (définition I.33).

Définition I.36. *Soient a et b deux agents de $\mathcal{A}$ et I un sous-ensemble de $\mathcal{A}$. On dit que l'agent b est partiellement coopératif vis-à-vis de l'agent a ssi l'ensemble des informations échangées de l'agent b vers l'agent a est exactement l'union de*

1. *l'ensemble des informations φ que b croit être maximalement utiles pour l'agent a*
2. *et de l'ensemble des informations de type « φ est une croyance distribuée de l'ensemble d'agents I » telles que l'agent b croit que φ est maximalement potentiellement utile pour l'agent a et telles qu'il existe un ensemble d'agents I tel que l'agent b croit que φ est une information distribuée dans l'ensemble I.*

Ceci peut être plus formellement représenté par la formule suivante :

$$\begin{aligned} Coop_f(b,a) \equiv \forall\varphi \in \mathcal{O} \quad & (Inf_{b,a}\varphi \leftrightarrow \exists Q \in \mathcal{O} Bm^Q_{a/b}\varphi) \\ & \wedge \exists I \subseteq \mathcal{A}\ (Inf_{b,a}DB_I\varphi \leftrightarrow \exists Q \in \mathcal{O}\ BPm^Q_{a/b}\varphi \wedge B_bDB_I\varphi) \end{aligned}$$

Ainsi, si l'agent b croit vraie une partie de l'information potentiellement utile pour l'agent a, et qu'il croit que le reste de cette information est jugée vraie par d'autres agents, alors il est coopératif en transmettant cette croyance à a.

Exemple I.13. Reprenons formellement l'exemple précédent. Soit w_O un monde tel que :

- L'agent b croit que l'agent a désire savoir si le nombre de personnes est supérieur à mille personnes.
  $\mathbb{M}, w_0 \models D_aBif_a\mathit{Plus1000}$
- L'agent b croit que l'information « Il y a 500 personnes dans la rue A et 700 personnes dans la rue B », modélisée par $\mathit{500RueA} \wedge \mathit{700RueB}$ fait partie des formules les plus potentiellement utiles pour l'agent a.
  $\mathbb{M}, w_0 \models BPm^{\mathit{Plus1000}}_{a/b}\mathit{500RueA} \wedge \mathit{700RueB}$ [28]
- L'agent b croit que l'information « Il y a 500 personnes dans la rue A » est vraie.
  $\mathbb{M}, w_0 \models B_b\mathit{500RueA}$
- L'agent b croit que l'agent c croit que l'information « Il y a 700 personnes dans la rue B »est vraie.
  $\mathbb{M}, w_0 \models B_bB_c\mathit{700RueB}$
- L'agent b croit donc que l'information « Il y a 500 personnes dans la rue A et 700 personnes dans la rue B » est une croyance distribuée des agents b et c.
  $\mathbb{M}, w_0 \models B_bDB_{\{b,c\}}\mathit{500RueA} \wedge \mathit{700RueB}$
- Si l'agent transmet l'information « Il y a 500 personnes dans la rue A et 700 personnes dans la rue B est une croyance distribuée des agents b et c », alors il est partiellement coopératif vis-à-vis de l'agent a pour l'information φ, la requête Q et l'ensemble d'agents I.

28. Cette formule est une notation, raccourci d'écriture de la formule $\mathit{500RueA} \wedge \mathit{700RueB}$ fait partie des informations les plus potentiellement utiles pour a selon b dans le monde w_0.

– Supposons que l'agent b n'ait aucune autre croyance sur les besoins de l'agent a. Dans ce cas, l'agent b est coopératif vis-à-vis de l'agent a s'il l'informe uniquement de l'information « Il y a 500 personnes dans la rue A et 700 personnes dans la rue B est une croyance distribuée des agents b et c ».

4.3 Coopération objective

Les trois définitions de la coopération que nous proposons sont subjectives car c'est un agent qui juge de l'utilité ou non de l'information qu'il transmet.
Le jugement de l'agent sur l'utilité de l'information pour un autre peut être juste ou erroné. C'est ce que Demolombe ([Dem04]) appelle la crédibilité de l'agent [29]. Que dire d'un agent (non-crédible) qui transmet des informations qu'il juge utiles mais qui ne le sont pas ? Dans ce cas, bien qu'étant subjectivement coopératif, nous voulons caractériser le fait que dans l'absolu, cet agent n'est pas coopératif : les informations qu'il transmet ne sont pas utiles.

Nous proposons donc dans cette partie une définition de la coopération qui ne prend pas en compte les croyances de celui qui transmet l'information. Nous appelons cette coopération la *coopération objective.*

Définition I.37 (Coopération objective pour une information et une requête)**.** *Soient a et b deux agents de $\mathcal{A}$. Soient φ et Q deux formules objectives. L'agent b est objectivement coopératif vis-à-vis de l'agent a pour l'information φ et la requête Q si et seulement si b informe a de φ si et seulement si l'information φ est utile pour l'agent a par rapport à Q.*

$$Coop_{obj}(b,a)_{\varphi,Q} \equiv Inf_{b,a}\varphi \leftrightarrow Um_a^Q\varphi$$

Comme pour les définitions précédentes de la coopération, celle-ci se place également dans un monde d'un modèle de la classe **C**.

On peut également généraliser cette définition.

Définition I.38 (Coopération objective)**.** *Soient a et b deux agents de $\mathcal{A}$. L'agent b est objectivement coopératif vis-à-vis de l'agent a si et seulement si l'ensemble des informations échangées de b vers a est exactement l'ensemble des informations maximalement utiles pour l'agent a.*

$$Coop_{obj}(b,a) \equiv \forall\varphi \in \mathcal{O}, Inf_{b,a}\varphi \leftrightarrow \exists Q \in \mathcal{O}, Um_a^Q\varphi$$

Un agent est donc objectivement coopératif s'il transmet les informations qui sont maximalement utiles pour les autres, indépendamment de son jugement.

Exemple I.14. Soient a et b deux agents. Soit w_0 un monde dans lequel :
– a désire savoir si sont train est en retard
$\mathbb{M}, w \vDash D_a Bif_a retard$
– l'agent a n'a pas d'autre besoin en information.
– l'information inc est utile pour l'agent a.
$\mathbb{M}, w \vDash U_a^{retard} inc$

29. Dans [Dem04], un agent est crédible pour une information si et seulement si s'il croit cette information, alors celle-ci est vraie.

– l'information *inc* est maximalement utile pour l'agent a dans le monde w.
– l'agent b informe uniquement a de l'information *inc*.
– l'agent b est donc objectivement coopératif avec l'agent a.

On peut remarquer que si les agents sont crédibles (c'est-à-dire si leur croyances sont vraies), alors les agents subjectivement coopératifs sont également objectivement coopératifs.

La figure 4.4 récapitule les différents ensembles qui interviennent dans les définitions de la coopération subjective (définition I.32) et de la coopération objective (définition I.38) objective. $Inf_{b,a}$ représente l'ensemble des informations échangées de b à a. Les intersections des ensembles $\mathcal{U}m_a^Q$ (respectivement $\mathcal{B}m_{a/b}^Q$) avec $Inf_{b,a}$ représentent les formules pour lesquelles l'agent a est objectivement coopératif respectivement subjectivement coopératif). L'intersection entre $\mathcal{U}m_a^Q$ et $\mathcal{B}m_{a/b}^Q$ est l'ensemble des informations utiles à propos desquelles les croyances de l'agent b sont justes. On peut donc voir cet ensemble comme l'ensemble des informations utiles pour a pour lesquelles b est crédible. Dans l'idéal, les trois ensembles présentés ici sont égaux : c'est-à-dire les informations envoyées sont exactement celles que l'agent croit être les plus utiles et sont exactement celles qui sont les plus utiles.

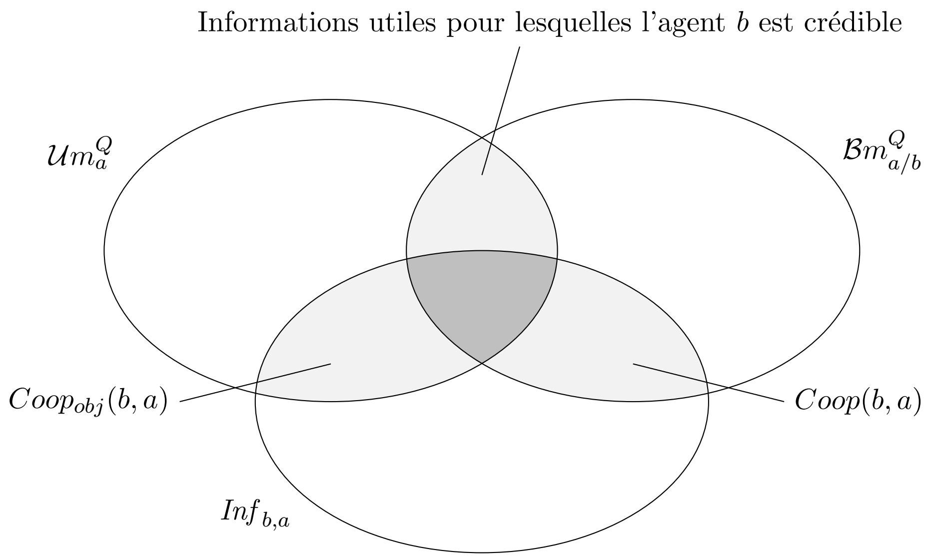

Figure 4.4 – Ensembles intervenant dans les définitions de la coopération

4.4 Comparaison avec le Principe de Coopération de Grice

Dans cette partie, nous comparons la définition de la coopération avec la définition informelle donnée par Grice dans son Principe de Coopération. Nous pouvons remarquer que la définition de Grice prend en compte les croyances de celui qui informe (voir état de l'art 1.1.2). Ainsi, nous jugeons que la comparaison entre la coopération subjective et le principe de coopération de Grice est plus intéressante que la comparaison entre la coopération objective et la coopération de Grice. Nous analysons donc en quoi la définition de la coopération subjective vérifie ou ne vérifie pas les différentes maximes.

4.4.1 Maxime de Quantité

La première sous-maxime de quantité est : *Fais en sorte que ta contribution soit aussi informative que nécessaire (pour les besoins de l'échange).*
Dans notre cas, le besoin sous-jacent à l'échange est la réponse au besoin en information du type « savoir si Q ». Il apparaît dans la notion d'utilité dans la formule $D_a Bif_a Q$. L'agent b transmet des informations φ telles que $B_b(B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))$. Cela signifie que l'agent b croit qu'à partir de φ, a pourra répondre à son besoin en information. Il croit donc que sa contribution est aussi informative que nécessaire pour la résolution du besoin.

La deuxième sous-maxime est *Fais en sorte que ta contribution ne soit pas plus informative que nécessaire.*
D'après notre définition, l'agent b n'échange que les informations qu'il croit être maximalement utiles pour l'agent a. Les informations les plus utiles sont les plus succinctes. Elles peuvent également être vues comme les moins informatives.

La définition de la coopération subjective est donc en accord avec la maxime de quantité.

4.4.2 Maxime de Qualité

La maxime de qualité est : *Fais en sorte que ta contribution soit vraie.*
L'agent b, s'il est coopératif suivant la définition I.32, ne transmet que des informations qu'il juge utiles, donc que des informations qu'il juge vraies. Ainsi, la maxime de qualité est respectée. Nous pouvons noter que le respect de la maxime de qualité correspond au caractère sincère de l'agent tel qu'il a été décrit dans [Dem04].

Remarquons que la maxime de qualité est divisée en deux sous-maximes (ne dis pas ce que tu sais être faux et ne dis ce pour quoi tu n'as pas de preuves). Dans ce travail, nous ne intéressons pas à la provenance des croyances des agents.

4.4.3 Maxime de Relation

La maxime de relation est *Sois pertinent.*

Comme nous l'avons déjà remarqué dans l'état de l'art (chapitre 1), les interprétations de cette maxime sont nombreuses et Grice reste relativement imprécis sur cette maxime.

Supposons que la maxime de relation réfère au moment où l'information est échangée. Elle peut alors être comprise comme *l'information doit être échangée au moment où elle est requise.* Pour être en accord avec cette maxime, il faudrait que la définition de la coopération ou la notion d'utilité compose avec la notion de temps. Dans notre travail, la notion de temps est implicite. En effet, nous nous plaçons à un instant donné et regardons quels sont les besoins des agents, leurs croyances, ... Les informations sont donc utiles (ou inutiles) et les agents sont coopératifs à cet instant donné. Ainsi, si l'information est échangée, c'est qu'elle est utile à cet instant. Notre travail est donc cohérent avec la maxime de relation prise dans ce sens. Néanmoins, nous reviendrons sur cette notion de temps dans les perspectives.

Interprétons maintenant la maxime de relation en terme d'approprié, de justifié. Dans ce cas, cette maxime signifie que l'information échangée doit être connectée au besoin de l'agent. Dans la définition que nous proposons de l'utilité, nous mettons en évidence le besoin de l'agent, ainsi que la connexion de celui-ci avec l'information utile. La définition de la coopération est donc en accord avec la maxime de relation interprétée de cette façon. Toutefois, comme nous le verrons dans la partie Limites (section 5.2), la modélisation de cette connexion peut poser quelques problèmes et pourrait être améliorée.

La comparaison de la définition avec cette maxime dépend énormément de l'interprétation de celle-ci. Étant donnée une interprétation différente, il faudrait vérifier que notre définition est bien cohérente.

4.4.4 Maxime de Manière

La maxime de manière est *Fais en sorte que ta contribution soit clairement exprimée.* Elle se divise en plusieurs sous-maximes.

Les deux premières sous-maximes sont *Evite d'être obscur* et *Evite l'ambiguïté.* Nous supposons que les informations échangées sont des formules de la logique. Nous supposons implicitement que les formules du langage sont comprises par tous les agents de l'ensemble $\mathcal{A}$. En particulier, nous supposons que les variables propositionnelles ne sont ni obscures ni ambiguës pour les différents agents. Ainsi, le cadre formel permet la cohérence de nos définitions avec les deux premières sous-maximes.

La sous-maxime suivante est *Sois bref.* Ainsi énoncée, cette sous-maxime fait référence à la forme des informations. En terme de formules, cette maxime adresse donc l'aspect syntaxique de celles-ci. Notre définition des informations les plus utiles (et donc notre définition de la coopération) ne tenant pas compte de la syntaxe des formules (voir conclusion intermédiaire de l'utilité, section 3.7), elle n'est pas en accord avec cette sous-maxime de Grice.

Finalement, la dernière sous-maxime est *Sois ordonné.* Pour certaines informations, l'ordre dans lequel les éléments qui la composent sont énoncés peut jouer un grand rôle pour la compréhension de celle-ci. Or, dans le cadre logique, certains connecteurs sont commutatifs ($\wedge$, $\vee$). Pour ceux-là, l'ordre n'a donc pas d'importance. La définition de la coopération subjective n'est pas en accord avec cette sous-maxime.

4.5 Conclusion intermédiaire

Comme nous l'avons fait pour l'utilité, nous pouvons vérifier l'adéquation de la définition que nous proposons avec les propriétés que nous nous sommes fixées comme cibles à la fin de l'état de l'art.

1. *un agent est coopératif si l'information qu'il transmet répond à un besoin de celui qui la reçoit ou tout du moins si elle répond à ce qu'il juge être un besoin de celui qui la reçoit.* Un agent subjectivement coopératif transmet uniquement des informations qu'il juge utiles et un agent objectivement coopératif transmet uniquement des informations utiles. Dans les deux cas, la propriété est respectée. Notons qu'il serait possible d'introduire des degrés de réponse d'une information à une requête. Ces degrés permettraient alors de définir des degrés de coopération de l'agent.
2. *un agent est coopératif s'il transmet des informations qui sont vraies ou tout du moins qu'il juge vraies.* Un agent subjectivement coopératif transmet uniquement des informations qu'il juge utiles donc vraies et un agent objectivement coopératif transmet uniquement des informations utiles donc vraies. Dans les deux cas, la propriété est également respectée.
3. *un agent est coopératif vis à vis d'un autre s'il anticipe sur les besoins de celui-ci.* Dans le cas de la coopération subjective, nous nous intéressons aux croyances de celui qui transmet l'information et notamment aux croyances qu'il a sur les besoins de l'autre agent. Nous ne nous intéressons pas à la génération de ces croyances. Cependant, les croyances des agents

sur les besoins des autres sont prises en compte dans la notion de coopération subjective. Dans le cas de la coopération objective, tous les besoins des agents sont pris en compte. Nous abordons le point de la génération de besoins dans la partie 5.3.1.

4. *un agent est d'autant plus coopératif qu'il transmet des informations faciles à interpréter.* La facilité d'interprétation des informations repose sur le fond (la signification) de l'information et sa forme. Au niveau de la signification de l'information, l'agent coopératif fait des échanges suffisamment informatifs mais pas trop informatifs (notre définition est en accord avec la maxime de quantité de Grice). Au niveau de la forme de la formule, nous avons vu que la définition de la coopération n'est pas complètement cohérente avec la maxime de Manière de Grice. Nous ne prenons donc pas en compte tous les paramètres de la facilité d'interprétation. Nous revenons sur ce point particulier dans la partie 5.2.2 de la conclusion.

Chapitre 5

Conclusion et perspectives

5.1 Travail réalisé

L'objectif de cette partie était de définir formellement les concepts d'utilité et de coopération pour des agents qui échangent des informations au sein d'un système.

Après avoir fait un état de l'art de quelques domaines dans lesquels ces notions étaient importantes, nous avons dressé une liste des éléments principaux qui doivent être présents dans la définition de ces concepts.

Nous avons commencé par poser le cadre formel : une logique multimodale propositionnelle dont les deux modalités sont le désir et la croyance. Nous avons proposé une axiomatique et une sémantique correspondante pour ces modalités.

Dans ce cadre formel, nous avons défini la notion d'utilité d'une information pour un besoin en information. Cette utilité est définie pour un agent et pour une requête. Elle contient trois éléments différents : le besoin en information de l'agent qui s'exprime en fonction de la requête, les croyances de l'agent qui permettent la connexion entre l'information et la requête, et la valeur de vérité de l'information. Nous avons mis en évidence de nombreuses propriétés des informations utiles et nous avons proposé une caractérisation formelle des informations les plus utiles pour un agent. Nous avons ensuite proposé une définition plus faible de l'utilité, que nous avons appelée utilité potentielle. Nous en avons fait une rapide étude. Finalement, nous avons généralisé le besoin en information et avons défini formellement l'utilité correspondante.

La plupart des propriétés cibles que nous voulions respecter pour la notion de l'utilité sont atteintes avec la définition que nous avons proposée. Nous revenons en détail sur les propriétés qui posent problème dans la section suivante.

Nous nous sommes intéressés dans un deuxième chapitre à la notion de coopération. En prenant en considération les croyances de celui qui transmet l'information, nous avons défini la coopération subjective. Un agent est alors coopératif s'il transmet les informations qu'il juge être les plus utiles pour celui qui la reçoit. Cette définition peut être étendue au cas où l'agent rapporte une information et au cas où cette information est distribuée entre plusieurs agents.

En ne tenant pas compte des croyances de celui qui transmet l'information, nous avons défini la coopération objective. Un agent est alors objectivement coopératif s'il transmet les informations qui sont les plus utiles pour celui qui la reçoit.

Finalement, nous avons comparé la coopération subjective avec les maximes de Grice. La définition que nous avons proposée est cohérente avec les deux premières maximes de Grice (Quantité et la Qualité). En ce qui concerne la maxime de Relation, la cohérence dépend de l'interprétation que nous donnons de la maxime. C'est avec la maxime de Manière avec laquelle notre définition est le moins en accord car nous nous sommes plus intéressés au sens des informations que leur forme.

Notons qu'un certain nombre de ces résultats ont fait l'objet de publications ([RC08, CR08c, RC09a, RC09d, RC09b, RC09c]).

5.2 Limites

Dans cette section, nous présentons les limites de notre modélisation de l'utilité et de la coopération. Pour chacune de ces limites, nous proposons des approches qui permettraient d'améliorer les caractérisations.

5.2.1 L'implication matérielle

L'implication matérielle et les paradoxes qui l'accompagnent représentent certainement un des facteurs les plus limitant de notre travail.

En effet, un des paradoxes liés à l'implication est qu'une formule fausse implique toute autre formule. Appliqué aux agents, cela signifie qu'à partir d'une formule qu'un agent croit fausse, il peut déduire n'importe quoi : si $B_a \neg\varphi$ alors pour tout ψ, $B_a(\varphi \to \psi)$. Dans notre cas, les formules qu'un agent croit fausses ne sont pas utiles pour lui (proposition I.3) [30].

Néanmoins, ce problème se retrouve dans les disjonctions et dans les implications. Par exemple, soit w un monde dans lequel

- inc est une information potentiellement utile par rapport à Q pour l'agent a
  $\mathbb{M}, w \vDash P_a^Q inc$
- inc est fausse
  $\mathbb{M}, w \vDash \neg inc$
- inc n'est donc pas une information utile pour Q dans w.
- $pleut$ est une information vraie
  $\mathbb{M}, w \vDash \neg pleut$
- l'agent a croit que $pleut$ est fausse
  $\mathbb{M}, w \vDash B_a \neg pleut$
- $pleut$ n'est donc pas une information utile pour Q dans w.

Alors l'information $inc \vee pleut$ (ou encore $\neg pleut \to inc$) est une information utile pour l'agent a par rapport à Q. En effet,

- $\mathbb{M}, w \vDash inc \vee pleut$ car $pleut$ est vraie.
- $\mathbb{M}, w \vDash B_a((inc \vee pleut) \to Q) \otimes B_a((inc \vee pleut) \to \neg Q)$ car inc est potentiellement utile et que $\neg pleut$ permet de déduire à la fois Q et $\neg Q$.

Cette propriété va à l'encontre de ce que l'on souhaite d'une caractérisation de l'utilité. En effet, dans l'information $pleut \vee inc$, c'est l'élément inc, qui est faux, qui permet à l'agent a de répondre à son besoin en information et c'est l'élément $pluie$, qui permet à a de déduire n'importe quoi, qui rend la disjonction vraie. La disjonction peut donc induire a en erreur et répondre à son besoin de manière erronée. Ce problème se répercute sur la définition de la coopération, cette dernière étant basée sur l'utilité.

Pour remédier à ce problèmes, de nombreuses approches sont possibles.

30. En effet, si un agent a un besoin en information « savoir si Q », une formule qu'il croit fausse lui permet de déduire Q. Cependant, une information qu'il croit fausse lui permet également de déduire $\neg Q$ et ne répond donc pas réellement à son besoin.

Une autre implication

Le problème venant des paradoxes liés à l'implication matérielle, utiliser une relation qui n'a pas ces paradoxes serait une solution. Comme nous l'avons vu dans l'état de l'art, le courant des logiques de la pertinence tente de définir des formalismes dans lesquels l'implication correspond exactement au « si ...alors ... » du langage naturel. Le problème de ces logiques est qu'elles se placent dans un cadre formel relativement complexe. Il faudrait donc définir un cadre formel qui permet à la fois de raisonner sur les croyances et désirs des agents et de travailler avec une implication plus réaliste ou tout du moins sans ces paradoxes.

La réduction à des conjonctions de littéraux

Afin d'éviter les problèmes avec les disjonctions ou les implications, une solution est la réduction de notre travail à des littéraux ou des conjonctions de littéraux. Les requêtes seraient donc des littéraux et les formules utiles soit des littéraux, soit des conjonctions de littéraux. Les formules qui s'écrivent sous la forme de disjonction ou d'implication ne serait donc plus des informations utiles.
La définition des informations les plus utiles serait également simplifiée. En effet, réduite aux conjonctions de littéraux, les explications minimales de Q pour a sont les impliquants premiers de Q dans la base de croyances de a.

Cette solution est relativement simple à mettre en place et offre une modélisation de l'utilité et de la coopération qui respecte les propriétés souhaitées.
Cependant, elle restreint énormément les échanges d'informations possibles entre agents coopératifs. En effet, si un agent est coopératif, il ne transmet que des littéraux ou des conjonctions de littéraux. Les informations de type « si ton train est un TGV, alors il part du quai 1. Si c'est un TEOZ, il part du quai 2 »ne peuvent pas être considérées comme utiles pour un agent qui cherche à savoir de quel quai part son train.

Croyances erronées de l'agent

Le problème que nous cherchons à résoudre est lié aux croyances erronées de l'agent. En effet, c'est en raisonnant sur ces croyances que nous tombons sur des propriétés contre-intuitives. Avant de déterminer quelles sont les informations utiles pour un agent par rapport à un besoin en information, il faudrait donc d'abord s'assurer que les croyances de ce dernier (qui servent dans la caractérisation de l'utilité) ne sont pas erronées. Pour cela, plusieurs approches sont possibles.

Tout d'abord, nous pourrions choisir de raisonner sur des connaissances ou sur des croyances objectives ([BS08]). Les connaissances imposent à toutes les formules de la base de l'agent d'être vraies. Les croyances objectives contraignent les croyances de l'agent sur le monde à être vraies. En terme de cadre formel, il faudrait donc ajouter l'axiome $(T_{obj}) : \vdash B_a\varphi \rightarrow \varphi$ pour φ formule objective [31] et modifier la propriété de sérialité des relations R_a ($a \in \mathcal{A}$) des modèles de la classe **C**

31. La différence avec la connaissance est que l'axiome T de la modalité épistémique s'applique à toutes les formules et non pas juste les formules objectives.

en une propriété d'o-sérialité ([32]). La limite de cette solution est que les échanges d'informations sont restreints aux informations vraies. En effet, ce sont des des formules objectives qui sont échangées (formules les plus utiles). Avec l'axiome (T_{obj}), les informations échangées sont donc vraies.

Une deuxième approche serait de considérer que les croyances de l'agent qui peuvent lui servir pour répondre à un besoin en information sont des croyances que celui-ci veut implicitement vérifier ou confirmer et sont donc sujettes à un besoin de vérification. Il faudrait de plus donner la priorité à ces besoins de vérification (priorité par rapport aux besoins en information). Les informations les plus utiles pour un agent seraient donc celles qui corrigent ses croyances ou tout du moins quelques-unes de ses croyances. Appliqué à la définition de la coopération, cela signifie que les agents coopératifs corrigent d'abord les croyances qu'ils jugent erronées des autres avant de leur transmettre les informations qui répondent à leur besoin en information. Nous développons la question de l'utilité pour ce type de besoin dans la section dédiée aux perspectives (section 5.3.4). Notons tout de même que ce besoin de vérification soulève de nombreuses questions. Par exemple, comment les agents peuvent-ils corriger les croyances des autres ? Faut-il leur fournir seulement l'information corrigée ou également leur fournir des preuves de cette information corrigée ? Un autre problème serait de modéliser le doute des agents sur leurs propres croyances. En effet, si un agent a besoin de vérifier une de ses croyances, alors c'est qu'il doute de celle-ci. Avec la modélisation que nous avons choisie pour la croyance, il n'est pas possible de douter de ses croyances. Une solution possible pourrait alors être l'introduction de degrés de croyance ([Lav07]).

5.2.2 Informations les plus utiles

La caractérisation des informations les plus utiles n'est valable que pour les formules objectives. De plus, comme nous l'avons vu en conclusion des chapitres 3 et 4, elle ne tient pas compte de l'aspect syntaxique des formules mais uniquement de leur signification. Il faudrait donc une « métrique » sur les formules générales qui permette de caractériser à la fois l'informativité (signification) d'une information et sa forme (information brève, simple, etc.).

De même, nous ne prenons pas en compte le processus de déduction que l'agent doit mettre en place pour connecter l'information utile à la requête. Or, une réponse directe à une requête Q (c'est-à-dire l'information Q ou l'information $\neg Q$) demande a priori moins d'effort d'interprétation à un agent que n'importe quelle autre réponse. Ainsi, plus une information est directe pour une requête, plus elle est utile.

La caractérisation des informations les plus utiles que nous présentons ici est une proposition qui se base sur le fait que les informations utiles ne doivent pas être trop informatives et donc que les informations les plus brèves sont les plus utiles. Il existe de nombreuses autres définitions d'informations les plus utiles. Par exemple, une toute autre approche pourrait être de considérer que les informations les plus utiles sont celles qui répondent à plusieurs besoins simultanément. Dans ce cas, le nombre de besoins auxquels l'information utile répond pourrait servir d'ordre ou de préordre. Quelle que soit l'approche utilisée, la définition des coopérations ne changent pas : elles correspondent toujours à l'échange des informations les plus utiles ou tout du moins jugées utiles.

32. La relation R_a est o-sérielle si et seulement si pour tout w de W, il existe un monde v tel que wR_av et tel que $\forall p \in \mathcal{V}, w \in V(p) ssiv \in V(p)$

Nous pourrions définir un degré de coopérativité en fonction du degré d'utilité des informations échangées. Plus l'information échangée est utile, plus l'agent est coopératif. Dans notre cas, nous pourrions donc définir une coopération de base qui correspond à l'échange des informations utiles et une coopération maximale lorsque les informations échangées sont les plus utiles.

5.2.3 Utilité partielle

Dans ce travail, les informations que nous caractérisons comme utiles sont des informations qui répondent complètement au besoin en information. Nous ne caractérisons pas les informations qui ne répondent que partiellement au besoin en information.
Une telle caractérisation peut être mise en place lorsque le besoin en information est généralisé. En effet, pour un besoin de type « savoir Q_1 ou ... ou Q_n ou s'ils sont tous faux », une information qui permet de déduire qu'une partie des Q_i sont faux est une information qui répond partiellement au besoin en information. Par exemple, si un agent veut savoir si son train part du quai 1 ou du quai 2 ou du quai 3 ou d'aucun des trois, alors l'information que son train ne part pas du quai 1 répond partiellement à son besoin en information.
Dans le cadre de la coopération (section 4.2.4), nous utilisons la notion de croyance distribuée dans la définition I.35. Dans ce cas, nous pourrions considérer que les agents possèdent individuellement une partie de l'information jugée utile, c'est-à-dire qu'ils possèdent une information jugée partiellement utile.
Une autre définition d'une information partiellement utile pourrait être une adaptation à notre cadre logique de la définition proposée par Lakemeyer ([Lak97]). En effet, il caractérise les informations utiles comme les informations qui ne répondent pas complètement au besoin mais qui contiennent des éléments nécessaires. Ceci peut être vu comme une définition de la partialité.

5.2.4 Passage en logique du premier ordre

Dans toute cette partie, nous utilisons une logique multimodale propositionnelle. Nous gagnerions énormément en expressivité en considérant une logique multimodale du premier ordre. Il faudrait alors prendre en compte un certain nombre de problèmes que nous aborderons dans le chapitre 7 de la deuxième partie de ce manuscrit.
Remarquons tout de même que nous pouvons simuler un certain nombre de requêtes du premier ordre lorsque celles-ci portent sur un nombre fini d'éléments. Par exemple, une requête du type « Quels sont les employés qui ont plus de 30 ans d'ancienneté ? », si l'on sait qu'il y a 4 employés a, b et c et que l'on modélise par 30_a (resp. 30_b, 30_c) le fait que a (resp. b et c) ait plus de 30 ans d'ancienneté, est équivalente à « Est ce que $\neg 30_a \wedge \neg 30_b \wedge \neg 30_c$, ou $30_a \wedge \neg 30_b \wedge \neg 30_c$, $\ldots$, ou $30_a \wedge 30_b \wedge 30_c$? » et correspond donc dans ce cas à un besoin en information généralisé.

5.2.5 Le temps

Comme nous l'avons vu dans la comparaison de nos travaux avec le Principe de Coopération (paragraphe 4.4), le temps est implicitement présent dans nos travaux. En effet, lorsque l'on se place dans un monde, on considère ce qui est vrai à un instant donné. Les informations échangées à cet instant sont les informations utiles à cet instant, etc.

Cependant, pour être plus réaliste dans la caractérisation de l'utilité et de la coopération, il faudrait faire intervenir le temps de manière plus explicite. En effet, les besoins en informations sont en général vrais sur une plage de temps. Les informations sont vraies pendant une certaine durée. Les croyances des agents évoluent avec le temps. Les informations sont donc utiles pour une durée qui dépend de toutes les dates et durées mentionnées ci-dessus. Pour cela, il serait possible d'ajouter des modalités temporelles à notre logique (voir par exemple [Gal08]).

5.3 Perspectives

5.3.1 Simplifications

La définition que nous donnons de la coopération subjective implique que les agents ont de nombreuses croyances les uns sur les autres et en particulier des croyances sur les désirs et sur les croyances des autres.

Il serait possible, étant données certaines hypothèses, de faciliter la génération de croyances des agents sur les états mentaux des autres.

Tout d'abord, les agents peuvent s'informer de leurs besoins respectifs : si l'agent a informe l'agent b qu'il désire savoir si Q, alors il est raisonnable de déduire que b croit que a a un tel besoin.

Il est également possible d'anticiper sur les besoin des agents. Si un agent a croit qu'un agent b a un certain type de besoin, alors a croit que b a également tel type de besoin. On retrouve alors des travaux réalisés dans le domaine des réponses coopératives (voir état de l'art, section 1.4).

Dans la plupart des systèmes, les agents ont des rôles ou des fonctions qui leur sont attribués. En fonction de leur rôle dans le système, il est possible de déterminer certains besoins des agents. Par exemple, un agent dont le rôle est de déterminer la météo du lendemain a besoin de savoir si la pression atmosphérique augmente ou diminue. De même, étant donné un rôle ou une fonction, il est possible de déduire un certain nombre de croyances des agents. Par exemple, pour un agent dont le rôle est médecin, on peut raisonnablement supposer qu'à partir d'un résultat sanguin, il saura déterminer si un patient est diabétique ou non.

5.3.2 Confiance

Nous ne nous sommes pas intéressés dans ce manuscrit à l'attitude de l'agent qui est informé par un autre d'une information jugée utile pour lui. En effet, les agents (subjectivement) coopératifs jugent de l'utilité de l'information mais peuvent avoir un jugement erroné. Dans ce cas, deux agents subjectivement coopératifs peuvent tout à fait transmettre des informations contradictoires au même agent. La question qui se pose alors, pour celui qui reçoit les deux informations, est de savoir quelle information croire ? Dans ce cadre, la relation avec le concept de confiance est à établir. L'agent va décider d'intégrer ou non l'information qu'on lui transmet s'il a confiance dans l'information et s'il a confiance dans celui qui la transmet. De très nombreux travaux on été menés sur ce concept, notamment des travaux basés sur des logiques modales et qui pourraient donc relativement facilement s'intégrer à nos travaux [Dem09, Dem04].

5.3.3 Groupe coopératif

Nous nous sommes intéressés à un échange d'information interpersonnel. Il serait intéressant d'étendre nos travaux à d'autres types de communication. Par exemple, que pourrait signifier qu'une information est utile pour un groupe d'agents ? Quelle serait la signification de « Le groupe d'agents juge que l'information est utile pour un autre groupe » ou de « Le groupe est coopératif vis-à-vis d'un autre groupe ». Encore une fois, il existe un état de l'art conséquent sur la notion de groupe dans les systèmes multi-agents (par exemple [Tuo95, Tuo92, HS08]).

5.3.4 Autres besoins

Finalement, ce travail est consacré à l'étude de l'utilité d'une information pour un besoin particulier des agents, le besoin en information. Or, ce n'est pas le seul type de besoin que les agents peuvent avoir.

Ils peuvent par exemple avoir besoin de vérifier certaines de leurs croyances. Dans ce cas, toute information vraie qui confirme ou contredit leurs croyances est une information utile. Avec le cadre formel que nous développons, la modélisation de l'utilité pour ce besoin n'est pas possible. En effet, dans notre cadre formel, les agents ne peuvent pas douter de leurs croyances (axiomes (B4) et (B5)). Afin d'introduire le doute des agents sur leurs propres croyances, il serait possible d'introduire un degré de croyance. Cette notion a été notamment étudiée dans le domaine de la révision de croyances ([LEA89]).

Un besoin plus complexe des agents pourrait être de compléter ou de vérifier leurs croyances (dans un domaine donné). Ainsi, toute information vraie qui est inconnue des agents ou qui corrige une croyance et qui est dans le domaine souhaité est une information utile pour ce besoin [33].

Plus généralement, pour un besoin quelconque, nous pouvons proposer une définition pseudo-formelle de l'utilité. Une information φ est utile pour un agent a par rapport à X si et seulement si $D_a X \wedge (\varphi \rightsquigarrow_a X) \wedge \varphi$, c'est-à-dire si et seulement si

- l'agent a a un besoin de type X : $D_a X$
- l'information φ est connectée à X pour a : $\varphi \rightsquigarrow_a X$
- et si l'information φ est vraie : φ

Quelle que soit la définition de l'utilité choisie, la définition de la coopération d'un agent pour ce besoin reste la même : un agent est (subjectivement) coopératif s'il transmet les informations qu'il juge être les plus utiles pour ce besoin.

33. Remarquons que dans ce cas, nous pouvons nous ramener à un besoin en information.

Deuxième partie

Politiques d'échange d'informations : cohérence et complétude

Introduction

En général, les agents d'un système ne sont pas libres dans leurs échanges d'information. Ces échanges sont en effet réglementés par une politique d'échange d'informations. Dans cette partie, nous nous intéressons à la modélisation des politiques d'échanges d'informations et des différentes propriétés qu'elles doivent respecter pour être fonctionnelles. Plus particulièrement, pour être applicables, les politiques d'échange d'information doivent être cohérentes. Informellement, une politique est cohérente si elle ne conduit pas à une contradiction. Dans la plupart des cas et plus particulièrement dans les politiques qui régissent des systèmes critiques, il est également souhaitable que les politiques d'échanges soient complètes, c'est-à-dire que dans toute situation, elles prescrivent le comportement des agents.

Les politiques d'échange d'informations peuvent être perçues comme des réglementations particulières : les réglementations qui spécifient quels échanges d'informations sont permis, interdits ou obligatoires et sous quelles conditions. L'étude des politiques d'échange peut donc se ramener à une application de l'étude des réglementations en général. Ainsi, dans cette partie, nous traitons tout d'abord le cas des réglementations générales. Nous étudions ensuite le cas particulier des politiques d'échange.

Dans un premier chapitre, nous réalisons un état de l'art de la littérature portant sur la cohérence et de la complétude des règles. Puis, dans le chapitre 7, nous décrivons le cadre formel sur lequel nous nous basons pour modéliser les réglementations et leurs propriétés (chapitre 8). Nous proposons ensuite une méthode pour raisonner avec des réglementations incomplètes dans le chapitre 9. Le chapitre 10 est consacré à l'application de nos résultats aux politiques d'échange d'informations. Finalement, nous concluons sur notre travail et proposons des perspectives sur celui-ci dans le chapitre 11.

Chapitre 6

État de l'art

Le raisonnement normatif et plus particulièrement l'étude des réglementations et des leurs propriétés a reçu beaucoup d'attention et ce dans de nombreux domaines.

Dans un premier temps, nous étudions la notion de cohérence d'une réglementation dans le domaine de l'Intelligence Artificielle. Puis, nous nous intéressons à la notion de complétude. Celle-ci a non seulement été étudiée en Intelligence Artificielle mais est également une problématique au cœur de domaines tels que le domaine juridique, le domaine de la théorie des contrats ou le domaine des bases de données.

6.1 Cohérence des réglementations

La cohérence des réglementations est une propriété sur laquelle de nombreux travaux ont été menés. Nous n'en donnons ici qu'un petit aperçu.

La cohérence a été définie dans le cadre des politiques de confidentialité ([BC93]), c'est à dire les politiques dans lesquelles les agents ont la permission ou l'interdiction de connaître une information. Deux politiques de confidentialité sont alors dites *cohérentes* si l'utilisateur n'a pas à la fois la permission par la première politique de connaître une formule et l'interdiction par la seconde de la connaître.
Cette définition est adaptée dans [CD97] pour les politiques de sécurité multi-niveaux. Des différences sont alors faites entre l'interdiction explicite et l'interdiction implicite, et entre la permission explicite et la permission implicite. L'interdiction explicite (resp. la permission explicite) est une interdiction (resp. une permission) qui apparaît en tant que telle dans la réglementation alors que l'interdiction implicite (resp. la permission implicite) est une interdiction (resp. une permission) déduite d'une absence de permission explicite (resp. d'interdiction explicite). Une réglementation est alors cohérente si elle ne conduit pas à la fois à la permission explicite de connaître une information et à l'interdiction explicite de la connaître. En d'autres termes, il faut que la permission explicite implique la permission implicite.

Dans [Cho99], les réglementations générales sont étudiées. Une réglementation est cohérente si et seulement s'il n'existe pas de situation dans laquelle la réglementation amène à une *contradiction normative* (un agent a à la fois la permission et l'interdiction de faire une action) ou dans laquelle la réglementation met un agent face à un *dilemme* (un agent est obligé d'effectuer une action et obligé d'effectuer son contraire). Étant exprimée dans une logique du premier ordre, la vérification de la cohérence d'une réglementation se ramène dans ce cas à un problème d'abduction.

6.2 Complétude et règles de fermeture

6.2.1 Intelligence Artificielle

La complétude des réglementations a été beaucoup moins étudiée dans les domaines de l'Informatique et de l'Intelligence Artificielle.

[BC93] propose une définition de complétude entre deux politiques de confidentialité : pour chaque information, l'agent doit soit avoir la permission de la connaître dans la première politique, soit avoir l'interdiction de la connaître dans la seconde.

On retrouve également une définition de complétude dans [CD97] pour les politiques de sécurité multi-niveaux. Une réglementation est dite complète si pour chaque information, la réglementation conduit à la permission explicite de la connaître ou à l'interdiction explicite de la connaître. En d'autres termes, il faut que la permission implicite implique la permission explicite. Dans ce travail, les auteurs utilisent des règles au méta-niveau qui permettent de compléter les réglementations. Les deux règles proposées correspondent à une approche permissive (tout ce qui n'est pas précisé par la réglementation est permis) et à une approche restrictive (tout ce qui n'est pas précisé par la réglementation est interdit).

6.2.2 Vides juridiques

On parle de *vide juridique* lorsqu'il existe un cas qui n'est pas du tout couvert par une règle ou par un ensemble de règles légales. Plus formellement, Alchourron et Bulygin ([AB71]) énoncent qu'il y a un vide juridique dans un système légal L si un acte p n'est ni obligatoire, ni interdit, ni permis dans une situation q. Le vide juridique peut donc être perçu comme une incomplétude du système légal.

Dans le domaine juridique, une différence est également faite entre les permissions explicites (ou fortes) et permissions implicites (ou faibles) [vW51]. Les permissions explicites sont exprimées dans les lois ou normes et les permissions implicites résultent en général de l'application d'une règle de fermeture indifféremment appelée *principe d'interdiction*, *sealing principle* ou encore *principe nulla poena* et stipule que « tout ce qui n'est pas interdit est permis » ([Roy97]).

Suivant les systèmes légaux, la considération de la complétude ou du vide juridique varie. En effet, la notion d'incomplétude peut être définie en terme d'absence de permission ou d'interdiction explicite (par exemple en loi civile). Dans les cas où la règle du sealing principle est partie intégrante du système légal (par exemple en loi criminelle), il ne peut y avoir d'incomplétude : dans les cas pour lesquels il n'y a ni interdiction explicite ni permission explicite, on peut appliquer la règle de fermeture et ainsi déduire une permission implicite.

6.2.3 Théorie des contrats

La notion de complétude est également très importante dans le domaine des contrats. Dans cet état de l'art, nous ne donnons qu'un bref aperçu de ce domaine et de la façon dont la complétude y est étudiée. En effet, même si les approches de la complétude sont intéressantes, elles sont en général basées sur les gains ou les utilités des différents partis du contrat. Ainsi, elles s'appliquent difficilement aux politiques d'échange d'informations.

En théorie des contrats, la notion de complétude ou d'incomplétude est assez controversée et il n'existe pas de définition précise [AF00]. Cela vient du fait que les contrats sont en général de nature très variées et que selon cette nature, l'incomplétude n'est pas perçue de la même manière. Une définition qui nous paraît relativement générique est la suivante : un contrat est dit *complet* s'il explicite une réponse ou une solution pour toutes les situations qui sont jugées réalistes par une cour [Seg99]. En général, les contractants souhaitent la complétude des contrats auxquels ils prennent part, un vide ou une incomplétude pouvant être exploitée négativement pour eux.

De façon à gérer les incomplétudes des contrats, de nombreuses méthodes ont été mises en place, notamment des méthodes de renégociation.

6.2.4 Bases de données

Dans le domaine des bases de données, la Closed World Assumption (CWA), formalisée par Reiter ([Rei78]) permet de compléter les bases de données. En effet, cette règle de fermeture stipule que ce qui n'est pas connu par la base de données est considéré comme faux pour celle-ci. Plus formellement, si $BD \nvdash l$ alors $BD \vdash \neg l$. Les bases de données dans laquelle la Closed World Assumption est appliquée sont complètes.

Toujours dans ce domaine, [Rei92] et [Dem99] ont remarqué que certaines contraintes d'intégrité exprimées sur une base de données sont des règles à propos de ce que la base de données devrait savoir (ou, pour le dire différemment, ce sont des règles portant sur ce qui devrait être déduit de la base de données). Par exemple, la contrainte d'intégrité « tout employé a un numéro de téléphone, un numéro de fax ou une adresse mail » exprime en fait que, pour tout employé contenu dans la base de données, celle-ci connaît son numéro de téléphone, son numéro de fax ou son adresse mail. Cela n'empêche pas que dans le monde réel, un employé de l'entreprise n'ait ni numéro de téléphone, ni numéro de fax, ni adresse mail… Comme précisé par Reiter [Rei92], cette contrainte d'intégrité exprime une sorte de *complétude locale* de la base de données. Des règles de fermeture (par exemple les défauts de Reiter) peuvent être utilisées pour compléter la base de données dans de tels cas. Par exemple, une des règles pourrait être que si la base de données ne contient pas les informations demandées pour un employé donné, alors on pourrait supposer que le numéro de téléphone de cet employé est celui de son département.

6.2.5 Conclusion

Alors que la cohérence a été définie dans [Cho99] pour des réglementations générales, il n'existe pas de tel travail pour la complétude des réglementations générales. Ainsi, il faut définir un cadre formel pour étudier ces deux notions. De plus, bien que la notion de règle de fermeture soit connue, et ce dans de nombreux domaines, elle n'a pas été formalisée en logique déontique. L'objectif de cette partie est donc de formaliser la complétude des réglementations générales en logique déontique et d'y intégrer la notion de règle de fermeture.

Chapitre 7

Cadre logique

Au cours de cette thèse, nous avons étudié la notion de complétude pour des réglementations particulières, les politiques d'échange d'informations dans un système multi-agents [CR07b]. Une définition de l'incomplétude de telles politiques a été proposée et une méthode de raisonnement avec de telles politiques a été définie. L'approche prise dans ce travail était prometteuse et nous l'avons étendue pour des réglementations plus générales dans [CR08a]. Le langage formel utilisé dans ces deux articles est la logique du premier ordre (FOL) suivant les idées développées dans [Cho99]. En particulier, les notions déontiques (obligation, permission, interdiction) sont représentées en utilisant des symboles de prédicats. Cela amène à effectuer une partition compliquée du langage. De plus, les notions déontiques sont classiquement représentées en utilisant une logique modale [vW51, Hil71]. C'est pourquoi nous utilisons dans cette partie une logique modale du premier ordre [FM99] pour exprimer les réglementations d'une manière plus élégante. Nous reformulons donc le travail de [CR08a] dans un cadre modal du premier ordre.

Le formalisme de base utilisé pour représenter des réglementations est SDL (*Standard Deontic Logic*), une logique modale propositionnelle [Che80]. SDL est une logique modale particulière qui permet de représenter les notions déontiques comme l'obligation [34]. En suivant les techniques présentées dans [FM99], nous étendons SDL en FOSDL (*First-Order Standard Deontic Logic*) pour pouvoir exprimer des réglementations plus complexes impliquant plusieurs agents. Les notions présentées dans ce qui suit seront illustrées sur des exemples à partir de la section 8.1.

7.1 Langage

L'alphabet de FOSDL est composé des ensembles suivants de symboles non logiques :

- un ensemble $\mathcal{P} = \{Q, R, \ldots\}$ de symboles de prédicats
- un ensemble $\mathcal{F} = \{f, g \ldots\}$ de symboles de fonctions
- une modalité O représentant l'obligation

Notons que l'ensemble des fonctions d'arité 0 est appelé *ensemble des constantes* et est noté $\mathcal{C}$.

Nous définissons également les symboles logiques suivants :

- un ensemble $\mathcal{V} = \{x, x_1, x_2, y \ldots\}$ de symboles de variables
- des connecteurs et quantificateurs : $\neg$, $\vee$, $\forall$, (et).

Pour des raisons techniques que nous expliciterons plus tard, nous distinguons deux types de prédicats : les d-prédicats et les f-prédicats. Les d-prédicats sont les prédicats qui peuvent

34. Il a été montré que SDL conduisait à un certain nombre de problèmes, comme le paradoxe de Ross, le dilemme de Jörgensen... Nous ne nous intéressons pas ici à la résolution de ces problèmes. On pourra consulter [MN06] pour une liste plus détaillée des problèmes posés par l'utilisation de SDL.

apparaître sous une modalité déontique. Les f-prédicats sont les prédicats qui nous serviront à décrire l'état du monde ainsi que les contraintes d'intégrité sur ce monde.

Définition II.1 (Termes)**.** *On définit un terme récursivement de la façon suivante :*
- *si x est une variable de $\mathcal{V}$, alors x est un terme.*
- *si c est une constante de $\mathcal{C}$, alors c est un terme.*
- *si $t_1, \ldots, t_n$ sont des termes et f une fonction d'arité n, alors $f(t_1, \ldots, t_n)$ est un terme.*

Définition II.2 (Formules de FOSDL)**.** *Les formules de FOSDL sont définies récursivement comme suit :*
- *si $t_1, \ldots, t_n$ sont des termes et Q un symbole de prédicat d'arité n, alors $Q(t_1, \ldots, t_n)$ est une formule de FOSDL.*
- *si φ est une formule de FOSDL, alors $O\varphi$ est une formule de FOSDL.*
- *si ψ_1 et ψ_2 sont des formules de FOSDL et x_1 un symbole de variable, alors $\neg\psi_1$, $\psi_1 \vee \psi_2$, $\forall x_1\ \psi_1$ sont des formules de FOSDL.*

On appelle *littéral positif* toute formule de la forme $Q(t_1, \ldots, t_n)$ et *littéral négatif* toute formule de la forme $\neg Q(t_1, \ldots, t_n)$ avec Q un symbole de prédicat et $t_1, \ldots, t_n$ des termes. Un *littéral* est un littéral positif ou un littéral négatif. Finalement, si Q est un d-prédicat (resp. un f-prédicat), alors $Q(t_1, \ldots, t_n)$ et $\neg Q(t_1, \ldots, t_n)$ sont des *d-littéraux* (resp. des f-littéraux).

Si ψ_1, ψ_2 et ψ_3 sont des formules de FOSDL et x_1 est un symbole de variable, nous définissons également les abréviations suivantes :
- $\psi_1 \wedge \psi_2 \equiv \neg(\neg\psi_1 \vee \neg\psi_2)$
- $\psi_1 \otimes \psi_2 \otimes \psi_3 \equiv (\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3) \vee (\neg\psi_1 \wedge \psi_2 \wedge \neg\psi_3) \vee (\neg\psi_1 \wedge \neg\psi_2 \wedge \psi_3)$
- $\psi_1 \rightarrow \psi_2 \equiv \neg\psi_1 \vee \psi_2$
- $\psi_1 \leftrightarrow \psi_2 \equiv (\neg\psi_1 \vee \psi_2) \wedge (\psi_1 \vee \neg\psi_2)$
- $\exists x_1\ \psi_1 \equiv \neg\forall x_1\ \neg\psi_1$

Définition II.3. *Les modalités pour la permission, notée P, et l'interdiction, notée F sont définies à partir de O comme suit :*

$$F\varphi \equiv O\neg\varphi$$
$$P\varphi \equiv \neg O\varphi \wedge \neg O\neg\varphi$$

Une formule de FOSDL sans modalités est dite *objective*. Les termes et formules de FOSDL sans symboles de variables sont dits *de base*. L'ensemble des termes de base de FOSDL est appelé univers de Herbrand HU.

Dans les formules $\forall\ x\ \varphi$ et $\exists\ x\ \varphi$, φ s'appelle la *portée* du quantificateur. Une occurrence d'une variable est *libre* si elle n'est dans la portée d'aucun quantificateur. Sinon, elle est *liée*.

Enfin, nous appelons une *substitution de base* toute fonction $\chi\ : \mathcal{V} \rightarrow HU$. Si $\varphi(x)$ est une formule de FOSDL avec une variable libre x, $\varphi(\chi(x))$ est la formule φ dans laquelle les occurrences libres de x ont été remplacées par $\chi(x)$.

7.1.1 Permission bilatérale

Notons que notre définition de la permission ne correspond pas à la définition usuelle de SDL. En effet, dans SDL, quelque chose est permis si sa négation n'est pas obligatoire. Cependant, il a été montré par des juristes [Gro06] que les cas où la permission est bilatérale (permission de faire et permission de ne pas faire) sont les seuls valides. Si une permission n'est pas bilatérale, alors elle implique l'obligation. En effet, si lorsque l'on parle de la permission que φ soit vraie (représenté par $\neg O \neg \varphi$) on n'impose pas également que $\neg\varphi$ soit permise (représenté par $\neg O\varphi$), alors on obtient $O\varphi$ soit l'obligation que φ soit vraie. Par exemple, si fumer est autorisé, alors ne pas fumer est également autorisé. Sinon, cela signifierait que fumer est obligatoire. Notre définition de la permission bilatérale correspond à la notion d'*optionalité* [MN06] (quelque chose est optionnel si et seulement si ni lui ni sa négation ne sont obligatoires). On remarquera finalement que les définitions que nous donnons de l'obligation, la permission et l'interdiction correspondent aux trois positions normatives qui sont définies dans la théorie des positions normatives de Kanger et Lindahl [Kan72, Lin77].

7.2 Sémantique

On reprend ici la sémantique développée dans [FM99] pour les logiques modales du premier ordre. La sémantique des logiques modales propositionnelles est classiquement définie en utilisant des modèles de Kripke [Kri63a, Kri63b]. Les modèles sont définis par un *cadre* $\langle \mathcal{W}, \mathcal{R} \rangle$, où $\mathcal{W}$ est un ensemble de mondes et $\mathcal{R}$ une relation d'accessibilité entre les mondes, et une valuation V entre les mondes et les lettres propositionnelles. Dans le cas du premier ordre, nous définissons des modèles en utilisant un cadre *augmenté* et une interprétation du premier ordre au lieu de V.

La sémantique des langages du premier ordre est fondée sur un ensemble de symboles (les *objets du discours*), appelé le *domaine*. Le domaine représente les objets sur lesquels les prédicats vont porter par opposition aux termes qui sont des notions purement mathématiques. Dans le cas d'une logique modale du premier ordre, nous devons choisir entre des cadres augmentés à domaine constant ou à domaine variable. Dans le premier cas, le domaine est fixé pour tous les mondes de $\mathcal{W}$, dans le second cas, chaque monde de $\mathcal{W}$ peut avoir son propre domaine. Nous choisissons ici un domaine constant. Comme les normes que nous étudions ne concernent que des éléments fixes, ce choix est assez intuitif [35].

Définition II.4. *Soient $\mathcal{W}$ un ensemble de mondes, $\mathcal{R}_O$ une relation sur $\mathcal{W}^2$ et $\mathcal{D}$ un ensemble non vide de symboles représentant le domaine, alors $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D} \rangle$ est appelé un cadre.*

Pour définir un modèle, nous devons définir une interprétation du premier ordre, ce qui est fait classiquement dans ce qui suit.

Définition II.5. *Une interprétation $\mathcal{I}$ dans un cadre $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D} \rangle$ est une application telle que :*
- *pour tout symbole de constante (ou fonction d'arité 0) de $\mathcal{C}$ et tout monde $w \in \mathcal{W}$, $\mathcal{I}(c, w)$ est un élément de $\mathcal{D}$.*
- *pour tout symbole de fonction n-aire f dans $\mathcal{F}$ et tout monde $w \in \mathcal{W}$, $\mathcal{I}(f, w)$ est une fonction de $\mathcal{D}^n$ dans $\mathcal{D}$.*
- *pour tout symbole de prédicat n-aire Q dans $\mathcal{P}$ et tout monde $w \in \mathcal{W}$, $\mathcal{I}(Q, w)$ est une relation sur $\mathcal{D}^n$.*

Définition II.6. *Un modèle $\mathcal{M}$ est une structure $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D}, \mathcal{I} \rangle$ où $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D} \rangle$ est un cadre et $\mathcal{I}$ une interprétation sur $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D} \rangle$.*

Enfin, nous utilisons une classe de cadres qui modélisent le comportement de l'opérateur O en contraignant la relation d'accessibilité $\mathcal{R}_O$.

35. Les domaines variables peuvent néanmoins être utiles. Par exemple, dans une logique modale doxastique du premier ordre, un agent peut apprendre l'existence d'un objet particulier, ou un nouvel objet peut apparaître. Notons toutefois qu'il est possible de recréer le concept du domaine variable avec un cadre à domaine constant, par exemple en introduisant un prédicat d'existence qui, informellement, appliqué à un objet hypothétique, renvoie vrai si et seulement si l'objet existe dans le domaine actuel. Le lecteur pourra se référer à [FM99] pour plus de détails à ce sujet.

Définition II.7. *Un modèle de FOSDL est un modèle* $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D}, \mathcal{I} \rangle$ *tel que* $\mathcal{R}_O$ *est sérielle*[36].

Nous définissons ensuite la notion de *valuation* qui associe les variables du langage aux éléments de $\mathcal{D}$:

Définition II.8 (Valuation)**.** *Soient* $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D}, \mathcal{I} \rangle$ *un modèle. Une valuation dans le modèle* $\mathcal{M}$ *est une fonction* $\mathcal{V} \to \mathcal{D}$ *qui associe à chaque variable libre* x *un élément de* $\mathcal{D}$.

Une valuation σ' *est une valuation* x*-variante d'une valuation* σ *si* σ *et* σ' *sont identiques sauf en* x.

Définition II.9 (Evaluation des termes)**.** *Soient* $\langle \mathcal{W}, \mathcal{R}_O, \mathcal{D}, \mathcal{I} \rangle$ *un modèle,* w *un monde de* $\mathcal{W}$ *et* σ *une valuation dans* $\mathcal{M}$*. À chaque terme* t*, nous associons une valeur dans un monde* w*, notée* $(\sigma \star \mathcal{I})(t, w)$ *comme suit :*

- *Si* x *est une variable libre,* $(\sigma \star \mathcal{I})(x, w) = \sigma(x)$.
- *Si* c *est un symbole de constante,* $(\sigma \star \mathcal{I})(t, w) = \mathcal{I}(c, w)$.
- *Si* f *est un symbole de fonction d'arité* n *et que* $t_1, \ldots, t_n$ *sont* n *termes alors* $(\sigma \star \mathcal{I})(f(t_1, \ldots, t_n)) = \mathcal{I}(f, w)((\sigma \star \mathcal{I})(t_1, w), \ldots, (\sigma \star \mathcal{I})(t_n, w))$

Tous les termes fermés sont localement rigides, c'est-à-dire que pour tout couple de mondes (w, v) *de* $\mathcal{W}^2$*, si* $w\mathcal{R}_O v$ *alors* $(\sigma \star \mathcal{I})(t, w) = (\sigma \star \mathcal{I})(t, v)$.

Notons que nous imposons que les termes fermés soient localement rigides, c'est-à-dire que les termes désignent les mêmes objets dans les mondes qui sont accessibles les uns pour les autres.

Au niveau de l'interprétation des prédicats, cela nous évite d'avoir à choisir la signification d'une formule telle que $O\ Q(c)$ où c est une constante : est-ce que cela signifie que « il est obligatoire que l'objet représenté par c *dans le monde courant* a la propriété Q »ou « il est obligatoire que *dans chaque monde accessible depuis le monde courant*, l'objet représenté par c a la propriété Q ». c étant un terme localement rigide, il désigne le même objet dans le monde courant et dans les mondes accessibles. Les deux significations possibles sont donc équivalentes. Dans notre cas, cela signifie que les objets du monde dans lequel la réglementation est définie sont identiques aux objets sur lesquels porte la réglementation. Nous évitons ainsi des détails techniques compliqués comme par exemple l'utilisation de l'abstraction pour les prédicats (voir [FM99] pour plus de détails).

La relation de satisfaisabilité $\models$ est définie comme suit :

Définition II.10. *Soient* $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}_O, \mathcal{D}, \mathcal{I} \rangle$ *un modèle de FOSDL,* w *un monde de* $\mathcal{W}$ *et* σ *une valuation sur* $\mathcal{D}$*. Alors :*

- *si* Q *est un symbole de prédicat* n*-aire et* $t_1, \ldots, t_n$ *sont des termes, alors* $\mathcal{M}, w \models_\sigma Q(t_1, \ldots, t_n)$ *si et seulement si* $\langle (\sigma \star \mathcal{I})(t_1, w), \ldots, (\sigma \star \mathcal{I})(t_n, w) \rangle \in \mathcal{I}(Q, w)$.
- *si* ψ *est une formule de FOSDL, alors* $\mathcal{M}, w \models_\sigma \neg\psi$ *si et seulement si* $\mathcal{M}, w \not\models_\sigma \psi$.
- *si* ψ_1 *et* ψ_2 *sont des formules de FOSDL, alors* $\mathcal{M}, w \models_\sigma \psi_1 \vee \psi_2$ *si et seulement si* $\mathcal{M}, w \models_\sigma \psi_1$ *ou* $\mathcal{M}, w \models_\sigma \psi_2$.
- *si* $O\psi$ *est une formule de FOSDL,* $\mathcal{M}, w \models_\sigma O\psi$ *si et seulement si pour tout* $v \in \mathcal{W}$ *tel que* $w\mathcal{R}_O v$*,* $\mathcal{M}, v \models_\sigma \psi$.
- *si* ψ *est une formule de FOSDL,* $\mathcal{M}, w \models_\sigma \forall x\ \psi$ *si et seulement si pour toute valuation* σ' x*-variant de* σ*,* $\mathcal{M}, w \models_{\sigma'} \psi$.

Soit ψ une formule fermée de FOSDL. Si pour une valuation σ on a $\mathcal{M}, w \models_\sigma \psi$, alors on a pour toute valuation σ, $\mathcal{M}, w \models_\sigma \psi$. On note alors dans ce cas $\mathcal{M}, w \models \psi$ et on dit que ψ est

36. I.e. tout monde de $\mathcal{W}$ a un successeur par $\mathcal{R}_O$. Ceci nous garantit que l'axiome DO est bien représenté par la relation d'accessibilité entre mondes.

satisfaite dans w. Si $\mathcal{M}, w \models \psi$ pour tout w de $\mathcal{W}$, on note $\mathcal{M} \models \psi$ et on dit que ψ est valide dans $\mathcal{M}$. Enfin, si $\mathcal{M} \models \psi$ pour tout modèle $\mathcal{M}$ de FOSDL, on note $\models \psi$ et on dit que ψ est valide.

7.3 Axiomatique

Nous allons maintenant définir un système axiomatique pour FOSDL en suivant l'approche proposée dans [FM99]. Dans ce qui suit, $\varphi(x)$ est une formule dans laquelle la variable x peut avoir des occurrences libres. On dira qu'une variable libre y est *substituable* à x dans $\varphi(x)$ s'il n'y a pas d'occurrence de x dans $\varphi(x)$ qui soit une sous-formule commençant par $\forall y$.

Les formules présentées sur la figure 7.1 sont des axiomes. Les règles d'inférence sont présentées sur la figure 7.2.

(Taut)	les tautologies de la logique propositionnelle
(OK)	$O(\varphi \to \psi) \to (O\varphi \to O\psi)$
(OD)	$O\varphi \to \neg O \neg \varphi$
(UnivDist)	$(\forall x)[\varphi \to \psi] \to [(\forall x\ \varphi) \to (\forall x\ \psi)]$
(Per)	$(\forall x)(\forall y)\ \varphi \leftrightarrow (\forall y)(\forall x)\ \varphi$
(UnivInst)	$(\forall y)[(\forall x)\ \varphi(x) \to \varphi(y)]$, avec y substituable à x dans $\varphi(x)$
(Bar1)	$O(\forall x\ \varphi) \to \forall x\ O\varphi$
(Bar2)	$\forall x\ O\varphi \to O(\forall x\ \varphi)$

Figure 7.1 – Schéma d'axiomes

(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$
(Gen)	$\dfrac{\varphi}{\forall x\ \varphi}$
(ONec)	$\dfrac{\varphi}{O\varphi}$

Figure 7.2 – Règles d'inférence

(OD) signifie que les obligations sont cohérentes ; (UnivDist) est la distributivité universelle ; (Per) est la permutation ; (UnivInst) est l'instanciation universelle ; (Bar1) et (Bar2) sont les formules de Barcan ; (MP) est la règle de Modus Ponens ; (Gen) est la règle de généralisation ; (ONec) est la règle de nécessitation pour l'obligation.

Les théorèmes de FOSDL sont les formules dont les formules (notées $\vdash \varphi$) qui peuvent être déduites des axiomes et des règles d'inférence.

Proposition II.1 (Validité et complétude)**.** *Le système précédent est valide et complet par rapport à la sémantique de FOSDL.*

Dans ce qui suit, $\bot$ représente toute formule qui est une contradiction et $\top$ représente toute formule qui est une tautologie.

Chapitre 8

Réglementations

8.1 Modélisation des réglementations

Dans cette section, nous définissons la notion de réglementation. Dans un premier temps, nous définissons la notion de règle, qui est le composant de base d'une réglementation. Dans cette définition, les règles ont une forme générale, en particulier, elles peuvent être conditionnelles.

Définition II.11. *Une règle est une formule de FOSDL de la forme* $\forall \overrightarrow{x}\ l_1 \vee \ldots \vee l_n$ *avec* $n \geq 1$ *telle que :*

1. l_n *est de la forme* $O\varphi$ *ou de la forme* $\neg O\varphi$ *où* φ *est de type* $Q(t_1, \ldots, t_n)$ *ou de type* $\neg Q(t_1, \ldots, t_n)$*,* Q *étant un d-prédicat et* $t_1, \ldots, t_n$ *des termes.*
2. $\forall i \in \{1, \ldots, n-1\}$*,* l_i *est un f-littéral.*
3. *si* x *est une variable dans* l_n*, alors* $\exists i \in \{1, \ldots, n-1\}$ *tel que* l_i *est un littéral négatif et contient la variable* x
4. $\forall \overrightarrow{x}$ *représente* $\forall x_1 \ldots \forall x_m$ *où* $\{x_1, \ldots, x_m\}$ *est l'ensemble des variables libres apparaissant dans* $l_1 \vee \ldots \vee l_{n-1}$*.*

Dans cette définition, les contraintes (1) et (2) permettent d'exprimer des règles de la forme « si telle condition est vraie alors quelque chose est obligatoire (resp. permis, interdit) ». En effet, les règles sont de la forme $l_1 \vee \ldots \vee l_{n-1} \vee O\varphi$, qui est équivalente à la forme $\neg l_1 \wedge \ldots \wedge \neg l_{n-1} \rightarrow O\varphi$. La contrainte (3) restreint les règles aux formules à champ restreint [37]. Enfin, les règles sont des *phrases*, c'est-à-dire des formules fermées, comme exprimé par la contrainte (4).

Remarquons également que nous restreignons dans la définition des règles la formule qui peut être obligatoire : seuls les d-littéraux peuvent être obligatoires ou non.

On écrira $\forall \overrightarrow{x}\ l_1 \vee \ldots \vee l_{n-1} \vee P\varphi$ comme un raccourci d'écriture pour les deux règles $\{\forall \overrightarrow{x}\ l_1 \vee \ldots \vee l_{n-1} \vee \neg O\varphi, \forall \overrightarrow{x}\ l_1 \vee \ldots \vee l_{n-1} \vee \neg O\neg\varphi\}$.

Définition II.12. *Une réglementation est un ensemble de règles.*

Considérons un exemple qui illustrera les notions développées dans les sections 8 et 9.

Exemple II.1. Nous considérons une réglementation qui régule le comportement d'un conducteur devant un feu tricolore. Le langage utilisé est défini comme suit :

37. Les formules à champ restreint sont un fragment décidable des formules domaine-indépendant dont on a prouvé qu'elles étaient les seules formules du premier ordre ayant une signification en modélisation [Dem82]. Remarquons en particulier que par définition du langage de FOSDL, toutes les variables apparaissant dans l_n sont libres dans l_n.

- *vert*, *orange*, *rouge*, *auto*, *camion*, *velo*, A et T sont des fonctions d'arité nulle, i.e. des constantes.
- x, y, z, i et t sont des variables.
- *Conducteur*(.) est un symbole de f-prédicat qui indique qu'un terme est un conducteur.
- FT(.) est un symbole de f-prédicat qui indique qu'un terme est un feu tricolore.
- *Couleur*(.,.) est un symbole de f-prédicat qui prend pour paramètres un feu tricolore et une couleur et indique la couleur du feu.
- *Vehicule*(.,.) est un symbole de f-prédicat qui prend pour paramètres un conducteur et le type de véhicule qu'il conduit.
- *Devant*(.,.) est un symbole de f-prédicat qui prend pour paramètre un conducteur et un feu tricolore et indique que le véhicule conduit par le conducteur est devant un feu.
- *Stop*(.,.) est un symbole de d-prédicat qui prend un conducteur et un feu tricolore en paramètres et indique que l'agent s'arrête devant le feu tricolore.

Considérons maintenant les trois règles :

(r_0) *« Quand un conducteur d'une voiture est devant un feu tricolore qui est rouge, il doit s'arrêter. »*

(r_1) *« Quand un conducteur d'une voiture est devant un feu tricolore qui est orange, il peut s'arrêter. »*

(r_2) *« Quand un conducteur d'une voiture est devant un feu tricolore qui est vert, il ne doit pas s'arrêter. »*

Ces règles sont modélisées par :

$$\begin{aligned}
(r_0) \quad & \forall x \forall t\ \mathit{Conducteur}(x) \wedge FT(t) \wedge \mathit{Vehicule}(x, \mathit{auto}) \wedge \mathit{Couleur}(t, \mathit{rouge}) \\
& \wedge \mathit{Devant}(x,t) \rightarrow O\mathit{Stop}(x,t) \\
(r_1) \quad & \forall x \forall t\ \mathit{Conducteur}(x) \wedge FT(t) \wedge \mathit{Vehicule}(x, \mathit{auto}) \wedge \mathit{Couleur}(t, \mathit{orange}) \\
& \wedge \mathit{Devant}(x,t) \rightarrow P\mathit{Stop}(x,t) \\
(r_2) \quad & \forall x \forall t\ \mathit{Conducteur}(x) \wedge FT(t) \wedge \mathit{Vehicule}(x, \mathit{auto}) \wedge \mathit{Couleur}(t, \mathit{vert}) \\
& \wedge \mathit{Devant}(x,t) \rightarrow F\mathit{Stop}(x,t)
\end{aligned}$$

On pourra vérifier facilement que ces formules sont bien des règles comme spécifié par la définition II.11.

8.2 Cohérence des réglementations

Nous définissons maintenant une première notion utile pour les réglementations, la *cohérence*. Intuitivement, nous dirons qu'une réglementation est cohérente si et seulement si nous ne pouvons pas dériver de la réglementation d'incohérences comme $O\mathit{Stop}(x,t) \wedge F\mathit{Stop}(x,t)$ en utilisant le système défini dans 7.3. La cohérence d'une réglementation est évaluée sous des *contraintes d'intégrité*, par exemple des contraintes physiques ou des contraintes du domaines.

Définition II.13 (Contraintes d'intégrité)**.** *Une contrainte d'intégrité est une formule objective fermée qui ne contient pas de d-prédicat. On note IC l'ensemble des contraintes d'intégrité.*

Nous définissons d'abord la cohérence d'une réglementation dans un *état du monde* particulier. Intuitivement, les états du monde sont des représentations syntaxiques des interprétations du premier ordre. Ils peuvent également être assimilés à des modèles de Herbrand classiques.

Définition II.14 (État du monde)**.** *Un état du monde s est un ensemble complet et cohérent de littéraux de base.*

Un état du monde est une représentation syntaxique d'une interprétation de Herbrand. Donc, pour tout symbole de prédicat n-aire Q, tous termes de base $t_1, \dots, t_n$ et tout état du monde s, soit $Q(t_1, \dots, t_n) \in s$ soit $\neg Q(t_1, \dots, t_n) \in s$ (c'est ce que nous appelons complet pour un état du monde). Dans ce qui suit, quand nous décrivons un état du monde, nous omettrons les littéraux négatifs pour plus de lisibilité.

Définition II.15. *Soient IC un ensemble de contraintes d'intégrité et s un état du monde. s est* ***cohérent*** *avec IC si et seulement si $s, IC \not\vdash \bot$.*

Définition II.16. *Soient ρ une réglementation, IC un ensemble de contraintes d'intégrité et s un état du monde cohérent avec IC. ρ est* ***cohérente par rapport à*** *IC* ***dans*** *s si et seulement si $\rho, IC, s \not\vdash \bot$.*

Exemple II.2. Reprenons l'exemple II.1. Considérons que IC contient deux contraintes :
(1) un feu tricolore a une et une seule couleur et cette couleur peut être verte, orange ou rouge,
(2) un conducteur conduit un et un seul type de véhicule

Ici, $IC = \{\forall t \ FT(t) \to Couleur(t, vert) \otimes Couleur(t, orange) \otimes Couleur(t, rouge),$
$\forall x \forall y \forall z \ Conducteur(x) \wedge Vehicule(x, y) \wedge Vehicule(x, z) \to y = z\}$ [38].

Soit s l'état du monde suivant : $\{Conducteur(A),\ FT(T), Devant(A, T),\ Vehicule(A, auto),$ $Couleur(T, rouge)\}$. s est tel que $s, IC \not\vdash \bot$.

Considérons une réglementation ρ qui contient les trois règles (r_0), (r_1) et (r_2).

Dans ce cas, $\rho, IC, s \not\vdash \bot$ (parce que le seul littéral déontique qui peut être déduit de ρ, IC et s est $OStop(A, T)$). Donc, ρ est cohérent par rapport à IC dans s.

Nous pouvons généraliser la définition de la cohérence à l'ensemble des états du monde possibles.

Définition II.17 (Cohérence d'une réglementation)**.** *Soient ρ une réglementation et IC un ensemble de contraintes d'intégrité. ρ est cohérent par rapport à IC si et seulement si pour tous les états du monde s tels que $s, IC \not\vdash \bot$ alors $\rho, IC, s \not\vdash \bot$.*

8.3 Complétude des réglementations

Informellement, une réglementation est complète dès qu'elle permet de contraindre le comportement d'un agent dans toute situation. On peut se demander si cette définition a du sens : est-ce qu'une réglementation doit prendre en compte toutes les situations possibles ? Nous jugeons que la réponse à cette question est non : la réglementation doit être complète pour certaines situations. Nous suggérons donc de définir une complétude partielle restreinte à deux formules φ et ψ : φ représente une situation particulière dans laquelle nous voulons évaluer la réglementation et ψ un symbole de prédicat gouverné par la réglementation. Nous voulons qu'une réglementation soit complète pour φ et ψ si et seulement si dans toute situation où φ est vraie, il est obligatoire (resp. permis, interdit) que ψ soit vraie.

Ceci conduit à la définition suivante :

38. L'introduction de l'égalité est faite dans le même esprit que dans [FM99].

Définition II.18. *Soient IC un ensemble de contraintes d'intégrité, ρ une réglementation cohérente par rapport à IC et s un état du monde cohérent avec IC. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral objectif, $\overrightarrow{x}$ représentant les variables libres dans $\varphi(\overrightarrow{x})$ et $\psi(\overrightarrow{x})$ signifiant que les variables libres de ψ sont un sous-ensemble de $\overrightarrow{x}$. ρ est* $\boldsymbol{(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))}$**-complète** ***par rapport à IC dans s pour*** $\vdash$ *si et seulement si pour toutes les substitutions de base χ telles que $s \vdash \varphi(\chi(\overrightarrow{x}))$:*

$$\begin{aligned} &\boldsymbol{\rho, s \vdash O\psi(\chi(\overrightarrow{x}))} \ \textit{ou} \\ &\boldsymbol{\rho, s \vdash F\psi(\chi(\overrightarrow{x}))} \ \textit{ou} \\ &\boldsymbol{\rho, s \vdash P\psi(\chi(\overrightarrow{x}))} \end{aligned}$$

Exemple II.3. Considérons l'état du monde s_0 suivant : $\{Conducteur(A), FT(T), Devant(A,T), Vehicule(A, auto), Couleur(T, rouge)\}$. Considérons ρ et IC définis dans l'exemple II.2.

s_0 est cohérent par rapport à IC et $\rho, s \vdash O(Stop(A,T))$.

Prenons $\varphi_0(x,t) \equiv FT(t) \wedge Conducteur(x) \wedge Devant(x,t)$ et $\psi_0(x,t) \equiv Stop(x,t)$.

$s_0, IC \vdash \varphi_0(A,T)$ et $\rho, IC, s_0 \vdash O(Stop(A,T))$. Donc ρ est $(\varphi_0(x,t), \psi_0(x,t))$-complète par rapport IC à s_0 pour $\vdash$.

Considérons maintenant $s_1 = \{Conducteur(A), FT(T), Devant(A,T), Vehicule(A, camion), Couleur(T, rouge)\}$. s_1 est cohérent avec IC.

$s_1, IC \vdash \varphi_0(A,T)$ mais $\rho, IC, s_1 \not\vdash O\psi_0(A,T)$, $\rho, IC, s_1 \not\vdash P\psi_0(A,T)$ et $\rho, IC, s_1 \not\vdash F\psi_0(A,T)$. Donc ρ est $(\varphi_0(x,t), \psi_0(x,t))$-incomplète par rapport à IC dans s_1 pour $\vdash$. En fait, aucune règle de la réglementation ne peut être appliquée car le véhicule n'est pas une voiture, mais un camion.

On pourrait se demander si SDL ne permet pas par nature d'obtenir des réglementations complètes. En effet, par définition des opérateurs P et F, si l'on considère une formule objective φ, alors $O\varphi$, $P\varphi$ et $F\varphi$ sont mutuellement exclusives et $O\varphi \vee P\varphi \vee F\varphi$ est une tautologie de SDL. On pourrait donc être amené à en conclure que SDL suffit par elle-même à éviter les réglementations incomplètes. Ce n'est pas le cas : en effet, si l'on considère une réglementation ρ, un état du monde s et une formule objective φ, on aura $\rho, s \vdash O\varphi \vee P\varphi \vee F\varphi$. Mais cela ne signifie pas pour autant que l'on est capable de dériver *soit* $\rho, s \vdash O\varphi$, *soit* $\rho, s \vdash P\varphi$, *soit* $\rho, s \vdash F\varphi$, ce qui nous intéresse précisément ici.

Les définitions précédentes sont généralisées comme suit :

Définition II.19 (Complétude d'une réglementation)**.** *Soient IC un ensemble de contraintes d'intégrité et ρ une réglementation. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral avec la même signification que dans la définition II.18. ρ est* $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète *par rapport à IC pour $\vdash$ ssi pour tout état du monde s cohérent avec IC, ρ est $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète par rapport à IC dans s pour $\vdash$.*

La complétude est un aspect important des réglementations. Dans une situation donnée, sans comportement stipulé, n'importe quel comportement peut être observé et les conséquences peuvent être très importantes, notamment dans les systèmes critiques. La question qui se pose alors est de savoir comment traiter les réglementations incomplètes.

Afin de pouvoir travailler avec une telle réglementation, on peut :

1. détecter les « trous » de la réglementation et les envoyer aux concepteurs de la réglementation pour qu'ils puissent la corriger ou
2. détecter les « trous » de la réglementation et appliquer sur ces trous des règles de complétion pour les corriger systématiquement.

La première solution peut être très difficile à mettre en œuvre car elle implique beaucoup de travail pour les concepteurs de la réglementation, c'est pourquoi nous choisissons la seconde solution.

Chapitre 9

Raisonner avec des réglementations incomplètes

Nous avons choisi ici de compléter systématiquement les « trous » de la réglementation en utilisant des règles par défaut. Ces règles seront appliquées lorsque l'on est incapable de déterminer si, dans une situation donnée, une proposition particulière est autorisée, interdite ou obligatoire. L'utilisation de ces règles nous permet d'obtenir une réglementation complète comme nous le montrerons plus loin.

9.1 Défauts pour la complétion des réglementations

Raisonner avec des informations incomplètes est un problème classique en logique et en Intelligence Artificielle : peut-on inférer quelque chose sur une information qui n'est pas présente dans une base de croyances ? Plusieurs approches ont été définies, mais nous nous intéressons à une en particulier, le raisonnement par défaut. Le principe du raisonnement par défaut est simple : si une information n'est pas contradictoire avec les informations qui peuvent être classiquement déduites de la base de croyances, alors on peut déduire une autre information de la base de croyances. Un exemple classique est le suivant : supposons qu'un agent croit que « *tous les oiseaux volent* », que « *les pingouins ne volent pas* » et que « *les pingouins sont des oiseaux* ». La représentation de cet ensemble de formules dans une logique du premier ordre est incohérent (un oiseau qui est également un pingouin vole et ne vole pas en même temps). En fait, la première règle « *tous les oiseaux volent* » est un défaut : « *si a est un oiseau et s'il n'est pas incohérent que a vole, alors a vole* »[39]. Si a est un pingouin, alors « a vole » ne peut pas être déduit et s'il ne peut pas être déduit que a est un pingouin, alors on peut en déduire que a vole.

La logique des défauts est une extension non-monotone de la logique du premier ordre introduite par Reiter [Rei80] pour formaliser le raisonnement par défaut. Nous suivons ici la présentation de cette logique donnée par Besnard [Bes89].

Un défaut d est une configuration $\dfrac{P : J_1, \ldots, J_n}{C}$ où $P, J_1, \ldots, J_n, C$ sont des formules du premier ordre. P est appelée le *pré-requis* de d, $J_1, \ldots, J_n$ la *justification* de d et C la *conséquence* de d. Une théorie de défauts $\Delta = (D, F)$ est composée d'un ensemble de formules fermées objectives F (les faits) et d'un ensemble de défauts. Un défaut est dit *normal* si sa justification et sa conséquence sont égales. Un défaut dont le pré-requis, la conséquence et la justification

39. Dans ce cas, l'information qui n'est pas contradictoire avec la base de croyances et la nouvelle information sont identiques.

sont des formules fermées est dit *fermé*.

Une théorie des défauts (D, F) peut être représentée par une *forme de surface* (D', F) à condition que

$$D = \{\frac{P(\overrightarrow{a}) : J_1(\overrightarrow{a}), \ldots, J_n(\overrightarrow{a})}{C(\overrightarrow{a})} \;:\; \frac{P(\overrightarrow{x}) : J_1(\overrightarrow{x}), \ldots, J_n(\overrightarrow{x})}{C(\overrightarrow{x})} \in D' \text{ et } \overrightarrow{a} \text{ est un terme de base}\}$$

et chaque élément de D' est de la forme $\frac{P(\overrightarrow{x}) : J_1(\overrightarrow{x}), \ldots, J_n(\overrightarrow{x})}{C(\overrightarrow{x})}$ où $P(\overrightarrow{x}), J_1(\overrightarrow{x}), \ldots, J_n(\overrightarrow{x})$, $C(\overrightarrow{x})$ sont des formules du premier ordre avec des variables libres apparaissant dans $\overrightarrow{x}$.

En utilisant des défauts, on obtient des *extensions*, c'est à dire des ensembles de formules qui sont déduites monotoniquement et non-monotoniquement de F. Soit $\Delta = (D, F)$ une théorie des défauts où les défauts ne contiennent que des formules closes, alors une extension de Δ est un ensemble de formules E vérifiant les conditions suivantes :

1. $F \subseteq E$
2. $Th(E) = E$ où $Th(E) = \{\varphi \;:\; E \vdash \varphi\}$
3. si $\frac{P : J_1, \ldots, J_n}{C}$ est un défaut de D, alors si $P \in E$ et J_1 est cohérent avec E, …, J_n est cohérent avec E, alors $C \in E$

Une théorie des défauts peut avoir plusieurs extensions ou aucune extension. Reiter a montré dans [Rei80] que si F est cohérent et si (D, F) a une extension, alors cette extension est cohérente. Il a également montré que toute théorie des défauts normale et fermée a au moins une extension.

Nous ne nous intéressons pas ici à la croyance d'un d-littéral ψ donnée, mais à la dérivation depuis une réglementation donnée du fait que ψ soit obligatoire, permise ou interdite (ces trois cas sont les seuls possibles à cause de l'axiome DO de O). Donc, si la réglementation est incomplète pour un d-littéral ψ (c'est-à-dire qu'on ne peut en déduire ni $O\psi$ ni $F\psi$ ni $P\psi$), alors elle ne peut être complétée qu'en supposant que $O\psi$ peut être déduite, ou $P\psi$, ou $F\psi$. Cela va nous conduire à la construction des trois ensembles de défauts présentés ci-dessous.

Dans ce qui suit, soient IC un ensemble de contraintes d'intégrité, ρ une réglementation cohérente avec IC et s un état du monde cohérent avec IC. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral vérifiant la définition II.18.

Définition II.20. *Soient $E_F(\overrightarrow{x})$, $E_P(\overrightarrow{x})$ et $E_O(\overrightarrow{x})$ trois formules objectives telles que leur ensemble respectif de variables libres est dans $\overrightarrow{x}$. On définit un ensemble de configurations comme suit :*

$$(DF_{\varphi,\psi}) \quad \frac{\varphi(\overrightarrow{x}) \wedge E_F(\overrightarrow{x}) \;:\; F\psi(\overrightarrow{x})}{F\psi(\overrightarrow{x})}$$

$$(DP_{\varphi,\psi}) \quad \frac{\varphi(\overrightarrow{x}) \wedge E_P(\overrightarrow{x}) \;:\; P\psi(\overrightarrow{x})}{P\psi(\overrightarrow{x})}$$

$$(DO_{\varphi,\psi}) \quad \frac{\varphi(\overrightarrow{x}) \wedge E_O(\overrightarrow{x}) \;:\; O\psi(\overrightarrow{x})}{O\psi(\overrightarrow{x})}$$

Nous notons $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ la théorie des défauts pour ρ, s, $\varphi(\overrightarrow{x})$ et $\psi(\overrightarrow{x})$ dont la forme de surface est donnée par $(\{DF_{\varphi,\psi}, DP_{\varphi,\psi}, DO_{\varphi,\psi}\}, \rho \cup s)$

Nous pouvons compléter une réglementation incomplète de telle sorte que $\psi(\overrightarrow{x})$ soit interdite $(DF_{\varphi,\psi})$, permise $(DP_{\varphi,\psi})$ ou obligatoire $(DO_{\varphi,\psi})$ en nous appuyant sur $E_F(\overrightarrow{x})$, $E_P(\overrightarrow{x})$ et $E_O(\overrightarrow{x})$. Nous pouvons maintenant définir une nouvelle relation d'inférence $\vdash_*$ en utilisant les règles par défaut.

Définition II.21. *Soit γ une formule de FOSDL. $\rho, s \vdash_* \gamma$ ssi $\gamma \in \bigcup E_{\Delta_{\rho,s}(\varphi(\overrightarrow{x}),\psi(\overrightarrow{x}))}$ où $\bigcup E_{\Delta_{\rho,s}(\varphi(\overrightarrow{x}),\psi(\overrightarrow{x}))}$ est l'union de toutes les extensions de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$.*

Revenons sur notre définition de $\vdash_*$. Classiquement, il y a deux façons de définir $\vdash_*$:
- *existentiellement*, c'est-à-dire que $\rho, s \vdash_* \gamma$ si et seulement si il y a *une* extension de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ qui contient γ
- *universellement*, c'est-à-dire que $\rho, s \vdash_* \gamma$ si et seulement si γ apparaît dans *toutes* les extensions de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$

Malheureusement, ces deux définitions ne nous satisfont pas. Si l'on utilise l'inférence existentielle, nous ne pouvons pas détecter de contradictions, car par exemple si $O\varphi$ est dans l'extension E_1 et $O\neg\varphi$ est dans l'extension E_2, on ne peut pas en dériver $O\varphi \wedge O\neg\varphi$ qui permettrait de détecter la contradiction. Si on utilise l'inférence universelle, nous n'obtiendrons pas une réglementation complète : s'il y a plus d'une extension, cela signifie que deux règles différentes concernant la même proposition φ peuvent être appliquées et dans ce cas, ni $O\varphi$, ni $P\varphi$, ni $F\varphi$ ne pourront être dérivées. La réglementation ne sera donc pas complète.

Pourquoi alors utiliser l'union de toutes les extensions de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$? Cela représente en fait l'ensemble qui contient toutes les obligations, permissions et interdictions qui peuvent être déduites de la réglementation (en utilisant des défauts ou pas). Tout agent appliquant la réglementation doit trouver le comportement à adopter dans cet ensemble.

On peut alors se demander si l'utilisation de l'union de toutes les extensions possibles peut mener à une contradiction. Mais, comme nous le verrons dans la section 9.2, une seule extension sera obtenue dans les cas qui nous intéressent.

L'étape suivante est de définir les conditions sous lesquelles la réglementation est complète et cohérente avec cette nouvelle inférence.

9.2 Cohérence et complétude de la réglementation complétée

Nous étendons tout d'abord les définitions II.17, II.18 et II.19 en utilisant $\vdash_*$ à la place de $\vdash$ dans ces définitions. Pour distinguer les nouvelles notions de cohérence et de complétude des anciennes, nous utiliserons $*$ comme préfixe (par exemple, nous écrirons « $*$-cohérence ») ou écrirons « pour $\vdash_*$ »(par exemple, nous écrirons « cohérent pour $\vdash_*$ »).

Définition II.22. *(Cohérence d'une réglementation pour $\vdash_*$) Soient ρ une réglementation, IC un ensemble de contraintes d'intégrité et s un état du monde cohérent avec IC. ρ est $*$-cohérente par rapport à IC dans s si et seulement si $\rho, IC, s \not\vdash_* \bot$.*

Définition II.23. *(Complétude d'une réglementation pour $\vdash_*$)*
Soient IC un ensemble de contraintes d'intégrité, ρ une réglementation cohérente par rapport à IC et s un état du monde cohérent avec IC. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral objectif, $\overrightarrow{x}$ représentant les variables libres dans φ et $\psi(\overrightarrow{x})$ signifiant que les variables libres de ψ sont un sous-ensemble de $\overrightarrow{x}$. ρ est $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète *par rapport à IC dans s pour $\vdash_*$ si et seulement si pour toutes les substitutions de base χ telles que $s \vdash \varphi(\chi(\overrightarrow{x}))$:*

$$\rho, s \vdash_* O\psi(\chi(\overrightarrow{x})) \text{ ou}$$
$$\rho, s \vdash_* F\psi(\chi(\overrightarrow{x})) \text{ ou}$$
$$\rho, s \vdash_* P\psi(\chi(\overrightarrow{x}))$$

Le résultat principal concernant la complétude et la cohérence de la réglementation obtenue en utilisant la théorie des défauts définie précédemment est présenté dans la proposition suivante.

Proposition II.2. *Considérons un ensemble de contraintes d'intégrité IC, une réglementation ρ cohérente avec IC et un état du monde s cohérent avec IC et tel que $\rho \cup s$ est cohérent. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral vérifiant la définition II.18 et $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ la théorie des défauts correspondante.*

Les propositions suivantes sont équivalentes :

1. *pour tout vecteur $\overrightarrow{d}$ de termes de base, si $s \vdash \varphi(\overrightarrow{d})$, $\rho, s \not\vdash O\psi(\overrightarrow{d})$, $\rho, s \not\vdash P\psi(\overrightarrow{d})$ et $\rho, s \not\vdash F\psi(\overrightarrow{d})$ (c'est-à-dire si ρ n'est pas $(\varphi(\overrightarrow{d}), \psi(\overrightarrow{d}))$-complète dans s), alors $s \vdash E_O(\overrightarrow{d}) \otimes E_P(\overrightarrow{d}) \otimes E_F(\overrightarrow{d})$.*
2. *ρ est cohérent et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s.*

Cette proposition caractérise les conditions nécessaires et suffisantes pour que les défauts puissent compléter de façon cohérente une réglementation incomplète. Plus précisément, cette proposition signifie que si chaque fois que la réglementation ne permet pas de dériver le comportement attendu alors un seul des E_i est vrai, alors les défauts complètent de façon cohérente la réglementation (un et un seul défaut sera appliqué pour une $\psi(\overrightarrow{d})$ particulière). On remarquera que la préservation de la cohérence d'une réglementation complétée est une propriété que l'on attend naturellement.

Remarquons que cette proposition devrait être modifiée si les contraintes d'intégrité IC pouvaient porter sur les d-prédicats. En effet, s'il y avait une contrainte du type $\psi(\overrightarrow{d}_1) \wedge \psi(\overrightarrow{d}_2) \rightarrow \neg\psi(\overrightarrow{d}_3)$, alors il serait ennuyeux d'inférer à la fois $O\psi(\overrightarrow{d}_1)$, $O\psi(\overrightarrow{d}_2)$ et $O\psi(\overrightarrow{d}_3)$ car la contrainte d'intégrité serait violée par la réglementation. Nous revenons sur cette particularité dans la conclusion de cette partie.

Exemple II.4. Considérons l'exemple précédent et $s_1 = \{Conducteur(A), FT(T), Devant(A,T), Vehicule(A, camion), Couleur(T, rouge)\}$.

ρ est incomplète dans s_1 pour $\varphi_0(x,t) \equiv Conducteur(A) \wedge FT(T) \wedge Devant(A,T)$ et $\psi_0(x,t) \equiv Stop(A,T)$ dans s_1.

Posons maintenant :

- $E_F(x,t) = Vehicule(x, camion) \wedge Couleur(t, vert)$,
- $E_P(x,t) = Vehicule(x, camion) \wedge Couleur(t, orange)$
- $E_O(x,t) = Vehicule(x, camion) \wedge Couleur(t, rouge)$.

On a alors $s_1 \vdash E_O(A,T) \wedge \neg E_F(A,T) \wedge \neg E_P(A,T)$. Donc ρ est cohérente et $(\varphi_0(x,t), \psi_0(x,t))$-complète pour $\vdash_*$ dans s_1.

Même si cette condition nécessaire et suffisante est intéressante en théorie, elle n'est pas réellement utile dans des cas concrets : pour vérifier que cette condition est satisfaite, nous devrions détecter chaque « trou » dans la réglementation. Cette détection est une opération que nous voulons éviter, car très coûteuse en terme de calculs. C'est pourquoi nous proposons des conditions plus générales qui sont suffisantes, mais non nécessaires, pour pouvoir compléter la réglementation. Nous présentons ainsi deux corollaires immédiats de la définition précédente.

Corollaire II.1. *Si*

$$s \vdash \forall \overrightarrow{x}\ (\varphi(\overrightarrow{x}) \rightarrow E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x}))$$

alors ρ est cohérente et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète par rapport à IC pour $\vdash_*$ dans s.

Exemple II.5. Considérons $s_2 = \{Conducteur(A), FT(T), Devant(A,T), Vehicule(A, velo), Couleur(T, rouge)\}$.

s_2 est cohérent avec IC. Considérons la réglementation définie dans l'exemple II.1.

Cette fois, posons :
- $E_F(x,t) = Couleur(t, vert)$,
- $E_P(x,t) = Couleur(t, orange)$
- et $E_O(x,t) = Couleur(t, rouge)$.

Dans ce cas, $s_2 \vdash E_O(A,T) \wedge \neg E_P(A,T) \wedge \neg E_F(A,T)$. Donc ρ est $*$-cohérente et $*$-complète pour $\varphi_0(x,t)$ et $\psi_0(x,t)$ dans s_2.
Mais on a également $s_1 \vdash E_O(A,T) \wedge \neg E_P(A,T) \wedge \neg E_F(A,T)$, donc ρ est $*$-cohérente et $(\varphi_0(x,t), \psi(x,t))$-complète pour $\vdash_*$ dans s_1. Ces E_i plus généraux permettent d'avoir une réglementation complète pour tout type de véhicule.

Corollaire II.2. *Si*
$$IC \vdash \forall \overrightarrow{x}\ (E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x}))$$
alors ρ est cohérent et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète par rapport à IC pour $\vdash_*$.

Exemple II.6. $IC \vdash \forall t\ Couleur(t, rouge) \otimes Couleur(t, vert) \otimes Couleur(t, orange)$.
Donc ρ est $*$-cohérente et $(\varphi_0(x,t), \psi_0(x,t))$-complète pour $\vdash_*$.
IC spécifie qu'un feu tricolore n'a qu'une et une seule couleur parmi *Rouge*, *Orange* et *Vert*. S'il y a un E_i pour chaque couleur, nous sommes sûrs que quelque soit la situation, nous pouvons appliquer un et un seul défaut s'il y a un « trou » dans la réglementation.

9.3 Cas particuliers

Une autre solution serait de prendre des E_i fixés. Par exemple, nous pourrions prendre un E_i égal à $\top$ et les deux autres à $\bot$. On peut distinguer trois cas.

Supposons que $E_F \equiv \top$, $E_P \equiv \bot$ et $E_O \equiv \bot$.
Dans ce cas, tout ce qui n'est pas spécifié comme obligatoire ou permis par la réglementation est interdit. Ce comportement strict peut être observé par exemple dans les réglementations qui régissent un système hautement sécurisé où chaque action doit être explicitement autorisé avant d'être exécutée ;

Supposons que $E_F \equiv \bot$, $E_P \equiv \top$ et $E_O \equiv \bot$.
Nous sommes ici dans la situation opposée : tout ce qui n'est pas obligatoire ou interdit est permis. Ce comportement « tolérant » peut être observé par exemple pour des réglementations pour certains systèmes faiblement sécurisés dans lesquels tout ce qui n'est pas obligatoire ou interdit est implicitement permis ; remarquons que ce cas correspond au *sealing principle*, mentionné dans l'état de l'art, qui stipule que tout ce qui n'est pas interdit est permis.

Supposons que $E_F \equiv \bot$, $E_P \equiv \bot$ et $E_O \equiv \top$.
Dans ce cas, chaque action qui n'est pas interdite ou permise doit obligatoirement être exécutée.

Chapitre 10

Politiques d'échange d'informations

Une politique d'échange d'informations est une réglementation qui contraint le comportement des agents d'un système multi-agents en ce qui concerne la communication des informations. Pour décrire de telles politiques, nous avons besoin de symboles de prédicats : *Agent*, *Info*, *Recoit*, *Theme*, *Manager*, *Employe* sont des f-prédicats et *Informe* est un d-prédicat.

- $Agent(x)$ signifie que x est un agent,
- $Info(i)$ signifie que i est une information,
- $Recoit(x, i)$ signifie que l'agent x reçoit l'information i,
- $Theme(i, t)$ signifie que l'information i traite du sujet t,
- $Manager(x)$ signifie que x est un manager,
- $Employe(x)$ signifie que x est un employé,
- $Informe(x, i, y)$ signifie que l'agent x informe l'agent y de l'information i.

Pour l'exemple avec lequel nous illustrons ce chapitre, nous définissons également les constantes a, b, i_1, *VerifMat* (qui signifie que le matériel est vérifié), *RiskExpl* (qui signifie qu'il y a un risque d'explosion), *Reunion* (qui signifie qu'il y a une réunion) et *MatHS* (qui signifie que le matériel est hors d'usage).

La cohérence de telles politiques est définie par la définition II.16 et leur complétude est définie en instanciant la définition II.18 avec les formules spécifiques suivantes :

$$\varphi(x, i, y) \equiv Agent(x) \wedge Info(i) \wedge Recoit(x, i) \wedge Agent(y) \wedge \neg(x = y)$$
$$\psi(x, i, y) \equiv Informe(x, i, y)$$

Cela conduit à la définition suivante :

Définition II.24. *Soient IC un ensemble de contraintes d'intégrité, s un état du monde cohérent avec IC et ρ une réglementation cohérente dans s par rapport à IC. ρ est complète par rapport à IC dans s pour $\vdash$ si et seulement si pour toute substitution de base χ telle que $s \vdash Agent(\chi(x)) \wedge Info(\chi(y)) \wedge Recoit(\chi(x), \chi(i)) \wedge Agent(\chi(y)) \wedge \neg(\chi(x) = \chi(y))$:*

$$\rho, s \vdash O Informe(\chi(x), \chi(i), \chi(y)) \ ou$$
$$\rho, s \vdash F Informe(\chi(x), \chi(i), \chi(y)) \ ou$$
$$\rho, s \vdash P Informe(\chi(x), \chi(i), \chi(y))$$

Les défauts sont donc les suivants :

$$(DF_{\varphi,\psi}) \quad \frac{\varphi(x,i,y) \wedge E_F(x,i,y) : FInforme(x,i,y)}{FInforme(x,i,y)}$$

$$(DP_{\varphi,\psi}) \quad \frac{\varphi(x,i,y) \wedge E_P(x,i,y) : PInforme(x,i,y)}{PInforme(x,i,y)}$$

$$(DO_{\varphi,\psi}) \quad \frac{\varphi(x,i,y) \wedge E_O(x,i,y) : OInforme(x,i,y)}{OInforme(x,i,y)}$$

Les résultats donnés dans la section 9 sont valides. En particulier, nous pouvons toujours définir trois cas :

- $E_F \equiv \top$, $E_P \equiv \bot$ et $E_O \equiv \bot$.
  Cela s'applique à des systèmes multi-agents hautement sécurisés dans lesquels toute action de communication doit être explicitement obligatoire ou permise pour être exécutée ;
- $E_F \equiv \bot$, $E_P \equiv \top$ et $E_O \equiv \bot$.
  Ce cas s'applique à des systèmes faiblement sécurisés dans lesquels toute action de communication qui n'est pas explicitement interdite est autorisée ;
- $E_F \equiv \bot$, $E_P \equiv \bot$ et $E_O \equiv \top$.
  Dans ce cas, à moins que cela ne soit explicitement mentionné, l'envoi d'information est obligatoire.

Pour illustrer cela, considérons l'exemple d'une entreprise dans laquelle il y a un manager et deux employés. Considérons la politique π_0 avec une seule règle précisant que « *les managers ne doivent pas informer les employés des informations concernant la vérification des équipements* ». Cette règle est modélisée par :

$$\forall x \forall i \forall y \ Manager(x) \wedge Employe(y) \wedge Recoit(x,i) \wedge Theme(i, VerifMat) \rightarrow FInforme(x,i,y)$$

Considérons $IC = \emptyset$ (il n'y a pas de contraintes d'intégrité). Soit s_0 l'état du monde $\{Agent(a), Agent(b), Manager(a), Employee(b), Info(i_1), Theme(i_1, RiskExpl), Recoit(a, i_1)\}$.

Dans cette situation, a est un manager et b un employé. a a reçu une information i_1 dont le sujet est « risque d'explosion ».

Comme π_0 ne contient qu'une et une seule règle et s_0 est cohérent avec IC, π_0 est cohérente dans s_0.

Nous avons $s_0 \vdash Agent(a) \wedge Info(i_1) \wedge Recoit(a, i_1) \wedge Agent(b) \wedge \neg(a = b)$ mais $\pi_0, s_0 \not\vdash O(Informe(a, i_1, b))$ et $\pi_0, s_0 \not\vdash P(Informe(a, i_1, b))$ et $\pi_0, s_0 \not\vdash F(Informe(a, i_1, b))$. Donc π_0 est incomplète pour $\vdash$.

L'incomplétude vient du fait que la politique contraint le comportement du manager si il ou elle reçoit une information à propos de la vérification des équipements, mais elle ne dit rien par rapport aux informations concernant les risques d'explosion.

Pour pouvoir compléter la politique précédente, nous pourrions prendre

- $E_F(x,y,i) \equiv Theme(i, VerifMat)$,
- $E_P(x,y,i) \equiv \bot$ et
- $E_O(x,y,i) \equiv Theme(i, RiskExpl)$.

Un tel choix oblige le manager à dire à ses employés les informations concernant les risques d'explosion. On peut vérifier que π_0 est cohérente et complète pour $\vdash_*$ dans s_0 pour $\varphi(x,i,y)$ et $\psi(x,i,y)$.

Considérons maintenant que IC contient la contrainte « *une information a un et un seul sujet et ce sujet peut être VerifMat, RiskExpl, Réunion ou MatHS* ». Prenons

$$
\begin{aligned}
E_F(x, y, i) &\equiv \mathit{Theme}(i, \mathit{VerifMat}) \vee \mathit{Theme}(i, \mathit{Meeting}) \\
E_P(x, y, i) &\equiv \mathit{Theme}(i, \mathit{MatHS}) \\
E_O(x, y, i) &\equiv \mathit{Theme}(i, \mathit{RiskExpl})
\end{aligned}
$$

on peut alors appliquer le corollaire II.2 pour conclure que π_0 est $*$-complète et $*$-cohérente pour $\varphi(x, i, y)$ et $\psi(x, i, y)$.

Chapitre 11

Conclusion et perspectives

11.1 Conclusion

Dans cette partie, nous nous sommes intéressés à l'analyse de la cohérence et de la complétude de réglementations qui peuvent exister dans des systèmes d'agents pour contraindre leur comportement.

Plus précisément, nous avons défini une logique déontique du premier ordre et avons montré comment exprimer une réglementation dans ce cadre. Nous avons alors donné une définition de la cohérence et de la complétude d'une réglementation. La complétude d'une réglementation est définie pour deux formules φ et ψ qui représentent la situation et l'action pour laquelle la réglementation doit être complète.

Nous nous sommes également intéressés aux réglementations incomplètes et avons proposé une méthode basée sur les défauts de Reiter pour compléter ces réglementations. Nous avons établi plusieurs résultats qui montrent quand ces défauts peuvent compléter de façon cohérente une réglementation.

Finalement, nous appliquons ces différents résultats aux réglementations particulières que sont les politiques d'échange d'informations. Nous proposons notamment des formules φ et ψ pour lesquelles l'étude de la complétude est pertinente dans le cas des politiques d'échange.

Notons que ce travail a donné lieu aux publications suivantes : [CR07b, CR07a, CR08a, CR08b, GRC09b, GRC09a, GRC10].

11.2 Limites

Le fait que les contraintes d'intégrité ne puissent pas porter sur les prédicats régis par la réglementation constitue la limite principale de ce travail. Par exemple, cela signifie que l'on ne peut pas contraindre (par IC) l'envoi d'information comme « À un moment donné, un et un seul agent envoie une information à un autre ». Pour pallier ce problème, il serait possible de considérer que les contraintes d'intégrité doivent être vraies dans tous les mondes et donc également dans les mondes « idéaux » selon la réglementations. En d'autres termes, nous pourrions rendre obligatoires les contraintes d'intégrité.

Nous avons développé ici un modèle très simple des notions déontiques en utilisant SDL et nombre de problèmes classiques en logique déontique ne sont pas traités ici : normes avec exceptions, contrary-to-duties, obligations collectives etc. Il faudrait donc définir une logique qui permettrait de traiter ces problèmes.

11.3 Perspectives

Ce travail peut donner lieu à de nombreuses extensions.

Tout d'abord, l'étude du lien formel existant entre la complétude introduite dans cette partie et la notion de complétude locale développée par Reiter[Rei92] et présentée dans l'état de l'art (6.2.4) constitue une extension intéressante de ce travail.

Puis, nous pourrions différencier les obligations (respectivement permissions, interdictions) explicites des obligations (respectivement permissions et interdictions) implicites. Les obligations explicites seraient celles qui apparaissent en tant que telles dans la réglementation, alors que les obligations implicites seraient celles qui sont déduites des règles par défauts. Un lien formel pourrait alors être établi entre la définition que nous proposons dans ce travail et la définition de la complétude selon [CD97].

Nous n'avons pas travaillé sur la complexité éventuelle de la recherche de la « bonne » formule φ pour pouvoir compléter localement une réglementation. Pour certains domaines d'application précis elle reste assez facile à déterminer : dans le cas des politiques d'échanges d'information présenté en section 10, il paraît naturel de choisir le fait qu'un agent reçoive une information. Cependant, il peut être difficile de choisir pour φ une formule intéressante, c'est à dire une formule permettant de couvrir suffisamment de cas intéressants.

Une question naturelle sur laquelle nous ne nous sommes pas penchés concerne la détection automatique des « trous » d'une réglementation et la génération automatique des règles de défauts permettant de compléter une réglementation. Pour le premier point, la tâche semble difficile : la définition que nous proposons de la complétude est en effet relative à une proposition donnée (un contexte). Pour détecter les trous d'une réglementation, il nous faudrait donc balayer tous les contextes possibles. Pour le second point, nous proposons des définitions très simples de E_O, E_F et E_P en utilisant seulement $\top$ et $\bot$. Ces règles ne permettent pas à un concepteur de réglementation de compléter plus finement la réglementation suivant les cas. Il faudrait pouvoir construire automatiquement, à partir des contextes amenant des incomplétudes dans la réglementation, des règles de défauts plus « intelligentes » que celles proposées précédemment.

Enfin, pour pouvoir travailler avec des réglementations plus générales, ce travail doit être étendu, en particulier en considérant des notions comme le temps et l'action. Comme montré dans [DBL06], la question du temps est très importante lorsque l'on modélise des réglementations et nous devrons considérer différents types de temps parmi lesquels, au moins, le temps de validité des normes et les *deadlines* imposées sur les obligations. Dans beaucoup d'exemples présentés ici, les prédicats concernés par les opérateurs déontiques représentent des actions (informer, stopper, ...). L'addition d'un opérateur modal dynamique et/ou temporel pourrait être intéressant. Nous obtiendrions alors une logique multimodale avec une très bonne expressivité. Par exemple, avec une logique modale temporelle de type LTL (Linear Temporal Logic), l'opérateur *until* $\mathcal{U}$ permettrait d'exprimer que certaines obligations (ou permissions ou interdictions) sont valables jusqu'à un certain point dans le temps. Par exemple, il est obligatoire pour l'agent x d'envoyer l'information i tant que x la juge valide s'exprimerait par $Odire(x,i,y)\mathcal{U}Valide(x,i)$. L'addition d'un opérateur dynamique permettrait quant à lui d'exprimer qu'après certaines actions, les obligations (ou permissions ou interdictions) peuvent être modifiées ou apparaître ou même disparaître. Par exemple, après avoir dit une information à un agent y, un agent x n'est plus obligé de la lui transmettre. $[dire(x,i,y)]\neg Odire(x,i,y)$.

On pourrait également utiliser la théorie des positions normatives pour pouvoir raisonner sur les actions comme présenté dans [JS92].

Troisième partie

Agent coopératif et agent obéissant

Introduction

En première partie de ce manuscrit, nous avons proposé des caractérisations d'agent coopératif. Un agent est coopératif envers un autre s'il lui transmet les informations qu'il juge être les plus utiles pour l'autre. Les agents coopératifs peuvent être perçus comme des agents qui suivent une politique de coopération. Cette politique préconise qu'il faut transmettre uniquement les informations les plus utiles pour les autres. Il est possible de voir cette politique de coopération comme une politique d'échange d'informations qui indique quels échanges sont obligatoires, interdits ou permis pour être coopératif.

Comme nous l'avons vu en partie II, les échanges d'information sont en général réglementés par une politique d'échange d'informations.

Les agents peuvent donc être à la fois soumis à la politique de coopération et à une politique d'échange d'informations. Évidemment, il est possible que les deux politiques préconisent des attitudes contradictoires vis-à-vis de l'échange d'une information. Par exemple, un agent, pour être coopératif, doit transmettre une information à un autre agent mais selon la politique d'échange d'information, cet échange est interdit. Dans ce cas, peut-on dire que l'agent n'est pas coopératif alors qu'il ne fait que respecter la politique d'échange à laquelle il est soumis ? Nous redéfinissons alors ce que signifie « un agent est coopératif vis-à-vis d'un autre »dans ce cas.

Remarquons que de nombreux travaux ont été réalisés sur la résolution de conflits entre des réglementations (par exemple [CC99]). Dans cette partie, nous ne proposons pas une nouvelle méthode de résolution de conflit mais nous analysons l'impact des conflits sur les définitions d'agent coopératif et d'agent obéissant.

Dans un premier temps, nous présentons un cadre qui nous permet de traiter à la fois l'aspect coopératif des agents (et l'utilité de l'information) et l'aspect déontique. Puis, nous modélisons les politiques d'échange d'informations dans ce cadre. Nous proposons deux politiques d'échange d'informations (une stricte et une faible) pour modéliser la coopération. Nous étudions ensuite le cas où l'agent est soumis à une politique ρ et à une politique d'échange représentant la coopération. Nous définissons alors le fait qu'un agent soit coopératif et qu'il soit obéissant dans ce contexte.

Chapitre 12

Cadre

Nous voulons un cadre dans lequel nous pouvons à la fois caractériser les informations qu'un agent juge être les plus utiles pour un autre et exprimer que les échanges des agents sont soumis à une politique d'échange d'informations. Le cadre formel développé dans la partie I est une logique modale propositionnelle dont les modalités sont la croyance et le désir. Dans la deuxième partie, le cadre utilisé pour la modélisation des réglementations est une logique déontique du premier ordre. Le cadre que nous utilisons ici est donc un hybride des deux cadres présentés précédemment. Néanmoins, nous ne voulons pas, dans cette partie, rentrer dans des détails techniques. C'est pourquoi nous présentons rapidement les modalités dont nous voulons disposer. Nous ne nous attardons pas sur l'axiomatique et la sémantique de ces modalités et préférons approfondir, de manière informelle, les résultats de cette partie.

Nous nous basons sur une logique multimodale propositionnelle. Nous ne représentons pas ici les différents éléments propositionnels du langage. Le lecteur pourra éventuellement se référer à la section 2.2.1.

Pour chaque agent a de $\mathcal{A}$, nous définissons une modalité B_a et une modalité $Recoit_a$ et pour chaque couple d'agents (a, b) nous définissons une modalité $Inf_{a,b}$.

Nous voulons représenter le fait que les agents ont des obligations (et des permissions et interdictions) qui peuvent provenir de politiques différentes. Plus précisément, nous supposons que les agents sont soumis à une politique d'échange d'information ρ, telle que nous en avons vues en deuxième partie de ce manuscrit. Ainsi, lorsqu'une modalité déontique est déduite à partir de la politique d'échange, nous l'indexons par ρ. Nous étudions également deux politiques qui représentent la coopération subjective (politique stricte et politique faible [40]). Lorsque la modalité déontique provient de la politique de coopération strict, nous l'indexons par S et lorsqu'elle provient de la politique de coopération faible, nous l'indexons par f. Pour chaque agent a de $\mathcal{A}$ et pour tout i de $\{\rho, S, f\}$, nous définissons donc une modalité O_a^i.

De même, nous ne redéfinissons pas les formules propositionnelles mais ne détaillons que les formules dans lesquelles apparaissent des modalités. Si φ est une formule et a et b deux agents de $\mathcal{A}$, alors

- $O_a^i\varphi$ est une formule qui se lit « L'agent a a l'obligation de φ par la politique i »
- $B_a\varphi$ est une formule qui se lit « L'agent a croit que φ »
- $Recoit_a\varphi$ est une formule qui se lit « L'agent a reçoit l'information φ »
- $Inf_{a,b}\varphi$ est une formule qui se lit « L'agent a informe l'agent b de φ »

Soient a et b deux agents. Soit $\mathcal{O}$ l'ensemble des formules objectives. Nous considérons l'en-

40. Notons que le nom de politique faible ne vient pas de ce que nous avons appelé coopération faible 4.2.3 mais du fait que la transformation de la coopération en politique d'échange est moins stricte.

semble $\mathcal{B}m_{a/b} = \{\varphi | \exists Q \in \mathcal{O}, Bm^Q_{a/b}\varphi\}$ qui est l'ensemble de toutes les formules que b juge être les plus utiles pour a, pour tout besoin qu'il puisse avoir. Nous supposons que cet ensemble existe et a été calculé. Ainsi, si φ est une formule de cet ensemble $\mathcal{B}m_{a/b}$, alors nous le notons $Bm_{a/b}\varphi$.

Les modalités de permission et d'interdiction sont définies en fonction de l'obligation de la façon suivante : Pour chaque agent a de $\mathcal{A}$ et pour tout i de $\{\rho, S, f\}$,

- $P^i_a\varphi \equiv \neg O^i_a\varphi \wedge \neg O^i_a\neg\varphi$
- $F^i_a\varphi \equiv O^i_a\neg\varphi$

12.1 Modélisation d'une politique d'échange

Dans cette section, nous modélisons une politique d'échange d'information dans le cadre propositionnel.

Définition III.1. *Une règle d'une politique d'échange est une formule de la forme $l_1 \vee \ldots \vee l_n$, avec $n > 1$ telle qu'il existe deux agents a et b de $\mathcal{A}$ pour lesquels on a :*

- *l_n est de la forme $O^i_a Inf_{a,b}\varphi$ ou de la forme $O^i_a\neg Inf_{a,b}\varphi$, φ étant une formule, i appartenant à l'ensemble $\{\rho, S, f\}$.*
- *il existe $j < n$ tel que $l_j = \neg Recoit_a\varphi$*
- *pour $j < n$, l_j est une formule.*

Comme dans la partie II, nous utilisons le raccourci d'écriture $l_1 \vee \ldots \vee l_{n-1} \vee P^i_a Inf_{a,b}\varphi$ à la place des deux règles $l_1 \vee \ldots \vee l_{n-1} \vee O^i_a Inf_{a,b}\varphi$ et $l_1 \vee \ldots \vee l_{n-1} \vee O^i_a\neg Inf_{a,b}\varphi$.

Une politique d'échange d'informations est un ensemble de règles d'échange.

Dans cette partie, nous considérons une politique d'échange particulière ρ que les agents doivent respecter. Nous supposons que cette politique est composée de trois règles qui sont : pour tout agent a et b de $\mathcal{A}$,

- $Recoit_b\varphi \wedge C_O \rightarrow O^\rho_b Inf_{b,a}\varphi$
- $Recoit_b\varphi \wedge C_P \rightarrow P^\rho_b Inf_{b,a}\varphi$
- $Recoit_b\varphi \wedge C_F \rightarrow F^\rho_b Inf_{b,a}\varphi$

Nous supposons que les trois formules C_O, C_F et C_P sont telles que $C_O \otimes C_F \otimes C_P$, c'est-à-dire qu'une et une seule de ces formules est vraie. Ainsi, la politique d'échange d'informations ρ est cohérente, c'est-à-dire qu'elle ne mène pas à une contradiction [41]. Elle est également *Recoit-Inf*-complète, car pour toute information φ telle que $Recoit_a\varphi$, au moins un des C_O, C_F, C_P est vraie donc au moins une règle peut être appliquée.

12.2 La coopération subjective : une politique particulière

12.2.1 Politique de coopération stricte

Dans cette section, nous récrivons la coopération subjective sous la forme d'une politique d'échange d'information. Nous ne nous intéressons qu'à la première définition de la coopération (coopération subjective, définition I.31 [42]).

41. Pour avoir une contradiction, il faudrait avoir à la fois C_O et C_F, ou C_O et C_P, ou C_P et C_F, ce qui n'est pas possible.

42. Nous utiliserons le terme de coopération à la place de coopération subjective.

Un agent coopératif envoie aux autres les informations qu'il juge être les plus utiles pour eux et n'envoie pas les autres informations. Être coopératif peut donc être perçu comme être respectueux d'une politique d'échange qui stipule que si l'information est jugée des plus utiles, alors c'est une obligation de la faire suivre, sinon c'est une interdiction de la faire suivre.

Formellement, pour tout agent a et b de $\mathcal{A}$

- $Recoit_b\varphi \wedge Bm_{a/b}\varphi \rightarrow O_b^S Inf_{b,a}\varphi$
- $Recoit_b\varphi \wedge \neg Bm_{a/b}\varphi \rightarrow F_b^S Inf_{b,a}\varphi$

Nous pouvons remarquer que cette politique est cohérente et *Recoit-Inf*-complète. En effet, pour toute information φ telle que $Recoit_b\varphi$ est vraie, on peut appliquer une et une seule des deux règles.

On peut remarquer que la permission n'intervient dans aucune des règles de cette politique.

12.2.2 Politique de coopération faible

Dans cette section, nous proposons une interprétation plus faible de la coopération. Plus précisément, nous divisons les informations que l'agent ne croit pas être des plus utiles en deux catégories : les informations que l'agent croit ne pas être utiles et les informations pour lesquelles l'agent ne sait pas si elles sont des plus utiles ou si elles sont inutiles. Dans cette deuxième catégorie, on trouve par exemple, si b est celui qui transmet l'information à a, des informations qui répondent à un besoin pour lequel b ne sait pas si c'est un besoin ou non de l'agent a ($\neg B_b D_a Bif_a Q \wedge \neg B_b \neg D_a Bif_a Q$).

Ainsi, si un agent juge que l'information fait partie des plus utiles pour un autre, alors il a l'obligation de la lui transmettre. S'il pense qu'elle est inutile pour l'autre, alors il a l'interdiction de la lui transmettre. S'il ne sait pas si cette information est utile ou non, alors il a la permission de la transmettre.

Nous introduisons donc l'ensemble $\mathcal{N}_{b/a} = \{\varphi | \forall Q \in \mathcal{O}, B_b \neg U_a^Q \varphi\}$ des informations que l'agent juge ne pas être utiles pour a pour n'importe quel besoin qu'il puisse avoir. Si φ est une information de cet ensemble, nous le notons $N_{b/a}\varphi$.

La politique de coopération faible est formalisée par la politique d'échange suivante : pour tout agent a et b de $\mathcal{A}$,

- $Recoit_b\varphi \wedge Bm_{a/b}\varphi \rightarrow O_b^f Inf_{b,a}\varphi$
- $Recoit_b\varphi \wedge N_{a/b}\varphi \rightarrow F_b^f Inf_{b,a}\varphi$
- $Recoit_b\varphi \wedge \neg Bm_{a/b}\varphi \wedge \neg N_{a/b}\varphi \rightarrow P_b Inf_{b,a}\varphi$

Cette fois encore, la politique est cohérente et *Recoit-Inf* complète.

La différence avec la première interprétation de la coopération en terme de politique d'échange est l'introduction d'une règle permissive qui correspond à la « marge de manœuvre » laissée à l'agent lorsqu'il ne sait pas juger de l'utilité d'une information pour un autre. Il est alors libre d'échanger cette information avec l'agent en question.

12.3 Agent en accord avec une politique d'échange

Nous caractérisons le fait qu'un agent soit en accord ou non avec une politique d'échange.

Définition III.2. *Soient a et b deux agents de $\mathcal{A}$. On dit que l'agent a est en accord avec une politique d'échange d'information i pour l'information φ (avec $i \in \{\rho, S, f\}$) si et seulement si :*

- $O_a^i Inf_{a,b}\varphi \rightarrow Inf_{a,b}\varphi$
- $F_a^i Inf_{a,b}\varphi \rightarrow \neg Inf_{a,b}\varphi$

Un agent est en accord avec une politique d'échange d'information pour une information φ si et seulement s'il transmet φ s'il a une obligation d'envoi par cette politique et ne la transmet pas s'il a une interdiction par cette politique.

Un agent a est donc considéré en désaccord avec une politique d'échange d'information pour l'information φ lorsqu'il transmet φ alors qu'il a une interdiction d'échange et lorsqu'il ne la transmet pas lorsqu'il a une obligation d'échange.

On remarquera qu'un agent ne peut pas être en désaccord avec les règles permissives. En effet, quelle que soit son attitude vis-à-vis de l'envoi d'une information pour laquelle il a une permission, celle-ci est autorisée par la politique.

Nous noterons $pol_a\varphi$ (resp. $coop^S_a\varphi$ et $coop^f_a\varphi$) le fait que l'agent a soit en accord avec la politique d'échange d'information ρ (resp. la politique de coopération stricte et la politique de coopération faible).

Nous pouvons remarquer que si l'agent est en accord avec la politique de coopération stricte alors il est en accord avec la politique de coopération faible. En effet, si l'agent croit que l'information n'est pas utile, alors il ne croit pas que cette information fasse partie des informations les plus utiles ($N_{a/b}\varphi \to \neg Bm_{a/b}\varphi$).

Chapitre 13

En accord avec la politique d'échange ρ et la politique de coopération

Les agents peuvent être soumis à plusieurs politiques. Nous considérons successivement le cas où les agents sont à la fois soumis à la politique d'échange ρ et à la politique de coopération stricte S puis le cas où ils sont soumis à la politique ρ et à la politique de coopération faible f.

13.1 Politique de coopération stricte

Dans cette section, nous supposons que les agents sont soumis à la politique d'échange d'informations ρ et à la politique de coopération stricte. Nous étudions tout d'abord le cas où l'agent donne la priorité à la politique d'échange d'information ρ et regardons dans quel cas nous pouvons qualifier l'agent de coopératif. Nous étudions ensuite le deuxième cas où l'agent donne la priorité à la politique de coopération. Finalement, nous étudions le cas général où a priori, aucune priorité n'est donnée.

13.1.1 Priorité pour la politique d'échange d'informations

On étudie le cas où lorsqu'il y a conflit, l'agent donne priorité à la politique d'échange d'informations, c'est-à-dire qu'il privilégie le fait d'être en accord avec la politique d'échange ρ au fait d'être en accord avec la politique de coopération. Par exemple, lorsque la politique d'échange d'informations prescrit une obligation d'échange, alors l'agent échange l'information, quelle que soit l'attitude prescrite par la politique de coopération. Si la politique de coopération prescrit également une obligation, alors l'agent est en accord avec cette politique (c'est-à-dire que l'on a $coop_a^S\varphi$), sinon il est en désaccord ($\neg coop_a^S\varphi$). Nous étudions les différents cas possibles et présentons le résultat sur la figure 13.1.

		Politique d'échange ρ		
		$O_b^\rho Inf_{a,b}\varphi$	$F_b^\rho Inf_{a,b}\varphi$	$P_b^\rho Inf_{a,b}\varphi$
Coopération Stricte	$O_b^S Inf_{a,b}\varphi$	$coop_a^S\varphi$	$\neg coop_a^S\varphi$	$\mathbf{coop_a^S}\varphi \leftrightarrow \mathbf{Inf_{a,b}}\varphi$
	$F_b^S Inf_{a,b}\varphi$	$\neg coop_a^S\varphi$	$coop_a^S\varphi$	$\mathbf{coop_a^S}\varphi \leftrightarrow \neg\mathbf{Inf_{a,b}}\varphi$

FIGURE 13.1 – Coopérativité stricte - Priorité pour la politique d'échange d'informations.

Lorsque la politique d'échange d'information donne la permission à l'agent d'échanger une information, celui-ci est libre d'échanger ou non cette information. Il peut alors choisir d'être coopératif ou de ne pas l'être. Dans ce cas précis, lorsque l'agent choisit d'être en accord avec la politique de coopération, nous disons qu'il est **strictement coopératif sous contrainte**.

13.1.2 Priorité pour la politique de coopération

On procède de même que dans la section précédente mais on donne cette fois la priorité à la politique de coopération. Cette fois, nous regardons donc si l'agent est en accord avec la politique d'échange d'information ρ, c'est-à-dire nous regardons si $pol_a\varphi$ ou si $\neg pol_a\varphi$. Le résultat est présenté sur le tableau 13.2.

	$O_b^\rho Inf_{a,b}\varphi$	$F_b^\rho Inf_{a,b}\varphi$	$P_b^\rho Inf_{a,b}\varphi$
$O_b^S Inf_{a,b}\varphi$	$pol_a\varphi$	$\neg pol_a\varphi$	$pol_a\varphi$
$F_b^S Inf_{a,b}\varphi$	$\neg pol_a\varphi$	$pol_a\varphi$	$pol_a\varphi$

FIGURE 13.2 – Coopérativité stricte - Priorité pour la politique de coopération.

Dans ce cas, on remarque que la politique de coopération impose complètement à l'agent l'accord ou le désaccord avec la politique d'échange d'information. L'agent n'a pas de marge de manœuvre et ne peut à aucun moment choisir d'être en accord avec la politique ρ.

13.1.3 Cas général

Les tableaux de la figure 13.3 synthétisent les deux cas étudiés précédemment. Comme nous ne donnons pas de priorité à aucune des deux politiques, nous étudions à chaque fois le cas où il échange l'information et le cas où il ne l'échange pas.

	$O_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^S\varphi$	$\neg pol_a\varphi \wedge \neg coop_a^S\varphi$
$F_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge \neg coop_a^S\varphi$	$\neg pol_a\varphi \wedge coop_a^S\varphi$

	$F_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^S Inf_{a,b}\varphi$	$\neg pol_a\varphi \wedge coop_a^S\varphi$	$pol_a\varphi \wedge \neg coop_a^S\varphi$
$F_b^S Inf_{a,b}\varphi$	$\neg pol_a\varphi \wedge \neg coop_a^S\varphi$	$pol_a\varphi \wedge coop_a^S\varphi$

	$P_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^S\varphi$	$pol_a\varphi \wedge \neg coop_a^S\varphi$
$F_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge \neg coop_a^S\varphi$	$pol_a\varphi \wedge coop_a^S\varphi$

FIGURE 13.3 – Coopérativité stricte - Cas général

Sur le troisième tableau, qui correspond au cas où la politique d'échange est permissive pour φ, on voit que l'agent est toujours en accord avec la politique ρ. Dans ce cas et dans ce cas seulement, la politique de coopération est indépendante de la politique ρ. Son choix de transmettre l'information va donc déterminer son accord ou son désaccord avec la politique de coopération. C'est ainsi que l'on peut caractériser le fait qu'un agent soit coopératif ou non.

En ce qui concerne la réciproque, il n'y a aucun cas ou l'agent peut choisir d'être en accord avec la politique d'échange d'informations indépendamment du fait d'être en accord avec la politique de coopération.

13.1.4 Coopératif et obéissant sous contrainte

Nous pouvons alors proposer une nouvelle définition du caractère coopératif d'un agent dans ce cas.

Définition III.3. *Lorsque l'agent est à la fois soumis à la politique de coopération strict et à une politique d'échange d'informations, il est dit* ***strictement coopératif sous contrainte*** *pour l'information φ si et seulement si :*

- *s'il donne la priorité à la politique de coopération, alors il est coopératif si et seulement s'il transmet φ si et seulement si $O_b^S Inf_{b,a}\varphi$.*
  Formellement, il faut que $Inf_{b,a}\varphi \leftrightarrow O_b^S Inf_{b,a}\varphi$ soit vrai.
- *sinon, lorsque la politique d'échange d'information ρ permet l'échange de φ alors l'agent transmet φ si et seulement si $O_b^S Inf_{b/a}\varphi$.*
  Formellement, il faut que $P_b^\rho Inf_{b,a}\varphi \rightarrow (O_b^S Inf_{b/a}\varphi \leftrightarrow Inf_{b,a}\varphi)$ soit vrai.

On peut également définir la notion d'obéissant sous contrainte :

Définition III.4. *Lorsque l'agent est à la fois soumis à la politique de coopération strict et à la politique d'échange d'informations, il est dit* ***strictement obéissant sous contrainte*** *si et seulement si il donne la priorité à la politique d'échange d'informations et qu'il est obéissant selon la définition qui avait été donnée sans présence de politique de coopération stricte.*

13.2 Politique de coopération faible

Nous supposons ici que les agents sont soumis à la politique d'échange d'information ρ et à la politique de coopération faible.

Nous pourrions, comme dans la section précédente, tout d'abord étudier le cas où l'agent donne priorité à la politique ρ puis le cas où il donne la priorité à la politique de coopération faible. Nous choisissons de ne présenter que le cas général, c'est-à-dire le cas où aucune priorité n'est donnée.

Ce cas est présenté sur la figure 13.4.

Dans le troisième tableau, nous retrouvons le même résultat que dans la section précédente : l'agent étant toujours en accord avec la politique ρ, il peut choisir d'être ou de ne pas être en accord avec la politique de coopération faible.

Contrairement à la politique de coopération stricte, la politique de coopération faible possède une règle permissive (qui correspond à la troisième ligne de chacun des tableaux ci-dessous). Dans le cas où c'est la politique de coopération qui est permissive, l'agent peut donc choisir d'être en accord ou en désaccord avec la politique d'échange ρ.

13.2.1 Coopératif et obéissant sous contrainte

Nous définissons alors les notions de *faiblement coopératif sous contrainte* et *faiblement obéissant sous contrainte*.

	$O_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^f Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$	$\neg pol_a\varphi \wedge \neg coop_a^f\varphi$
$F_b^f Inf_{a,b}\varphi$	$pol_a\varphi \wedge \neg coop_a^f\varphi$	$\neg pol_a\varphi \wedge coop_a^f\varphi$
$P_b^f Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$	$\neg pol_a\varphi \wedge coop_a^f\varphi$

	$F_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^f Inf_{a,b}\varphi$	$\neg pol_a\varphi \wedge coop_a^f\varphi$	$pol_a\varphi \wedge \neg coop_a^f\varphi$
$F_b^f Inf_{a,b}\varphi$	$\neg pol_a\varphi \wedge \neg coop_a^f\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$
$P_b^f Inf_{a,b}\varphi$	$\neg pol_a\varphi \wedge coop_a^f\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$

	$P_b^\rho Inf_{a,b}\varphi$	
	$Inf_{a,b}\varphi$	$\neg Inf_{a,b}\varphi$
$O_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$	$pol_a\varphi \wedge \neg coop_a^f\varphi$
$F_b^S Inf_{a,b}\varphi$	$pol_a\varphi \wedge \neg coop_a^f\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$
$P_b^f Inf_{a,b}\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$	$pol_a\varphi \wedge coop_a^f\varphi$

FIGURE 13.4 – Coopérativité faible - Cas général

Définition III.5. *Lorsque l'agent est à la fois soumis à la politique de coopération faible et à une politique d'échange d'informations, il est dit* ***faiblement coopératif sous contrainte*** *pour l'information φ si et seulement si :*

- *s'il donne la priorité à la politique de coopération, alors il est coopératif si et seulement si il est en accord avec la politique de coopération.*
  Formellement, il faut que $O_b^f Inf_{b,a}\varphi \to Inf_{b,a}\varphi \wedge F_b^f Inf_{b,a}\varphi \to \neg Inf_{b,a}\varphi$ soit vrai.
- *sinon, lorsque la politique d'échange d'information ρ permet l'échange de φ alors l'agent transmet φ lorsqu'il a une obligation de le faire selon la politique de coopération et s'il ne la transmet pas lorsqu'il a une interdiction de le faire.*
  Formellement, il faut que $P_b^\rho Inf_{b,a}\varphi \to (O_b^f Inf_{b/a}\varphi \to Inf_{b,a}\varphi \wedge F_b^f Inf_{b/a}\varphi \to \neg Inf_{b,a}\varphi)$ soit vrai.

On peut également définir la notion d'obéissant sous contrainte :

Définition III.6. *Lorsque l'agent est à la fois soumis à la politique de coopération faible et à une politique d'échange d'informations, il est dit* ***faiblement obéissant sous contrainte*** *pour l'information φ si et seulement si :*

- *s'il donne la priorité à la politique ρ, alors il est coopératif si et seulement si il est en accord avec celle-ci.*
  Formellement, il faut que $(O_b^\rho Inf_{b,a}\varphi \to Inf_{b,a}\varphi) \wedge (F_b^\rho Inf_{b,a}\varphi \to \neg Inf_{b,a}\varphi)$ soit vrai.
- *sinon, lorsque la politique de coopération faible permet l'échange de φ alors l'agent transmet φ lorsqu'il a une obligation de le faire selon la politique d'échange ρ et s'il ne la transmet pas lorsqu'il a une interdiction de le faire.*
  Formellement, il faut que $P_b^f Inf_{b,a}\varphi \to (O_b^\rho Inf_{b/a}\varphi \to Inf_{b,a}\varphi) \wedge (F_b^\rho Inf_{b/a}\varphi \to \neg Inf_{b,a}\varphi)$ soit vrai.

Chapitre 14

Conclusion

Dans cette partie, nous avons étudié la coopération comme une politique d'échange d'information particulière que les agents d'un système seraient tenus de respecter. Or, les agents d'un système sont en général déjà soumis à une voire plusieurs politiques d'échanges d'information. Toutes ces politiques pouvant être en conflit, nous avons proposé, dans ce cas, une définition du caractère coopératif d'un agent. Nous proposons également une définition du caractère obéissant (aux politiques autres que les politiques de coopération) d'un agent.

Nous avons donné deux interprétations possibles de la coopération en terme de politique d'échange. Il en existe bien évidemment beaucoup d'autres. Par exemple, nous pourrions décider d'utiliser une autre définition de la coopération (voir 4.3).

Dans cette partie, nous ne sommes pas rentrés dans les détails techniques. Pour être tout à fait rigoureux, il faudrait définir précisément les modalités et surtout les schémas d'axiome et règles d'inférence qui leur sont associés et définir une sémantique adéquate. Afin de garder l'expressivité que nous avions en partie II, il faudrait étendre notre travail sur l'utilité et la coopérativité à une logique modale du premier ordre. Ce point a été abordé dans la section 5.2.

Pour la résolution du conflit entre les politiques d'échange, nous avons utilisé une méthode relativement simple de priorité sur les politiques. Il serait possible d'utiliser une autre méthode que celle-ci.

Finalement, nous avons considéré une définition relativement simple du fait d'être en accord avec une politique. Cette définition peut cependant poser certains dilemmes, surtout quand les obligations portent sur des actions d'agent ([MN04]). Par exemple, il n'est pas possible de déléguer des obligations à d'autres agents sans violer la politique. Une extension de ce travail serait donc d'utiliser une définition plus complexe de l'accord avec une politique.

Quatrième partie

Le mot de la fin...

Dans ce travail de thèse, nous nous sommes intéressés à deux aspects des agents échangeant des informations dans un système. Le premier aspect est le caractère coopératif des agents. Nous sommes appuyés sur une définition de l'utilité d'une information pour un agent et avons proposé plusieurs définitions d'un agent coopératif. Bien que cette modélisation soit limitée sur certains points, nous obtenons un modèle qui répond à un certain nombre de propriétés que l'on souhaite avoir pour une définition de la coopération.

Dans un deuxième temps, nous avons étudié les politiques qui réglementent les échanges d'informations dans les systèmes d'agents. Nous avons vu que ces politiques doivent être cohérentes et complètes pour être fonctionnelles. Nous avons proposé des définitions de ces propriétés ainsi qu'une méthode pour raisonner avec des politiques incomplètes.

Finalement, dans une troisième partie, nous montrons que ces deux aspects d'un système multiagents ne sont pas indépendants l'un de l'autre. En effet, la définition d'un agent coopératif n'est pas la même lorsque nous prenons en compte la politique d'échange d'informations qu'il doit respecter. Ceci montre que les différents aspects d'un système multiagents, bien que pouvant être étudiés en détail indépendamment les uns des autres, doivent être replacés dans le contexte du système et de ses autres paramètres pour pouvoir prendre tout leur sens.

Nous ne reprenons pas dans cette dernière partie les différentes limites et perspectives que nous avons déjà développées dans les différents chapitres de conclusion (chapitres 5, 11 et 14).

L'ensemble de ce travail est très théorique. En effet, afin d'exprimer les différentes notions que nous avons abordées, nous avons utilisé un formalisme qui offrait une très grande expressivité, des logiques modales propositionnelle et du premier ordre, mais dont l'implémentation informatique est limitée. La principale perspective de ce travail serait donc de trouver un équilibre entre expressivité et complexité pour pouvoir travailler sur des applications concrètes de cette théorie.

Preuves

Preuves partie I

Théorème . *I.1* *Soit a un agent de $\mathcal{A}$ et φ une formule.*
Alors $\vdash D_a\varphi \to \neg B_a\varphi$.

Démonstration. Soit φ une formule.

1. $\vdash B_a(\top \leftrightarrow \varphi) \to (D_a\top \leftrightarrow D_a\varphi)$ - (DRE)
2. $\vdash B_a(\top \to \varphi) \wedge B_a(\varphi \to \top) \to (D_a\top \to D_a\varphi) \wedge (D_a\varphi \to D_a\top)$ - (BK) $(B_a(\varphi_1 \wedge \varphi_2) \leftrightarrow (B_a\varphi_1 \wedge B_a\varphi_2))$ et 1.
3. $\vdash B_a\varphi \wedge B_a\top \to (\neg D_a\top \vee D_a\varphi) \wedge (\neg D_a\varphi \vee D_a\top)$ - 2.
4. $\vdash B_a\varphi \to \neg D_a\varphi$ - (BNec), (DUE) et 3.
5. $\vdash D_a\varphi \to \neg B_a\varphi$ - 4.

□

Théorème . *I.2* *Soit φ une formule. φ est un théorème de notre logique ssi φ est valide dans la classe* $\mathbf{C}$.

Démonstration. **Validité**

Montrons que les axiomes de la logique sont valides dans $\mathbf{C}$ et que les règles d'inférence conservent la validité.

Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$. Dans l'ensemble de la preuve, on considère que a est un agent de $\mathcal{A}$. On s'intéressera donc à une relation d'accessibilité R_a et une fonction de voisinage N_a.

Montrons que les schémas d'axiome et règles d'inférence sont valides

– **Règle (MP)**. On se réfère à [Che80] pour la preuve que ces règles conservent la validité.

– **Axiomes (BD), (B4), (B5) et règle (BNec)**. On se réfère à [Che80] pour la preuve que les axiome (BK), (BD), (B4) et (B5) sont valide pour la classe des modèles sériels, transitifs et euclidiens et que la règle (BNec) conserve la validité dans ces modèles.

– **Règle (DUE)**. Montrons que la règle (DUE) $\dfrac{\vdash \varphi}{\vdash \neg D_a\varphi}$ conserve la validité.
Soit φ une formule valide dans la classe de modèles $\mathbf{C}$. Soit $\mathbb{M}$ un modèle de la classe $\mathbf{C}$.
Pour tout monde w de W, on a $\mathbb{M}, w \vDash \varphi$. On a donc $(\varphi)^{\mathbb{M}} = W$.
$\mathbb{M}$ est unit-exclusif donc $\forall w \in W, W \notin N_a(w)$.
On a donc, $\forall w \in W, (\varphi)^{\mathbb{M}} \notin N_a(w)$
ce qui implique $\forall w \in W, \mathbb{M}, w \nvDash D_a(\varphi)$
ou encore $\forall w \in W, \mathbb{M}, w \vDash \neg D_a(\varphi)$.
La formule $\neg D_a(\varphi)$ est donc satisafaite dans le modèle $\mathbb{M}$. On en déduit que la formule $\neg D_a(\varphi)$ est valide dans la classe $\mathbf{C}$

– **Axiome (D4)**. Soient $w \in W$ et φ une formule du langage. Supposons que $\mathbb{M}, w \vDash D_a\varphi$ et montrons que $\mathbb{M}, w \vDash B_aD_a\varphi$.
Soit $w' \in R_a[w]$. $\mathbb{M}$ est introspectif donc $N_a(w) = N_a(w')$.

$$\begin{aligned}\mathbb{M}, w \vDash D_a\varphi &\Rightarrow (\varphi)^{\mathbb{M}} \in N_a(w)\\ &\Rightarrow (\varphi)^{\mathbb{M}} \in N_a(w')\\ &\Rightarrow \mathbb{M}, w' \vDash D_a\varphi\end{aligned}$$

On a donc $\forall w' \in R_a[w]$, $\mathbb{M}, w' \vDash D_a\varphi$. Cela signifie que $\mathbb{M}, w \vDash B_aD_a\varphi$.

– **Axiome (D5)**. Soient $w \in W$ et φ une formule du langage. Supposons que $\mathbb{M}, w \vDash \neg D_a\varphi$ et montrons que $\mathbb{M}, w \vDash B_a\neg D_a\varphi$.
  $\mathbb{M}$ est introspectif donc $\forall w \in R_a[w]$, $N_a(w) = N_a(w')$. On a donc :

$$\begin{aligned}\mathbb{M}, w \vDash \neg D_a\varphi &\Rightarrow (\varphi)^{\mathbb{M}} \notin N_a(w)\\ &\Rightarrow \forall w' \in R_a[w], (\varphi)^{\mathbb{M}} \notin N_a(w')\\ &\Rightarrow \forall w' \in R_a[w], Non(\mathbb{M}, w' \vDash D_a\varphi)\\ &\Rightarrow \forall w' \in R_a[w], \mathbb{M}, w' \vDash \neg D_a\varphi\\ &\Rightarrow \mathbb{M}, w \vDash B_a\neg D_a\varphi\end{aligned}$$

– **Axiome (DRE)**. Soient $w \in W$, φ_1 et φ_2 deux formules du langage. Supposons que $\mathbb{M}, w \vDash B_a(\varphi_1 \leftrightarrow \varphi_2)$ et montrons que $\mathbb{M}, w \vDash D_a\varphi_1 \leftrightarrow D_a\varphi_2$.

  Soit $w' \in R_a[w]$. Si $w' \in (\varphi_1)^{\mathbb{M}}$, alors cela signifie que $\mathbb{M}, w' \vDash \varphi_1$. D'après l'hypothèse, on a donc $\mathbb{M}, w' \vDash \varphi_2$. Cela signifie que $\forall w' \in R_a[w]$, si $w' \in (\varphi_1)^{\mathbb{M}}$ alors $w' \in (\varphi_2)^{\mathbb{M}}$. Les rôles de φ_1 et φ_2 étant interchangeables, on a donc $\forall w' \in R_a[w]$, $w' \in (\varphi_1)^{\mathbb{M}}$ ssi $w' \in (\varphi_2)^{\mathbb{M}}$. Cela signifie que $R_a[w] \cap (\varphi_1)^{\mathbb{M}} = R_a[w] \cap (\varphi_2)^{\mathbb{M}}$.

  Or $\mathbb{M}$ est un modèle introspectif. On en déduit que $(\varphi_1)^{\mathbb{M}} \in N_a(w)$ ssi $(\varphi_2)^{\mathbb{M}} \in N_a(w)$. On a donc $\mathbb{M}, w \vDash D_a\varphi_1$ ssi $\mathbb{M}, w \vDash D_a\varphi_2$. C'est équivalent à $\mathbb{M}, w \vDash D_a\varphi_1 \leftrightarrow D_a\varphi_2$.

Nous avons montré que tous les axiomes sont valides dans la classe **C** et que les règles d'inférence conservent la validité.

Complétude

Nous utilisons ici l'argument basé sur le modèle canonique. En effet, pour prouver que le système d'axiomes et les règles d'inférence est complet pour la classe de modèles **C**, il est suffisant de prouver que **C** contient un modèle canonique $\mathbb{M}_C$. Dès lors, si une formule φ est valide dans **C**, alors elle est vraie dans $\mathbb{M}_C$ et est donc théorème de notre logique ([Che80]). Notre but ici est donc de construire un modèle canonique $\mathbb{M}_C$ du système de preuves et de montrer qu'il appartient à la classe **C**. [43]

Construction du modèle canonique

On construit donc un modèle $\mathbb{M}_C =< W_C, R_C, N_C, V_C >$ tel que :

– W_C est la collection des ensembles de formules maximaux consistants pour notre logique.

– Pour tout ensemble de formules Γ et pour tout agent a de $\mathcal{A}$, $\Gamma/B_a = \{\varphi | B_a\varphi \in \Gamma\}$. On note $W^{\Gamma}_{C_a}$ le plus grand sous-ensemble de W_C dont tous les ensembles de formules maximaux consistants contiennent Γ/B_a.

– Pour chaque agent a de $\mathcal{A}$, R_{C_a} est une relation d'accessibilité sur W^2_C telle que $\forall w, w' \in W_C$, $wR_{C_a}w'$ ssi $w' \in W^w_{C_a}$ (autrement dit ssi w' est un ensemble maximal consistant

43. Cette preuve est une généralisation de la preuve dont les grandes lignes sont présentées dans [SSLR07].

contenant $\{\varphi | B_a\varphi \in w\}$ [44]). On note R_C l'ensemble des relations R_{C_a}

– Pour tout $p \in \mathcal{V}$, $V_C(p) = \{w | w \in W_C \text{ et } p \in w\}$

– Pour toute formule φ du langage, $|\varphi| = \{w | w \in W_C \text{ et } \varphi \in w\}$

– Pour tout agent a de $\mathcal{A}$, N_{C_a} est une fonction de W_C sur $2^{2^{W_C}}$ telle que pour tout $w \in W_C$, $N_{C_a}(w)$ est un ensemble de mondes qui contient tous les ensembles $|\varphi|$ tels que $D_a\varphi \in w$.

Pour chaque $w \in W_C$ et pour chaque agent a de $\mathcal{A}$, on note $\mathcal{U}_{w_a}$ la collection des sous-ensembles de $W^w_{C_a}$ qui ne sont pas de la forme $|\varphi| \cap W^w_{C_a}$. Autrement dit, $\mathcal{U}_{w_a} = \{X \cap W^w_{C_a} | \forall\varphi X \neq |\varphi|\}$.

Pour $Y \in \mathcal{U}_{w_a}$, on définit l'ensemble $\mathcal{W}_Y = \{|\varphi| \cap W^w_{C_a} | D_a\varphi \in w\}$.

On peut finalement caractériser l'appartenance à N_{C_a} de la façon suivante :

pour tout $X \subseteq W_C$, pour tout $w \in W_C$ alors : – si X est de la forme $|\varphi|$ alors $X \in N_{C_a}(w)$ ssi $D_a\varphi \in w$; – sinon, on pose $Y = X \cap W^w_{C_a}$ (on note que $Y \in \mathcal{U}_{w_a}$) et $X \in N_{C_a}(w)$ ssi $Y \in \mathcal{W}_Y$.

Il faut maintenant montrer que $\mathbb{M}_C$ est sériel, transitif, euclidien, unit-exclusif et introspectif.

Preuve que $\mathbb{M}_C$ est un modèle de la classe C

– **Sérialité** Pour montrer que $\mathbb{M}_C$ est sériel, il faut montrer que pour tout agent a de $\mathcal{A}$, pour tout $w \in W_{C_a}$, l'ensemble $W^w_{C_a}$ est non-vide. Il suffit de montrer que l'ensemble $w/B_a = \{\varphi | B_a\varphi \in w\}$ est consistant [45].
Supposons que w/B_a soit inconsistant. Cela signifie que l'on peut déduire $\perp$ de cet ensemble, donc qu'il existe des formules A_1, ... , A_n telles que
$\vdash A_1 \wedge ... \wedge A_n \rightarrow \perp$
En appliquant (BK), on obtient :
$\vdash B_aA_1 \wedge ... \wedge B_aA_n \rightarrow B_a\perp$
En appliquant (BD), on obtient :
$\vdash B_aA_1 \wedge ... \wedge B_aA_n \rightarrow \neg B_a\top$
w contient B_aA_1, ... B_aA_n donc $B_a\perp$ est déductible de w. Or, $\top$ est une tautologie. On peut donc appliquer (BNec) et ainsi dériver que $\vdash B_a\top$.
$\neg B_a\top$ et $B_a\top$ sont donc tous deux déductibles de w, ce qui est une contradiction.

– **Transitivité**. Soit a un agent de $\mathcal{A}$. Montrons que pour tous les monds w_1, w_2 et w_3 de W_C, si $w_1R_{C_a}w_2$ et $w_2R_{C_a}w_3$ alors $w_1R_{C_a}w_3$.
Supposons que $w_1R_{C_a}w_2$ et $w_2R_{C_a}w_3$. Cela signifie que $\{\varphi | B_a\varphi \in w_1\} \subseteq w_2$ et $\{\varphi | B_a\varphi \in w_2\} \subseteq w_3$.

Soit φ une formule telle que $B_a\varphi \in w_1$.

$$\begin{array}{lll} B_a\varphi \in w_1 & \Rightarrow & B_aB_a\varphi \in w_1 \text{ (axiome (B4) et } w_1 \text{ maximal consistant)} \\ & \Rightarrow & B_a\varphi \in w_2 \\ & \Rightarrow & \varphi \in w_3 \end{array}$$

44. On peut remarquer que $W^w_{C_a} = R_{C_a}[w]$. Nous utiliserons indifféremment les deux notations.

45. En effet, on peut utiliser le lemme de Lindenbaum [Che80] : Si Γ est un ensemble consistant pour un système Σ alors il existe Δ tel que $\Gamma \subseteq \Delta$ et tel que Δ soit un ensemble Σ maximal consistant

On a donc $\{\varphi | B_a\varphi \in w_1\} \subseteq w_3$.

– **Euclidéanité**. Soit a un agent de $\mathcal{A}$. Montrons que pour tous les mondes w_1, w_2 et w_3 de W_C, si $w_1 R_{C_a} w_2$ et $w_1 R_{C_a} w_3$ alors $w_2 R_{C_a} w_3$.
On suppose que $w_1 R_{C_a} w_2$ et $w_1 R_{C_a} w_3$, c'est à dire que $\{\varphi | B_a\varphi \in w_1\} \subseteq w_2$ et $\{\varphi | B_a\varphi \in w_1\} \subseteq w_3$.
Soit φ telle que $B_a\varphi \in w_2$.
– Si $B_a\varphi \in w_1$ alors d'après l'hypothèse, $\varphi \in w_3$
– sinon, comme w_1 est maximal consistant, on a $\neg B_a\varphi \in w_1$. En appliquant l'axiome (5), on obtient que $B_a \neg B_a\varphi \in w_1$. D'après l'hypothèse $w_1 R_{C_a} w_2$, cela signifie que $\neg B_a\varphi \in w_2$. C'est en contradiction avec $B_a\varphi \in w_2$ et notre supposition que w_2 est maximal consistant.
On a donc pour tout φ telle que $B\varphi \in w_2$, $\varphi \in w_3$, ce qui signifie que $\{\varphi | B\varphi \in w_2\} \subseteq w_3$.

– **Unit-exclusif** Montrons que pour tout $w \in W_C$ et pour tout agent a de $\mathcal{A}$, $W_C \notin N_{C_a}(w)$.
Soit $w \in W_C$ et soit a un agent de $\mathcal{A}$. w est maximal consistant pour le système de preuve qui contient (UE). On a donc $\neg D_a(\top) \in w$.

$$\begin{array}{lll} \neg D_a(\top) \in w & \Rightarrow & D_a(\top) \notin w \\ & \Rightarrow & |\top| \notin N_{C_a}(w) \\ & \Rightarrow & W_C \notin N_{C_a}(w) \end{array}$$

On a donc $\forall w \in W_C$, $W_C \notin N_{C_a}(w)$.

– **Introspectif-(1)**. Montrons que pour tout agent a, pour tout $w_1, w_2 \in W_C$ tels que $w_1 R_{C_a} w_2$, $N_{C_a}(w_1) = N_{C_a}(w_2)$.

Lemme 1. *Soient deux mondes w_1 et w_2 de W_C et a un agent de $\mathcal{A}$. Si $w_1 R_{C_a} w_2$ alors $W_{C_a}^{w_1} = W_{C_a}^{w_2}$ et $\{|\varphi| \cap W_{C_a}^{w_1} | D_a\varphi \in w_1\} = \{|\varphi| \cap W_{C_a}^{w_2} | D_a\varphi \in w_2\}$.*

Démonstration. Soient deux mondes w_1 et w_2 de W_C tels que $w_1 R_{C_a} w_2$. Par transitivité et euclidéanité de R_{C_a}, on a $R_{C_a}[w_1] = R_{C_a}[w_2]$ ou encore $W_c^{w_1} = W_c^{w_2}$.
Soit φ telle que $D_a\varphi \in w_2$. Supposons que $D_a\varphi \notin w_1$. Alors $\neg D_a\varphi \in w_1$ donc $B_a \neg D_a\varphi \in w_1$. $w_1 R_{C_a} w_2$ donc $\neg D_a\varphi \in w_2$. C'est en contradiction avec l'hypothèse $D_a\varphi \in w_2$. On a donc $D_a\varphi \in w_1$. Cela prouve que $\{|\varphi| \cap W_C^{w_2} | D_a\varphi \in w_2\} \subseteq \{|\varphi| \cap W_C^{w_1} | D_a\varphi \in w_1\}$.
Soit φ telle que $D_a\varphi \in w_1$. Alors on a $B_a D_a\varphi \in w_1$ donc $D_a\varphi \in w_2$. Cela prouve que $\{|\varphi| \cap W_C^{w_1} | D_a\varphi \in w_1\} \subseteq \{|\varphi| \cap W_C^{w_2} | D_a\varphi \in w_2\}$.
On a donc $\{|\varphi| \cap W_C^{w_1} | D_a\varphi \in w_1\} = \{|\varphi| \cap W_C^{w_2} | D_a\varphi \in w_2\}$. □

Soit X un sous-ensemble de W_C et a un agent de $\mathcal{A}$.

– Supposons qu'il existe φ tel que $X = |\varphi|$.

  – Supposons que $X \in N_{C_a}(w_1)$. Par définition de N_{C_a}, on a donc $D_a\varphi \in w_1$. w_1 est maximal consistant pour le système de preuve qui contient l'axiome (D4), on a donc $B_a D_a\varphi \in w_1$. Or $w_1 R_{C_a} w_2$. On a donc $D_a\varphi \in w_2$. Cela signifie que $|\varphi|(= X) \in N_{C_a}(w_2)$.

- Supposons que $X \in N_{C_a}(w_2)$. Par définition de N_{C_a}, on a donc $D_a\varphi \in w_2$.
  - si $D_a\varphi \in w_1$ alors dans ce cas $X \in N_{C_a}(w_1)$.
  - sinon $\neg D_a\varphi \in w_1$. En utilisant l'axiome (D5), on obtient $B_a\neg D_a\varphi \in w_1$. Or, $w_1 R_{C_a} w_2$ donc $\neg D_a\varphi \in w_2$, ce qui est en contradiction avec $D_a\varphi \in w_2$.

  On a donc $X \in N_{C_a}(w_1)$.

  Donc, dans ce cas, $X \in N_{C_a}(w_1) \Leftrightarrow X \in N_{C_a}(w_2)$.

- **Sinon**. Supposons qu'il n'existe pas φ tel que $X = |\varphi|$.
  $w_1 R_{C_a} w_2$ donc $W_c^{w_1} = W_c^{w_2}$ (lemme 1). On pose $Y_1 = X \cap W_c^{w_1}$ et $Y_2 = X \cap W_c^{w_2}$. On a $Y_1 = Y_2$.
  $\mathcal{W}_{Y_2} = \{|\varphi| \cap W_c^{w_1} | D_a\varphi \in w_2\}$ et $\mathcal{W}_{Y_1} = \{|\varphi| \cap W_c^{w_2} | D_a\varphi \in w_1\}$. D'après le lemme 1, ces deux ensembles sont égaux.
  On a donc
  $$\begin{array}{rcl} X \in N_{C_a}(w_1) & \Leftrightarrow & Y_1 \in \mathcal{W}_{Y_1} \\ & \Leftrightarrow & Y_2 \in \mathcal{W}_{Y_2} \\ & \Leftrightarrow & X \in N_{C_a}(w_2) \end{array}$$

Quelle que soit la forme de X, on a montré que $X \in N_{C_a}(w_1) \Leftrightarrow X \in N_{C_a}(w_2)$. Autrement dit $N_{C_a}(w_1) = N_{C_a}(w_2)$.

- **Introspectif-(2)**. Montrons que pour tout $w \in W_C$ et $X_1, X_2 \subseteq W_C$ tels que $X_1 \cap R_{C_a}[w] = X_2 \cap R_{C_a}[w]$, on a $X_1 \in N_{C_a}(w) ssi X_2 \in N_{C_a}(w)$.
  Soient $w \in W_C$ et $X_1, X_2 \subseteq W_C$ tels que $X_1 \cap R_{C_a}[w] = X_2 \cap R_{C_a}[w]$.
  - **Si il existe φ_1 et φ_2 telles que $X_1 = |\varphi_1|$ et $X_2 = |\varphi_2|$.**
    Supposons qu'il existe deux formules φ_1 et φ_2 telles que $X_1 = |\varphi_1|$ et $X_2 = |\varphi_2|$.
    Soit $w' \in W_C$.
    - Soit $w \in R_{C_a}(w')$ tel que $\varphi_1 \in w$. On a donc $w \in X_1 \cap R_{C_a} w'$. Par hypothèse, cela signifie que $w \in X_2 \cap R_{C_a} w'$. On a donc $\varphi_2 \in w$ et $w R_{C_a} w'$.
    - on montre de même que si $w \in R_{C_a}(w')$ et que $\varphi_2 \in w$ alors $\varphi_1 \in w$ et $w R_{C_a} w'$.

    Pour w' tel que $w R_{C_a} w'$, on a donc $\varphi_1 \in w' \Leftrightarrow \varphi_2 \in w'$. Autrement dit, $\varphi_1 \leftrightarrow \varphi_2 \in w'$ car w' est maximal conxistant. Cela implique que $B_a(\varphi_1 \leftrightarrow \varphi_2) \in w$.
    Par l'axiome (DRE), on en déduit que $D_a\varphi \leftrightarrow D_a\varphi_2 \in w$. C'est équivalent à $D_a\varphi_1 \in w \Leftrightarrow D_a\varphi_2 \in w$, ou encore $X_1 \in N_{C_a}(w) \Leftrightarrow X_2 \in N_{C_a}(w)$.

  - **Si** $X_1 = |\varphi_1|$ **et** $X_2 \neq |\varphi|$.
    Supposons qu'il existe φ_1 tel que $X_1 = |\varphi_1|$ et qu'il n'existe pas de formule φ tel que $X_2 \neq |\varphi|$.
    On pose $Y_2 = X_2 \cap R_{C_a}[w]$. On a $Y_2 \in \mathcal{U}_{w_a}$ et $\mathcal{W}_{Y_2} = \{|\varphi| \cap W_{C_a}^{w} | D_a\varphi \in w\}$.
    - Premier cas : $X_1 \in N_{C_a}(w)$. $X_1 = |\varphi_1|$. Par définition de N_{C_a}, $D_a\varphi_1 \in w$. On a donc $|\varphi_1| \cap R_{C_a}[w] \in \mathcal{W}_{Y_2}$. Or, $|\varphi_1| \cap R_{C_a}[w] = X_1 \cap R_{C_a}[w] = X_2 \cap R_{C_a}[w] = Y_2$. On a donc $Y_2 \in \mathcal{W}_{Y_2}$, c'est à dire $X_2 \in N_{C_a}(w)$.
    - Deuxième cas : $X_1 \notin N_{C_a}(w)$. Cela signifie que $|\varphi_1| \notin N_{C_a}(w)$, ou encore $D_a\varphi_1 \notin w$. Si $Y_2 \in \mathcal{W}_{Y_2}$, alors cela signifie qu'il existe φ_0 telle que $Y_2 = |\varphi_0| \cap R_{C_a}[w]$ et telle que $D_a\varphi_0 \in w$. Or, $Y_2 = |\varphi_1| \cap R_{C_a}[w]$. Donc, $|\varphi_1| \cap R_{C_a}[w] = |\varphi_0| \cap R_{C_a}[w]$. On a vu dans le cas précédent que cela impliquait $B_a(\varphi_0 \leftrightarrow \varphi_1) \in w$ et donc (axiome (DRE)) $D_a\varphi_1 \leftrightarrow D_a\varphi_0 \in w$. Or ici, nous avons $D_a\varphi_1 \notin w$ et $D_a\varphi_0 \in w$. C'est une

contradiction et on a donc $Y_2 \notin \mathcal{W}_{\mathcal{Y}_\in}$, c'est à dire $X_2 \notin N_{C_a}(w)$.

Nous avons montré que $X_1 \in N_{C_a}(w) ssi X_2 \in N_{C_a}(w)$.

Les deux ensembles X_1 et X_2 ont ici un rôle symétrique. On ne traite donc pas le cas où $X_1 \neq |\varphi|$ et $X_2 = |\varphi_2|_{\mathcal{S}}$.

– **Si pour tout φ, $X_1 \neq |\varphi|$ et $X_2 \neq |\varphi|$.**
On pose $Y_1 = X_1 \cap R_{C_a}[w]$ et $Y_2 = X_2 \cap R_{C_a}[w]$. On a $Y_1 = Y_2$ et $\mathcal{W}_{Y_1} = \mathcal{W}_{Y_2}$. On peut donc en déduire que $Y_1 \in \mathcal{W}_{Y_1}$ ssi $Y_2 \in \mathcal{W}_{Y_2}$, c'est à dire $X_1 \in N_{C_a}(w)$ ssi $Y_2 \in N_{C_a}(w)$.

Le modèle canonique $\mathbb{M}_C$ est donc introspectif-2.

Preuve de complétude.
Le modèle canonique $\mathbb{M}_C$ est un modèle de **C** : il est sériel, transitif, euclidien, unit-exclusif et introspectif.
De par la définition d'un modèle canonique, on a $(\varphi)^{\mathbb{M}_C} = |\varphi|$. En effet,

$$\begin{aligned} w \in (\varphi)^{\mathbb{M}_C} &\Leftrightarrow \mathbb{M}_C, w \vDash \varphi \\ &\Leftrightarrow \varphi \in w \text{ (car } w \text{ est maximal consistant pour notre logique).} \\ &\Leftrightarrow w \in |\varphi| \text{ (par définition)} \end{aligned}$$

Soit φ une formule valide dans la classe de modèle **C**. Alors φ est également satisfaite dans le modèle $\mathbb{M}_C$ qui appartient à la classe **C**. Cela signifie que pour tout $w \in W_C$, $\mathbb{M}_C, w \vDash \varphi$. Autrement dit, $(\varphi)^{\mathbb{M}_C} = W_C$. Donc $|\varphi| = W_C$, c'est à dire que φ est satisfaite dans tous les ensembles maximaux consistants pour notre logique. φ est donc un théorème de notre logique.

□

Corollaire . *I.1 Le système de preuve étant valide pour la classe* **C** *qui est non-vide (voir exemple I.1), on peut en déduire que le système est consistant.*

Démonstration. C'est une proposition de [HC96]. □

Proposition . ***I.1***

$$\vdash D_a Bif_a Q \to \neg Bif_a Q$$

Démonstration. Montrons tout d'abord que $\vdash D_a Bif_a Q \to \neg B_a Q$.

1. $\vdash B_a Q \to B_a B_a Q$ (axiome (B4))
2. $\vdash B_a Q \to B_a(B_a Q \vee B_a \neg Q)$ (1. et axiome (BK))
3. $\vdash D_a Bif_a Q \to \neg B_a(B_a Q \vee B_a \neg Q)$ (théorème I.1)
4. $\vdash (D_a Bif_a Q \wedge B_a Q) \to \perp$ (2., 3. et axiome (BD))
5. $\vdash D_a Bif_a Q \to \neg B_a Q$ (4.)

Dans cette preuve, Q et $\neg Q$ ont un rôle symétrique. Nous pouvons donc également prouver que $\vdash D_a Bif_a Q \to \neg B_a \neg Q$. □

Proposition . ***I.2***

$$\vdash U_a^Q \varphi \to \neg B_a \varphi$$

Démonstration. 1. $\vdash U_a^Q \varphi \to D_a Bif_a Q$ (définition de l'utilité)

2. $\vdash U_a^Q \varphi \to \neg B_a Q \wedge \neg B_a \neg Q$ (1. et prop. I.1)
3. $(B_a(\varphi \to Q) \wedge \neg B_a(\varphi \to \neg Q) \wedge B_a \varphi) \to B_a Q$ (axiome (BK))
4. $(B_a(\varphi \to \neg Q) \wedge \neg B_a(\varphi \to Q) \wedge B_a \varphi) \to B_a \neg Q$ (axiome (BK))

5. $(B_a\varphi \wedge (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))) \to (B_aQ \vee B_a\neg Q)$ (3. et 4.)

6. $(U_a^Q\varphi \wedge B_a\varphi) \to \bot$ (2. et 5.)

7. $U_a^Q\varphi \to \neg B_a\varphi$ (6.)

□

Proposition . *I.3*

$$\vdash U_a^Q\varphi \to \neg B_a\neg\varphi$$

Démonstration. 1. $\vdash B_a\neg\varphi \to B_a(\neg\varphi \vee Q)$ (axiome (BK))

2. $\vdash B_a\neg\varphi \to B_a(\varphi \to Q)$ (1.)

3. $\vdash B_a\neg\varphi \to B_a(\varphi \to \neg Q)$ (axiome (BK))

4. $\vdash B_a\neg\varphi \to (B_a(\varphi \to Q) \wedge B_a(\varphi \to \neg Q))$ (2. et 3.)

5. $\vdash (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q)) \to \neg(B_a(\varphi \to Q) \wedge B_a(\varphi \to \neg Q))$ (définition de $\otimes$)

6. $\vdash U_a^Q\varphi \to (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))$ (définition de l'utilité)

6. $\vdash (U_a^Q\varphi \wedge B_a\neg\varphi) \to \bot$ (4., 5. et 6.)

7. $\vdash U_a^Q\varphi \to \neg B_a\neg\varphi$ (6.)

□

Corollaire . *I.2* *Si une information φ est utile pour un agent a par rapport à une requête Q, alors φ n'est ni une tautologie, ni une contradiction.*

Démonstration. Si φ est une tautologie alors pour tout monde w, $\mathbb{M}, w \vDash \neg B_a\varphi$. D'après la proposition I.2, on a donc pour tout w, $\mathbb{M}, w \vDash \neg U_a^Q\varphi$.
Si φ est une contradiction alors pour tout monde w, $\mathbb{M}, w \vDash \neg B_a\neg\varphi$. D'après la proposition I.3, on a donc pour tout w, $\mathbb{M}, w \vDash \neg\neg U_a^Q\varphi$.

□

Proposition . *I.4* *Soit $*$ un opérateur de révision de croyances satisfaisant les postulats AGM (postulats 1 à 4) [AGM85]. Bel_a représente l'ensemble des croyances de l'agent a dans un monde w et $Bel_a * \varphi$ l'ensemble des croyances de l'agent après avoir été révisé par φ en utilisant l'opérateur $*$. Si $U_a^Q\varphi$ est satisfaite dans w alors soit $Q \in Bel_a * \varphi$ soit $\neg Q \in Bel_a * \varphi$.*

Démonstration. Commençons par rappeler les postulats AGM. Soit Bel_a la base de croyances de l'agent a (nous supposons que la base est fermée pour la déduction). On note $Bel_a * \varphi$ la révision de Bel_a par φ et $Bel_a + \varphi$ la fermeture déductive de $Bel_a \cup \{\varphi\}$. Les postulats AGM sont :

(G*1) $Bel_a * \varphi$ est une base de croyances (un ensemble fermé pour la déduction)
(G*2) $\varphi \in Bel_a * \varphi$
(G*3) $Bel_a * \varphi \subset Bel_a + \varphi$
(G*4) si $\neg\varphi \notin Bel_a$ alors $Bel_a + \varphi \subset Bel_a * \varphi$

Supposons $U_a^Q\varphi$ soit satisfaite dans w. Alors, $\neg B_a\varphi$ est également satisfaite dans w (proposition I.2). Cela signifie que $\varphi \notin Bel_a$. Ainsi, réviser Bel_a par φ revient à étendre Bel_a par φ. Cela signifie donc que $\varphi \in Bel_a * \varphi$.
De $U_a^Q\varphi$ satisfaite dans w, on peut déduire soit $B_a(\varphi \to Q)$ satisfaite dans w soit $B_a(\varphi \to \neg Q)$

satisfaite dans w. Ces deux formules sont également dans l'ensemble $Bel_a * \varphi$. Ainsi, on peut déduire soit que $Q \in Bel_a * \varphi$ soit que $\neg Q \in Bel_a * \varphi$. $\square$

Proposition . *I.5*

$$\vdash D_a Bif_a Q \to U_a^Q Q \otimes U_a^Q \neg Q$$

Démonstration. La formule $U_a^Q Q \otimes U_a^Q \neg Q$ est équivalente à $(U_a^Q Q \vee U_a^Q \neg Q) \wedge \neg(U_a^Q Q \wedge U_a^Q \neg Q)$. Nous montrons donc un premier temps que $\vdash \neg(U_a^Q Q \wedge U_a^Q \neg Q)$.

1. $\vdash U_a^Q Q \wedge U_a^Q \neg Q \to Q \wedge \neg Q$
2. $\vdash U_a^Q Q \wedge U_a^Q \neg Q \to \bot$
3. $\vdash \neg(U_a^Q Q \wedge U_a^Q \neg Q)$

Nous montrons ensuite que $\vdash D_a Bif_a Q \to U_a^Q Q \vee U_a^Q \neg Q$.

1. $\vdash B_a(Q \to Q)$ (BNec)
2. $\vdash \neg B_a \neg Q \to \neg B_a(Q \to \neg Q)$
3. $\vdash \neg B_a \neg Q \to B_a(Q \to Q) \wedge \neg B_a(Q \to \neg Q)$ (1. et 2.)
4. $\vdash \neg B_a \neg Q \to B_a(Q \to Q) \otimes \neg B_a(Q \to \neg Q)$ (3.)
5. $\vdash \neg B_a \neg Q \to B_a(Q \to Q) \otimes \neg B_a(Q \to \neg Q)$ (4.)
6. $\vdash \neg B_a Q \to B_a(\neg Q \to \neg Q) \otimes \neg B_a(\neg Q \to Q)$ (5. dans lequel on a remplacé Q par $\neg Q$)
7. $\vdash \neg U_a^Q Q \to \neg D_a Bif_a Q \vee \neg(B_a(Q \to Q) \otimes B_a(Q \to \neg Q)) \vee \neg Q$ (définition I.10)
8. $\vdash D_a Bif_a Q \to \neg B_a Q \wedge \neg B_a \neg Q$ (proposition I.1)
9. $\vdash \neg U_a^Q Q \wedge D_a Bif_a Q \to \neg Q$ (5., 7. et 8.)
10. $\vdash \neg U_a^Q \neg Q \wedge D_a Bif_a Q \to Q$ (6., 7. et 9.)
11. $\vdash \neg U_a^Q Q \wedge \neg U_a^Q \neg Q \wedge D_a Bif_a Q \to \neg Q \wedge Q$ (9. et 10.)
12. $\vdash D_a Bif_a Q \to U_a^Q Q \vee U_a^Q \neg Q$ (11.)

$\square$

Proposition . *I.6*

$$\vdash B_a(Q_1 \leftrightarrow Q_2) \to (U_a^{Q_1} \varphi \leftrightarrow U_a^{Q_2} \varphi)$$

Démonstration.

1. $\vdash B_a Q_1 \wedge B_a(Q_1 \leftrightarrow Q_2) \to B_a Q_2$ (axiome (BK))
2. $\vdash B_a \neg Q_1 \wedge B_a(Q_1 \leftrightarrow Q_2) \to B_a \neg Q_2$ (axiome (BK))
3. $\vdash Bif_a Q_1 \wedge B_a(Q_1 \leftrightarrow Q_2) \to Bif_a Q_2$ (1. et 2.)
4. $\vdash Bif_a Q_2 \wedge B_a(Q_1 \leftrightarrow Q_2) \to Bif_a Q_1$ (3. et symétrie de Q_1 et Q_2)
5. $\vdash B_a(Q_1 \leftrightarrow Q_2) \to (Bif_a Q_1 \leftrightarrow Bif_a Q_2)$ (3. et 4.)
6. $\vdash (Bif_a Q_1 \leftrightarrow Bif_a Q_2) \to (B_a Bif_a Q_1 \leftrightarrow B_a Bif_a Q_2)$ (axiome (B4))
7. $\vdash B_a(Q_1 \leftrightarrow Q_2) \to (B_a Bif_a Q_1 \leftrightarrow B_a Bif_a Q_2)$ (5. et 6.)

8. $\vdash B_a(Q_1 \leftrightarrow Q_2) \to (D_a Bif_a Q_1 \leftrightarrow D_a Bif_a Q_2)$ (7. et axiome (DRE))

9. $\vdash B_a(Q_1 \leftrightarrow Q_2) \land B_a(\varphi \to Q_1) \to B_a(\varphi \to Q_2)$ (axiome (BK))

10. $\vdash B_a(Q_1 \leftrightarrow Q_2) \land \neg B_a(\varphi \to \neg Q_1) \to \neg B_a(\varphi \to \neg Q_2)$ (axiome (BK))

11. $\vdash B_a(Q_1 \leftrightarrow Q_2) \land (B_a(\varphi \to Q_1) \otimes \neg B_a(\varphi \to \neg Q_1)) \to (B_a(\varphi \to Q_2) \otimes \neg B_a(\varphi \to \neg Q_2))$ (9. et 10.)

12. $\vdash B_a(Q_1 \leftrightarrow Q_2) \to ((B_a(\varphi \to Q_1) \otimes \neg B_a(\varphi \leftrightarrow \neg Q_1)) \to (B_a(\varphi \to Q_2) \otimes \neg B_a(\varphi \to \neg Q_2)))$ (11. et symétrie de Q_1 et Q_2)

13. $B_a(Q_1 \leftrightarrow Q_2) \to (U_a^{Q_1}\varphi \leftrightarrow U_a^{Q_2}\varphi)$ (8. et 12.)

□

Proposition . *I.7*

$$\vdash U_a^Q\varphi \leftrightarrow U_a^{\neg Q}\varphi$$

Démonstration. 1. $\vdash Bif_a Q \leftrightarrow Bif_a \neg Q$

2. $\vdash B_a Bif_a Q \leftrightarrow B_a Bif_a \neg Q$ (1. et axiome (B4))

3. $\vdash D_a Bif_a Q \leftrightarrow D_a Bif_a \neg Q$ (2. et axiome (DRE))

4. $\vdash B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q) \leftrightarrow B_a(\varphi \to \neg Q) \otimes B_a(\varphi \to Q)$

5. $\vdash U_a^Q\varphi \leftrightarrow U_a^{\neg Q}\varphi$ (3. et 4.)

□

Proposition . *I.8*

$$\vdash \neg(\varphi_1 \land \varphi_2) \to \neg(U_a^Q\varphi_1 \land U_a^Q\varphi_2)$$

Démonstration. 1. $U_a^Q\varphi_1 \land U_a^Q\varphi_2 \land \varphi_1 \land \varphi_2$ (définition I.10)

2. $\vdash \neg(\varphi_1 \land \varphi_2) \to \neg(U_a^Q\varphi_1 \land U_a^Q\varphi_2)$ (1.)

□

Proposition . *I.9*

$$\vdash \neg B_a U_a^Q\varphi$$

Démonstration. 1. $\vdash B_a U_a^Q\varphi \to B_a\varphi$ (définition I.10)

2. $\vdash B_a U_a^Q\varphi \to \neg U_a^Q\varphi$ (1. et proposition I.2)

3. $\vdash U_a^Q\varphi \to \neg B_a U_a^Q\varphi$ (2.)

4. $\vdash B_a U_a^Q\varphi \to B_a \neg B_a U_a^Q\varphi$ (3., axiomes (BNec) et (BK))

5. $\vdash B_a U_a^Q\varphi \to B_a B_a U_a^Q\varphi$ (axiome (B4))

6. $\vdash B_a U_a^Q\varphi \to \bot$ (4., 5. et axiome (BD))

7. $\vdash \neg B_a U_a^Q\varphi$ (6.)

□

Proposition . *I.10*

$$\vdash (\varphi_1 \leftrightarrow \varphi_2) \land B_a(\varphi_1 \leftrightarrow \varphi_2) \to (U_a^Q\varphi_1 \leftrightarrow U_a^Q\varphi_2)$$

Démonstration. 1. $\vdash U_a^Q\varphi_1 \to D_aBif_aQ$ (définition I.10)

2. $\vdash U_a^Q\varphi_1 \to \varphi_1$ (définition I.10)
3. $\vdash U_a^Q\varphi_1 \wedge (\varphi_1 \leftrightarrow \varphi_2) \to \varphi_2$ (2.)
4. $\vdash B_a(\varphi_1 \leftrightarrow \varphi_2) \wedge B_a(\varphi_1 \to Q) \to B_a(\varphi_2 \to Q)$ (axiome (BK))
5. $\vdash B_a(\varphi_1 \leftrightarrow \varphi_2) \wedge B_a(\varphi_2 \to \neg Q) \to B_a(\varphi_1 \to \neg Q)$ (axiome (BK))
6. $\vdash B_a(\varphi_1 \leftrightarrow \varphi_2) \to (\neg B_a(\varphi_1 \to \neg Q) \to \neg B_a(\varphi_2 \to \neg Q))$ (5.)
7. $\vdash B_a(\varphi_1 \leftrightarrow \varphi_2) \to (B_a(\varphi_1 \to Q) \wedge \neg B_a(\varphi_1 \to \neg Q) \to (B_a(\varphi_2 \to Q) \wedge \neg B_a(\varphi_2 \to \neg Q)))$ (4. et 6.)
8. $\vdash B_a(\varphi_1 \leftrightarrow \varphi_2) \to (B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q) \to (B_a(\varphi_2 \to Q) \otimes B_a(\varphi_2 \to \neg Q)))$ (7.)
9. $\vdash (B_a(\varphi_1 \leftrightarrow \varphi_2) \wedge (\varphi_1 \leftrightarrow \varphi_2)) \to (U_a^Q\varphi \to U_a^Q\varphi_2)$ (1., 3. et 8.)
10. $\vdash (B_a(\varphi_1 \leftrightarrow \varphi_2) \wedge (\varphi_1 \leftrightarrow \varphi_2)) \to (U_a^Q\varphi \leftrightarrow U_a^Q\varphi_2)$ (9.)

□

Proposition . *I.11*

$$\vdash U_a^Q\varphi_1 \wedge \varphi_2 \wedge \neg B_a(\neg(\varphi_1 \wedge \varphi_2)) \to U_a^Q(\varphi_1 \wedge \varphi_2)$$

Démonstration. 1. $\vdash U_a^Q\varphi_1 \wedge \varphi_2 \to \varphi_1 \wedge \varphi_2$ (définition I.10)

2. $\vdash U_a^Q\varphi_1 \to D_aBif_aQ$ (définition I.10)
3. $\vdash B_a(\varphi_1 \to Q) \wedge B_a((\varphi_1 \wedge \varphi_2) \to Q)$ (axiome (BK))
4. $\vdash B_a(\varphi_1 \to Q) \wedge B_a((\varphi_1 \wedge \varphi_2) \to \neg Q) \to B_a\neg(\varphi_1 \wedge \varphi_2)$ (axiomes (BK) et (BNec))
5. $\vdash B_a(\varphi_1 \to Q) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to \neg B_a((\varphi_1 \wedge \varphi_2) \to \neg Q)$ (4.)
6. $\vdash B_a(\varphi_1 \to Q) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to B_a((\varphi_1 \wedge \varphi_2) \to Q) \wedge \neg B_a((\varphi_1 \wedge \varphi_2) \to \neg Q)$ (3. et 5.)
7. $\vdash B_a(\varphi_1 \to \neg Q) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to B_a((\varphi_1 \wedge \varphi_2) \to \neg Q) \wedge \neg B_a((\varphi_1 \wedge \varphi_2) \to Q)$ (6. et symétrie de Q et $\neg Q$)
8. $\vdash (B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q)) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to (B_a((\varphi_1 \wedge \varphi_2) \to Q) \otimes \neg B_a((\varphi_1 \wedge \varphi_2) \to \neg Q))$ (6. et 7.)
9. $\vdash U_a^Q\varphi_1 \wedge \varphi_2 \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to U_a^Q(\varphi_1 \wedge \varphi_2)$ (1., 2. et 8.)

□

Proposition . *I.12*

$$\vdash U_a^Q\varphi_1 \wedge U_a^Q\varphi_2 \wedge \neg B_a(\neg(\varphi_1 \wedge \varphi_2)) \to U_a^Q(\varphi_1 \vee \varphi_2)$$

Démonstration. 1. $\vdash U_a^Q\varphi_1 \wedge U_a^Q\varphi_2 \wedge (\varphi_1 \vee \varphi_2)$ (définition I.10)

2. $\vdash U_a^Q\varphi_1 \to D_aBif_aQ$ (définition I.10)
3. $\vdash B_a(\varphi_1 \to Q) \wedge B_a(\varphi_2 \to Q) \leftrightarrow B_a(\varphi_1 \vee \varphi_2 \to Q)$ (axiomes (BK) et (BNec))
4. $\vdash B_a(\varphi_1 \to \neg Q) \wedge B_a(\varphi_2 \to \neg Q) \leftrightarrow B_a(\varphi_1 \vee \varphi_2 \to \neg Q)$ (axiomes (BK) et (BNec))

5. $\vdash B_a(\varphi_2 \to \neg Q) \to B_a(Q \to \neg\varphi_2)$

6. $\vdash B_a(\varphi_2 \to \neg Q) \wedge B_a(\varphi_1 \to Q) \to B_a(\varphi_1 \to \neg\varphi_2)$ (5. et axiome (BK))

7. $\vdash B_a(\varphi_2 \to \neg Q) \wedge B_a(\varphi_1 \to Q) \to B_a\neg(\varphi_1 \wedge \varphi_2)$ (6.)

8. $\vdash B_a(\varphi_2 \to \neg Q) \wedge B_a(\varphi_1 \to Q) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to \bot$ (7.)

9. $\vdash B_a(\varphi_1 \to \neg Q) \wedge B_a(\varphi_2 \to Q) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to \bot$ (8. et symétrie de φ_1 et φ_2)

10. $\vdash (B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q)) \wedge (B_a(\varphi_2 \to Q) \otimes B_a(\varphi_2 \to \neg Q)) \to ((B_a(\varphi_1 \to Q) \wedge B_a(\varphi_2 \to Q)) \vee (B_a(\varphi_1 \to Q) \wedge B_a(\varphi_2 \to \neg Q)) \vee (B_a(\varphi_1 \to \neg Q) \wedge B_a(\varphi_2 \to Q)) \vee (B_a(\varphi_1 \to \neg Q) \wedge B_a(\varphi_2 \to \neg Q)))$ (définition de $\otimes$)

11. $\vdash (B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q)) \wedge (B_a(\varphi_2 \to Q) \otimes B_a(\varphi_2 \to \neg Q)) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to (B_a((\varphi_1 \vee \varphi_2) \to Q) \vee B_a((\varphi_1 \vee \varphi_2) \to \neg Q))$ (3., 4., 8. et 9.)

12. $\vdash (B_a((\varphi_1 \vee \varphi_2) \to Q) \wedge B_a((\varphi_1 \vee \varphi_2) \to \neg Q)) \to B_a\neg\varphi_1 \wedge B_a\neg\varphi_2$

13. $\vdash B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q) \to \neg B_a\varphi_1$ (définition de $\otimes$)

14. $\vdash (B_a(\varphi_1 \to Q) \otimes B_a(\varphi_1 \to \neg Q)) \wedge (B_a(\varphi_2 \to Q) \otimes B_a(\varphi_2 \to \neg Q)) \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to (B_a((\varphi_1 \vee \varphi_2) \to Q) \otimes B_a((\varphi_1 \vee \varphi_2) \to \neg Q))$ (10., 11. et 12.)

15. $\vdash U_a^Q\varphi_1 \wedge U_a^Q\varphi_2 \wedge \neg B_a\neg(\varphi_1 \wedge \varphi_2) \to U_a^Q(\varphi_1 \vee \varphi_2)$ (1., 2. et 13.)

$\square$

Proposition . *I.13* *Soient Q et φ des formules objectives. Si φ appartient à l'ensemble $\mathcal{U}_a^Q$ (informations utiles dans w), alors φ est soit une explication de Q dans w, soit une explication de $\neg Q$ dans w.*

Démonstration. Soit φ une information utile pour l'agent a par rapport à Q dans le monde w. Par la définition de l'utilité (definition I.10), on a donc $\mathbb{M}, w \models B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q)$.
Supposons que $\mathbb{M}, w \models B_a(\varphi \to Q) \wedge \neg B_a(\varphi \to \neg Q)$. D'après la proposition I.3, on a $\mathbb{M}, w \models \neg B_a\neg\varphi$. φ est donc une explication de Q dans w.
Si $\mathbb{M}, w \models B_a(\varphi \to \neg Q) \wedge \neg B_a(\varphi \to Q)$, on montre que φ est une explication de $\neg Q$ dans w.
φ est donc une explication de Q ou une explication de $\neg Q$ dans w. $\square$

Proposition . *I.14* *Soit a un agent de $\mathcal{A}$, soient φ et Q deux formules objectives. Si φ est une formule utile pour l'agent a par rapport à sa requête Q alors il existe une explication minimale de Q ou une explication minimale de $\neg Q$ qui mentionne une variable propositionelle de φ. Autrement dit π_φ est L-pertinent pour Q ou pour $\neg Q$.*

Démonstration. Soit φ une explication utile pour l'agent a dans le monde w. D'après la proposition I.13, φ est donc une explication de Q ou une explication de $\neg Q$.
Si φ est une explication minimale de Q ou une explication minimale de $\neg Q$, alors il existe une explication minimale de Q ou de $\neg Q$ qui mentionne une variable de π_φ (car φ mentionne une variable de π_φ). On a donc π_φ L-pertinent pour Q dans w ou π_φ L-pertinent pour $\neg Q$ dans w.
Sinon, alors il existe une explication ψ de Q ou une explication de $\neg Q$ telle que $DepLit(\psi) \subset DepLit(\varphi)$. $DepLit(\psi) \subset DepLit(\varphi)$ implique que ψ mentionne une variable propositionnelle de π_φ. Il existe donc une explication minimale de Q ou de $\neg Q$ qui mentionne une variable de π_φ.
On a donc π_φ L-pertinent pour Q dans w ou π_φ L-pertinent pour $\neg Q$ dans w. $\square$

Corollaire . *I.3* *Soit a un agent de $\mathcal{A}$, soient φ et Q deux formules objectives. Supposons que φ appartient à $\mathcal{U}m_a^Q$, c'est à dire que φ fait partie des informations les plus utiles pour l'agent a par rapport à sa requête Q (dans un monde w). Alors non seulement il existe une explication minimale de Q ou une explication minimale de $\neg Q$ dans w qui mentionne une variable propositionnelle de φ, mais φ est une des explications minimales de Q ou de $\neg Q$ dans w.*

Démonstration. La preuve de cette proposition est déjà présente dans la preuve de la proposition I.14. En effet, le deuxième cas de cette preuve (cas où φ n'est pas une explication minimale) n'est pas considéré ici à cause des hypothèses. □

Proposition . *I.15*

$$\vdash D_a Bif_a Q \to P_a^Q Q \wedge P_a^Q \neg Q$$

Démonstration.

1. $\vdash D_a Bif_a Q \to \neg B_a Q \wedge \neg B_a \neg Q$ (proposition I.1)
2. $\vdash B_a(Q \to Q)$ (axiome (BNec))
3. $\vdash \neg B_a \neg Q \to \neg B_a(Q \to \neg Q)$
4. $\vdash \neg B_a \neg Q \to (B_a(Q \to Q) \wedge \neg B_a(Q \to \neg Q))$ (2. et 3.)
5. $\vdash \neg B_a \neg Q \to (B_a(Q \to Q) \otimes B_a(Q \to \neg Q))$ (4.)
6. $\vdash \neg B_a Q \to (B_a(Q \to Q) \otimes B_a(Q \to \neg Q))$ (5. et symétrie de Q et $\neg Q$)
7. $\vdash D_a Bif_a Q \to P_a^Q Q$ (1. et 5.)
8. $\vdash D_a Bif_a Q \to P_a^Q \neg Q$ (1. et 6.)

□

Proposition . *I.16*

$$\vdash P_a^Q \varphi \to \neg B_a \varphi \wedge \neg B_a \neg \varphi$$

Démonstration. La preuve de cette proposition est identique aux preuves des propositions I.2 et I.3. En effet, dans ces deux preuves, le fait que φ soit une information vraie n'intervient pas. □

Proposition . *I.17*

$$\vdash P_a^Q \varphi \to B_a P_a^Q \varphi$$

Démonstration.

1. $\vdash D_a Bif_a Q \to B_a D_a Bif_a Q$ (axiome (D4))
2. $\vdash B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q) \to B_a(B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q))$ (axiomes (B4) et (B5))
3. $\vdash P_a^Q \varphi \to B_a P_a^Q \varphi$

□

Proposition . *I.18* *Soit a un agent de $\mathcal{A}$. Soit $\mathbb{M}$ un modèle de la classe* **C** *et soit w un monde de ce modèle. Soit n un entier naturel supérieur ou égal à 1. Soient Q_1, $\ldots Q_n$ n formules mutuellement exclusives pour l'agent a dans le monde w. On note* $\mathbf{Q}$ *l'ensemble $\{Q_1, ...Q_n\}$. On a alors :*

$$\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n} Q_i \to \left[Us_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{\mathbf{Q}} \varphi\right]$$

Démonstration. La preuve de cette proposition est divisée en trois parties. Nous montrons tout d'abord l'équivalence des parties de la formule qui portent sur le besoin en information. Puis, nous montrons l'équivalence des parties de la formule qui portent sur les croyances de l'agent. Finalement, nous traitons le cas $n = 1$ à part. Dans toute cette preuve, on considère que a est un agent de $\mathcal{A}$, que n est un entier supérieur ou égal à 2, que $Q_1, \ldots, Q_n$ sont n formules objectives mutuellement exclusives pour l'agent a et que $\mathbf{Q}$ représente l'ensemble l'ensemble $\{Q_1, ...Q_n\}$.

Montrons que $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow (D_a (\bigvee_i B_a Q_i) \leftrightarrow D_a (\bigvee_i Q_i \vee \bigwedge_i B_a \neg Q_i))$.

1. $\mathbb{M}, w \vDash \bigwedge_i B_a \neg Q_i \leftrightarrow B_a \bigwedge_i \neg Q_i$ (axiome (BK) et règle (BNec))
2. $\mathbb{M}, w \vDash \bigwedge_i B_a \neg Q_i \leftrightarrow B_a \neg \bigvee_i Q_i$ (1.)
3. Or, $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow \neg B_a \neg \bigvee_i Q_i$ (axiome (BD))
4. $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow \neg \bigwedge_i B_a \neg Q_i$ (2. et 3.)

   Soient α, β et γ trois formules telles que $\mathbb{M}, w \vDash \alpha \rightarrow \neg \beta$. On a
5. $\mathbb{M}, w \vDash \alpha \rightarrow (\gamma \leftrightarrow (\gamma \vee \beta))$

   En prenant $\alpha = B_a \bigvee_i Q_i$, $\beta = \bigwedge_i B_a \neg Q_i$ et $\gamma = \bigvee_i B_a Q_i$, on a :
6. $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow (\bigvee_i B_a Q_i \leftrightarrow (\bigvee_i B_a Q_i \vee \bigwedge_i B_a \neg Q_i))$ (4. et 5.)
7. $\mathbb{M}, w \vDash B_a(B_a \bigvee_i Q_i \rightarrow (\bigvee_i B_a Q_i \leftrightarrow (\bigvee_i B_a Q_i \vee \bigwedge_i B_a \neg Q_i)))$ (règle (BNec))
8. Or, $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow B_a B_a \bigvee_i Q_i$ (axiome (B4))
9. $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow B_a (\bigvee_i B_a Q_i \leftrightarrow (\bigvee_i B_a Q_i \vee \bigwedge_i B_a \neg Q_i)))$ (7., 8. et axiome (BK))
10. $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow (D_a (\bigvee_i B_a Q_i) \leftrightarrow D_a (\bigvee_i Q_i \vee \bigwedge_i B_a \neg Q_i))$ (9. et axiome (DRE))

Montrons que $\mathbb{M}, w\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow ((\bigvee_i B_a(\varphi \rightarrow Q_i) \otimes \bigwedge_i B_a(\varphi \rightarrow \neg Q_i)) \leftrightarrow \bigotimes_i B_a(\varphi \rightarrow Q_i))$.

1. $\vdash (\bigotimes_i B_a(\varphi \rightarrow Q_i) \wedge B_a \neg \varphi) \rightarrow (\bigvee_i (B_a(\varphi \rightarrow Q_i) \wedge \bigwedge_{i \neq j} \neg B_a(\varphi \rightarrow Q_i) \wedge \bigwedge_{i \neq j} B_a(\varphi \rightarrow Q_i)))$ (définition de $\otimes$, axiome (BK), et pour toute formule α, $\mathbb{M}, w \vDash B_a \neg \varphi \rightarrow B_a(\varphi \rightarrow \alpha)$)
2. $\mathbb{M}, w \vDash \bigotimes_i B_a(\varphi \rightarrow Q_i) \rightarrow \neg B_a \neg \varphi$ (1. et axiome (BD))
3. $\mathbb{M}, w \vDash \bigotimes_i B_a(\varphi \rightarrow Q_i) \rightarrow (\neg B_a \neg \varphi \wedge \bigvee_i B_a(\varphi \rightarrow Q_i))$ (2.)
4. Soient i et j deux entiers différents et inférieurs ou égaux à n (possible car $n \geq 2$). On a $\mathbb{M}, w \vDash (B_a(\varphi \rightarrow (Q_i \wedge Q_j)) \wedge B_a \neg (Q_i \wedge Q_j)) \rightarrow B_a \neg \varphi$ (axiome (BK))
5. $\mathbb{M}, w \vDash B_a \neg (Q_i \wedge Q_j)$ car Q_i et Q_j sont mutuellement exclusives pour l'agent a.
6. $\mathbb{M}, w \vDash (B_a(\varphi \rightarrow Q_i) \wedge B_a(\varphi \rightarrow Q_j) \wedge \neg B_a \neg \varphi) \rightarrow \bot$ (4. et 5.)
7. $\mathbb{M}, w \vDash (B_a(\varphi \rightarrow Q_i) \wedge \neg B_a \neg \varphi) \rightarrow \neg B_a(\varphi \rightarrow Q_j)$ (6.)
8. $\mathbb{M}, w \vDash (\bigvee_i B_a(\varphi \rightarrow Q_i) \wedge \neg B_a \neg \varphi) \rightarrow (\bigvee_i (B_a(\varphi \rightarrow Q_i) \wedge \bigwedge_{i \neq j} \neg B_a(\varphi \rightarrow Q_j)))$ (généralisation de 7. pour tout i)

9. $\mathbb{M}, w \models (\bigvee_i B_a(\varphi \to Q_i) \wedge \neg B_a \neg\varphi) \to \bigotimes_i B_a(\varphi \to Q_i)$ (8.)

10. $\boxed{\mathbb{M}, w \models \bigotimes_i B_a(\varphi \to Q_i) \leftrightarrow (\neg B_a \neg\varphi \wedge \bigvee_i B_a(\varphi \to Q_i)}$ (3. et 9.)

11. $\mathbb{M}, w \models (\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i)) \leftrightarrow ((\bigvee_i B_a(\varphi \to Q_i) \wedge \neg \bigwedge_i B_a(\varphi \to \neg Q_i)) \vee (\neg \bigvee_i B_a(\varphi \to Q_i) \wedge \bigwedge_i B_a(\varphi \to \neg Q_i)))$

12. $\mathbb{M}, w \models \neg \bigvee B_a(\varphi \to Q_i) \leftrightarrow \wedge \bigwedge_i \neg B_a(\varphi \to Q_i)$

13. pour tout i, $\mathbb{M}, w \models \neg B_a(\varphi \to Q_i) \to \neg B_a \neg\varphi$

14. $\mathbb{M}, w \models \bigwedge_i(\neg B_a(\varphi \to Q_i) \wedge B_a(\varphi \to \neg Q_i)) \to (\bigwedge_i B_a(\varphi \to \neg Q_i) \wedge \neg B_a \neg\varphi)$ (3.)

15. $\mathbb{M}, w \models \bigwedge_i B_a(\varphi \to \neg Q_i) \leftrightarrow B_a(\varphi \to \bigwedge_i \neg Q_i)$ (axiome (BK))

16. $\mathbb{M}, w \models \bigwedge_i B_a(\varphi \to \neg Q_i) \leftrightarrow B_a(\varphi \to \neg \bigvee_i Q_i)$ (5.)

17. $\mathbb{M}, w \models (\bigwedge_i B_a(\varphi \to \neg Q_i) \wedge B_a \bigvee_i Q_i) \to B_a \neg\varphi$ (6. et axiome (BK))

18. $\mathbb{M}, w \models (\bigwedge_i(\neg B_a(\varphi \to Q_i) \wedge B_a(\varphi \to \neg Q_i)) \wedge B_a \bigvee_i Q_i) \to \perp$ (4., 7. et axiome (BD))

19. $\mathbb{M}, w \models ((\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i)) \wedge B_a \bigvee_i Q_i) \leftrightarrow ((\bigvee_i B_a(\varphi \to Q_i) \wedge \neg \bigwedge_i B_a(\varphi \to \neg Q_i) \wedge B_a \bigvee_i Q_i)$ (1., 2. et 18.)

20. $\mathbb{M}, w \models B_a \bigvee_i Q_i \to ((\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i)) \leftrightarrow ((\bigvee_i B_a(\varphi \to Q_i) \wedge \neg \bigwedge_i B_a(\varphi \to \neg Q_i))))$

21. $\mathbb{M}, w \models B_a \neg\varphi \to \bigwedge_i(\varphi \to \neg Q_i)$ (axiome (BK))

22. $\mathbb{M}, w \models \neg \bigwedge_i(\varphi \to \neg Q_i) \to \neg B_a \neg\varphi$ (21.)

23. $\mathbb{M}, w \models (B_a \bigvee_i Q_i \wedge \bigwedge_i B_a(\varphi \to \neg Q_i)) \to B_a \neg\varphi$ (axiome (BK))

24. $\mathbb{M}, w \models B_a \bigvee_i Q_i \to (\neg B_a \neg\varphi \to \neg \bigwedge_i B_a(\varphi \to \neg Q_i))$ (23.)

25. $\mathbb{M}, w \models B_a \bigvee_i Q_i \to (\neg B_a \neg\varphi \leftrightarrow \neg \bigwedge_i B_a(\varphi \to \neg Q_i))$ (22. et 24.)

26. $\mathbb{M}, w \models B_a \bigvee_i Q_i \to ((\neg B_a \neg\varphi \wedge \bigvee_i B_a(\varphi \to Q_i)) \leftrightarrow (\neg \bigwedge_i B_a(\varphi \to \neg Q_i) \wedge \bigvee_i B_a(\varphi \to Q_i))$ (25.)

27. $\boxed{\mathbb{M}, w \models B_a \bigvee_i Q_i \to ((\neg B_a \neg\varphi \wedge \bigvee_i B_a(\varphi \to Q_i)) \leftrightarrow ((\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i))))}$ (20. et 26.)

28. $\boxed{\mathbb{M}, w \models B_a \bigvee_i Q_i \to ((\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i)) \leftrightarrow \bigotimes_i B_a(\varphi \to Q_i))}$ (10. et 27.)

A ce stade, nous avons montré que pour $n \geq 2$, on a bien

$$\mathbb{M}, w \models B_a \bigvee_{i=1}^{n} Q_i \to \left[\left(D_a(\bigvee_{i=1}^{n} B_a Q_i) \wedge \bigotimes_{i=1}^{n} B_a(\varphi \to Q_i) \wedge \varphi \right) \leftrightarrow U_a^{\mathbf{Q}} \varphi \right]$$

Nous étudions maintenant le cas $n = 1$. Dans ce cas, l'hypothèse $B_a \bigvee_i Q_i$ peut se réduire à $B_a Q_1$, on peut également réduire $\bigvee_i Q_i$ à Q_1 et $\bigwedge_i Q_i$ à Q_1.

1. $\mathbb{M}, w \models D_a B_a Q_1 \to \neg B_a B_a Q_1$ (théorème I.1)

2. $\mathbb{M}, w \models B_a Q_1 \to B_a B_a \neg Q_1$ (axiome (B4))

3. $\mathbb{M}, w \models (B_a Q_1 \wedge D_a B_a Q_1) \to \perp$ (1. et 2.)

4. $\mathbb{M}, w \models B_a Q_1 \to \neg D_a B_a Q_1$ (3.)

5. $\mathbb{M}, w \models B_a Q_1 \to \neg(D_a B_a Q_1 \wedge B_a(\varphi \to Q_1) \wedge \varphi)$ (4.)

6. $\mathbb{M}, w \models D_a(B_a Q_1 \vee B_a \neg Q_1) \to \neg B_a Q_1$ (proposition I.1)

7. $\mathbb{M}, w \models B_a Q_1 \to \neg D_a(B_a Q_1 \vee B_a \neg Q_1)$ (6.)

8. $\mathbb{M}, w \models B_a Q_1 \to \neg U_a^{\mathbf{Q}} \varphi$ (7.)

9. $\boxed{\mathbb{M}, w \models B_a Q_1 \to ((D_a B_a Q_1 \wedge B_a(\varphi \to Q_1) \wedge \varphi) \leftrightarrow U_a^{\mathbf{Q}} \varphi) \text{ (5. et 8.)}}$

$\square$

Corollaire . *I.4* *Soit a un agent de $\mathcal{A}$. Soit n un entier naturel supérieur ou égal à 1. Soient $Q_1, \ldots Q_n$ n formules mutuellement exclusives pour l'agent a. On note $\mathbf{Q}$ l'ensemble $\{Q_1, ...Q_n\}$. On a alors :*

$$\textit{Si } \vdash \bigvee_{i=1}^{n} Q_i \textit{ alors } \mathbb{M}, w \models Us_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{\mathbf{Q}} \varphi$$

Démonstration. Supposons $\vdash \bigvee_{i=1}^{n} Q_i$.

1. $\vdash \bigvee_i Q_i$ (hypothèse)

2. $\vdash B_a(\bigvee_i Q_i)$ (1. et règle (BNec))

3. pour tout w, $\mathbb{M}, w \models B_a(\bigvee_i Q_i) \to (Us_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{\mathbf{Q}} \varphi)$ (proposition I.18)

4. $\mathbb{M}, w \models Us_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{\mathbf{Q}} \varphi$ (2. et 3.)

$\square$

Corollaire . *I.5* *Soient φ une formule et Q une formule objective. La formule φ est utile pour l'agent a par rapport à l'ensemble de requêtes $\{Q, \neg Q\}$ si et seulement si elle est utile pour l'agent a par rapport à sa requête Q ou encore si et seulement si elle est utile pour l'agent par rapport à l'ensemble de requêtes $\{Q\}$.*
Formellement, cela signifie que

$$\vdash U_a^{Q} \varphi \leftrightarrow U_a^{\{Q, \neg Q\}} \varphi \leftrightarrow U_a^{\{Q\}} \varphi$$

Démonstration. Montrons dans un premier temps que $\vdash U_a^{Q} \varphi \leftrightarrow U_a^{\{Q, \neg Q\}}$.

1. $\vdash Q \vee \neg Q$

2. $\vdash Us_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{\mathbf{Q}} \varphi$ (1. et corollaire I.4)

3. $\vdash Us_a^{\mathbf{Q}} \leftrightarrow (D_a(B_a Q \vee B_a \neg Q) \wedge (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q)) \wedge \varphi)$ (définition de $Us_a^{\mathbf{Q}}$)

4. $\vdash Us_a^{\mathbf{Q}} \leftrightarrow U_a^{Q} \varphi$ (définition de l'utilité (définition I.10) et 3.)

5. $\vdash U_a^{\mathbf{Q}} \varphi \leftrightarrow U_a^{Q} \varphi$ (2. et 4.)

Montrons maintenant que $\vdash U_a^{\{Q\}} \varphi \leftrightarrow U_a^{Q} \varphi$.
Il suffit d'appliquer la définition de l'utilité (définition I.10) avec $n = 1$. En effet, $U_a^{\{Q\}}$ signifie que $D_a(B_a Q \vee B_a \neg Q) \wedge (B_a(\varphi \to Q) \otimes B_a(\varphi \to \neg Q)) \wedge \varphi$. C'est la définition de $U_a^{Q} \varphi$. $\square$

Proposition . *I.19*

$$\vdash D_a\left(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i\right) \to \bigwedge_{i=1}^{n} \neg B_aQ_i \wedge \bigvee_{i=1}^{n} \neg B_a\neg Q_i$$

Démonstration. Montrons que pour tout entier k, $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg B_aQ_k$. Soit k un entier compris entre 1 et n.

1. $\vdash B_aQ_k \to B_aB_aQ_k$ (axiome (B4))
2. $\vdash B_aQ_k \to \bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i$ (calcul propositionnel)
3. $\vdash B_aB_aQ_k \to B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (2. et axiome (BK))
4. $\vdash B_aQ_k \to B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (1. et 3.)
5. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (théorème I.1)
6. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \wedge B_aQ_k \to \bot$ (4. et 5.)
7. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg B_aQ_k$ (6.)

Montrons que $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg \bigwedge_i B_a\neg Q_i$

1. $\bigwedge_i B_a\neg Q_i \to B_a \bigwedge_i \neg Q_i$ (axiome (BK))
2. $B_a \bigwedge_i \neg Q_i \to B_aB_a \bigwedge_i \neg Q_i$ (axiome (B4))
3. $B_aB_a \bigwedge_i \neg Q_i \to B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (axiome (BK))
4. $\bigwedge_i B_a\neg Q_i \to B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (1. et 3.)
5. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg B_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (théorème I.1)
6. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \wedge \bigwedge_i B_a\neg Q_i \to \bot$ (4. et 5.)
7. $\vdash D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \to \neg \bigwedge_i B_a\neg Q_i$ (6.)

□

Proposition . *I.20*

$$\vdash U_a^{\mathbf{Q}}\varphi \to \neg B_a\varphi$$

Démonstration.

1. $\vdash U_a^{\mathbf{Q}}\varphi \to D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (définition de l'utilité I.22)
2. $\vdash U_a^{\mathbf{Q}}\varphi \to (\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i))$ (définition de l'utilité I.22)
3. $\vdash (B_a\varphi \wedge \bigvee_i B_a(\varphi \to Q_i)) \to \bigvee_i B_aQ_i$ (axiome (BK))
4. $\vdash (B_a\varphi \wedge \bigwedge_i B_a(\varphi \to \neg Q_i) \to \bigwedge_i B_a\neg Q_i$ (axiome (BK))
5. $\vdash (B_a\varphi \wedge (\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i))) \to (\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (3. et 4.)
6. $\vdash (U_a^{\mathbf{Q}}\varphi \wedge B_a\varphi) \to \neg(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (2. et 5.)
7. $\vdash U_a^{\mathbf{Q}}\varphi \to (\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)$ (1. et proposition I.19)
8. $\vdash (U_a^{\mathbf{Q}}\varphi \wedge B_a\varphi) \to \bot$ (6. et 7.)
9. $\vdash U_a^{\mathbf{Q}}\varphi \wedge \neg B_a\varphi$ (8.)

□

Proposition . *I.21*

$$\vdash U_a^{\mathbf{Q}}\varphi \to \neg B_a\neg\varphi$$

Démonstration. 1. $\vdash U_a^{\mathbf{Q}}\varphi \to (\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i))$ (définition de l'utilité I.22)

2. $\vdash B_a\neg\varphi \to (\bigvee_i B_a(\varphi \to Q_i) \wedge \bigwedge_i B_a(\varphi \to \neg Q_i))$ (axiome (BK))

3. $\vdash B_a\neg\varphi \to \neg(\bigvee_i B_a(\varphi \to Q_i) \otimes \bigwedge_i B_a(\varphi \to \neg Q_i))$ (2.)

4. $\vdash U_a^{\mathbf{Q}}\varphi \to \neg B_a\neg\varphi$ (1. et 3.)

□

Proposition IV.1. *I.22*

$$\vdash \left(\bigotimes_{i=1}^{n} Q_i \wedge \bigwedge_{i=1}^{n} \neg B_a(\neg Q_i) \wedge D_a(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i)\right) \to \bigotimes_{i=1}^{n} U_a^{\mathbf{Q}}Q_i$$

Démonstration. Montrons tout d'abord que
$\vdash (\bigvee_i Q_i \wedge \bigwedge_i \neg B_a(\neg Q_i) \wedge D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)) \to \bigvee_i U_a^{\mathbf{Q}}Q_i$. Soit k un entier compris entre 1 et n.

1. $\vdash \neg U_a^{\mathbf{Q}}Q_k \to (\neg Q_k \vee \neg D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \vee \neg(\bigvee_i B_a(Q_k \to Q_i) \otimes \bigwedge_i B_a(Q_k \to \neg Q_i)))$ (définition de I.22)

2. $\vdash \bigvee_i B_a(Q_k \to Q_i)$ (règle (BNec))

3. $\vdash \neg B_a\neg Q_k \leftrightarrow \neg B_a(Q_k \to \neg Q_k)$

4. $\vdash \neg B_a\neg Q_k \to \neg \bigwedge_i B_a(Q_k \to \neg Q_i)$ (3.)

5. $\vdash \neg B_a\neg Q_k \to (\bigvee_i B_a(Q_k \to Q_i) \otimes \bigwedge_i B_a(Q_k \to \neg Q_i))$ (2. et 4.)

6. $\vdash (\bigotimes_i Q_i \wedge \bigwedge_i \neg B_a(\neg Q_i) \wedge D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \wedge \neg U_a^{\mathbf{Q}}Q_k) \to (\neg Q_k \wedge \bigvee_i Q_i)$ (1. et 5.)

7. $\vdash (\bigotimes_i Q_i \wedge \bigwedge_i \neg B_a(\neg Q_i) \wedge D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i) \wedge \bigwedge_i \neg U_a^{\mathbf{Q}}Q_i) \to (\bigwedge_i \neg Q_i \wedge \bigvee_i Q_i)$ (conjonction de 6. pour tout k)

8. $\vdash (\bigotimes_i Q_i \wedge \bigwedge_i \neg B_a(\neg Q_i) \wedge D_a(\bigvee_i B_aQ_i \vee \bigwedge_i B_a\neg Q_i)) \to \bigvee_i U_a^{\mathbf{Q}}Q_i$ (7.)

Montrons maintenant que pour tout j et k compris entre 1 et n, on a $\vdash \bigotimes_i Q_i \to \neg(U_a^{\mathbf{Q}}Q_j \wedge U_a^{\mathbf{Q}}Q_k)$. Soient j et k deux entiers compris entre 1 et n.

1. $\vdash (U_a^{\mathbf{Q}}Q_j \wedge U_a^{\mathbf{Q}}Q_k) \to (Q_j \wedge Q_k)$ (définition de I.22)

2. $\vdash \bigotimes_i Q_i \to \neg(Q_j \wedge Q_k)$

3. $\vdash \bigotimes_i Q_i \to \neg(U_a^{\mathbf{Q}}Q_j \wedge U_a^{\mathbf{Q}}Q_k)$ (1. et 2.)

□

Proposition . *I.23*

$$\vdash \neg B_aU_a^{\mathbf{Q}}\varphi$$

Démonstration. 1. $\vdash \neg B_aU_a^{\mathbf{Q}}\varphi \to B_a\varphi$ (définition de I.22).

2. $\vdash \neg B_aU_a^{\mathbf{Q}}\varphi \to B_aB_a\varphi$ (axiome (B4))

2. $\vdash U_a^{\mathbf{Q}}\varphi \rightarrow \neg B_a\varphi$ (proposition I.20)

3. $\vdash B_a U_a^{\mathbf{Q}}\varphi \rightarrow B_a\neg B_a\varphi$ (2., règle (BNec) et axiome (BK))

4. $\vdash \neg B_a U_a^{\mathbf{Q}}\varphi$ (2., 3. et axiome (BD))

□

Proposition . *I.24*

$$\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \rightarrow (U_a^{\{Q_1,\dots Q_{n+1}\}}\varphi \leftrightarrow U_a^{\{Q_1,\dots Q_n\}}\varphi)$$

Démonstration. Pour montrer cette proposition, nous démontrons tout d'abord le lemme suivant :

Lemme IV.1.

$$\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \rightarrow (D_a(\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} Q_i) \leftrightarrow D_a(\bigvee_{i=1}^{n+1} B_aQ_i)$$

Démonstration. 1. $\mathbb{M}, w \vDash B_a \bigvee_i Q_i \rightarrow (U_a^{\{Q_1,\dots,Q_n\}}\varphi \leftrightarrow Us_a^{\{Q_1,\dots Q_{n+1}\}}\varphi)$ (proposition I.18)

2. $\mathbb{M}, w \vDash \bigvee_{i=1}^{n+1} Q_i \wedge \bigwedge_{i=1}^{n} \neg Q_i \rightarrow Q_{n+1}$

3. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \wedge B_a(\bigwedge_{i=1}^{n} \neg Q_i) \rightarrow B_aQ_{n+1}$ (2. et (BNec))

4. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \wedge \bigwedge_{i=1}^{n}(B_a\neg Q_i) \rightarrow B_aQ_{n+1}$ (3.)

5. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \wedge (\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i)) \rightarrow (\bigvee_{i=1}^{n} B_aQ_i \vee B_aQ_{n+1})$ (4.)

6. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow ((\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i)) \rightarrow \bigvee_{i=1}^{n+1} B_aQ_i)$ (5.)

7. $\mathbb{M}, w \vDash B_aQ_{n+1} \rightarrow \bigwedge_{i=1}^{n} B_aQ_i$ car $\forall i$, Q_{n+1} et Q_i sont mutuellement exclusives dans w pour a.

8. $\mathbb{M}, w \vDash \bigvee_{i=1}^{n+1} B_aQ_i \rightarrow \bigvee_{i=1}^{n} B_aQ_i \vee B_aQ_{n+1}$

9. $\mathbb{M}, w \vDash \bigvee_{i=1}^{n+1} B_aQ_i \rightarrow \bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n} B_a\neg Q_i$ (7. et 8.)

10. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow ((\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i)) \leftrightarrow \bigvee_{i=1}^{n+1} B_aQ_i)$ (6. et 9.)

11. $\mathbb{M}, w \vDash B_aB_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow B_a((\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i)) \leftrightarrow \bigvee_{i=1}^{n+1} B_aQ_i)$ (10. et (BNec))

12. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow B_aB_a(\bigvee_{i=1}^{n+1} Q_i)$ (B4)

13. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow B_a((\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i)) \leftrightarrow \bigvee_{i=1}^{n+1} B_aQ_i)$ (10., 12. et (BK))

14. $\mathbb{M}, w \vDash B_a(\bigvee_{i=1}^{n+1} Q_i) \rightarrow D_a((\bigvee_{i=1}^{n} B_aQ_i \vee \bigwedge_{i=1}^{n}(B_a\neg Q_i))) \leftrightarrow D_a(\bigvee_{i=1}^{n+1} B_aQ_i))$ (13. et (DRE))

□

Montrons ensuite le lemme suivant :

Lemme IV.2.

$$\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \rightarrow ((\bigvee_{i=1}^{n} B_a(\varphi \rightarrow Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \rightarrow Q_i)) \leftrightarrow (\bigotimes_{i=1}^{n+1} B_a(\varphi \rightarrow Q_i))$$

Démonstration. 1. $\mathbb{M}, w \vDash \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i) \leftrightarrow B_a(\varphi \to \bigwedge_{i=1}^{n} \neg Q_i)$ (axiome (BK))

2. $\mathbb{M}, w \vDash \bigvee_{i=1}^{n+1} Q_i \to (\bigwedge_{i=1}^{n} \neg Q_i \leftrightarrow Q_{n+1})$ car pour tout i, Q_{n+1} et Q_i sont exclusives dans w pour a

3. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to B_a(\bigwedge_{i=1}^{n} \neg Q_i \leftrightarrow Q_{n+1})$ ((BNec), (BK) et 2.)

4. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to (B_a(\varphi \to \bigwedge_{i=1}^{n} \neg Q_i) \to B_a(\varphi \to Q_{n+1}))$ (3. et (BK))

5. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)) \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \vee B_a(\varphi \to Q_{n+1}))$ (4.)

6. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)) \to \bigvee_{i=1}^{n+1} B_a(\varphi \to Q_i)$ (5.)

7. Soient j et k deux entiers inférieurs ou égaux à $n+1$,
$\mathbb{M}, w \vDash (B_a(\varphi \to Q_j) \wedge B_a(\varphi \to Q_k)) \to B_a(\varphi \to (Q_j \wedge Q_k))$ (axiome (BK))

8. $\mathbb{M}, w \vDash B_a(\varphi \to (Q_j \wedge Q_k)) \to B_a\neg\varphi$ car Q_j et Q_k sont exclusives dans w pour a.

9. $\mathbb{M}, w \vDash B_a\neg\varphi \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \wedge \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i))$ (axiome (BK))

10. $\mathbb{M}, w \vDash B_a\neg\varphi \to \neg(\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i))$ (9.)

11. pour tout j et k inférieurs ou égaux à $n-1$, on a donc
$\mathbb{M}, w \vDash (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)) \to \neg(B_a(\varphi \to Q_j) \wedge B_a(\varphi \to Q_k))$ (7., 8. et 10.)

12. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)) \to \bigotimes_{i=1}^{n+1} B_a(\varphi \to Q_i)$ (6. et 11.)

13. $\mathbb{M}, w \vDash \bigotimes_{i=1}^{n+1} B_a(\varphi \to Q_i) \to \bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes B_a(\varphi \to Q_{n+1})$ (définition de $\bigotimes$)

14. $\mathbb{M}, w \vDash \bigvee_{i=1}^{n+1} B_a Q_i \to (\bigotimes_{i=1}^{n+1} B_a(\varphi \to Q_i) \to \bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes B_a(\varphi \to \bigwedge_{i=1}^{n} Q_i))$ (3. et 13.)

15. $\mathbb{M}, w \vDash B_a \bigvee_{i=1}^{n+1} Q_i \to (\bigvee_{i=1}^{n} B_a(\varphi \to Q_i) \otimes \bigwedge_{i=1}^{n} B_a(\varphi \to \neg Q_i)) \leftrightarrow \bigotimes_{i=1}^{n+1} B_a(\varphi \to Q_i)$ (12. et 14.)

□

Pour montrer la proposition I.23, il suffit finalement d'utiliser les deux lemmes que nous venons de montrer.

□

Proposition . ***I.25*** *Le système de preuves complété avec l'axiome (InfRE) est valide et complet pour la classe de modèles étendue* **C**.

Démonstration. **Validité**
Montrons que la règle d'inférence conserve la validité.
Soit φ et ψ deux formules telles que $\varphi \leftrightarrow \psi$ soit une formule valide dans la classe **C**. Soit $\mathbb{M}$ un modèle de **C**.
On a $\forall w, \mathbb{M}, w \vDash\to (\varphi \leftrightarrow \psi)$. Cela signifie que $(\varphi)^{\mathbb{M}} = (\psi)^{\mathbb{M}}$. Pour tout w, on a donc $(\varphi)^{\mathbb{M}} \in N_{a,b}^{inf}(w)$ ssi $(\psi)^{\mathbb{M}} \in N_{a,b}^{inf}(w)$. Donc, pour tout w, $\mathbb{M}, w \vDash Inf_{a,b}\varphi \leftrightarrow Inf_{a,b}\psi$, c'est à dire que $Inf_{a,b}\varphi \leftrightarrow Inf_{a,b}\psi$ est satisfaite dans $\mathbb{M}$. Ceci est vrai pour tout modèle de la classe **C** donc $Inf_{a,b}\varphi \leftrightarrow Inf_{a,b}\psi$ est valide dans la classe $\mathbb{C}$.

Complétude
Nous étendons la construction du modèle canonique $\mathbb{M}_C$.
Pour tout couple d'agents (a, b) de $\mathcal{A}^2$, $N_{C_{a,b}}$ est une fonction de W_C sur $2^{2^{W_C}}$ telle que pour tout $w \in W_C$, $N_{C_{a,b}}(w)$ est un ensemble de mondes qui contient tous les ensembles $|\varphi|$ tels que $Inf_{a,b}\varphi \in w$.
Soit X un sous-ensemble de W_C. Si $X = |\varphi|$ alors $X \in N_{C_{a,b}}(w)$ ssi $\mathbb{M}, w \vDash Inf_{a,b}\varphi$. Sinon, $X \notin N_{C_{a,b}}(w)$. On se réfère à la preuve du Chellas pour la complétude [Che80].

□

Proposition . *I.26* *La théorie de la preuve complétée avec le schéma d'axiomes et la règle d'inférence est valide et complète pour la classe de modèles* **C** *étendue avec l'ensemble de relations* R_I^{DB}.

Démonstration. Montrons que les différents axiomes sont valides et que les règles d'inférence conservent la validité.

- La validité des axiomes (DBK), (DB4) et (DB5) dérivent de la validité des axiomes (BK), (B4) et (B5).
- La conservation de la validité par la règle (DBNec) dérive de la conservation de la validité de (BNec).
- Montrons que l'axiome (DB-Int) est valide. Soit $\mathbb{M}$ un modèle et w un monde de ce modèle. Soit $\mathbb{M}$ un modèle et w un monde de ce modèle. Les deux relations $R_{\{a\}}^{DB}$ et R_a sont égales. Ainsi, tout monde v en relation avec w selon R_a l'est également selon $R_{\{a\}}^{DB}$ et réciproquement. On a donc $\mathbb{M}, w \vDash B_a\varphi \leftrightarrow DB_{\{a\}}\varphi$.
- Montrons que l'axiome (DB-Mon) est valide. Soit $\mathbb{M}$ un modèle et w un monde de ce modèle. Soient I et J deux sous-ensembles de $\mathcal{A}$ tels que $J \subseteq I$. Supposons que $\mathbb{M}, w \vDash DB_J\varphi$.
  Soit v un monde tel que $wR_I^{DB}v$. Cela signifie que pour tout a de I, wR_av. Comme $J \subseteq I$, on a donc pour tout a de J, wR_av, c'est à dire que $wR_J^{DB}v$. Or, $\mathbb{M}, w \vDash DB_J\varphi$. On a donc $\mathbb{M}, v \vDash \varphi$.
  Pour tout v tel que $wR_I^{DB}v$, on a $\mathbb{M}, v \vDash \varphi$. Cela signifie que $\mathbb{M}, w \vDash DB_I\varphi$. L'axiome (DB-Mon) est donc valide.

Pour la complétude, nous utilisons la méthode du modèle canonique. Nous construisons, pour chaque sous-ensemble I de $\mathcal{A}$, une relation d'accessibilité $R_{C_I}^{DB}$ sur W_C^2 telle que pour tout w, w', $wR_{C_I}^{DB}w'$ ssi pour tout a de I, on a $wR_{C_a}w'$.
Nous ne supposons pas que les relations de type $R_{C_I}^{DB}$ aient des propriétés spécifiques (comme l'introspection ou la transitivité). Par construction, le modèle canonique est donc un modèle de la classe **C**. □

Preuves partie II

Proposition . *II.1* *Le système précédent est valide et complet par rapport à la sémantique de FOSDL.*

Démonstration. La preuve est donnée dans [FM99]. La partie validité est relativement aisée (il faut prouver que les axiomes sont valides et et que les règles d'inférence conservent la validité). La partie complétude demande plus de travail. Elle passe par la construction d'un modèle canonique, comme nous l'avons fait dans la preuve du théorème I.2. $\square$

Proposition . *II.2* *Considérons un ensemble de contraintes d'intégrité IC, une réglementation ρ cohérente avec IC et un état du monde s cohérent avec IC et tel que $\rho \cup s$ est cohérent. Soient $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral vérifiant la définition II.18 et $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ la théorie des défauts correspondante.*

Les propositions suivantes sont équivalentes :

1. *pour tout vecteur $\overrightarrow{a}$ de termes de base, si $s \vdash \varphi(\overrightarrow{a})$, $\rho, s \nvdash O\psi(\overrightarrow{a})$, $\rho, s \nvdash P\psi(\overrightarrow{a})$ et $\rho, s \nvdash F\psi(\overrightarrow{a})$ (i.e. ρ n'est pas $(\varphi(\overrightarrow{a}), \psi(\overrightarrow{a}))$-complète dans s), alors $s \vdash E_O(\overrightarrow{a}) \otimes E_P(\overrightarrow{a}) \otimes E_F(\overrightarrow{a})$.*
2. *ρ est cohérent et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s.*

Démonstration. Dans ce qui suit, nous considérons un ensemble de contraintes d'intégrité IC, une réglementation ρ cohérente par rapport à IC et un état du monde s cohérent avec IC et tel que $\rho \cup s$ est cohérent. Soit $\varphi(\overrightarrow{x})$ une formule objective et $\psi(\overrightarrow{x})$ un d-littéral vérifiant la définition II.18 et $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ la théorie des défauts correspondante.

La preuve de la proposition se fait en deux étapes.

$1 \Rightarrow 2$) L'idée de la preuve est de construire une extension particulière de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$, et de montrer ensuite que cette extension est unique, cohérente et telle que ρ est $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$.

Nous définissons l'ensemble suivant de formules :

- $E_0 = \rho \cup s$
- $E_1 = Th(E_0) \cup \{\Box\psi(\overrightarrow{a}) \ : \ E_0 \vdash \psi(\overrightarrow{a}) \wedge E_\Box(\overrightarrow{a})$, $E_0 \cup \{\Box\psi(\overrightarrow{a})\}$ est cohérent, $\Box \in \{O, P, F\}$ et $\overrightarrow{a}$ est un terme de base$\}$
- $E = Th(E_1)$

1. E est une extension de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ par construction (cf. définition de la section 9.1)
2. montrons que ρ est cohérente pour $\vdash_*$ dans s et que E est la seule extension de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$. Supposons que ρ ne soit pas cohérente pour $\vdash_*$ dans s. Comme $\rho \cup s$ est cohérent, la seule source d'incohérence provient des modalités qui ont été ajoutées par les défauts.

   Cela signifie qu'il existe des d-littéraux de base $\alpha_1, \ldots, \alpha_n$ et des formules de base ne contenant que des f-littéraux $L_0, \ldots L_n$ tels que $L_0, L_1 \vee O\alpha_1, \ldots, L_k \vee O\alpha_k, L_{k+1} \vee \neg O\alpha_{k+1}, \ldots, L_n \vee \neg O\alpha_n$ soit inconsistant [46 47]. L_0 est la conjonction des formules de l'état du monde s, pour tout $i > 0$, $L_i \vee O\alpha_i$ (ou $L_i \vee \neg O\alpha_i$) est soit une règle de ρ soit une déduction des défauts auquel cas $L_i = \bot$.

46. Nous pouvons nous ramener à des littéraux de base car de tout ensemble inconsistant, on peut extraire un sous-ensemble inconsistent de formules de base.
47. Nous ne considérons qu'une partie des modalités. En effet, $F\neg\alpha$ est équivalent à $O\neg\alpha$ et $P\alpha$ est équivalent à $\neg O\alpha \wedge \neg O\neg\alpha$

Cela implique que l'ensemble $\{O\alpha_1, \ldots, O\alpha_k, \neg O\alpha_{k+1}, \ldots, \neg O\alpha_n\}$ est inconsistant. D'après [FndC85], cela signifie qu'il existe deux entiers i et j tels que $1 \leq i \leq k$, $k+1 \leq j \leq n$ et $\alpha_i = \alpha_j$.
Cela signifie qu'il existe deux d-littéraux de base tels que $\rho, s \vdash_* O\alpha_j \wedge \neg O\alpha_i$.
Supposons que $\rho, s \vdash O\alpha_i$. Alors, comme ρ est cohérente dans s, on a $\rho, s \nvdash \neg O\alpha_i$. Cela signifie que $\neg O\alpha_i$ a été déduite en utilisant un des défauts (notammenent le défaut d_P appliqué à $\overrightarrow{a}$). Or, pour pouvoir appliquer le défaut, il faut que $P\psi(\overrightarrow{a})$ soit cohérent avec ρ, s ce qui n'est manifestement pas le cas. De même, si on suppose que $\rho, s \vdash \neg O\alpha_i$, on montre qu'il n'est pas possible d'utiliser un défaut pour déduire $O\alpha_i$. Cela signifie donc que les deux formules $O\alpha_i$ et $\neg O\alpha_i$ ont été déduites en utilisant deux défauts.
Cela signifie donc qu'il existe $\overrightarrow{a}$ tel que $\Box_1\psi(\overrightarrow{a})$ et $\Box_2\psi(\overrightarrow{a})$ ont été déduites en utilisant deux défauts d_1 et d_2. Donc $s \vdash \varphi(\overrightarrow{a})$, $s \vdash E_{\Box_1}(\overrightarrow{a})$ et $s \vdash E_{\Box_2}(\overrightarrow{a})$. Comme ni $O\psi(\overrightarrow{a})$, ni $F\psi(\overrightarrow{a})$ ni $P\psi(\overrightarrow{a})$ ne peuvent être déduites avec $\vdash$ de ρ, s, on peut appliquer la partie (1) de la proposition et dériver $s \vdash (E_{\Box_1}(\overrightarrow{a}) \wedge \neg E_{\Box_2}(\overrightarrow{a})) \vee (\neg E_{\Box_1}(\overrightarrow{a}) \wedge E_{\Box_2}(\overrightarrow{a}))$, ce qui est contradictoire avec l'application des défauts d_1 et d_2.
Il n'est donc pas possible de dériver une contradiction de $\rho \cup s$ en utilisant $\vdash_*$, donc ρ est cohérent pour $\vdash_*$ dans s.
Remarquons que cela garantit qu'il n'existe qu'une seule extension de $\Delta_{\rho,s}(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$ (car sinon les extensions amèneraient une contradiction) et que cette extension est E.

3. supposons que ρ n'est pas $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s. Dans ce cas, ρ n'est également pas $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash$ dans s. Donc il existe un vecteur de termes de base $\overrightarrow{a}$ tel que $s \vdash \varphi(\overrightarrow{a})$, $\rho, s \nvdash O\psi(\overrightarrow{a})$, $\rho, s \nvdash P\psi(\overrightarrow{a})$ et $\rho, s \nvdash F\psi(\overrightarrow{a})$. Mais dans ce cas, $s \vdash E_O(\overrightarrow{a}) \otimes E_P(\overrightarrow{a}) \otimes E_F(\overrightarrow{a})$, donc il existe $\Box \in \{O, P, F\}$ tel que $s \vdash E_\Box(\overrightarrow{a})$. En effet, s est complet et cohérent pour les littéraux objectifs. Comme $E_\Box(\overrightarrow{a})$ est une formule objective, soit elle soit sa négation peuvent être déduites de s.
Il existe également un défaut d de la forme $\dfrac{\varphi(\overrightarrow{a}) \wedge E_\Box(\overrightarrow{a}) \ : \ \Box\psi(\overrightarrow{a})}{\Box\psi(\overrightarrow{a})}$.
Dans ce cas, la seule raison pour laquelle d pourrait ne pas être appliqué est le cas où $\Box\psi(\overrightarrow{a})$ est en contradiction avec $\rho \cup s$. Cela signifie que $\rho \cup s \vdash \Box_1\psi(\overrightarrow{a})$ avec $\Box_1 \neq \Box$ et $\Box_1 \in \{O, P, F\}$. Or, ρ est $(\varphi(\overrightarrow{a}), \psi(\overrightarrow{a}))$-incomplète pour $\vdash$ dans s.
Le défaut d peut donc être appliqué. Donc $\rho, s \vdash_* \Box\psi(\overrightarrow{a})$. Donc ρ est $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s.

$2 \Rightarrow 1$) Supposons que ρ est cohérente et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s. Soit $\overrightarrow{a}$ un vecteur de termes de base tel que $s \vdash \varphi(\overrightarrow{a})$, $\rho, s \nvdash O\psi(\overrightarrow{a})$, $\rho, s \nvdash P\psi(\overrightarrow{a})$ et $\rho, s \nvdash F\psi(\overrightarrow{a})$. Comme ρ est $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète pour $\vdash_*$ dans s, il existe $\Box \in \{O, P, F\}$ tel que $\rho, s \vdash_* \Box\psi(\overrightarrow{a})$. Donc il existe un défaut d qui a été appliqué pour dériver $\Box\psi(\overrightarrow{a})$ et à cause du prérequis de d, $s \vdash E_\Box(\overrightarrow{a})$. Donc $s \vdash E_O(\overrightarrow{a}) \vee E_F(\overrightarrow{a}) \vee E_P(\overrightarrow{a})$. Considérons $\Box_1$ et $\Box_2$ les deux autres modalités dans $\{O, P, F\}$ différentes de $\Box$. Comme ρ est cohérent avec $\vdash_*$ dans s, on ne peut dériver ni $\Box_1\psi(\overrightarrow{a})$ ni $\Box_2\psi(\overrightarrow{a})$ de ρ, s en utilisant $\vdash_*$. Donc les défauts permettant de dériver $\Box_1\psi(\overrightarrow{a})$ et $\Box_2\psi(\overrightarrow{a})$ ne peuvent pas être appliqués. Comme leurs justifications sont vraies, leurs prérequis sont faux. Donc ni $s \nvdash E_{\Box_1}(\overrightarrow{a})$ ni $s \nvdash E_{\Box_2}(\overrightarrow{a})$ ne sont vrais. Donc $s \vdash E_O(\overrightarrow{a}) \otimes E_P(\overrightarrow{a}) \otimes E_F(\overrightarrow{a})$. $\Box$

Corollaire . *II.1* *Si*

$$s \vdash \forall \overrightarrow{x} \ (\varphi(\overrightarrow{x}) \rightarrow E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x}))$$

alors ρ est cohérente et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète par rapport à IC pour $\vdash_*$ dans s.

Démonstration. La preuve est immédiate : il suffit de remarquer que $s \vdash \forall \overrightarrow{x}\ \varphi(\overrightarrow{x}) \to E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x})$ implique que la première partie de la proposition II.2 est vraie. □

Corollaire . ***II.2*** *Si*

$$IC \vdash \forall \overrightarrow{x}\ (E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x}))$$

alors ρ est cohérent et $(\varphi(\overrightarrow{x}), \psi(\overrightarrow{x}))$-complète par rapport à IC pour $\vdash_*$.

Démonstration. Là encore, la preuve est immédiate : $IC \vdash \forall \overrightarrow{x}\ (E_O(\overrightarrow{x}) \otimes E_F(\overrightarrow{x}) \otimes E_P(\overrightarrow{x}))$ implique que la première partie de la proposition II.2 est vraie. □

Bibliographie

[AB71] C.E. Alchourròn and E. Bulygin. *Normative Systems.* Springer, Wien, 1971.

[AB75] Alan R. Anderson and Nuel D. Belnap. *Entailment : The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton, 1975.

[AF00] Lucas Anderlini and Leonardo Felli. Bounded rationality and incomplete contracts. Technical report, London School of Economics - Suntory and Toyota International Centres for Economics and Related Disciplines (LSE STICERD), 2000.

[AGM85] Carlos E. Alchourròn, Peter Gärdenfors, and David Makinson. On the logic of theory change : partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50 :510–530, 1985.

[Aus62] John Langshaw Austin. *How to do things with words.* Oxford : Clarendon, 1962.

[Bar94] Carol L. Barry. User-defined relevance criteria : An exploratory study. *Journal of the American Society for Information Science*, 45 (3) :149–159, 1994.

[BC93] P. Bieber and F. Cuppens. Expression of confidentiality policies with deontic logic. In *Deontic logic in computer science : normative system specification*, pages 103–121. John Wiley and Sons, 1993.

[Bes89] P. Besnard. *An introduction to default logic.* Springer-Verlag, 1989.

[Bor03] Pia Borlund. The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10) :913–925, 2003.

[Bra87] Michael Bratman. *Intentions, Plans, and Practical Reason.* Harvard University Press, 1987.

[Bre95] Philippe Bretier. *La communication orale coopérative : contribution à la modélisation logique et à la mise en œuvre d'un agent rationnel dialoguant.* PhD thesis, Université de Paris 13, Villetaneuse, France, 1995.

[BS08] Philippe Balbiani and Pablo Seban. Logique de mise à jour des croyances objectives. In *IAF'08 : Actes des Journées Fondamentales de l'Intelligence Artificielle 2008*, Paris, 2008.

[Car98] Robyn Carston. Informativeness, relevance and scalar implicature. In Robyn Carston and Seiji Uchida, editors, *Relevance Threory : Applications and Implications*, Pragmatics and Beyond New Series, pages 179–236. John Benjamins, Amsterdam, 1998.

[CC99] L. Cholvy and F. Cuppens. Reasoning about norms provided by conflicting regulations. In P. McNamara and H. Prakken, editors, *Norms, logics and information systems : new studies in deontic logic and computer science*, volume 49 of *Frontiers in artificial intelligence and applications*, pages 247–262. IOS Press, Amsterdam, 1999.

[CD86] Laurence Cholvy and Robert Demolombe. Querying a rule base. In *Expert Database Conf.*, pages 477–485, 1986.

[CD89] Frédéric Cuppens and Robert Demolombe. How to recognize interesting topics to provide cooperative answering. *Inf. Syst.*, 14(2) :163–173, 1989.

[CD91] Frédéric Cuppens and Robert Demolombe. Extending answers to neighbour entities in a cooperative answering context. *Decis. Support Syst.*, 7(1) :1–11, 1991.

[CD97] F. Cuppens and R. Demolombe. A modal logical framework for security policies. In *Lectures Notes in Artificial Intelligence*, volume 1325, page 1997. Springer, 1997.

[CdD02] Brahim Chaib-draa and Robert Demolombe. L'interaction comme champ de recherche. *Information-Interaction-Intelligence*, 2002. Numéro hors série sur l'interaction.

[CDJM01] Brahim Chaib-Draa, Imed Jarras, and Bernard Moulin. Systèmes multiagents : Principes généraux et applications. In Jean-Pierre Briot and Yves Demazeau, editors, *Agent et systèmes multiagents*. HermÃ¨s, 2001.

[Che80] B. F. Chellas. *Modal logic, an introduction.* Cambridge University Press, 1980.

[Che04] Jean-Pierre Chevallet. Modélisation logique pour la recherche d'information. *Les systèmes de recherche d'information*, pages 105–138, 2004.

[Cho99] L. Cholvy. Checking regulation consistency by using SOL-resolution. In *International Conference on Artificial Intelligence and Law*, pages 73–79, 1999.

[CL90] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artif. Intell.*, 42(2-3) :213–261, 1990.

[CR07a] Laurence Cholvy and Stéphanie Roussel. Raisonner avec une politique d'échange d'information incomplètes. In *Actes des Journées Nationales de L'IA Fondamentale - IAF'07*, Grenoble, 2007.

[CR07b] Laurence Cholvy and Stéphanie Roussel. Reasoning with incomplete information exchange policies. In K. Mellouli, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 9th European Conference, ECSQARU'07*, number 4724 in Lecture Notes in Artificial Intelligence, pages 683–694. Springer-Verlag, 2007.

[CR08a] Laurence Cholvy and Stéphanie Roussel. Consistency and completeness of regulations. In *Proceedings of the third International Workshop on Normative Multiagent Systems (NORMAS'08)*, pages 51–65, 2008.

[CR08b] Laurence Cholvy and Stéphanie Roussel. Raisonner avec une réglementation incomplète : cas des politiques d'échange d'informations. In *Proceedings of Reconnaissance de Formes et Intelligence Artificielle - RFIA'08*, Varna - Bulgaria, Janvier 2008.

[CR08c] Laurence Cholvy and Stéphanie Roussel. Toward agent-oriented relevant information. In *Proceedings of Artificial Intelligence : Methodology, Systems, Applications - AIMSA'08*, Varna, Bulgaria, 2008.

[DBL06] R. Demolombe, P. Bretier, and V. Louis. Norms with deadlines in dynamic deontic logic. In Gerhard Brewka, Silvia Coradeschi, Anna Perini, and Paolo Traverso, editors, *Proceedings of ECAI 2006, 17th European Conference on Artificial Intelligence*, pages 751–752. IOS Press, 2006.

[DdC00] Robert Demolombe and Luis Fariñas del Cerro. La modélisation logique dans les nouveaux systèmes d'information et de communication. *Techniques et sciences informatiques*, 19(1) :165–174, 2000.

[Dem82] R. Demolombe. Syntactical characterization of a subset of domain independent formulas. *Journal of the Association for Computer Machinery*, 39(1) :71–94, 1982.

[Dem99] R. Demolombe. Database validity and completeness : another approach and its formalisation in modal logic. In Enrico Franconi and Michael Kifer, editors, *Proc. of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99)*, pages 11–13. CEUR-WS.org, 1999.

[Dem04] Robert Demolombe. Reasonig about trust : a formal logical framework. In *Proc. 2d International Conference iTrust*, 2004.

[Dem09] Robert Demolombe. La confiance dans les relations entre agents : essai de modélisation en logique modale. In *Proc. Journées Francophones sur la Modélisation Formelle de l'Interaction*, 2009.

[DP98] James P. Delgrande and Jeff Pelletier. A formal analysis of relevance. *Erkenntnis*, 49(2) :137–147, 1998.

[DR02] J. Michael Dunn and Greg Restall. Relevance logic. In G. M. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic*, volume 6, pages 1–128. Kluwer, Dordrecht, second edition, 2002.

[End72] H.B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, New York, 1972.

[Flo07] Luciano Floridi. Understanding epistemic relevance. *Erkenntniss*, 69 (1) :69–92, 2007.

[FM99] M. Fitting and R. L. Mendelsohn. *First-order modal logic.* Kluwer Academic, 1999.

[FndC85] Luis Fariñas del Cerro. Resolution logic model. *Logique et Analyse*, (110-11) :153–172, 1985.

[Gal08] Antony Galton. Temporal logic. Stanford Encyclopedia of Philosophy, `http://plato.stanford.edu/entries/logic-temporal/`, 2008.

[GGM92] Terry Gaasterland, Parke Godfrey, and Jack Minker. An overview of cooperative answering. *Journal of Intelligent Infomation Systems*, 1(2) :123–157, 1992.

[GM87] Annie Gal and Jack Minker. Informative and cooperative answers in databases using integrity constraints. In *Natural Language Understanding and Logic Programming Workshop*, pages 277–300, 1987.

[GRC09a] Christophe Garion, Stéphanie Roussel, and Laurence Cholvy. How to complete regulations in multi-agent systems. In *Proceedings of Intelligent Agent Technology (IAT)*, pages 285–288, 2009.

[GRC09b] Christophe Garion, Stéphanie Roussel, and Laurence Cholvy. Une logique modale pour raisonner sur la cohérence et la complètude de réglementations. In Nicolas Maudet, Pierre-Yves Schobbens, and Marc Guyomard, editors, *Actes des Cinquièmes Journées Francophones Modèles Formels de l'Interaction (MFI'09)*, pages 147–158, Lannion, june 2009. Imprimerie de l'Université de Rennes I.

[GRC10] Christophe Garion, Stéphanie Roussel, and Laurence Cholvy. Une logique modale pour raisonner sur la cohérence et la complètude de réglementations. In *Revue RSTI - Revue d'Intelligence Artificielle*, volume 24, pages 267–290. Lavoisier, 2010.

[Gri75] H. P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics*, volume 3, pages 41–58. New York : Academic Press, 1975.

[Gro99] Jeroen Groenendijk. The logic of interrogation : classical version. In Tanya Matthews and Devon Strolovitch, editors, *SALT IX : Semantics and Linguistic Theory*, pages 109–126, Ithaca, 1999. Cornell University Press.

[Gro06] C. Groulier. *Normes permissives et droit public.* PhD thesis, Université de Limoges, 2006. Available on `http://www.unilim.fr/scd/theses/accesdoc.html`. In French.

[GSV96] Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In S Lapin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–213. Blackwell, Oxford, 1996.

[HC96] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic.* Routledge, September 1996.

[Hil71] R. Hilpinen, editor. *Deontic logic.* Reidel Publishing Company, 1971.

[HL02] Andreas Herzig and Dominique Longin. A logic of intention with cooperation principles and with assertive speech acts as communication primitives. In *Proc. AAMAS 2002*, pages 920–927. ACM Press, 2002.

[HL09] Andreas Herzig and Emiliano Lorini. Logic of individual and collective intentionality. Cours préséntć Ã ESSLLI'09, Juillet 2009.

[Hor84] Laurence R. Horn. Toward a new taxonomy for pragmatic inference : Q-based and R-based implicature. In *Meaning, Form and Use in context (GURT 84)*, pages 11–42, 1984.

[HS08] Andreas Herzig and François Schwarzentruber. Properties of logics of individual and group agency. In *Proceedings of Advances in Modal Logic 2008 - AiML'08*, 2008.

[JS92] Andrew J. I. Jones and Marek J. Sergot. Formal specification of security requirements using the theory of normative positions. In Yves Deswarte, Gérard Eizenberg, and Jean-Jacques Quisquater, editors, *Proceedings of the Second European Symposium on Research in Computer Security (ESORICS'92)*, pages 103–121, 1992.

[Kan72] S. Kanger. Law and logic. *Theoria*, (38), 1972.

[Kri63a] Saul Kripke. A semantical analysis of modal logic i, normal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 9 :63–96, 1963.

[Kri63b] Saul Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16 :83–94, 1963.

[Lak97] Gerhard Lakemeyer. Relevance from an epistemic perspective. *Artif. Intell.*, 97(1-2) :137–167, 1997.

[Lal96] Mounia Lalmas. *Theories of information and uncertainty for the modelling of information retrieval : An application to situation theory and Demster-Shafer's theory of evidence.* PhD thesis, Université de Glasgow, 1996. Thèse de doctorat.

[Lav07] Noël Laverny. Logique doxastique graduelle. In *Actes des Quatrièmes Journées Francophones Modèles Formels de l'Interaction (MFI'07)*, 2007.

[LEA89] LEASOMBE. Logique des défauts. In *Raisonnement sur des informations incomplètes en intelligence artificielle.* Teknea, Marseille, 1989.

[Lev00] Stephen C. Levinson. *Presumptive meaning : The theory of generalized conversational implicature.* Language, Speech and Communication. MIT Press, Cambridge, MA, 2000.

[Lin77] L. Lindahl. *Position and Change - a Study in Law and Logic.* Number 112 in Synthese Library. D. Reidel, 1977.

[LLM03] Jérôme Lang, Paolo Liberatore, and Pierre Marquis. Propositional independence - formula-variable independence and forgetting. *Journal of Artificial Intelligence Research*, 18 :391–443, 2003.

[LR08] Dominique Luzeaux and Jean-René Ruault. *Ingénierie des systèmes de systèmes - Méthodes et outils*. ic2 informatique et systèmes. Hermes Science Publications, 2008.

[Min98] Jack Minker. An overview of cooperative answering in databases. In *FQAS '98 : Proceedings of the Third International Conference on Flexible Query Answering Systems*, pages 282–285, London, UK, 1998. Springer-Verlag.

[Miz98] Stefano Mizzaro. How many relevances in information retrieval ? *Interacting with Computers*, 10(3) :303–320, 1998.

[MN04] Paul Mac Namara. Agential obligation as non-agential personal obligation plus agency. *Journal of Applied Logic*, 2(1) :117–152, 2004.

[MN06] Paul Mac Namara. Deontic logic. Stanford Encyclopedia of Philosophy, `http://plato.stanford.edu/entries/logic-deontic/`, 2006.

[Moe07] Jacques Moeschler. *Speech and Language Engineering*, chapter 1. Rajman m. and Pallota V., 2007.

[MR06] Jacques Moeschler and Anne Reboul. 3 - "compréhension, pragmatique et argumantation". In *Compréhension des langues et des interactions*, pages 113–119. Hermès, sabah, g. edition, 2006.

[Nie92] Jian-Yun Nie. Towards a probabilistic modal logic for semantic-based information retrieval. In N Belkin, Ingwersen P, and Pejtersen A M, editors, *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 140–151. ACM Press, juin 1992.

[Par93] Taemin Kim Park. The nature of relevance in information retrieval : An empirical study. *The Library Quarterly*, 63(No 3) :318–351, 1993.

[Pri01] Graham Priest. *An Introduction to Non-Classical Logic*. Cambridge University Press, 2001.

[RC08] Stéphanie Roussel and Laurence Cholvy. Vers une pertinence orientée-agent des informations. In *Actes des Journées Nationales de L'IA Fondamentale - IAF'08*, Paris, 2008.

[RC09a] Stéphanie Roussel and Laurence Cholvy. Cooperative interpersonal communication and relevant information. In *ESSLLI Workshop on Logical Methods for Social Concepts*, Bordeaux, 2009.

[RC09b] Stéphanie Roussel and Laurence Cholvy. A definition of agent-oriented relevance in modal logic. *The Information - Interaction - Intelligence (I3) Journal*, 29(2), 2009.

[RC09c] Stéphanie Roussel and Laurence Cholvy. A formal characterization of relevant information in multi-agent systems. In *NATO/RTO Symposium, Information management - Exploitation, IST087*, Stockholm, october 2009. Imprimerie de l'Université de Rennes I.

[RC09d] Stéphanie Roussel and Laurence Cholvy. Une définition en logique modale de la pertinence orientée-agent. In Nicolas Maudet, Pierre-Yves Schobbens, and Marc Guyomard, editors, *Actes des Cinquièmes Journées Francophones Modèles Formels de l'Interaction (MFI'09)*, pages 279–287, Lannion, june 2009. Imprimerie de l'Université de Rennes I.

[Rei78] R. Reiter. On closed world databases. In J. Minker J.-M. Nicolas H. Gallaire, editor, *Logic and Databases.* Plenum Publications, 1978.

[Rei80] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2), 1980.

[Rei92] R. Reiter. What should a database know? *Journal of Logic Programming*, 14(1,2) :127–153, 1992.

[Rei99] Jane Reid. A new task oriented paradigm for information retrieval : Implications for evaluation of information retrieval systems. In *CoLIS 3, third international conference on the Conceptions of Library and Information Science*, 1999.

[RG91] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a bdi-architecture. In James Allen, Richard Fikes, and Erik Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann publishers Inc. : San Mateo, CA, USA, 1991.

[RM98] Anne Reboul and Jacques Moeschler. Pertinence. In *Vocabulaire des sciences cognitives*, pages 305–307. PUF, Paris, 1998.

[Roy97] L. M. M. Royakkers. Giving permission implies giving choice. In *International Workshop on Database and Expert Systems Applications*, pages 198–203, Los Alamitos, CA, USA, 1997. IEEE Computer Society.

[Sad94] David Sadek. Attitudes mentales et fondement du comportement coopératif. In *Systèmes coopératifs : de la modélisation à la conception*, pages 93–117. Pavard, B, octares edition, 1994.

[Sad96] David Sadek. Le dialogue homme-machine : de l' ergonomie des interfaces à l' agent intelligent dialoguant. In *Nouvelles interfaces hommemachine*, pages 277–321. Lavoisier Editeur, Arago 18, 1996.

[Sad04] David Sadek. De nouvelles perspectives pour l'ergonomie des interactions personne-machine : dialogue naturel et agents intelligents. In *Actes du workshop sur l'ergonomie et l'informatique avancée (Ergo'IA)*, 2004.

[Sar96] T. Saracevic. Relevance reconsidered. In *Information science : Integration in perspectives.Proceedings of the Second Conference on Conceptions of Library and Information Science, Copenhagen (Denmark)*, 1996.

[SBP97] David Sadek, Philippe Bretier, and Franck Panaget. Artimis : Natural dialogue meets rational agency. In *Prooceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1030–1035, Nagoya, Japon, 1997.

[Sch94] L. Schamber. Relevance and information behavior. *Annual Review of Information Science and Technology*, 29 :3–48, 1994).

[Seg99] Ilya Segal. Complexity and renegotiation : A foundation for incomplete contracts. *Review of Economic Studies*, volume(66) :57–82, 1999.

[SEN90] L. Schamber, M. B. Eisenberg, and M.S. Nilan. A re-examination of relevance : toward a dynamic, situational definition. *Information Processing & Management*, 26(6) :755–776, 1990.

[SG87] Devika Subramanian and Michael R. Genesereth. The relevance of irrelevance. In *IJCAI*, pages 416–422, 1987.

[SSLR07] Kaile Su, Abdul Sattar, Han Lin, and Marks Reynolds. A modal logic for beliefs and pro attitudes. In *AAAI*, pages 496–501, 2007.

[SW86] Dan Sperber and Deirdre Wilson. *Relevance : communication and cognition.* Harvard University Press, Cambridge, MA, USA, 1986.

[SW04] Dan Sperber and Deirdre Wilson. Relevance theory. In G. Ward and L. Horn, editors, *Handbooks of Pragmatics*, pages 607–632. Oxford : Blackwell, 2004.

[tCS07] Balder ten Cate and Chung-Chieh Shan. Axiomatizing Groenendijk's logic of interrogation. In M. Aloni, A. Butler, and P. Dekker, editors, *Questions in dynamic semantics*, CRiSPI, pages 63–82. Elsevier, 2007.

[Tri04] André Tricot. La prise de conscience du besoin d'information : une compétence documentaire fantôme ? [en ligne : http ://docs-docs.free.fr/modules.php ?name=Sections&op=viewarticle&artid=55], 2004.

[Tuo92] Raimo Tuomela. Group beliefs. *Synthese*, 91(3) :285–318, June 1992.

[Tuo95] Raimo Tuomela. *The importance of us : a philosophical study of basic social notions.* Stanford University Press, 1995.

[vW51] G. H. von Wright. Deontic logic. *Mind*, 60 :1–15, 1951.

[Woo00] Michael Wooldridge. *Reasoning about rational agents.* The MIT Press, Cambridge, Massachusetts, 2000.

Apports de la logique mathématique pour la modélisation de l'information échangée dans des systèmes multiagents interactifs

Cette thèse s'intéresse aux systèmes multiagents dans lesquels les agents échangent des informations de façon à atteindre un but commun.
Lors de l'échange d'informations entre agents, il est souhaitable que les agents aient un comportement coopératif, c'est-à-dire qu'ils ne se transmettent que les informations qui leur sont utiles. Nous définissons donc en logique modale quelles sont les informations utiles pour un agent étant donné un besoin en information de celui-ci. A partir de ces travaux, nous travaillons sur la notion d'agent coopératif et en proposons plusieurs définitions.
Les échanges des agents sont en général réglementés par une politique d'échange d'informations. Pour qu'une telle politique soit efficace, il faut qu'elle soit cohérente et complète. Nous formalisons ces deux notions en logique déontique pour des réglementations générales. Nous proposons ensuite une méthode pour raisonner avec des réglementations incomplètes et appliquons nos résultats aux politiques d'échange d'informations.
Finalement, nous analysons dans quelle mesure il est possible d'être à la fois coopératif et obéissant à une politique d'échange d'informations.

Mots clés français : Intelligence Artificielle - systèmes multiagents - logique modale - besoin en information - information utile – agent coopératif - réglementation cohérente et complète

Taking advantage of mathematical logic for modelling information exchanged in interactive multiagent systems

This thesis deals with multiagent systems in which agents have to exchange information in order to achieve a global goal.
When agents exchange pieces of information, cooperative behaviour is generally expected, meaning that agents should only exchange pieces of information they think useful for the others. We give a definition in modal logic of useful pieces of information for an agent in respect of some information need he has. Then, we give several definitions of a cooperative agent.
In most systems, information exchanges are regulated by an information exchange policy. Such a policy should be consistent and complete. We formalize those two notions for a general regulation within a deontic logic. Then, we suggest a method to work with incomplete regulations and we apply our results to information exchange policies.
Finally, we analyze to what extent it is possible for an agent to be at the same time cooperative and obedient to an information exchange policy.

Keywords : Artificial Intelligence - multiagents systems - modal logic - information need - useful piece of information - cooperative agent - consistent and complete regulation