Formalisation logique de préférences qualitatives pour la sélection de la réaction d'un agent rationnel dialoguant

Gautier Meyer

Université de Paris XI France Télécom R&D

Rapport de thèse
au 28 juillet 2006

de

Gautier MEYER

Formalisation logique de préférences qualitatives pour la sélection de la réaction d'un agent rationnel dialoguant

présenté le 7 juillet 2006 en vue de l'obtention du titre de

Docteur en informatique
de
l'université Paris XI

devant le jury suivant :

Jean-Paul SANSONNET	LIMSI-CNRS, Orsay	(Directeur de thèse)
Vincent LOUIS	FT-R&D, Lannion	(Co-encadrant)
Jérôme LANG	IRIT, CNRS, Toulouse	(Rapporteur)
Pierre MARQUIS	LIMSI-CNRS, Orsay	(Rapporteur)
Gérard LIGOZAT	LIMSI-CNRS, Orsay	(Président du jury)
David SADEK	FT-R&D, Lannion	(Examinateur)
Robert DEMOLOMBE	ONERA-CERT, Toulouse	(Invité)

Thèse préparée au sein du laboratoire TECH/EASY de France Télécom R&D à Lannion

Remerciements

Ce document est sans conteste le résultat d'un travail personnel. Néanmoins, il n'existerait pas sans le concours de tous ceux qui ont croisé ma route récemment. C'est pourquoi je souhaite exprimer ici ma reconnaissance à toutes les personnes qui ont contribué à cet aboutissement tant professionnel que personnel.

Je ne saurais pourtant ici exprimer toute ma gratitude envers mes deux encadrants de thèse : Vincent Louis et Jean-Paul Sansonnet. Durant ces quarante deux mois, ils ont su me soutenir et me diriger judicieusement. Vincent, je te serai encore longtemps reconnaissant de m'avoir transmis une partie de ton esprit critique et ton goût pour le travail bien fait. Sans eux, ce document n'aurait pas cette qualité. Jean-Paul, sans ton expérience cette thèse serait sûrement encore dans les cartons. Ton insistance n'a pas été vaine et m'a, entre autres, permis de dépasser ma timidité pour produire et présenter des résultats.

De façon pratique, ces travaux de thèse n'auraient jamais vu le jour sans le concours bienveillant de France Télécom R&D. Aussi j'adresse des remerciements tout particuliers à David Sadek, Philippe Bretier et Patrice Soyer pour m'avoir accordé leur confiance en m'accueillant respectivement dans leur pôle de recherche, équipe et laboratoire et pour m'avoir permis de bénéficier de toutes les ressources nécessaires.

Mes remerciements vont aussi aux membres du jury pour l'intérêt marqué qu'ils ont manifesté à l'égard de mon travail, tant au travers de leurs remarques et compliments sur le document qu'au travers de leurs questions lors de la soutenance. Plus particulièrement, je souhaite remercier Jérôme Lang et Pierre Marquis qui m'ont fait l'honneur d'accepter d'être les rapporteurs ainsi que pour la qualité de leurs rapports. Je remercie également Robert Demolombe pour sa lecture rigoureuse du document et les remarques pertinentes qu'il a faites lors de la soutenance.

Ma gratitude va bien sûr aussi à toutes les autres personnes que j'ai eu la chance de rencontrer et de côtoyer tout au long de ces mois de travail de façon plus ou moins officieuse et qui m'ont tant apporté. Ce sont les membres de ma famille, mes collègues, mes amis. Je ne prendrais pas le risque de les citer ici ; les énumérer tous serait absolument impossible tant ils sont nombreux. J'espère néanmoins qu'ils se reconnaîtront et qu'ils ne me tiendront pas rigueur de ne pas leur adresser un mot personnel ici. Parmi eux, je voudrais plus particulièrement remercier mes nombreux relecteurs officieux (en particulier celle pour qui le français n'est pas la langue maternelle) ainsi que tous ceux qui m'ont permis de rester motivé et m'ont apporté chaleur et joie de vivre.

à tous ceux à qui je dois une part de mes préférences . . . merci.

Introduction

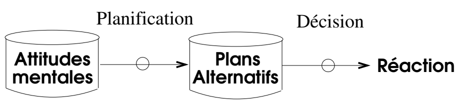
Contexte de la thèse

Le paradigme agent propose un point de vue attrayant pour modéliser les systèmes informatiques complexes de façon abstraite avec des notions intuitives. Dans le cas qui nous intéresse, les agents constituant de tels systèmes sont formalisés à partir de notions primitives appelées *attitudes mentales* telles que la croyance, le désir, l'intention, etc. [Rao and Georgeff, 1991b][Sadek, 1991] et mis en œuvre par des mécanismes d'inférence automatique [Bretier, 1995]. Ce parallèle entre la façon de formaliser une application et la représentation que se fait tout un chacun d'entités animées permet, par exemple, de concevoir et de réaliser des systèmes de dialogue avancés ayant une propension naturelle à communiquer, tels que ceux mis en œuvre par la technologie *Artimis* [Sadek et al., 1997][Sadek and Mori, 1998][Sadek, 1999].

Problématique

Un agent rationnel est avant tout une entité qui perçoit son environnement et agit dessus suivant le principe de rationalité (c'est-à-dire en cherchant à satisfaire ses intentions, en accord avec ses connaissances). Malheureusement, la mise ne œuvre de ce principe ne spécifie pas en général complètement les réactions d'un agent. Ceci est particulièrement gênant dans les situations complexes où il a plusieurs motivations (ou intentions) et/ou plusieurs moyens de les satisfaire (plans d'action). Par exemple, dans le cadre d'une interaction multimodale, un agent assistant qui fournit un itinéraire à un utilisateur doit choisir entre une modalité vocale (en énonçant les instructions) ou graphique (en montrant un plan) selon le contexte (cas de plusieurs plans satisfaisant la même intention).

C'est pourquoi les auteurs de [Haddawy and Hanks, 1993][Louis, 2002] ont affirmé la nécessité de faire explicitement le lien entre l'état mental et la réaction de l'agent. Par la suite, Louis dans [Louis, 2002] a proposé de diviser ce lien en deux phases distinctes : une phase de planification [Allen et al., 1991] et une de décision [Doyle and Thomason, 1999]. Ces deux phases se distinguent par leurs objectifs : la première vise à générer les réactions



Lien général entre état mental et réactions

potentielles de l'agent (i.e. les alternatives) tandis que la seconde cherche à ordonner ces dernières pour choisir effectivement la réaction à mettre en œuvre. La phase de décision requiert la modélisation de la *désirabilité* de chaque alternatives : les préférences de l'agent.

Dans de nombreux domaines, une fonction d'utilité est souvent utilisée pour appréhender la notion de préférence et donc pour mettre en oeuvre la phase de décision. L'utilité de chaque alternative est calculée et celle ayant la valeur la plus forte est choisie comme réaction. C'est une méthode très précise pour représenter les connaissances sur la désirabilité des alternatives. De plus, elle permet de prendre en compte la nature stochastique de l'environnement (notion d'utilité espérée) et est développée depuis longtemps dans le contexte de la théorie de la décision. Malheureusement, ces représentations numériques souffrent d'inconvénients (difficulté pratique pour obtenir les informations nécessaires, précision souvent inutile, représentation non homogène avec la description interne d'un agent, représentation peu intuitive) qui les rendent peu adaptées à une modélisation cognitive d'agents.

Notre proposition

Afin de contourner ces limitations, nous choisissons de baser le classement des alternatives sur une représentation logique des préférences (voir [Lang, 2004] et [Coste-Marquis et al., 2004] pour une comparaison non exhaustive de quelques représentations à base d'ordonnancement). De plus, afin de coller au plus près au sens commun nous supposons que les informations sur la désirabilité des alternatives sont des comparaisons qui sont spécifiées par points de vue éventuellement contradictoires. AU sein d'un même point de vue, ces comparaisons vérifient les principes completifs d'expansion, de transitivité, de *Ceteris Paribus*, et sont telles que, par défaut, deux propriétés quelconques sont jugées indifférentes.

La plupart des formalismes logiques ([von Wright, 1972], [Doyle and Wellman, 1994], [Boutilier, 1994]) ne formalisant pas de façon satisfaisante l'hypothèse *Ceteris Paribus*, nous proposons notre propre définition. Celle-ci se base sur l'idée que les informations sur la désirabilité des alternatives doivent-être utilisées comme des arguments[1] pour comparer les alternatives entre elles. Plus précisément, l'idée directrice de notre approche consiste à scinder la phase de spécification de la préférence en trois étapes (voir la figure suivante). Dans un premier temps, les données initiales de chaque point de vue sont interprétées selon l'hypothèse *Ceteris Paribus*, ce qui génère, pour chaque point de vue, une *préférence partielle primitive*. Cette phase dite de génération permet d'exprimer de façon naturelle et intuitive les données initiales. Dans un deuxième temps, chaque préférence partielle primitive est étendue (au-delà de l'hypothèse *Ceteris Paribus*) en une *préférence partielle étendue*. Cette phase dite « d'extension » permet de départager un plus grand nombre d'alternatives. Dans un troisième temps, la *préférence globale* est construite, à partir de toutes les préférences partielles étendues spécifiées en parallèle, via un mécanisme d'élection. Cette phase dite « d'agrégation » permet de gérer des contradictions entre points de vue.

Plan du document

Dans le premier chapitre, nous présentons les systèmes cognitifs d'agents et en particulier le formalisme proposé par Rao et Georgeff ainsi que celui proposé par Sadek. Par la suite, nous mettons en évidence le fait que fréquemment ces formalismes ne permettent pas

[1]Dans ce document, le terme « argument » est utilisé pour son sens intuitif. Il ne réfère en aucun cas à une définition formelle telle que celles que l'on trouve dans les théories de l'argumentation.

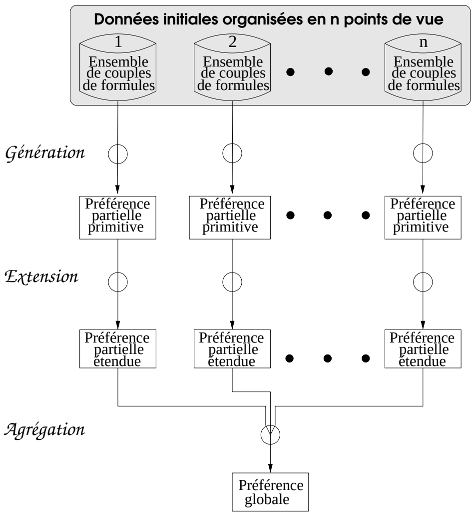


Schéma de construction des préférences

de (1) départager des alternatives que le sens commun départagerait et surtout de (2) spécifier clairement la réaction que le système met en oeuvre. Ce faisant nous présentons notre vision du cycle d'interaction de l'agent avec son environnement et ainsi affirmons que ces deux limitations sont des conséquences du manque d'une phase explicite de décision dans ces formalismes. Par la suite nous présentons les enrichissements que Louis a proposés au formalisme de Sadek afin de mettre explicitement en œuvre cette phase de décision. Nous terminons ce chapitre en constatant la nécessité de nouvelles informations pour mettre en œuvre cette phase de décision : les préférences.

Dans le second chapitre nous faisons un état de l'art de *la théorie de la décision*. Plus particulièrement, après avoir présenté rapidement cette théorie ainsi que les formalismes couramment utilisés basés sur des représentations numériques, nous exposons les formalismes logiques proposés par von Wright, Doyle et Wellman, Boutilier ainsi que celui des CP-nets. Nous montrons alors que les méthodes basées sur des représentations qualitatives sont plus prometteuses que celles basées sur des représentations quantitatives dans l'optique d'être utilisées au sein de formalismes logiques d'agents rationnels et en particulier afin de mettre en oeuvre des systèmes de dialogue.

Dans le troisième chapitre nous donnons un aperçu de notre proposition à savoir formaliser logiquement le concept de préférence. Plus précisément, après avoir donné les raisons d'une nouvelle modélisation logique de la préférence, nous exposons les hypothèses sur lesquelles se base notre travail puis donnons le schéma général de notre formalisation. Nous concluons ce chapitre en regardant les travaux connexes afin de mettre en évidence l'intérêt de notre travail.

Dans le quatrième chapitre nous donnons quelques repères pour mieux appréhender notre formalisme. Plus précisément, (1) nous faisons un rappel de la logique des prédicats du premier ordre ; (2) nous définissons la notation $\phi+\overline{\psi}$ ainsi que les concepts d'alternative, de préférence et d'indifférence ; (3) nous exposons un exemple que nous reprendrons au travers des chapitres suivants afin d'illustrer notre proposition.

Dans le cinquième chapitre nous formalisons la notion à la base de notre proposition : la **préférence partielle primitive**. Celle-ci permet de spécifier de façon intuitive les préférences d'un agent. Après avoir formalisé de façon originale l'hypothèse *Ceteris Paribus*, nous exposons la syntaxe, la sémantique et l'axiomatique d'un nouvel opérateur logique modal (noté $\succeq_i$) permettant de décrire n'importe qu'elle relation de préférence partielle primitive. Par la suite, après avoir exposé les propriétés remarquables de cet opérateur, nous reprenons l'exemple introduit dans le chapitre précédent afin d'illustrer l'apport d'un tel opérateur.

Dans le sixième chapitre nous formalisons la notion de **préférence partielle étendue**. Cette dernière permet d'utiliser les informations décrites avec l'opérateur associé de préférence partielle primitive pour comparer les alternatives parmi lesquelles l'agent a à faire un choix. Après avoir formalisé ce que nous appelons l'hypothèse *Ceteris Imparibus*, nous exposons la syntaxe, la sémantique et l'axiomatique d'un second nouvel opérateur logique modal (noté $\geq_i$) permettant de décrire n'importe qu'elle relation de préférence partielle étendue. Par la suite, après avoir exposé les propriétés remarquables de cet opérateur, nous reprenons l'exemple introduit dans le quatrième chapitre afin d'illustrer l'apport d'un tel opérateur.

Dans le septième chapitre nous formalisons la notion de **préférence globale**. Cette dernière permet de gérer les contradictions entre les diverses préférences partielles étendues qui peuvent être définies. Après avoir exposé la syntaxe et la sémantique d'un troisième nouvel opérateur logique modal (noté $\geq$) permettant de décrire n'importe qu'elle relation de préférence globale, nous remarquons que la gestion des contradictions dépend de la politique d'agrégation choisie. C'est pourquoi nous exposons quelques exemples de politiques ainsi que leur axiomatisation puis discutons du choix d'une politique adaptée. Par la suite, nous proposons une voie pour utiliser cette relation de préférence globale afin de déterminer l'alternative à mettre en œuvre. Comme pour les deux chapitres précédents, nous terminons ce chapitre en reprenant l'exemple introduit dans le quatrième chapitre.

Dans le huitième et dernier chapitre nous proposons deux extensions possibles de notre formalisme. La première consiste en son intégration au sein du formalisme logique sur lequel se base la théorie de l'interaction proposée par Sadek. La seconde consiste en l'utilisation, dans une phase de décision, de préférences conjointement avec d'autres informations sur la désirabilité des alternatives. Cette dernière est à considérer comme un travail préliminaire et considère deux voies différentes : voir l'utilisation de ces informations sur la désirabilité des alternatives comme des filtres ou comme des arguments.

En annexe nous exposons les preuves relatives à notre formalisation de l'hypothèse *Ceteris Paribus* ainsi que celles relatives à l'opérateur de préférence partielle primitive et de préférence partielle étendue. Nous y exposons aussi quelques propriétés supplémentaires de la notation $\phi{+}\overline{\psi}$ ainsi qu'un rappel sur les relations binaires et sur le théorème d'Arrow.

Préambule i
Remerciements . . . i
Introduction . . . ii
Sommaire . . . vi
Table des figures . . . viii

I Problématique : agents rationnels et prise de décision **1**

1 Contexte de la thèse : agents rationnels **3**
1.1 Modèles cognitifs d'agents . . . 4
1.2 Une limitation : le choix de la réaction de l'agent . . . 20

2 Prise de décision **31**
2.1 La Théorie de la Décision . . . 32
2.2 Modélisations numériques des préférences . . . 36
2.3 Modélisations logiques des préférences . . . 42
2.4 Conclusion . . . 58

II Fondements pour des préférences adaptées aux agents **63**

3 Notre proposition dans les grandes lignes **65**
3.1 Pourquoi proposer une nouvelle modélisation logique des préférences ? . . . 65
3.2 Hypothèses de notre travail . . . 67
3.3 Le modèle proposé dans les grandes lignes . . . 70
3.4 Travaux connexes et intérêts de l'approche . . . 73

4 Langages, notations et exemple utilisés **77**
4.1 Le langage des prédicats du premier ordre $\mathscr{L}$ . . . 77
4.2 La notation $\phi{+}\overline{\psi}$ . . . 83
4.3 Les concepts importants . . . 87
4.4 Exemple de scénario . . . 89

III Formalisation des préférences en LPPO **93**

5 Préférences partielles primitives **95**
5.1 L'hypothèse *Ceteris Paribus* . . . 96
5.2 La relation de *préférence partielle primitive* G_i . . . 101
5.3 Syntaxe de l'opérateur $\succeq_i$ . . . 101
5.4 Sémantique de l'opérateur $\succeq_i$ . . . 102
5.5 Axiomatisation de l'opérateur $\succeq_i$ . . . 104
5.6 Propriétés remarquables . . . 107
5.7 Exemple (suite 1) . . . 109

6 Préférences partielles étendues **111**
6.1 L'hypothèse *Ceteris Imparibus* . . . 112
6.2 Syntaxe de l'opérateur $\geq_i$ . . . 114
6.3 Sémantique de l'opérateur $\geq_i$ . . . 115
6.4 Axiomatisation de l'opérateur $\geq_i$ . . . 116
6.5 Propriétés remarquables . . . 118
6.6 Exemple (suite 2) . . . 119

7 Préférence globale et choix de la réaction **123**
7.1 La relation de *préférence globale G* . . . 124
7.2 Syntaxe de l'opérateur $\geq$ . . . 125
7.3 Sémantique de l'opérateur $\geq$ . . . 125
7.4 Politiques d'agrégation . . . 127
7.5 Axiomatisation des politiques . . . 130
7.6 Discussion sur les politiques agrégation . . . 133
7.7 Choix de la réaction . . . 135
7.8 Exemple (Fin) . . . 137

IV Extensions et conclusion générale **141**

8 Extensions possibles de notre formalisme **143**
8.1 Vers l'intégration de notre proposition à un formalisme d'agents . . . 143
8.2 Utilisation conjointe de différents types de désirabilités . . . 149

9 Conclusion générale du document **157**
9.1 Contexte et problématique de la thèse . . . 157
9.2 Notre proposition . . . 158
9.3 Contributions . . . 159
9.4 Perspectives . . . 160

Références **163**

Table des matières **179**

Annexes **187**

A Preuves **189**
A.1 Preuves relatives à la notation $CP(a, b, \phi, \psi)$ . . . 189
A.2 Preuves relatives à la préférence partielle primitive . . . 191
A.3 Preuves relatives à la préférence partielle étendue . . . 194

B Divers **197**
B.1 Quelques propriétés supplémentaires de $\phi{+}\overline{\psi}$ . . . 197
B.2 Rappels sur les relations . . . 198
B.3 Le théorème d'impossibilité d'Arrow . . . 201

Table des figures

1.1 Vue simplifiée des attitudes mentales d'un agent BDI 6
1.2 Schéma général de fonctionnement d'un agent BDI 7
1.3 Schéma d'un arbre temporel CTL 10
1.4 Modèle du temps utilisé par Sadek 14
1.5 Schéma général et simplifié de l'évolution des formules du modèle de Sadek 18
1.6 Cas 1 : une intention, un plan pour la réaliser 21
1.7 Cas 2 : une intention, deux plans pour la réaliser 21
1.8 Cas 3 : deux intentions, un plan différent pour réaliser chacune d'elles . . 22
1.9 Lien général et simplifié entre état mental et réaction 25

2.1 Exemple de représentation matricielle d'un problème de décision 38
2.2 Exemple de représentation avec un arbre de décision 38
2.3 Exemple de représentation avec une matrice d'utilité 39
2.4 Exemple de CP-net 55
2.5 Graphe de préférences correspondant au CP-net précédant 56
2.6 Dialogue élicitant des valeurs d'utilité 60
2.7 Dialogue élicitant des comparaisons 60

3.1 Schéma d'organisation des informations nécessaires à une décision 69
3.2 Construction des préférences. 72

7.1 Schéma regroupant les trois types de préférences partielles primitives . . . 138
7.2 Schéma de la préférence globale avec la politique « Majorité » 138
7.3 Schéma de la préférence globale avec la politique « Lex » 139

8.1 Un arrangement de trois mécanismes permettant de déterminer la réaction 152
8.2 1er schéma d'agrégation pour différents types d'informations de désirabilité 154
8.3 2nd schéma d'agrégation pour différents types d'informations de désirabilité 155

B.1 Graphe représentant une relation binaire quelconque 200

Première partie

Problématique : agents rationnels et prise de décision

Chapitre 1

Contexte de la thèse : agents rationnels

Cette thèse s'inscrit dans la lignée des travaux effectués à France Télécom par Sadek, Bretier et Louis [Sadek, 1991][Bretier, 1995][Louis, 2002]. Ceux-ci visent à développer un modèle cognitif d'agents rationnels permettant entre autre de mettre en œuvre des systèmes de dialogue [Sadek et al., 1997][Sadek and Mori, 1998][Sadek, 1999]. C'est pourquoi dans ce premier chapitre nous présentons les modèles cognitifs d'agents rationnels, leurs intérêts ainsi que leurs limites.

Plus précisément, dans la première section de ce chapitre nous présentons le contexte de notre travail à savoir le paradigme agent et plus particulièrement les modèles cognitifs d'agents rationnels dits BDI (Belief, Desire, Intention). Ces modèles se fondent entre autres sur les travaux de Mc. Carthy et Hayes qui ont proposé d'associer aux systèmes informatiques des qualités mentales comme cela est souvent fait pour les êtres animés [McCarthy and Hayes, 1969] [McCarthy, 1979]. L'un des intérêts de ces modèles est d'ailleurs de chercher à spécifier le comportement des agents grâce à leurs attitudes mentales, telles que les connaissances, les désirs (ou souhaits ou préférences) et les intentions. Par la suite, après avoir présenté la formalisation logique proposée par Rao et Georgeff dans [Rao and Georgeff, 1991b], nous présenterons celle proposée par Sadek dans [Sadek, 1991]. C'est plus particulièrement sur cette dernière (et sur les extensions proposées par Louis dans [Louis, 2002]) que s'appuient nos travaux.

Dans la seconde section de ce chapitre nous montrons que, dans le cas général, ces modèles d'agents ne permettent pas de (1) départager des alternatives que le sens commun départagerait et surtout de (2) spécifier clairement la réaction que le système met en oeuvre. Ce faisant nous présentons notre vision du cycle d'interaction de l'agent avec son environnement et ainsi affirmons que ces deux limitations sont des conséquences du manque d'une phase explicite de décision. Par la suite nous présentons les enrichissements que Louis a proposés au formalisme de Sadek afin de mettre explicitement en œuvre cette phase de décision. Nous terminons ce chapitre en constatant la nécessité de nouvelles informations pour mettre en œuvre cette phase de décision : les préférences.

1.1 Modèles cognitifs d'agents

Dans cette section nous présentons dans les grandes lignes les modèles cognitifs d'agents (parfois désignés à tort sous le nom de « modèles BDI ») et plus particulièrement le modèle proposé par Sadek dans [Sadek, 1991]. Nous attirons l'attention du lecteur sur le fait que les informations présentées ici ne se veulent pas être exhaustives. Elles sont là uniquement pour situer notre travail et donner au lecteur les outils pour le comprendre.

1.1.1 La notion d'agent

Avant d'aller plus loin, nous devons donner un sens plus précis au terme *agent*. Celui-ci désignant étymologiquement une entité qui agit sur son environnement, tout programme peut, dans un certain sens, être vu comme un agent. C'est une des raisons pour laquelle de nombreuses définitions parfois contradictoires peuvent être rencontrées dans la littérature (voir [Franklin and Graesser, 1996] ou [Wooldridge, 1999]). Pour notre part, nous nous basons sur les travaux de Russell et Norvig qui définissent dans [Russell and Norvig, 2003] un agent par les trois qualités suivantes :

- un agent perçoit son environnement, s'adapte à ses changements et agit en fonction.
- un agent persiste dans le temps et peut donc percevoir sa propre dynamique et celle de son environnement.
- un agent évolue de façon autonome : il est capable d'apprendre afin de compenser le fait que ses croyances initiales sont partielles et incomplètes.

Pour Gasser, toutes ces caractéristiques reposent sur le concept d'*action persistante* (en anglais *the structured persistent action*)[Briot and Gasser, 1998]. Ce dernier indique que, au lieu de demander à l'agent d'exécuter certaines actions précises, le développeur spécifie des actions plus abstraites à effectuer sur une durée plus ou moins longue ; l'agent a alors un degré de flexibilité sur la manière de les réaliser. Ainsi, un peu comme un objet manipule ses variables, un agent manipule son comportement. Pour les auteurs de [Briot and Gasser, 1998] et de [Jennings and Wooldridge, 1998], les agents sont en fait un nouveau paradigme de programmation. Après le langage machine, l'assembleur, la programmation structurée (procédurale et déclarative), la programmation objet, vient la programmation par agents. L'idée sous-jacente est de développer des systèmes informatiques grâce à des « briques » (les agents) qui intègrent nativement les trois caractéristiques citées plus haut. Ce n'est donc plus au programmeur qu'incombe la gestion des réussites et des échecs des modules du système mais aux modules du système eux-mêmes (pourvu qu'ils sachent ce qu'ils doivent faire, s'ils ont réussi à le faire ou pas, et qu'ils aient à leur disposition différentes méthodes pour le faire). Remarquons ici qu'un nombre de plus en plus grand de systèmes complexes sont appréhendés comme des ensembles d'entités individuelles indépendantes et autonomes. Par conséquent, le paradigme agent semble être une abstraction adaptée pour concevoir et mettre en oeuvre un bon nombre d'applications d'aujourd'hui comme les applications de dialogues et les applications réparties.

Mais comment spécifier le comportement d'un agent ? Comment programmer commodément une telle structure afin qu'elle donne l'impression d'être autonome dans la réalisation de la tâche qui lui a été assignée ? Une des voies possibles est de faire un parallèle entre la façon dont le comportement rationnel d'une personne peut être décrit et la façon d'indiquer à un système informatique comment se comporter. C'est ce que nous exposons dans cette section.

1.1.2 Les modèles cognitifs d'agents

a) Fondements historiques et philosophiques

Dans les années 70, Mc. Carthy et Hayes ont proposé d'associer aux systèmes informatiques des qualités mentales comme cela est fait pour les êtres humains [McCarthy and Hayes, 1969] [McCarthy, 1979]. L'idée sous-jacente est de décrire l'état interne ainsi que le comportement d'un système informatique comme peuvent l'être ceux d'un être humain. Ceci pousse à affecter aux systèmes informatiques des attitudes mentales comme par exemple des croyances, des intentions ou des obligations.

On ne s'intéresse pas ici à la question philosophique de savoir si une machine peut ou non avoir des attitudes mentales avec la même signification que cela aurait pour un humain. Cependant, il apparaît utile de considérer une machine comme si elle avait des attitudes mentales pour modéliser son comportement rationnel (i.e. suivant des règles logiques). En effet, il est courant pour tout un chacun (pour un être humain) de conceptualiser des animaux, des machines ou des objets (en particulier technologiques) comme des entités ayant des attitudes mentales (voir par exemple [Shoham and Cousins, 1994]). Cela se justifie d'autant plus par le fait que, personne ne connaît les attitudes mentales de ses interlocuteurs ni même ne sait s'ils en ont, ils sont simplement supposés en avoir au vu de leur comportement *intelligent*.

Ainsi, par exemple, Newell a proposé de spécifier un système informatique à base de la notion de « croyance »[Newell, 1980]. Par la suite, de nombreux chercheurs en Intelligence Artificielle ont éprouvé le besoin d'utiliser des attitudes mentales supplémentaires (voir entre autres [Bratman, 1987], [Cohen and Levesque, 1990b], [Rao and Georgeff, 1991b], [Sadek, 1991], [Konolige and Pollack, 1993] [Brazier et al., 1996], [van der Torre and Weydert, 1998], [Burkhard et al., 1998], [Georgeff et al., 1999], [van der Hoek et al., 1999], [Padgham and Lambrix, 2000], [Thangarajah et al., 2002], [Dastani et al., 2002a], [Amgoud et al., 2003]). Aujourd'hui, le terme *BDI* est communément utilisé comme raccourci pour désigner les modèles décrivant le fonctionnement d'un agent en utilisant le concept d'attitude mentale et en particulier en utilisant des Croyance, des Désir et des Intention.

Notons qu'originellement le terme *BDI* fut introduit dans [Bratman et al., 1988]. Il désignait alors une architecture bien particulière permettant de modéliser des agents aux ressources limitées. Cette dernière a été désignée plus tard par Pollack sous le nom de IRMA pour « Intelligent Resource-Bounded Machine Architecture » afin de réserver le terme *BDI* pour désigner l'ensemble des modèles cognitifs d'agents [Georgeff et al., 1999, p6]. Ce modèle s'appuie sur l'hypothèse originellement proposée par Bratman dans [Bratman, 1987] selon laquelle les *intentions* jouent, en pratique, un rôle important dans le raisonnement d'un agent humain. Pour Bratman c'est un moyen de réduire le nombre d'alternatives à prendre en compte :

> « The formation of intentions and the commitments thereby entailed are seen as a mechanism - possibly one among many - for constraining the set of options about which an agent must reason.[Georgeff et al., 1999] »

Plus particulièrement, Bratman soutient qu'un agent rationnel (i.e un agent qui met en œuvre à chaque instant la réaction qu'il juge la meilleure) fonde son raisonnement sur ses intentions en mettant de côté les comportements qui sont en conflit avec ses intentions. Ceci est nécessaire car un agent ne peut considérer toutes les alternatives qui s'offrent à lui séparément et donc toutes les évaluer. Dans cette approche, l'intention peut être vue comme un moyen pour l'agent de focaliser son raisonnement et ainsi d'acquérir une certaine stabilité dans son comportement. Ceci est à rapprocher du fait que pour créer des agents, différents « constituants » doivent être assemblés. En particulier, comme les auteurs de [Rao and Georgeff, 1995] le font remarquer, un agent a besoin d'un mécanisme pour sélectionner la réaction à mettre en œuvre. Au moment de l'apparition des modèles BDI (fin des années 80), c'est la *Théorie de la décision* qui fournit les mécanismes les plus développés pour mettre en œuvre une phase de sélection. Malheureusement, ces derniers ne sont généralement pas bien adaptés aux situations où l'environnement peut changer de façon impromptue car le coût de calcul des fonctions d'utilité (sur lesquelles ils se basent) est très élevé. Dans les modèle BDI, ce sont les désirs et surtout les intentions qui jouent ce rôle de sélection. Remarquons enfin que dans [Rao and Georgeff, 1991a], les auteurs montrent qu'une modélisation BDI d'un système est une représentation, parmi d'autres, de l'arbre de décision de ce système.

b) Le modèle *BDI* (aperçu)

L'objectif d'un modèle type « *BDI* » est de décrire et de spécifier le comportement d'un agent (i.e. la production d'une sortie/réaction en fonction d'un(e) stimulus/entrée) uniquement en termes de Croyances (B), de Désirs (D) et d'Intentions (I) (voir la figure 1.1).

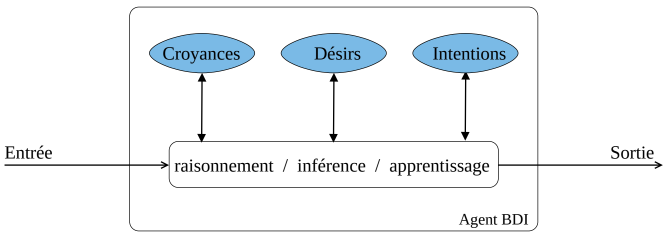


FIG. 1.1 – Vue simplifiée des attitudes mentales d'un agent BDI

À cet effet, il est nécessaire de définir les liens que les attitudes mentales entretiennent entre elles ainsi que le liens entre ces attitudes et l'environnement de l'agent. Les croyances représentent la partie *informative* des données : c'est le modèle du monde qu'a l'agent. Les désirs représentent la partie *motivationnelle* des données : c'est ce vers quoi l'agent désire faire évoluer le monde. Enfin, Les intentions représentent la partie *délibérative* des données : c'est ce que l'agent s'engage à mettre en œuvre afin de réaliser ses désirs. Notons que l'intention n'est pas une attitude mentale définie uniquement à partir de croyances et de désirs.

Les interactions entre ces attitudes suivent le schéma selon lequel les observations entraînent des croyances, qui entraînent des désirs, qui, à leur tour, entraînent des intentions qui, vont alors entraîner la réaction de l'agent (voir [Wooldridge, 1999, pp29-35]) . Plus précisément, les croyances dérivent des observations de l'agent et de ses anciennes croyances ; les désirs dérivent des croyances et de ses anciennes intentions ; les intentions sont un sous-ensemble consistant des désirs de l'agent (en se basant sur ses croyances) ; la réaction de l'agent se base uniquement sur les intentions de l'agent (voir la figure 1.2).

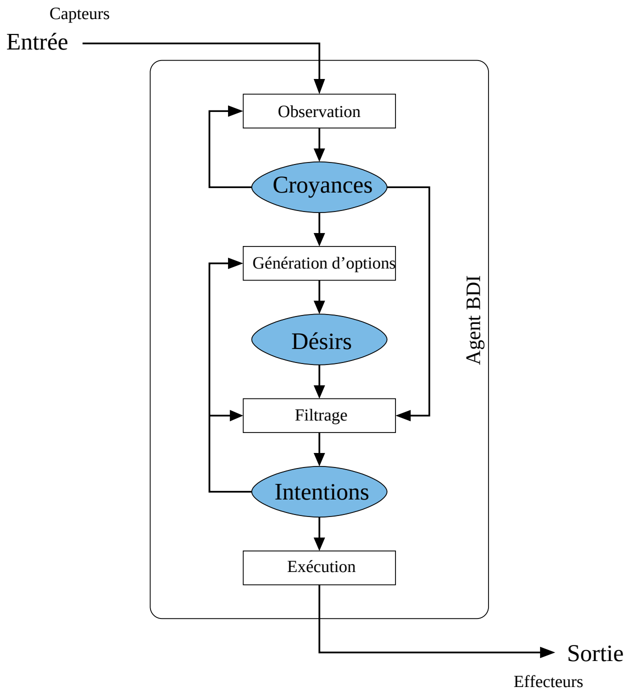


FIG. 1.2 – Schéma général de fonctionnement d'un agent BDI ([Wooldridge, 1999])

Pour Bratman, l'intention joue en pratique un rôle important dans le raisonnement et permet *in fine* de relier les réactions aux observations en disposant d'une théorie du passage à l'acte des agents (voir [Bratman, 1987]). C'est pourquoi le concept d'intention est au centre de cette approche :

– Les intentions dirigent le raisonnement puisque l'agent agit afin de les satisfaire. En particulier elles concernent le futur.
– Les intentions permettent d'adapter le raisonnement de l'agent à ses ressources limitées en restreignant la phase de délibération. En effet, d'une part, les intentions focalisent le raisonnement de l'agent, et, d'autre part, les plans (comme intentions) ne sont pas très détaillés[1] afin de pourvoir être adaptés à la précision limitée des

[1] La seule chose imposée aux plans en tant qu'intentions est que l'action qui débute le plan doit être

connaissances de l'agent.
- Les intentions ont vocation à être réalisées. Par conséquent, l'agent peut raisonner en considérant que, dans le futur, elle le seront (i.e. anticiper). Par exemple, si j'ai l'intention d'aller à Paris, je peux prendre pour acquis que dans un futur proche je serai à Paris pour planifier dès maintenant la suite de mon séjour.
- Les intentions persistent dans le temps. En particulier, suite à un échec l'agent tentera une autre action pour satisfaire l'intention. Une intention peut toutefois être abandonnée. C'est en particulier le cas si les raisons qui ont produit cette intention ont disparu, ou s'il apparaît qu'il est impossible de la satisfaire.

L'une des problématiques les plus importantes est relative à ce dernier point. Il s'agit de gérer la dualité entre la dimension proactive et la dimension réactive de l'agent. La difficulté est de bien savoir doser la persistance des intentions afin qu'elles permettent de focaliser le raisonnement de l'agent tout en pouvant être modifiées si l'évolution des croyances de l'agent le rend nécessaire. En d'autres termes, il faut trouver le juste équilibre entre la réactivité de l'agent et la stabilité de son comportement dirigé vers les buts.

L'intérêt d'une telle approche est multiple. Elle permet tout d'abord de produire des modélisations de systèmes informatiques faciles à comprendre puisqu'elle fait le parallèle entre le raisonnement humain et l'exécution du programme. En particulier, la prévision du comportement du système est plus intuitive. De plus, une telle approche rend possible et naturelle la modélisation des autres agents et en particulier du programme lui-même. Il est en effet possible d'avoir des croyances sur les croyances, des croyances sur les désirs, ... Enfin, la production du code ainsi que son évolution sont facilitées.

Cette architecture peut être vue comme « la base » de bon nombre de mises en œuvre actuelles d'agents logiciels rationnels. Ces mises en œuvre (en particulier PRS[2] [Georgeff and Ingrand, 1989, Ingrand et al., 1992], dMARS [D'Inverno et al., 1997, D'Inverno et al., 2004], JAM[3] [Huber, 1999], JACK[4] [Howden et al., 2001], Jadex[5] [Pokahr et al., 2003] utilisent souvent des buts (goals) au lieu des désirs, plus abstraits. Souvent les désirs d'un agent peuvent être en conflit entre eux et donc ne peuvent pas mener directement à une action. De leur côté, les buts forment toujours un ensemble consistant et donc motivent toujours directement l'exécution d'un plan. Notons toutefois que des techniques existent afin de gérer les conflits entre désirs (voir par exemple [Amgoud et al., 2003]).

c) La formalisation de Rao et Georgeff du modèle *BDI*

À la fin des année 80, Bratman a cherché à modéliser philosophiquement un agent humain rationnel. En particulier, il a insisté sur le rôle de l'intention dans le raisonnement [Bratman, 1987] et proposé une architecture d'agents basée sur les concepts de croyance,

déterminée avec exactitude du point de vue des effecteurs de l'agent. C'est une hypothèse très courante ; elle est par exemple faite par Louis pour définir ce qu'il appelle des *stratégies* [Louis, 2002, p79].

[2] http://www.ai.sri.com/~prs/

[3] http://www.marcush.net/IRS/irs_downloads.html

[4] http://www.agent-software.com/shared/products/

[5] http://vsis-www.informatik.uni-hamburg.de/projects/jadex/

de désir et d'intention [Bratman et al., 1988]. Aujourd'hui, ces travaux sont considérés comme les fondements du modèle BDI. Parmi les nombreux travaux dans ce domaine, ceux de Rao et Georgeff sont particulièrement intéressants à considérer car ils proposent une formalisation logique du modèle BDI. C'est à ce titre que nous présentons ici rapidement leurs travaux. Le lecteur intéressé par plus de détails pourra se reporter, entre autres, à [Rao and Georgeff, 1991b], [Rao and Georgeff, 1991a] et [Rao and Georgeff, 1995].

Pour Rao et Georgeff, un plan est un ensemble d'actions décrites plus ou moins précisément et permettant de faire évoluer le monde d'une situation à une autre. Il peut être soit juste connu (c'est alors une *recette* i.e. un *savoir-faire*), soit en exécution ou, du moins, sur le point de l'être (c'est alors une *intention*). Comme mentionné précédemment, la modélisation proposée par Rao et Georgeff dans [Rao and Georgeff, 1991b] se base sur trois attitudes mentales : les Croyances, les Buts (à la place des Désirs) et les Intentions. L'ensemble des croyances constituent le modèle du monde qu'a l'agent. Les croyances peuvent donc porter sur des informations très diverses (plan, état des objets du monde, croyances des autres agents, ...), être inexactes et/ou incomplètes, mais forment toujours un ensemble consistant. Les buts représentent les motivations / objectifs de l'agent : tout ce que l'agent souhaite mettre en œuvre ou réaliser. Notons que le terme de « but » employé ici indique que les motivations de l'agent ne sont pas contradictoires entre elles ; le terme désir étant généralement employé lorsque ce n'est pas le cas. Enfin, les Intentions désignent les plans que l'agent s'engage (envers lui-même) à réaliser. Les intentions sont le résultat du raisonnement de l'agent et sont maintenues tant qu'elles ne sont pas réalisées ou jugées comme impossibles. Elles forment un ensemble consistant et permettent à l'agent d'avoir un comportement sur le long terme. Notons que dans cette approche, l'intention est une attitude primitive et ne se réduit donc pas à une combinaison uniquement constituée de croyances et de buts. Ce sont des plans partiels que l'agent s'engage à mettre en œuvre afin de réaliser ses buts.

Rao et Georgeff formalisent leur modèle à l'aide d'une logique multimodale. Plus particulièrement, ils étendent une logique propositionnelle de style CTL (voir par exemple [Emerson and Srinivasan, 1989] et [Katoen, 2004]) avec trois opérateurs modaux : BEL, GOAL, INTEND. Ces derniers représentent respectivement les croyances, les buts et les intentions de l'agent. En particulier, cette formalisation permet de représenter des intentions sur des croyances, des croyances sur des croyances, des croyance sur des désirs futurs et ainsi de suite.

La sémantique de leur logique est basée sur les mondes possibles de Kripke [Kripke, 1963]. De façon similaire à la logique CTL, dans leur formalisation, chaque situation de l'univers est représentée par un monde possible. Ces derniers sont les nœuds d'un arbre temporel (appelé par la suite « arbre CTL ») dont les branches ont un sens et représentent les actions permettant de transformer l'univers d'une situation en une autre (voir la figure 1.3). Cette structure en arbre permet de représenter les options qui s'offrent à l'agent dans chaque situation : ce sont les branches au départ de chaque nœud. À chacun des trois nouveaux opérateurs modaux (i.e. BEL, GOAL, INTEN) est associé une relation d'accessibilité sur les mondes possibles notée respectivement B, G, I. Un agent a la croyance p (respectivement le but ou l'intention p) dans le monde w si tous les mondes reliés à w par

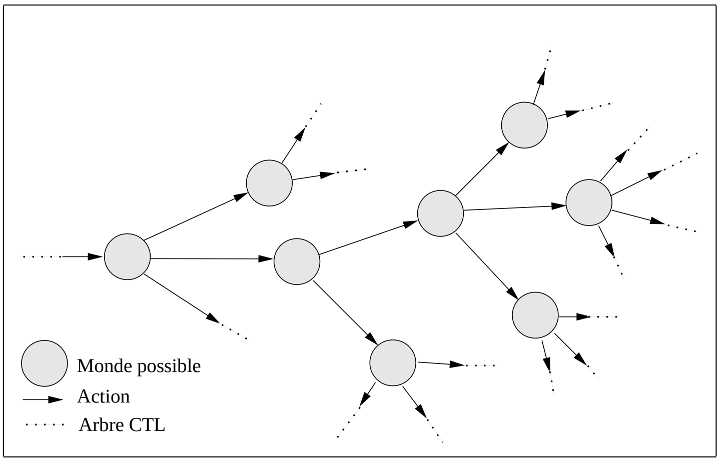


FIG. 1.3 – Schéma d'un arbre temporel CTL

la relation d'accessibilité B (respectivement G ou I) vérifient la propriété p. Ainsi, chaque monde w est relié à un ensemble de mondes par la relation d'accessibilité B. Intuitivement, ils correspondent aux mondes que l'agent pense comme possiblement être le monde actuel si le monde actuel est le monde w. De la même façon, chaque monde w est relié à un ensemble de mondes par la relation d'accessibilité G ainsi qu'à un ensemble de mondes par la relation d'accessibilité I. Intuitivement, ils correspondent respectivement aux mondes que l'agent veut atteindre et aux mondes que l'agent s'engage a essayer d'atteindre lorsque le monde actuel est le monde w.

Afin de capturer les interactions désirées entre les différentes attitudes mentales, Rao et Georgeff imposent à leur logique des axiomes, ainsi que des propriétés sur et entre les relations d'accessibilité (voir [Rao and Georgeff, 1991b]). Parmi ceux-ci, notons en particulier que :

- L'opérateur BEL est KD45 (i.e. il est clos sous l'implication logique, consistant et vérifie la propriété d'introspection positive et négative).
- Les opérateurs GOAL et INTEND sont KD (i.e. ils sont clos sous l'implication logique et consistants).
- Les trois opérateurs vérifient la règle de nécessitation : l'agent croie, désire et souhaite réaliser toutes les formules valides.
- $INTEND(p) \Rightarrow BEL(INTEND(p))$ et $GOAL(p) \Rightarrow BEL(GOAL(p))$ sont des axiomes : l'agent s'introspecte au sujet de ses buts et de ses intentions.
- $GOAL(p) \Rightarrow BEL(p)$ est un axiome : l'agent croit réalisable chacun de ses buts (propriété de réalisme [Cohen and Levesque, 1990b]) et même croit que quelqu'un peut les réaliser (propriété de réalisme fort) en imposant une autre condition sur les relations d'accessibilité : pour tout monde w accessible par la relation de croyance, il existe un monde accessible par la relation de but qui fait partie d'un sous-arbre « CTL » de l'arbre « CTL » dont fait partie le monde w. L'inverse est faux : un agent n'est pas obligé d'avoir pour but des faits qu'il croirait inévitables. Il en est de

même pour les intentions.
- Si e est une action primitive alors $done(e) \Rightarrow BEL(done(e))$ est un axiome : à chaque fois que l'action e est réalisée, l'agent pense que l'action e vient effectivement d'être faite (observation).
- Si e est une action primitive alors $INTEND(does(e)) \Rightarrow does(e)$ est un axiome : Si un agent à l'intention de mettre en œuvre une action primitive alors elle est mise en œuvre. Si un agent a, dans une situation, plusieurs actions primitives alternatives, il ne pourra pas agir tant qu'il n'en aura pas délibérément choisi une. Une façon de traiter ce problème est de considérer la phase de délibération comme une action primitive. Une autre approche est de modifier l'axiome ci-dessus de sorte à mettre en œuvre une action au hasard. Nous verrons dans la suite du document qu'il est possible de formaliser une phase de décision et que pour cela, il faut introduire de nouvelles informations (voir chapitre 3).
- $INTEND(p) \Rightarrow inevitable \diamond (\neg INTEND(p))$ est un axiome : un agent ne doit pas reporter sine die ses intentions.

Remarquons enfin que dans [Rao and Georgeff, 1995, p5], les auteurs évoquent d'autres axiomatisations possibles et insistent sur le fait qu'elles peuvent être adaptées à certains cas. En particulier, ils ont défini de cette manière dans [Rao and Georgeff, 1991b, p11] différentes stratégies d'engagement qu'un agent peut adopter (« blind », « single minded », et « open minded »). Informellement, un agent est dit avoir une stratégie « blind » s'il perd son intention uniquement lorsqu'il la sait réalisée ; un agent est dit avoir une stratégie « single minded » s'il la perd aussi lorsqu'il pense qu'elle n'est plus réalisable ; un agent est dit avoir une stratégie « open minded » si en plus il perd son intention lorsqu'elle ne fait plus partie de ses buts.

1.1.3 Le cas de la théorie de l'interaction rationnelle

Dans cette section nous introduisons le modèle d'agent sous-jacent à *la théorie de l'interaction rationnelle* proposée par Sadek dans [Sadek, 1991]. Afin de reprendre les modifications proposées par Louis dans [Louis, 2002] et d'exposer par la suite nos travaux avec plus de clarté, les notations et les définitions peuvent différés quelque peu des travaux originaux. En particulier, puisque nous ne considérons qu'un seul agent nous n'indicerons pas les opérateurs représentant ses attitudes mentales (voir par la suite) ; dans la théorie originelle ils sont indicés afin de différencier les agents. Le lecteur intéressé par un exposé plus complet de la théorie de l'interaction rationnelle pourra se reporter aux ouvrages [Sadek, 1991], [Bretier, 1995] et [Louis, 2002].

a) Introduction

L'une des activités du laboratoire *EASY* (**E**nrichissement des **S**ervices pour un **A**ccès **S**imple à l'**I**nformation) de France Telecom R&D est de concevoir et de mettre en œuvre des systèmes de dialogue humain-ordinateur conviviaux. Ceux-ci permettent, par exemple, de passer un ordre en bourse, de rechercher de l'information sur Internet ou de naviguer au sein de l'aide en ligne d'un logiciel avec une grande simplicité. À cet effet, l'équipe *ADN* (**A**gents intelligents **D**ialogues **N**aturels) développe la technologie *Artimis* (**A**gent **R**ationnel fondé sur une **T**héorie de l'**I**nteraction mise en œuvre par un **M**oteur d'**I**nférence **S**yntaxique) [Sadek et al., 1997][Sadek and Mori, 1998][Sadek, 1999]. Cette technologie d'Agents Intelligents est fondée sur *la théorie de l'interaction* développée par Sadek au début des années 90 [Sadek, 1991]. Cette théorie s'appuie entre autres sur les travaux philosophiques de Bratman en science cognitive ainsi que sur ceux d'Austin et de Searle sur le langage. Nous avons rapidement présenté au début de ce document les travaux de Bratman. Nous n'exposerons pas les travaux d'Austin et de Searle dans ce chapitre. Notons seulement que pour ces deux derniers, « communiquer c'est agir ». Plus précisément, ils ont défini le concept d'acte de langage et montré que la communication peut s'appréhender comme une théorie particulière de l'action (voir [Austin, 1962] et [Searle, 1969]).

Plus précisément, Sadek propose, d'une part, de modéliser les agents par leurs attitudes mentales (croyances, choix, intentions, incertitudes, ...) et, d'autre part, de leur conférer une propension "naturelle" à dialoguer. Cette capacité intrinsèque à dialoguer, et plus généralement à communiquer, est une conséquence « mécanique » de la spécification de l'évolution des attitudes mentales des agents lorsqu'ils interagissent avec leur environnement. En effet, pour Sadek un agent dialogue car il est *intelligent*. Pour nous d'ailleurs, un des apports majeurs de Sadek est d'intégrer à un modèle formel d'agent rationnel tout un ensemble relativement exhaustif d'actes communicatifs et ainsi donner à ces actes une sémantique précise.

Dans cette partie de document, nous nous focalisons sur le modèle logique d'agent proposé par Sadek dans [Sadek, 1991] et raffiné par la suite par Bretier dans [Bretier, 1995] et Louis dans [Louis, 2002]. Ce modèle repose d'une part sur deux attitudes mentales primitives : la croyance et le choix, et d'autre part sur la formalisation de la dynamique de l'univers avec une logique dynamique. Sur cette base sont construites des attitudes mentales non primitives comme l'incertitude ou l'intention. Remarquons ici que la méthodologie de construction de cette attitude reprend en partie celle proposée par Cohen et Levesque dans [Cohen and Levesque, 1990a]. Enfin, la notion d'équilibre rationnel permet de relier toutes les attitudes mentales entre elles et avec les observations et les réactions de l'agent. Comme nous le verrons plus en détail par la suite, ce modèle est formalisé dans une logique multimodale quantifiée des attitudes mentales et de l'action (noté $\mathscr{L}$). Schématiquement, Sadek définit chaque nouvelle notion de son modèle en enrichissant progressivement une logique du premier ordre.

b) Attitudes mentales primitives : B et C

La modélisation proposée par Sadek se base seulement sur deux attitudes mentales primitives : la Croyance et le Choix. La première représente la dimension informative de l'agent tandis que la seconde représente la dimension motivationnelle de l'agent.

La Croyance : Une proposition constitue une croyance d'un agent si celui-ci considère que cette proposition est vraie. La croyance est l'attitude mentale par laquelle un agent dispose d'un modèle de son environnement. Le système de croyance d'un agent est donc le moyen qui lui permet de conserver et de mettre à jour ce modèle, au fur et à mesure de son évolution [Sadek et al., 1996]. La croyance est modélisée (de façon classique) en enrichissant le langage $\mathscr{L}$ avec l'opérateur modal B. la notation $B(\phi)$ signifie que « la propriété ϕ est une conséquence logique des croyances de l'agent ». B est un opérateur modal normal à axiomatique KD45 quantifiée ; l'agent a donc, entre autres, les propriétés d'omniscience logique, de consistance entre ses croyances, d'introspection positive et négative. L'opérateur modal B est interprété via les mondes possibles de Kripke par une relation d'accessibilité (noté R_B) : un monde w vérifie la propriété $B(\phi)$ si et seulement si tous les mondes accessibles de w via la relation R_B vérifient la propriété ϕ. Cet opérateur étant KD45, la relation d'accessibilité qui lui est associée est donc sérielle, transitive et euclidienne.

Notons que l'opérateur B permet de spécifier les propriétés que l'agent croit implicitement être vérifiées. Ceci pose une problème aussi philosophique que pratique : un agent (humain ou pas) peut-il avoir conscience de toutes les conséquences logiques de ses croyances ? comment stoker l'ensemble de ces croyances puisqu'elles sont potentiellement infinies ? Afin de remédier à ce problème, Sadek introduit dans son modèle le concept d'*attention*. Celui-ci permet de distinguer les croyances explicites (celles auxquelles l'agent est attentif) de celles qui sont implicites. Nous ne présenterons pas plus en détail cette notion. Le lecteur intéressé peut se reporter à [Sadek, 1991, pp 42-48], [Sadek, 1992] et [Louis, 2002, p16].

Le Choix : le Choix capture la composante motivationnelle de l'agent (i.e. le point commun entre désirs, préférences, buts, ...). Plus particulièrement, c'est, dans la théorie de Sadek, la notion primitive à l'origine de l'ensemble des attitudes mentales qui motivent la réaction de l'agent (but à réaliser, but persistant, intention, etc.). Cette notion est modélisée en enrichissant le langage $\mathscr{L}$ avec l'opérateur modal C. La notation $C(\phi)$ signifie que « l'agent préfère que la propriété ϕ soit vérifiée à l'instant présent ». C est aussi un opérateur modal normal défini par une axiomatique KD45 quantifiée. De la même manière que pour la croyance, l'opérateur modal C est interprété via les mondes possibles de Kripke par une relation d'accessibilité (notée R_C) : un monde w vérifie la propriété $C(\phi)$ si et seulement si tous les mondes accessibles de w via la relation R_C vérifient la propriété ϕ. Notons que cette relation d'accessibilité est, comme pour la relation d'accessibilité relative à la croyance, sérielle, transitive et euclidienne.

Il est à remarquer que le choix est une notion de très bas niveau. En particulier, elle concerne le monde actuel et non pas, a priori, les mondes futurs. Or, pour qu'un agent puisse agir volontairement, il est nécessaire qu'il veuille transformer l'état actuel de l'univers pour obtenir dans le futur un état différent. Une motivation ne peut donc mener à l'action que si elle vise à ce que les mondes futurs soient différents du monde actuel. Nous verrons par la suite lorsque nous présenterons le concept d'intention de Sadek, comment cette dimension peut être associée aux motivations de l'agent.

c) Modélisation de la dynamique de l'univers

Modèle dynamique utilisé : Sadek modélise la dynamique de l'univers de façon très classique en se basant sur le concept d'*événement primitif*. En particulier, pour lui, l'occurrence d'un événement primitif (et seulement lui) transforme l'univers en le faisant évoluer d'un état à un autre. Deux événements primitifs ne peuvent se dérouler en même temps. Durant l'occurrence d'un événement, l'univers n'est pas observable. L'occurrence d'un événement n'est donc pas directement observable. Le temps résulte de l'occurrence des événements et est donc implicite dans cette modélisation. Enfin, la dynamique de l'univers respecte les contraintes suivantes (voir la figure Figure 1.4) :

- Déterminisme : un même événement ne peut transformer un état de l'univers qu'en au plus un seul autre.
- Non parallélisme : un état donné ne peut être transformé en un autre état donné que par au plus un événement.
- Futur ramifié : il n'y a pas de limite au nombre d'événements qui peuvent transformer un état donné.
- Passé linéaire : un même état ne peut résulter que de l'occurrence d'au plus un événement.

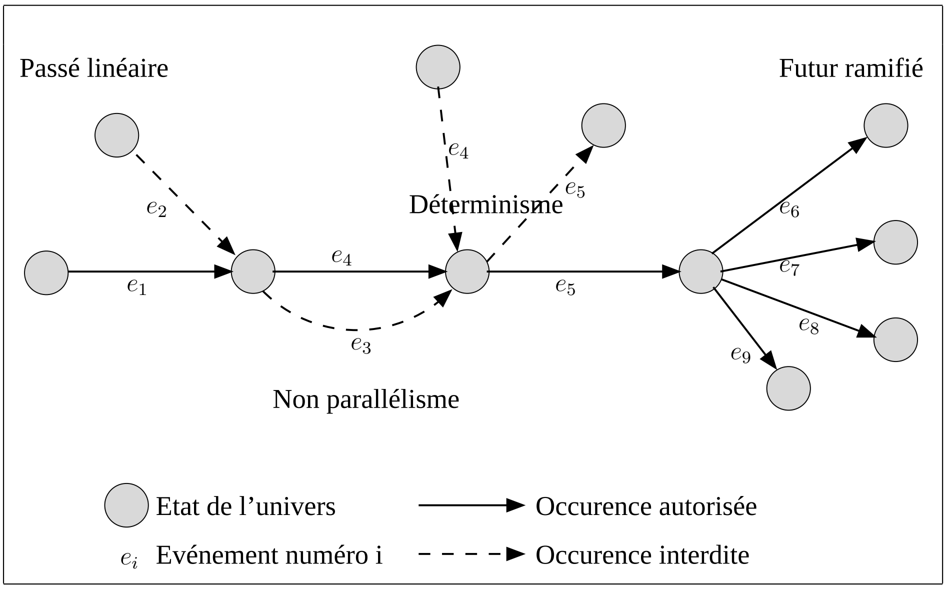


FIG. 1.4 – Modèle du temps utilisé par Sadek

La structure temporelle (arborescente) utilisée par Sadek est la même que celle utilisée par Rao et Georgeff dans [Rao and Georgeff, 1991b]. Néanmoins sa formalisation en est (fort) différente. Rao et Georgeff s'inspirent d'une logique de type CTL (Computation Tree Logic) là où Sadek choisit une logique dynamique [Pratt, 1976] [Harel, 1984].

Formalisation de la dynamique : Plus précisément, la dynamique de l'univers est formalisée en introduisant dans le langage pour chaque événement primitif « e » les opérateurs modaux $Fait(e, -)$ et $Faisable(e, -)$. La formule $Fait(e, \phi)$ signifie que l'évènement primitif e vient juste d'être accompli, avant quoi la propriété ϕ était vraie. De son côté, la formule $Faisable(e, \phi)$ signifie que l'évènement primitif e est réalisable, après quoi la propriété dénotée par ϕ pourra être vraie. L'emploi du verbe « être » pour se référer au passé et du verbe « pouvoir » pour se référer au futur s'explique par les contraintes de passé linéaire et de futur ramifié.

La sémantique de ces opérateurs est formalisée en introduisant pour chaque événement primitif « e », une relation d'accessibilité « R_e » sur les mondes possibles de Kripke. La notation wR_ew' indique que l'occurrence de l'événement e transforme le monde w en w'. Si la formule $Fait(e, \phi)$ est vérifiée dans le monde w_0 (i.e. on a $w_0 \vDash Fait(e, \phi)$) alors il existe un unique monde w qui vérifie la formule ϕ (i.e. on a $w \vDash \phi$) et que l'événement e transforme en w_0 (i.e. on a wR_ew_0). De plus, il n'existe alors pas d'événement « e' » (différent de e) et de monde w' (différent de w) tels que $w'R_{e'}w_0$. Si la formule $Faisable(e, \phi)$ est vérifiée dans le monde w_0 alors il existe un monde w tel que w_0R_ew et tel que $w \vDash \phi$. Notons l'existence d'un événement particulier : l'événement vide (noté $<>$) [Louis, 2002, p30]. Celui-ci relie tout monde possible à lui-même et est donc associé à la relation identité (i.e. pour tout monde w, on a $wR_{<>}w$). Notons aussi que la notation $Possible(\phi)$ est l'abréviation syntaxique de la formule $\exists e,\ Faisable(e, \phi)$.

Les plans : Les événements peuvent être combinés entre eux. Pour plus de clarté nous parlerons alors de plans[6] : tous les événements sont des plans ; si p_1 et p_2 sont des plans alors les constructions $p_1; p_2$ (appelée séquence) et $p_1|p_2$ (appelée alternative) sont des plans. Cette définition ne permet pas de considérer des plans qui s'exécutent en parallèle. Tout au plus, une incertitude sur l'ordre d'exécution des plans p_1 et p_2 peut être représentée de la manière suivante : $p = \big((p_1; p_2)|(p_2; p_1)\big)$. Pour énoncer des faits sur les plans, les modalités $Fait$ et $Faisable$ peuvent être composées afin de traiter les plans. Par exemple : $Fait((e; e'), \phi)$ est équivalent à $Fait(e', Fait(e, \phi))$. De plus, le prédicat $Debute(p_1, p_3)$ est introduit afin d'indiquer qu'il existe un plan p_2 tel que le plan $p_1; p_2$ est identique à p_3.

d) Attitudes mentales composites et équilibre rationnel

Sur la base des modèles de croyance, de choix et d'action présentés précédemment, Sadek construit un ensemble d'attitudes mentales composites dont l'*incertitude* et l'*intention*. Il est intéressant de remarquer que l'incertitude est formellement définie de façon sémantique alors que l'intention est formellement définie de façon syntaxique.

L'incertitude : Le concept d'incertitude revêt une importance particulière dans la modélisation d'agents ayant des perceptions « approximatives » de leur environnement. De

[6]Dans [Sadek, 1991], [Bretier, 1995] et [Louis, 2002] ce que nous appelons des « plans » sont dénommés des « expressions d'actions ».

tels agents doivent en effet se représenter leur environnement comme incertain. Le concept d'incertitude est appréhendé ici de façon globale, c'est-à-dire sans distinguer les différents degrés d'incertitude, même qualitatifs, qu'un agent peut éventuellement adopter à l'égard d'une proposition. Un agent est dit incertain d'une proposition s'il ne croit pas qu'elle est vraie mais qu'il est plus probable que ce soit le cas plutôt que le contraire, autrement dit, s'il a des raisons de douter de sa véracité en ayant davantage de raisons de croire qu'elle est vraie, que de croire qu'elle ne l'est pas [Sadek et al., 1996]. L'Incertitude est modélisée en enrichissant le langage $\mathscr{L}$ de l'opérateur U. La notation $U(\phi)$ signifie que « l'agent n'est pas tout à fait sûr que la propriété ϕ est vérifiée, mais en est plus sûr que du contraire ». Cet opérateur est défini de manière sémantique à partir de la relation d'accessibilité R_B relative à l'opérateur B. Plus précisément, la formule $U(\phi)$ indique que, dans la majorité absolue (mais pas la totalité) des mondes accessibles via la relation R_B, la formule ϕ est vérifiée. Notons que l'incertitude vérifie entre autres les propriétés d'introspection : $\models U(\phi) \Rightarrow B(U(\phi))$ et $\models \neg U(\phi) \Rightarrow B(\neg U(\phi))$. De plus, l'agent peut utiliser ses croyances pour raisonner sur ses incertitudes (et vice-versa) et ainsi obtenir de nouvelles incertitudes. En particulier les schémas d'axiomes suivants sont valides : $\models U(\phi) \wedge B(\phi \Rightarrow \psi) \Rightarrow U(\psi) \vee B(\psi))$ et $\models U(\phi \Rightarrow \psi) \wedge B(\phi) \Rightarrow U(\psi))$. Pour plus de détails, le lecteur intéressé pourra consulter [Louis, 2002, p19-23].

L'intention : L'intention se différencie des autres attitudes mentales sur plusieurs aspects. D'abord, elle entretient un lien privilégié avec la notion d'action et constitue, par ce fait même, l'attitude par laquelle un agent peut déterminer et contrôler son évolution. Par ailleurs, l'intention n'est pas un concept sémantiquement primitif comme c'est le cas dans la majorité des approches de type BDI. C'est un concept composite qui est défini à partir d'autres attitudes mentales et du modèle de la dynamique de l'univers. Un agent a l'intention de réaliser une proposition lorsqu'il (1) choisit d'évoluer vers les futurs où cette proposition, qu'il pense actuellement fausse, sera réalisée, (2) s'engage à maintenir ce choix jusqu'à ce que la proposition soit réalisée ou qu'il ait conclu à l'impossibilité de sa réalisation, et (3) choisit d'agir dans le sens de ce choix [Sadek et al., 1996].

Sadek formalise l'intention en s'inspirant sur le plan méthodologique de [Cohen and Levesque, 1990a]. Nous ne détaillons pas ici ce processus de construction. Le lecteur intéressé pourra se reporter à [Sadek, 1991, pp103-110], [Sadek, 1992] et [Bretier, 1995]. Notons toutefois que l'Intention est modélisée en enrichissant le langage $\mathscr{L}$ de l'opérateur I. La notation $I(\phi)$ se lit : «l'agent a l'intention que, dans le futur, la propriété ϕ soit vérifiée» et signifie, de façon informelle, que

1. l'agent croit que la propriété ϕ n'est pas actuellement vérifiée.
2. l'agent désire de façon persistante que cette formule devienne vérifiée dans le futur, c'est-à-dire, tant qu'il croit que ϕ n'est pas vérifiée et/ou qu'il croit qu'il est possible que ϕ soit vérifiée dans le futur. Remarquons qu'aucune contrainte n'est spécifiée quant à la durée pendant laquelle la formule ϕ doit être vérifiée dans le futur. Cela peut être pour un moment éphémère comme pour toujours.
3. l'agent désire que tout plan conduisant à la réalisation de la propriété ϕ soit réalisé.

L'opérateur I est défini de façon syntaxique (voir [Louis, 2002, p. 41]). Il n'est pas normal et ne peut donc être interprété par une relation d'accessibilité sur les mondes possibles. Nous ne présentons pas ici sa définition. Notons cependant que l'intention a, entre autres, les caractéristiques suivantes :

– Les intentions de l'agent sont consistantes entre elles :

$$I(\phi) \Rightarrow \neg I(\neg\phi)$$

– Comme pour toutes les autres attitudes mentales, l'agent s'introspecte sur ces intentions ; il est conscient de ses intentions :

$$I(\phi) \Leftrightarrow B(I(\phi))$$

$$\neg I\phi \Leftrightarrow B(\neg I(\phi))$$

– L'intention ne peut porter que sur des propriétés qui sont actuellement fausses d'après les croyances de l'agent :

$$I(\phi) \Rightarrow B(\neg\phi)$$

– L'intention présuppose l'existence du désir de voir la propriété se réaliser dans le futur.

$$I(\phi) \Rightarrow C(\exists p, Faisable(p, \phi))$$

L'intention de Sadek diffère de celle de Rao et Georgeff. Pour Rao et Georgeff un agent s'engage à mettre en œuvre lui-même ses intentions alors que pour Sadek, un agent s'engage à mettre en œuvre tout ce qui mène à la réalisation de ses intentions. À cet égard, l'intention de Sadek se rapproche plutôt d'un intermédiaire entre les notions de but et d'intention du modèle de Rao et Georgeff.

L'équilibre rationnel : Pour Sadek, les attitudes mentales de l'agent entretiennent de nombreuses relations entre elles. Nous avons vu quelques-unes de ces relations dans les paragraphes précédents consacrés à l'intention et à l'incertitude. À chaque fois, elles avaient la particularité de relier des attitudes composites et surtout d'être obtenues par construction. Pour Sadek, il existe aussi des relations entre les attitudes mentales primitives. C'est ce qu'il appelle l'équilibre rationnel qu'il modélise en introduisant quelques axiomes supplémentaires. Nous ne les exposons pas ici mais notons que dans cette théorie, par exemple :

– Les Choix et les Croyances de l'agent sont consistants entre eux (Contrainte de réalisme) :

$$B(\phi) \Rightarrow \neg C(\neg\phi)$$

– L'agent connaît ses Choix et ses non-Choix (Introspection des Choix) :

$$C(\phi) \Leftrightarrow B(C(\phi)) \quad \text{et} \quad \neg C(\phi) \Leftrightarrow B(\neg C(\phi))$$

– L'agent pense que le monde évoluera conformément à ce qu'il projette (Futur auto-réflexif) :

$$B\neg Faisable(e, \neg(B(\phi) \Rightarrow \phi))$$

– Si un agent désire qu'une propriété ϕ se réalise, il désire alors aussi croire que cette propriété est réalisée :

$$C(\exists p, Faisable(p, \phi)) \Rightarrow C\exists p', Faisable(p', B\phi))$$

e) L'agent et son environnement :

Le langage logique dont nous venons de donner un aperçu permet de décrire l'état mental d'un agent. Afin de spécifier un comportement pour cet agent, il est nécessaire de déterminer les relations qu'entretiennent les attitudes mentales entre elles et avec l'environnement de l'agent. En effet, l'agent peut être vu comme une base de formules[7]. Celle-ci évolue par le biais de deux mécanismes : *l'observation* et la *déduction logique* sur la base des axiomes de la théorie (voir la figure 1.5). Cette base permet d'engendrer la réaction de l'agent via un mécanisme algorithmique externe à la théorie. Plus précisément, un mécanisme algorithmique parallèle basé sur l'ensemble des formules (initiales et déduites) permet de déterminer et de mettre en œuvre la réaction de l'agent.

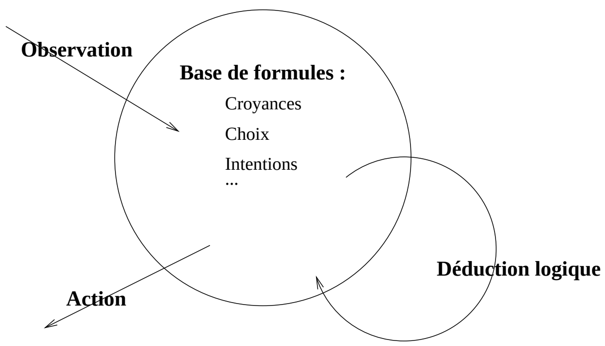


FIG. 1.5 – Schéma général et simplifié de l'évolution des formules du modèle de Sadek

L'observation : Pour Sadek, le phénomène d'observation est le moyen par lequel l'agent perçoit physiquement les changements de son environnement. C'est donc un des moyens par lesquels l'agent acquiert de nouvelles connaissances. Ce phénomène est modélisé en introduisant dans le langage $\mathscr{L}$ les prédicats *Observe* et *Realise* (voir [Sadek, 1991, pp145-148], [Bretier, 1995, pp45-49] et [Louis, 2002, pp42-44]). La notation $Observe(o)$ indique que l'agent vient juste d'observer l'entité observable o via l'un de ses capteurs. La notation

[7]Cette base de formules contient tous les types de formules décrites plus haut : des Croyances, des Choix, des Intention, ...

$Realise(o, p)$ indique que la réalisation du plan p est une interprétation possible de l'observation de l'entité observable o. Pour Sadek, l'observation repose sur les quatre postulats suivants :

1. Dès lors qu'un *événement primitif* non vide se produit dans l'environnement de l'agent, ce dernier observe quelque chose :

$$[\exists o, Observe(o)] \Leftrightarrow [\exists e \neq <>, Fait(e, \top)]$$

2. L'observation est une expérience sûre dans le sens où l'agent sait s'il a ou n'a pas observé quelque chose :

$$\forall o,\ [Observe(o) \Leftrightarrow B(Observe(o))]$$

3. L'observation peut être non conforme à la réalité ; le schéma suivant est satisfiable :

$$Observe(o) \wedge \forall p, (Realise(o, p) \Rightarrow \neg Fait(p, \top))$$

4. La cohérence des observations faites par un agent est une condition nécessaire et suffisante à la réussite subjective de son interaction avec son environnement.

La réaction de l'agent : La réaction de l'agent est spécifiée dans la théorie de l'interaction rationnelle par une axiomatique ainsi que par un mécanisme algorithmique. D'un côté, l'axiomatique construit la *chaîne causale d'intentions* à partir des intentions initiales de l'agent tout en restant homogène au reste de la théorie. De l'autre côté, l'algorithmique externe à la théorie exploite la chaîne causale d'intentions et la structure de son inférence pour déterminer le plan d'action correspondant à mettre en œuvre par l'agent.

Plus précisément, Sadek introduit dans sa théorie des axiomes de planification qui permettent à l'agent de dériver de nouvelles intentions (appelées aussi sous-buts) afin d'obtenir une intention directement réalisable par un plan. Ces déductions exploitent les intentions déjà déduites ainsi que les actions que l'agent a à sa disposition. De façon schématique, ces axiomes stipulent que sous certaines conditions, si l'agent a l'intention de ϕ, alors il a également l'intention de ψ. En particulier, si l'agent à l'intention que la propriété ϕ soit réalisée, alors il a l'intention que l'un des plans menant à sa réalisation soient aussi réalisé : pour toute formule ϕ et pour l'ensemble des plans p_i tel que p_i a pour effet ϕ et tel que l'agent ne « choisit » pas comme non possible son accomplissement :

$$\text{Si } \neg C \neg Possible(Fait(p_i)), \quad \text{alors} \quad I(\phi) \Rightarrow I(Fait(p_1|p_2|...|p_n))$$

Cet ensemble d'intentions (la première et toutes celles déduites) est appelé la *chaîne causale d'intentions*. C'est sur cette *chaîne causale d'intentions* que repose l'algorithmique externe qui détermine la réaction de l'agent à chaque instant. Plus précisément, ce mécanisme externe détermine l'ordre des actions à effectuer en se basant sur l'ordre des déductions durant le raisonnement et simplifie le plan qui en résulte. Dès à présent nous pouvons donc entrevoir une limitation de ce formalisme. Alors que Sadek s'est efforcé de définir l'ensemble de la mécanique interne de l'agent de façon logique et homogène, l'étape finale est repoussée à un mécanisme externe. En conséquence et de façon paradoxale, le formalisme logique de Sadek ne permet pas d'identifier en général la réaction effective de l'agent. Nous reviendrons plus en détail sur ce problème dans la section suivante.

1.2 Une limitation : le choix de la réaction de l'agent

La section précédente nous a permis d'exposer dans les grandes lignes les modèles cognitifs d'agents. Ceux-ci sont particulièrement attrayants car ils proposent de modéliser chaque entité d'un système informatique global de façon abstraite mais néanmoins intuitive grâce à des attitudes mentales comme la croyance ou le désir. Plus particulièrement, nous avons exposé le modèle proposé dans [Rao and Georgeff, 1991b] ainsi que le modèle d'agent à la base de la théorie de l'interaction rationnelle [Sadek, 1991]. Dans ces deux modèles, l'intention (même si elle a des sens différents) joue un rôle important dans la dynamique de l'état mental de l'agent en lui permettant de focaliser son raisonnement [Bratman, 1987]. À cet égard, ces deux approches permettent de modéliser des agents aux ressources limitées.

Comme nous le verrons dans cette section, ces modélisations sont (telles quelles) inadaptées pour traiter de façon satisfaisante les situations plus complexes que celles où l'agent a une seule intention et un seul plan la réalisant. Aussi, après avoir explicité la raison que nous pensons être à l'origine de cette inadaptation, nous verrons comment Louis propose de résoudre ce problème dans [Louis, 2002]. Enfin nous conclurons en soulignant que la mise en oeuvre de la proposition de Louis nécessite des informations supplémentaires : des préférences.

1.2.1 Situations adaptées aux formalismes classiques

Ces modèles permettent de spécifier (de façon déterministe) des systèmes où chaque agent a, à chaque instant, une seule intention et un unique savoir-faire lui indiquant comment la satisfaire (voir le schéma de la figure 1.6). Plus précisément, en reprenant le formalisme de Sadek, cela implique qu'à chaque instant et pour chaque agent il existe un seul couple (ϕ, p) tel que l'état mental de cet agent vérifie la formule $I(\phi) \wedge B(Faisable(p, \phi))$. Dans de telles situations la réaction de cet agent est, sans ambiguïté, de mettre en oeuvre ce savoir-faire et donc ici le plan p.

Considérons par exemple un agent fournissant des noms et adresses de restaurants. Si à la suite d'une interaction, il a l'intention de donner à un utilisateur l'adresse d'un restaurant servant du poisson et si il sait seulement que le restaurant « le Goéland » sert du poisson et qui connaît son adresse, alors, il indiquera de façon naturelle à l'utilisateur l'adresse du restaurant « le Goéland ». L'agent aura cette réaction car ainsi il satisfera son intention (c'est pourquoi il est rationnel) et surtout car cet agent n'a pas d'autre savoir-faire à sa disposition pour satisfaire son intention (c'est pourquoi il est déterministe). Ainsi, c'est parce que l'agent connaît un unique savoir-faire répondant à son intention, qu'il agit sans ambiguïté (à un défaut de ses effecteurs près) en le mettant en oeuvre.

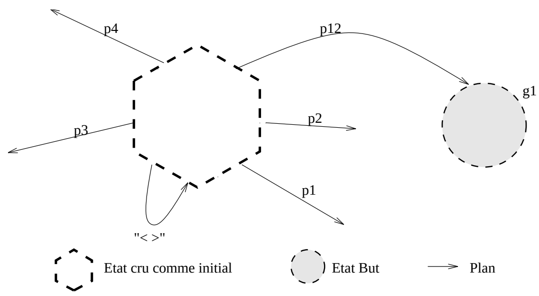


Fig. 1.6 – Cas 1 : une intention, un plan pour la réaliser

1.2.2 Situations non adaptées aux formalismes classiques

Si les formalismes logiques classiques d'agents rationnels sont attrayants, c'est surtout dans l'optique de modéliser des applications complexes. Or, qui dit applications complexes, entend généralement par là des applications dans lesquelles la situation précédente est rare. Plus précisément, il est a priori certain que l'agent connaîtra plusieurs plans satisfaisants son intention (voir la figure 1.7) ou aura plusieurs intentions et la connaissance d'au moins un plan pour satisfaire chacune d'elles (voir la figure 1.8). Par exemple, il est probable que notre agent dans la situation précédente connaisse d'autres restaurants servant du poisson ainsi que leurs adresses. Que doit-il faire ? Quelle adresse doit-il communiquer à l'utilisateur ? Nous sommes forcés de constater que les formalismes actuels et donc les

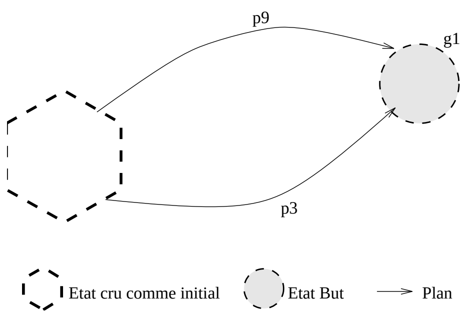


Fig. 1.7 – Cas 2 : une intention, deux plans pour la réaliser

implémentations actuelles ne peuvent traiter correctement de telles situations. Par exemple, dans *Jadex*, un agent peut avoir plusieurs buts mais il les traite indépendamment les uns des autres. Plus précisément, pour chacun de ses buts, un agent Jadex détermine un plan qui permet de le réaliser. Ces derniers sont alors sélectionnés à tour de rôle pour être exécutés étape par étape pour avoir une impression de parallélisme (coroutinage). Comme chaque plan est déterminé indépendamment des autres et est exécuté par étapes ceci peut entraîner des blocages ou de réaliser de nombreuses fois la même chose.

Un système informatique, quelle que soit la façon dont il est modélisé, ne peut mettre en œuvre qu'une unique réaction à chaque instant. Les modèles cognitifs proposent pour

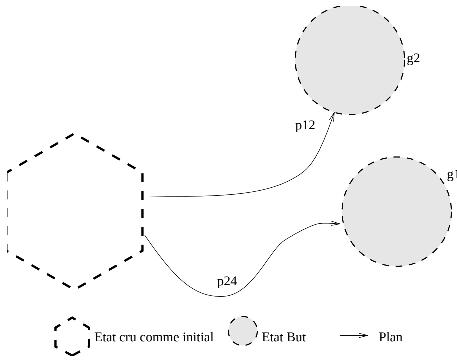


FIG. 1.8 – Cas 3 : deux intentions, un plan différent pour réaliser chacune d'elles

leur part et avant tout d'appréhender chaque agent comme une entité qui perçoit son environnement et qui agit sur celui-ci selon le principe de rationalité : il met en œuvre à chaque instant la réaction qu'il juge la meilleure étant donné ses connaissances. Dans les modèles que nous avons présentés dans la section 1.1 et plus généralement dans de nombreux modèles cognitifs d'agents, la mise en œuvre du principe de rationalité reprend l'interprétation de Newell :

> « If an agent has knowledge that one of its action will lead to one of its goals, then the agent will select that action. » [Newell, 1980, p.102]

Par conséquent, dans ces modèles, les réactions possibles ont le même intérêt à être réalisées (du point de vue de l'agent) dés lors qu'elles satisfont chacune au moins une des intentions de l'agent. Il en résulte que l'agent peut avoir plusieurs réactions possibles à sa disposition et ne pas savoir les départager. Ces modèles ne permettent donc pas toujours de spécifier complètement la réaction de l'agent à un instant donné (voir par exemple [Doyle, 1994]). De plus, comme l'agent doit, quoi qu'il arrive, mettre en œuvre une réaction, dans ces cas là c'est un mécanisme externe à l'agent qui détermine précisément la réaction de l'agent. Par conséquent, en plus de ne pas pouvoir toujours « choisir » jusqu'au bout sa réaction, l'agent ne « sait » pas forcément toujours précisément quelle réaction il met en œuvre.

Pourtant, dans de nombreuses situations ils est important que l'agent puisse choisir précisément sa réaction et ainsi savoir explicitement ce qu'il fait. En particulier, il semble a priori intéressant de départager des alternatives qui satisfont des intentions différentes, et/ou qui sont chacune meilleure que les autres au regard de points de vue différents. C'est fréquemment le cas lorsque l'agent a plusieurs motivations (ou intentions) et/ou connaît plusieurs moyens de les satisfaire (plans). Par exemple, dans le cadre d'une interaction multimodale, un agent qui a l'intention de fournir un itinéraire à un utilisateur doit aussi choisir par exemple entre la modalité vocale (en énonçant les instructions) et la modalité graphique (en affichant un plan). Sur la base des formalismes présentés dans la section 1.1, ces deux réactions ne peuvent être départagées car elles répondent toutes les deux à son intention.

Pour nous, cette limitation provient du fait que l'intention, même si elle guide le raisonnement de l'agent (et donc le motive à agir), ne définit pas a priori sans ambiguïté la réaction à mettre en oeuvre. Ce n'est pas un problème dû à une mauvaise définition de l'intention. C'est un problème intrinsèque aux modèles qui voient dans l'intention le moyen de départager *in fine* les réactions de l'agent. En effet, il faut se rappeler qu'« originellement », l'intention n'a été introduite que comme un moyen de réduire le nombre d'alternatives à prendre en compte afin de rendre le raisonnement faisable :

> « The formation of intentions and the commitments thereby entailed are seen as a mechanism - possibly one amongst many - for constraining the set of options about which an agent must reason.» [Bratman, 1987]

Par conséquent, l'intention est une heuristique (parmi d'autres possibles) permettant de filtrer les alternatives. Ceci est nécessaire car un agent ne peut considérer l'ensemble des plans imaginables. Cependant, bien des fois, ce n'est pas suffisant pour déterminer explicitement et sans ambiguïté la réaction de l'agent. Plus précisément, pour nous ce problème provient du manque d'une phase de décision explicite dans ces formalismes.

1.2.3 Notre vision du cycle d'interaction d'un agent avec son environnement

Comme toute application doit produire, à chaque instant un résultat unique, un agent, à chaque instant, (1) doit agir sur son environnement en fonction de ses observations et (2) ne peut mettre en œuvre qu'une seule réaction[8]. Plus précisément, un agent rationnel doit faire correspondre à toute observation une unique réaction adaptée.

Les intentions d'un agent et plus généralement ses buts créent une motivation pour agir. Il va de soi que sans intention, pas de réaction. Par contre, l'opposé n'est pas vrai, c'est-à-dire que l'intention n'est pas une réaction. Par exemple, si un agent ne sait pas comment satisfaire une intention, il ne peut rien entreprendre pour y répondre et doit attendre de trouver une « idée » débloquante. À l'autre extrême, si un agent connaît plusieurs plans différents pour satisfaire une intention, comme il ne peut pas tous les exécuter, il doit en choisir un.

Les plans, lorsqu'ils sont connus de l'agent, lui fournissent des solutions pour tenter d'atteindre ses buts. Ce sont des « savoir-faires » ou des « recettes ». Ceux-ci peuvent faire partie des connaissances initiales de l'agent, ou bien dériver d'un raisonnement. De plus, ils peuvent tout aussi bien être des plans qui ne réalisent pas ses propres buts que des plans qui les réalisent.

Enfin, un agent agit concrètement en appliquant un plan qu'il sélectionne parmi tous ses savoir-faires qui réalisent un ou plusieurs de ses buts suivant un critère d'optimalité donné.

Le cycle d'interaction d'un agent cognitif sur son environnent peut donc être décomposé en quatre phases : l'observation, l'adoption d'Intentions, la planification et la décision.

[8] Une réaction peut être plus ou moins compliquée. En particulier elle peut être constituée de plusieurs actions en parallèle.

- L'*observation* est le mécanisme qui permet à l'agent de modifier ses croyances et d'en acquérir de nouvelles. C'est ce qui permet donc de déterminer la situation courante.
- L'*adoption d'Intentions* permet à l'agent de déterminer ses motivations et plus particulièrement les situations qu'il s'engage à atteindre. C'est grâce à cette étape que l'agent peut focaliser l'utilisation de ses ressources.
- La *planification* permet à l'agent de déterminer les façons qu'il estime pouvoir lui permettre de satisfaire ses intentions à partir notamment de ses savoir-faires. Ce sont les *réactions possibles*.
- Enfin, la *décision* permet à l'agent de départager parmi toutes les réactions possibles précédemment identifiées, « la meilleure » et donc de déterminer quelle réaction effectivement mettre en œuvre.

À notre connaissance, dans la littérature, ce sont surtout les deux premières phases qui ont été traitées dans les travaux en rapport avec les modèles cognitifs d'agents. La planification et la décision y sont alors souvent laissées à des mécanismes externes, lorsqu'elles existent. À part dans [Brafman and Tennenholtz, 1997], la phase de décision d'un agent n'est généralement pas considérée en tant que telle. Pourtant, Bratman, Israel et Pollack dans l'introduction de [Bratman et al., 1988] ont insisté sur le fait qu'une architecture d'agent doit lui permettre entre autres de comparer et d'évaluer les alternatives qui s'offrent à lui :

> « When we attempt to build artificial agents that are capable of making rational decisions, we likewise need to provide them with techniques for evaluating the options they encounter. »[Horty and Pollack, 1998]

Pollack a même remarqué dans [Georgeff et al., 1999] que des idées comme *la maximisation de l'utilité* seraient intéressantes à considérer pour la mise en œuvre d'une phase de décision dans un formalisme logique d'agent rationnel. Il en a été de même pour Horvitz qui, dix ans plus tôt, a insisté dans [Horvitz et al., 1988]sur la nécessité d'introduire une phase de décision dans les systèmes experts.

1.2.4 Le formalisme de [Louis, 2002]

La théorie de l'interaction proposée par Sadek présente les mêmes limites. En effet, comme nous l'avons remarqué au point (e) de la section 1.1.3, dans cette théorie, contrairement aux observations qui lui arrivent de ses capteurs, l'action (ou la réaction) que produit un agent sur ses effecteurs n'est pas spécifiée entièrement. Plus précisément, il manque un lien (au sein de la théorie logique) entre les plans calculés à partir des intentions de l'agent et l'action qu'il entreprend effectivement sur son environnement. En particulier, quel plan choisir s'il en existe plusieurs pour satisfaire un même but ? Ou si l'agent poursuit deux buts distincts en même temps ? Même question si aucun plan ne peut être trouvé. De plus, du fait même du mécanisme externe à la théorie, l'agent ne sait pas en général quelle est sa réaction ! C'est pour cela que Louis dans [Louis, 2002] a proposé des évolutions de la théorie de Sadek afin de définir explicitement la façon dont la réaction de l'agent est déterminée et ainsi lui permettre un raisonnement plus fin sur l'action.

Pour Louis, la réaction de l'agent est l'expression de son engagement à faire *immédiatement* un ensemble d'actions qu'il juge susceptibles de mener à la satisfaction d'une de ses intentions. Pour lui, la réaction d'un agent se détermine de façon naturelle et univoque en fonction de son état mental. En particulier, une théorie logique modélisant un agent rationnel doit spécifier un raisonnement qui fait correspondre une unique réaction adaptée à toute observation. De plus, pour lui il est utile de faire explicitement et nettement la distinction entre les *buts*[9], les *plans* et la *réaction* d'un agent rationnel. Ceci lui permet de faire apparaître deux processus complémentaires : le calcul de plan lui-même à partir des buts (i.e. la planification [Allen et al., 1991]), et la sélection de la réaction à produire sur l'environnement en fonction des plans disponibles (i.e. la décision [Doyle and Thomason, 1999][Kast, 2002]). Pour lui, le calcul de plans et le calcul de la réaction (ce qu'il appelle une *stratégie*) sont proprement découplés (voir la figure 1.9). Alors que dans [Louis, 2002]

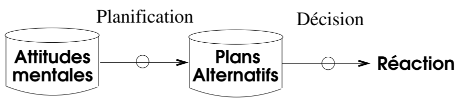


FIG. 1.9 – Lien général et simplifié entre état mental et réaction

il s'occupe surtout du premier mécanisme, nous cherchons pour notre part dans cette thèse à traiter le second.

À cet effet, Louis dans [Louis, 2002, pp77-84] introduit la notion de stratégie[10] dans la théorie de l'interaction rationnelle. Elle s'inspire des travaux de Singh [Singh, 1994] et van der Hoek [van der Hoek et al., 1999]. Plus précisément, Louis ajoute le prédicat « Agit » au vocabulaire de la théorie de l'interaction et en complète son axiomatique en conséquence.

a) La notion de Stratégie

Pour Louis, la stratégie d'un agent rationnel est un plan qu'il choisit parmi tous les plans qu'il sait satisfaire un but en fonction d'un critère d'optimalité. Pour formaliser ceci, il propose de spécifier explicitement dans le formalisme de la théorie de l'interaction rationnelle l'action que l'agent entreprend effectivement en introduisant le prédicat « Agit ». Ce dernier caractérise le plan que l'agent entreprend. Plus précisément, la formule $Agit(p)$ signifie que « l'agent est censé[11] agir sur son environnement selon le plan p ». En d'autres termes, $Agit(p)$ signifie que l'agent transmet à ses effecteurs le plan p.

[9]Notons ici que le mécanisme proposé par Louis est basé sur la notion plus générique de but et non d'intention.

[10]On pourrait croire, étant donné que l'intention de Sadek a un sens d'engagement « pratique » plus faible que celle de Rao et Georgeff, que les stratégies de Louis correspondent aux intention de Rao et Georgeff. Il n'en est rien car la stratégie de Louis indique que l'agent *met* en oeuvre le plan p alors que l'intention de Rao et Georgeff indique que l'agent *va tout faire pour mettre* en oeuvre le plan p.

[11]Le terme « est censé agir », légèrement plus faible que « agit », permet de tenir compte des cas où les effecteurs de l'agent présentent un dysfonctionnement qui l'empêche d'exécuter convenablement la stratégie de l'agent.

Cette vision des choses rappelle naturellement le prédicat *Observe* qui rend compte de la perception de l'agent. De même que la formule $Observe(o)$ indique l'objet observation en provenance de ses capteurs, la formule $Agit(p)$ indique l'objet réaction à destination de ses effecteurs. Le prédicat *Agit* boucle ainsi la chaîne d'interaction de l'agent avec son environnement.

En schématisant un peu, si l'observation o induit l'intention $I(\phi)$, et que l'agent sait que le plan p est faisable et peut mener à la satisfaction de ϕ et que de plus, le fait que l'agent mette en œuvre le plan p n'est pas en contradiction avec le reste de l'état mental, alors l'agent mettra en œuvre le plan p (i.e. la formule $Agit(p)$ est vraie). Comme nous le verrons à la fin de la prochaine section, s'il existe plusieurs plans de ce type alors le formalisme proposé par Louis permet à l'agent de mettre en oeuvre celui qu'il trouve le moins mauvais. Remarquons ici que Louis utilise des défauts (au sens de Reiter [Reiter, 1980]) dans la phase consistant à choisir la réaction à mettre en oeuvre : l'agent met en œuvre la réaction la moins mauvaise sachant que la réaction vide et toujours réalisable et est plus mauvaise que n'importe quelle réaction susceptible de mener à la satisfaction d'une de ses intentions.

b) Axiomatisation

Pour Louis, la stratégie d'un agent respecte de nombreuses propriétés. En particulier, un agent a toujours une (et une seule) stratégie dont le premier pas est déterministe. Afin de formaliser cette vision, Louis introduit toute une série d'axiomes relatifs au prédicat *Agit* dans la théorie de l'interaction.

Existence d'une stratégie : Un agent est avant tout en interaction permanente avec son environnement. Il observe sans arrêt son environnement afin de pouvoir agir dessus à tout moment. Côté perception, cela se manifeste par l'existence d'une observation chaque fois que se produit un évènement (voir le point (e) de la section 1.1.3). Symétriquement, côté action, cela se manifeste par l'existence permanente d'une réaction sélectionnée par l'agent, même si cette réaction traduit une attitude passive telle qu'une simple attente. Formellement, cela signifie que le prédicat Agit doit être vérifié par au moins un plan exécutable dans tout modèle de la théorie d'agents rationnels. C'est le plan que l'agent commande d'exécuter à ses effecteurs à un instant donné. Un agent doit tout le temps savoir ce qu'il est en train de faire, même si cela se réduit à ne rien faire. Dans le cas contraire, si aucune stratégie ne peut être dérivée, cela correspond à une situation dans laquelle l'agent perd le contrôle de son évolution. C'est pour cela que par défaut le plan vide est la stratégie de l'agent.

$$\top : Agit(<>) \rightarrow Agit(<>)$$

Déterminisme des stratégies : À chaque instant l'agent a une unique stratégie et celle-ci est exécutable, ce qui signifie que son premier pas doit être déterministe. Une stratégie est donc soit l'évènement vide $<>$, soit un plan de la forme $e; p$, ou e désigne un évènement primitif non vide et p un plan quelconque :

$$Agit(a) \wedge Agit(b) \Rightarrow (a = <> \wedge b = <>) \vee (\exists e,\ e \neq <> \wedge Debute(e, a) \wedge Debute(e, b))$$

Attitude mentale : L'adoption d'une stratégie rationnellement optimale par un agent est un processus mental subjectif sûr, au même titre que l'observation (voir le point (e) de la section 1.1.3). L'action qu'un agent entreprend doit être identique à l'action qu'il pense entreprendre. L'axiome suivant impose cet accord entre état mental et réaction :

$$Agit(a) \Leftrightarrow BAgit(a)$$

Stratégies et savoir-faires : Un agent rationnel sélectionne sa stratégie parmi un ensemble de plans qu'il juge réalisables au sens de l'opérateur $Faisable$:

$$Agit(a) \Rightarrow BFaisable(a)$$

Stratégies et buts : Un agent sélectionne une stratégie pour tenter de satisfaire ses buts. Plus précisément, dès que l'agent pense qu'un plan réalise potentiellement un de ses buts, il l'adopte par défaut comme stratégie. Si son raisonnement l'amène par la suite à préférer une autre stratégie, il abandonnera sa stratégie par défaut au profit de cette nouvelle solution :

$$CPossible(\phi) \wedge BFaisable(a, \phi) \wedge (\exists e, e \neq <> \wedge Debute(e, a)) : Agit(a) \rightarrow Agit(a)$$

Unicité d'une stratégie : Étant donné qu'un agent a une unique stratégie à chaque instant, il ne peut y avoir plusieurs défauts de la forme précédente déclenchés simultanément. Par conséquent, la théorie résultante risque d'admettre plusieurs extensions, chacune correspondant au choix d'une stratégie différente. Ce type de situation revient, en somme, à repousser la détermination de l'action d'un agent à l'extérieur du formalisme. Pour l'éviter, il faut que le concepteur d'agents s'efforce de spécifier le critère d'optimalité rationnelle sur lequel se fonde la décision de l'agent, par une famille d'axiomes que Louis baptise O (O pour calcul d'Optimalité). Chacun de ces axiomes est destiné à éliminer de façon catégorique les stratégies jugées non optimales pour que les défauts correspondants restent bloqués. Cette famille doit au moins contenir l'axiome suivant, qui gère la balance entre la stratégie vide, imposée par défaut dans le cas où aucun plan connu de l'agent ne coïncide avec ses buts, et une stratégie déclenchée par défaut si elle amène à la satisfaction d'un but. Formellement, pour toute formule ϕ non équivalente à $\top$ et tout plan non vide a :

$$CPossible(\phi) \wedge BFaisable(a, \phi) \wedge (\exists e, e \neq <> \wedge Debute(e, a)) \Rightarrow \neg Agit(<>)$$

Les autres axiomes se basent sur un ordonnancement des plans, noté $<$, tel que la formule $a < b$ indique que le plan a est jugé moins mauvais que le plan b. Ils spécifient que s'il existe deux plans et que l'un est moins bon que l'autre, alors le moins bon n'est pas une stratégie. Plus précisément, ces axiomes sont de la forme : Pour toutes formules ϕ et ψ différentes de $\top$

$$B\left[\left(\begin{array}{ll} & CPossible(\phi) \wedge Faisable(a,\phi) \\ \wedge & CPossible(\psi) \wedge Faisable(b,\psi) \\ \wedge & (\exists e, e \neq <> \wedge Debute(e,a)) \\ \wedge & (\exists e, e \neq <> \wedge Debute(e,b)) \\ \wedge & (a < b) \end{array}\right) \Rightarrow \neg Agit(b)\right] \tag{1.1}$$

1.2.5 Conclusion

La plupart des modèles logiques d'agent ont comme inconvénient majeur de ne pas formaliser entièrement la réaction de l'agent à chaque instant. Il en résulte qu'un agent ainsi modélisé est, la plupart du temps, incapable de « choisir » entre des alternatives que le sens commun a peu de mal à départager et surtout ne sait pas explicitement quelle réaction il est en train de mettre en œuvre.

C'est pourquoi, Louis a développé dans [Louis, 2002] des moyens entièrement logiques permettant de calculer explicitement le plan que l'agent met en oeuvre (i.e. la réaction de l'agent). Son travail se base entre autres sur l'idée qu'il faut distinguer nettement les notions de *but* (qui recouvre celle d'intention), de *savoir-faire* (i.e. les plans connus de l'agent) et de *réaction* (i.e. l'action physiquement exécutée) et qu'il faut donc dissocier la phase de planification de celle de décision. Plus précisément, Louis a enrichi la théorie des agents rationnels de Sadek avec entre autres (1) une méthode logique permettant de calculer des plans par raisonnement dans l'espace des états via une fonction de régression, ainsi qu'avec (2) la notion de stratégie et un jeu associé d'axiomes afin d'établir un cadre formel pour exprimer la sélection de la réaction de l'agent. Nous ne présentons pas dans ce document la méthode logique de calcul de plans. C'est néanmoins à notre avis l'apport majeur de Louis. En ce qui concerne la sélection de la réaction de l'agent proprement dite (i.e. la phase de décision) nous venons de présenter le cadre formel proposé par Louis. Il y laisse néanmoins libre la spécification « précise » de l'ordonnancement des plans en remarquant qu'elle est indispensable.

L'ordonnancement des plans ne peut se faire qu'en ayant des informations supplémentaires. C'est ce que nous appelons les *préférences* de l'agent. La question est de savoir comment modéliser ces préférences et surtout comment les obtenir. Pour notre part, nous ne pensons pas qu'il faille chercher les informations nécessaires dans les intentions ou dans toute autre composante de l'agent de type motivationnelle. Plus précisément, à la manière de Doyle, nous pensons que les préférences ne peuvent pas être obtenues en raisonnant sur les motivations de l'agent.

> « One cannot define the notion of preference and utility purely in terms of beliefs and goals, for these are independent notions ». [Doyle, 1994, p.5]

Pour nous, les préférences sont d'une autre nature et ont un autre intérêt. Alors que les motivations de l'agent lui permettent de focaliser son raisonnement, les préférences lui permettent d'identifier, de son point de vue, la (ou une des) meilleure(s) réaction(s). Par conséquent, pour qu'un agent soit rationnel il faut aussi qu'il ait des préférences sur les résultats possibles de ses actions, et qu'il agisse de manière à optimiser les résultats en accord avec ses préférences [Doyle, 1994].

Notons ici que la notion de préférence (en particulier à travers le concept d'utilité) a déjà été abordée dans la littérature relative aux modèles cognitifs d'agent. Néanmoins, elle sert souvent à définir des notions relatives aux motivations de l'agent (i.e. désirs, buts et intentions) [Dastani et al., 2002b] [Lang et al., 2002] [Haddawy and Hanks, 1993]. Nous verrons dans les sections suivantes comment les préférences sont modélisées dans la littérature (voir chapitre 2) et comment elles peuvent être modélisées afin d'être adaptées à un agent rationnel (voir chapitre 3).

Chapitre 2

Prise de décision

Le chapitre précédent nous a permis de présenter les modèles logiques d'agents rationnels basés sur le concept d'attitudes mentales. Plus particulièrement, nous y avons présenté les formalisations proposées par Rao et Georgeff dans [Rao and Georgeff, 1991b] et par Sadek dans [Sadek, 1991]. Par la suite, nous avons remarqué que ces formalismes ne permettent de traiter de façon satisfaisante que les situations où l'agent a une intention et un seul savoir-faire pour la satisfaire. Nous avons alors présenté la proposition de Louis dans [Louis, 2002] pour surpasser cette limitation. Plus précisément, Louis a introduit une phase de décision explicite dans le formalisme de Sadek afin de déterminer précisément la réaction de l'agent au sein du formalisme logique d'agent.

Cette phase de décision se base sur un ordonnancement des alternatives que l'agent peut mettre en oeuvre. Malheureusement, Louis, dans sa proposition, a laissé à de futurs travaux la spécification de cet ordonnancement. Celui-ci ne peut se baser que sur des informations qui ne sont pas déjà présentes dans ces formalismes : les préférences [Doyle, 1994]. C'est pourquoi, afin de définir de manière homogène aux théories un tel ordonnancement, il est nécessaire de définir de manière homogène aux théories, les outils permettant de représenter et de manipuler ces préférences. Aussi, nous allons présenter dans ce chapitre les travaux visant à représenter et manipuler des préférences : les travaux concernant la prise de décision.

Plus précisément, après avoir remarqué dans la section 2.1 qu'une décision s'inscrit dans le temps et plus particulièrement dans ce qui est usuellement appelé *le processus de décision*, nous remarquerons que les informations nécessaires à une telle étape peuvent être représentées numériquement ou logiquement. Par la suite nous présentons dans la section 2.2 les deux modélisations numériques de la préférence sur lesquelles sont basés les formalismes généralement employés pour mettre en oeuvre une décision. Puis, nous présenterons dans la section 2.3 les modélisations logiques de la préférence. Enfin, nous conclurons ce chapitre en montrant pourquoi ces dernières modélisations sont plus adaptées aux formalismes logiques d'agents rationnels que ne le sont les modélisations numériques.

Notons dès à présent que les sections 2.1 et 2.2 s'inspirent de manière substantielle de [Hansson, 1994], de [Kast, 2002], ainsi que de [Domshlak, 2002, pp9-30], [Ha, 2001, pp4-23]

[Slantchev, 2005], [Bouyssou and Vincke, 2003], [Fishburn, 1970] et [Fishburn, 1999].

2.1 La Théorie de la Décision

Depuis Condorcet (1743-1794), de nombreuses études ont été menées dans l'optique de décrire, d'évaluer, de comprendre, de prévoir, de faciliter et plus récemment d'automatiser les prises de décisions. La philosophie, l'économie, les mathématiques et plus récemment l'informatique sont autant de sciences qui ont contribué à ce développement. Il en résulte que nombreuses sont les idées et techniques que l'on retrouve dans des champs aussi variés que la biologie, les sciences sociales ou la politique.

Même s'il n'existe toujours pas de consensus sur la façon dont une décision doit être prise ainsi que sur la façon dont elle est prise par un humain, nous allons dans cette section décrire rapidement comment une décision est en générale appréhendée. Plus précisément, dans un premier temps, nous allons voir qu'une décision est généralement considérée comme une étape d'un processus plus large. Dans un second temps nous préciserons les hypothèses et les objectifs d'une telle étape. Enfin nous insisterons sur le fait que les informations nécessaires à cette étape peuvent être décrites de deux façons différentes : soit numériquement, soit logiquement.

2.1.1 Le processus de décision

Le langage commun a coutume de concevoir la décision comme un phénomène instantané impliquant le plus souvent un individu isolé et se concrétisant généralement par le choix d'une solution parmi un ensemble de solutions envisageables. Pourtant, un examen rapide du concept de décision montre que les choix à effectuer s'inscrivent dans le temps et nécessitent entre autres des périodes de définition des objectifs et de recherche d'informations. Puisque prendre une décision de façon réfléchie n'est pas instantané, il est donc naturel d'intégrer cette action au sein d'un processus : le processus de décision. Pour Condorcet déjà, la prise de décision (dans une assemblée) se fait en trois étapes successives : (1) la détermination des règles de choix, des choix à effectuer et des conséquences des décisions possibles ; (2) la circonscription des choix à effectuer et l'évaluation fine de chaque décision possible grâce à la discussion, et enfin (3) le choix proprement dit. Plus récemment, des chercheurs ont opposé à cette approche séquentielle une approche parallèle. En effet, l'expérience montre que même si un processus de décision peut se scinder en plusieurs phases, celles-ci ne sont pas organisées aussi simplement [Eberhard, 1972]. Aussi, par exemple, les auteurs de [Mintzberg et al., 1976] ont proposé de voir ce processus comme un ensemble de cycles constitué des trois étapes suivantes :

- la phase d'identification :
  - identification du problème et des moyens d'action,
  - identification des informations nécessaires à la résolution du problème.
- la phase de développement :
  - recherche d'alternatives déjà connues et répondant au problème,

- construction de nouvelles solutions au problème (planification, adaptation ...),
- la phase de sélection :
  - élimination des alternatives sub-optimales,
  - évaluation des alternatives restantes (jugement intuitif, analyse, négociation ...),
  - mise en oeuvre de l'alternative retenue.

Remarquons ici que le schéma d'évolution d'agent que nous avons proposé dans la section 1.2.3 peut très facilement s'apparenter à un cycle de la vision proposée par les auteurs de [Mintzberg et al., 1976].

Bien qu'expérimentalement la phase de développement est plus importante (en temps) que les phases d'identification et de sélection, pour bon nombre de chercheurs c'est la phase de sélection et en particulier l'évaluation qui est le point central de la décision. Aussi c'est le véritable objet de la *théorie de la décision*.

2.1.2 Qu'est-ce que la Théorie de la Décision ?

La Théorie de la Décision vise à expliciter les mécanismes permettant, pour une situation donnée, de choisir la « meilleure » réaction parmi un ensemble d'alternatives mutuellement exclusives. Plus précisément, présupposant des connaissances (philosophiques, politiques, scientifiques, ...) préalables sur l'environnement et sur les alternatives possibles, elle cherche à répondre aux questions suivantes :

- Comment choisir la réaction qui correspond le mieux à des connaissances données ?
- Parmi ces connaissances, quelles sont celles pertinentes pour faire un choix donné ?
- Comment savoir s'il n'est pas possible de trouver une meilleure réaction que la meilleure déjà exhibée ? En particulier, si A et B sont deux ensembles d'alternatives et si A est un sous-ensemble de B, comment savoir si l'alternative choisie dans B est la même que celle choisie dans A ?
- Comment faire un choix lorsque les connaissances utilisées sont incertaines ?
- Comment faire un choix à plusieurs et en particulier comment agréger des préférences possiblement contradictoires ?

Cette théorie ne concerne donc pas le problème de définir les objectifs ni même les alternatives et leurs conséquences. Elle cherche seulement à exprimer ces informations de façon précise afin de les utiliser le plus efficacement possible. Plus particulièrement, elle suppose que les alternatives et leurs conséquences sont données, qu'il existe des règles définissant ce qui est « bon » et ce qui est « mauvais » ainsi que ce qui est « mieux » et ce qui est « pire ». Le terme de *préférences* est usuellement utilisé pour désigner de telles règles.

Pour répondre aux cinq points cités précédemment, la théorie de la décision offre une riche collection de techniques et de procédures analytiques pour spécifier des préférences et les introduire dans un phase de décision. Fréquemment, on distingue les problèmes de décision sous certitude, ceux sous risque, et ceux sous incertitude.

- Décider sous certitude signifie choisir parmi un ensemble d'alternatives qui mènent chacune à une unique conséquence. Le choix parmi les alternatives est donc équivalent au choix parmi les différentes conséquences.

- Décider sous risque signifie choisir parmi un ensemble d'alternatives qui peuvent mener chacune à plusieurs conséquences dont les probabilités d'occurrence sont connues. Plus précisément, les résultats possibles de la mise en oeuvre de chaque alternative sont conditionnels à la situation effective de l'univers qui peut être dans différents états mais dont les probabilités d'existence sont connues. Le choix parmi les alternatives est donc équivalent au choix parmi différentes distributions des conséquences.
- Décider sous incertitude signifie choisir parmi un ensemble d'alternatives qui peuvent mener chacune à plusieurs conséquences dont les probabilités d'occurrence ne sont pas connues.

2.1.3 Représentations des préférences

Une décision peut être apparentée à un choix fait par un « décideur » parmi un ensemble d'alternatives mutuellement exclusives. Or, dans la majorité des problèmes de décision, les alternatives ne sont pas toutes aussi attractives. Une bonne décision est donc de choisir une alternative qui est, par exemple, meilleure que toutes les autres ou au moins aussi bonne que chacune des autres. Comme nous l'avons présenté précédemment, c'est la phase d'évaluation qui permet de déterminer de telles alternatives. Pour ce faire, il est néanmoins nécessaire de disposer d'informations indiquant l'« attractivité » des alternatives. Ce sont ce que l'on appelle les préférences du « décideur ». Ces informations peuvent être de deux natures a priori fondamentalement différentes. Elles peuvent être soit des jugements *absolus* soit des jugements *relatifs*.

Dans le premier cas, ces informations s'expriment en positionnant chaque alternative sur une échelle ou en les comparant à des niveaux de référence. Elles indiquent alors la valeur intrinsèque (du point de vue de l'« attractivité ») des différentes alternatives. Ceci est usuellement modélisé par des valeurs numériques représentant pour chaque alternative son « attractivité intrinsèque » ; chaque alternative étant considérée indépendamment des autres. Elles sont souvent qualifiées de représentations quantitatives.

Dans le second cas, ces informations s'expriment en utilisant des notions comparatives comme par exemple « meilleur » et « plus mauvais ». Elles décrivent alors « directement » le résultat des comparaisons (du point de vue de l'« attractivité ») des alternatives entre elles. Usuellement, ceci est modélisé avec des relations binaires dites de *comparaison* ou de *préférence*. Ces relations sont entre les alternatives et indiquent par exemple à chaque fois que l'une est « meilleure » que l'autre. Elles sont souvent qualifiées de représentations qualitatives ou logiques.

a) Représentation des préférences par valeurs numériques (utilité)

Afin de formaliser les préférences du décideur, des valeurs numériques sont souvent utilisées. Lorsque c'est le cas, une valeur numérique est assignée à chaque alternative. Cette valeur représente le niveau de « désirabilité » subjective intrinsèque de l'alternative considérée. C'est ce que l'on appelle l'*utilité* de l'alternative. Tout se passe comme si une transformation intermédiaire entre alternatives et valeurs d'utilité était réalisée afin de simplifier le choix.

Afin d'effectuer une décision sur la base de ces valeurs il est nécessaire de définir un critère de choix comme lorsqu'une représentation qualitative est utilisée. Un critère courant de décision est de choisir une des alternatives qui a la plus grande valeur d'*utilité*.

Souvent, la valeur des différences entre valeurs d'utilité a du sens. Par exemple une alternative qui a une valeur d'utilité deux fois plus grande qu'une autre alternative sera deux fois plus préférée. c'est ce qui est appelé l'hypothèse de *cardinalité*. Il est aussi d'usage d'utiliser une fonction dite d'*utilité* (noté u) pour associer une utilité à chaque alternative. La notion $u(a)$ dénote alors l'utilité de l'alternative a. Comme l'ensemble des valeurs d'utilité est alors manipulé au travers de cette fonction, cette dernière est dite représenter le comportement du décideur.

Il existe de nombreux outils mathématiques performants pour manipuler les représentations numériques ou quantitatives et en particulier pour maximiser les fonctions. C'est pourquoi, l'approche quantitative est la plus présente et cela même si elle peut paraître moins intuitive que l'approche qualitative .

b) Représentation des préférences par relations de comparaison

Dans les situations courantes, les préférences sont communément exprimées à base de comparaisons. C'est en particulier le cas lorsque sont utilisées des expressions comme « meilleur que », « pire que », « aussi bon que », « au moins aussi bon que ». Ces dernières ont en point commun de mettre en relation deux objets qui peuvent, a priori, être choisis. C'est pourquoi les préférences sont couramment modélisées et décrites par des relations binaires[1] sur l'ensemble des alternatives. Par exemple, on peut décider que, pour tout couple d'alternatives (a, b), la notation aRb dénote le fait que choisir l'alternative a est au moins aussi préférable que choisir l'alternative b.

Afin de déterminer la ou une meilleure alternative, la donnée d'un ensemble de préférences n'est pas suffisant ; il faut aussi savoir comment utiliser ces informations. Un critère de décision est donc nécessaire. Un rapide tour d'horizon de la littérature montre, à titre d'exemple, que les critères suivants sont communément utilisés :

- la meilleure alternative est celle qui est préférée à toutes les autres $(\forall b,\ aRb)$;
- une des meilleures alternatives est une alternative a telle qu'il n'existe pas d'alternative b qui lui soit préférée $(\nexists b,\ aRb)$;
- une des meilleures alternatives est une alternative a qui est au moins aussi bonne que toutes les autres $(\nexists b,\ bRa \wedge \neg(aRb))$.

Cependant, si aucune contrainte n'est imposée aux relations de préférence, il se peut que, quel que soit le critère utilisé, il n'existe pas de meilleure alternative. Comme ce n'est pas le cas si la relation de préférence est transitive, pour beaucoup, la notion centrale à la base de la préférence est la notion d'ordre (préférer un objet à un autre revient, en un sens, voir le premier devant l'autre) car ainsi le choix est toujours possible. Aussi, il est courant

[1] Voir l'annexe B.2.3 pour un rappel sur les relations binaires.

de faire l'hypothèse que ces relations sont *plus ou moins* transitives. Ainsi, les relations de préférence classiques sont des ordres totaux (i.e. des relations complètes, antisymétriques et transitives) aussi dits « linear orders » en anglais, des pré-ordres totaux (i.e. des relations complètes et transitives) aussi dits « weak orders » en anglais, des semi-ordres (i.e. des relations complètes, de Ferrers et semi-transitives), des ordres d'intervalle (i.e. des relations complètes et semi-transitives), des ordres partiels (i.e. des relations réflexives, antisymétriques et transitives), des pré-ordres partiels (i.e. des relations réflexives et transitives) aussi dits « quasi-orders » ou « preorders » en anglais (voir [Bouyssou and Vincke, 2003]). Pour plus de détails, voir [Roubens and Vincke, 1985] et [Öztürk et al., 2005] (incluant entre autres un rappel complet sur les relations binaires).

c) Rapport entre représentations logiques et numériques

Sous certaines conditions, les représentations des préférences à base de relations binaires de comparaison sont équivalentes aux représentations numériques. En particulier, un résultat fondamental indique que si l'ensemble des alternatives (noté Alt) est énumérable (fini ou infini) et si R est un pré-ordre total alors il existe une fonction u de Alt dans $\Re$ qui représente numériquement la relation R :

$$\forall a, b \in Alt^2, \qquad aRb \quad \text{si et seulement si} \quad u(a) \geq u(b) \tag{2.1}$$

Dans un tel cas, la fonction u est unique à une transformation strictement croissante près. Ce résultat permet d'associer le concept de maximisation à celui de transitivité.

Pour de nombreux chercheurs, la représentation fondée sur des jugements relatifs est à la base de toute modélisation des préférences et donc aussi à la base de la seconde représentation. Plus précisément, pour beaucoup, la préférence est avant tout une notion qualitative qui peut être représentée de façon numérique sous certaines conditions. C'est pourquoi il existe de nombreux travaux qui s'attachent à fonder les représentations numériques sur les représentations qualitatives (voir [Fishburn, 1999]). Nous ne prétendons pas traiter cette problématique mais nous remarquons que ce sont des représentations toutes aussi utiles l'une que l'autre car elles sont toutes les deux fort employées et cela sans faire référence (implicitement ou explicitement) l'une à l'autre (voir [Mousseau, 2003, p.4]). Par exemple, les notes que donne un professeur aux copies qu'il corrige sont des jugements absolus. D'un autre côté, le client d'un restaurant qui choisit les plats qu'il va manger utilise plutôt des informations relatives.

2.2 Modélisations numériques des préférences

Dans la littérature et en particulier dans les réalisations pratiques, il est souvent d'usage d'utiliser une représentation numérique des préférences. C'est pourquoi, nous exposons ici les travaux fondateurs dans le domaine. Plus précisément, dans cette partie de chapitre, nous présentons le concept d'*utilité espérée* formalisé par von Neuman et Morgenstern dans [von Neumann and Morgenstern, 1947] ainsi que la théorie de l'*utilité multi-attributs* (abrégée par « MAUT » en anglais) proposée dans [Keeney and Raiffa, 1976].

2.2.1 L'utilité espérée : la décision avec risque

Formellement, le résultat d'une décision est une alternative à mettre en oeuvre (celle par exemple qui a la plus grande valeur d'utilité ou qui est préférée à toutes les autres). Toutefois, comme un choix est fait dans l'optique de faire évoluer le monde, une alternative ne peut être dissociée de l'*effet* de sa mise en oeuvre. Dans de nombreuses situations, l'effet de la mise en oeuvre d'une alternative peut être ambivalent/multiple. Cela provient du fait que les réactions ne sont pas forcément déterministes ou que de nombreux facteurs hors de portée du décideur (les scénarii) peuvent influencer leur résultat. Certains de ces facteurs sont connus du décideur, d'autres pressentis et d'autres encore totalement ignorés. Afin de modéliser ce phénomène, il est courant d'associer à chaque alternative un ensemble d'effets possibles et de supposer que les préférences du décideur sur les différentes alternatives sont aussi fonction des effets résultants de leur mise en oeuvre.

Ceci permet d'envisager différentes stratégies pour choisir l'alternative à mettre en oeuvre : choisir l'alternative qui a le meilleur effet possible, l'alternative qui a le moins mauvais ou encore l'alternative qui a en moyenne le meilleur effet.

Rappelons ici que si tous les facteurs externes sont connus, alors c'est un problème de choix déterministe. La décision ne fait intervenir aucune incertitude : il n'y a qu'un seul état résultant pour chaque alternative (de probabilité égale à 1). Si tous les facteurs ne sont pas connus, alors c'est un problème de choix sous incertitudes. Dans cette dernière situation, le fait de choisir une alternative implique que l'effet résultant est parmi un certain ensemble d'effets possibles (un par scénario). La fréquence d'occurrence de chaque scénario, lorsqu'elle est connue, est très fréquemment modélisée avec des probabilités. En effet les probabilités sont un outil naturel pour décrire et analyser les situations d'incertitude ou d'ignorance (au sujet de certains facteurs pertinents) auxquelles fait face un décideur.

a) Représentations classiques des problèmes de décision

Afin de faciliter l'appréhension des problèmes de décision, deux représentations sont fréquemment utilisées : les *matrices* de décision et les *arbres* de décision.

La représentation matricielle est un standard de représentation des problèmes de décision. À chaque problème est associée une matrice qui résume les trois composantes suivantes : les alternatives, les scénarii possibles et les conséquences prévisibles (voir la figure (2.1)). *Les alternatives* sont les possibilités qui s'offrent au décideur. Pour des raisons de commodité, dans la théorie de la décision, cet ensemble est généralement considéré comme fini et tel que les alternatives sont mutuellement exclusives : ex. au lieu de faire un choix entre A et B on choisit entre $A \wedge \neg B$ et $B \wedge \neg A$. *Les scénarii* représentent la façon dont le monde va évoluer et ceci, quelque soit l'alternative choisie. *Les conséquences* sont les résultats de la mise en oeuvre de chaque alternative pour chaque scénario possible. S'il n'y a pas d'incertitude sur le résultat des réactions possibles (action non déterministe ou facteur externe) la matrice de décision est simplement un tableau.

Une autre manière de décrire un problème de décision consiste à représenter les éléments décrits précédemment sur un arbre dit de décision. Les alternatives sont représentées par

Scénarii Alternatives	Il pleut	Il ne pleut pas
prendre le parapluie	Habits secs et bagages lourds	Habits secs et bagages lourd
ne pas prendre le parapluie	Habits humides et bagages légers	Habits secs et bagages légers

Fig. 2.1 – Exemple de représentation matricielle d'un problème de décision

les branches de l'arbre. Les conséquences des alternatives selon les états possibles du monde sont représentées quant à elles par les nœuds et les feuilles de l'arbre (voir 2.2). C'est une

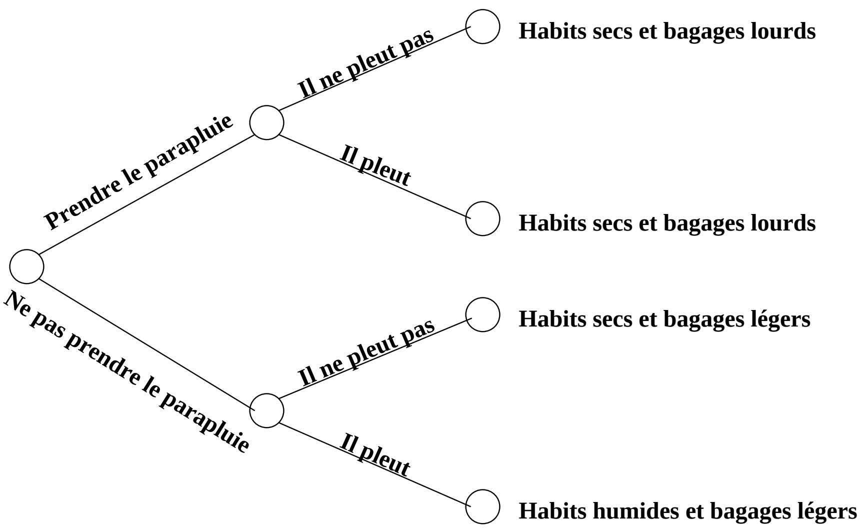


Fig. 2.2 – Exemple de représentation avec un arbre de décision

représentation très utilisée car elle permet de représenter intuitivement une séquence de prises de décision en concaténant les arbres des décisions mises en jeu.

Si les préférences du décideur sont représentées par des valeurs numériques, plus qu'aux alternatives elles mêmes, c'est aux effets que des utilités sont alors associées en premier lieu. Généralement, l'utilité d'une alternative est différente suivant son effet. Aussi, afin d'utiliser toujours les représentations graphiques présentées précédemment, il est donc d'usage de remplacer les descriptions verbeuses des conséquences par des valeurs d'utilité dans la matrice. L'exemple de la figure (2.1) devient par exemple celui de la figure (2.3).

b) Utilité espérée

L'utilité espérée d'une alternative représente la moyenne des différentes utilités de cette alternative suivant les différentes scénarii possibles, pondérés par leur probabilité d'occurrence. Formellement, si $\{1, \ldots, n\}$ représente l'ensemble des n scénarii possibles et si, pour

Alternatives \ Scénarii	Il pleut	Il ne pleut pas
prendre le parapluie	15	15
ne pas prendre le parapluie	0	18

FIG. 2.3 – Exemple de représentation avec une matrice d'utilité

tout scénario i, qui a la probabilité d'occurrence p_i, $u_i(x)$ représente l'utilité de l'alternative x si le scénario i advient, alors l'utilité espérée de l'alternative x est :

$$u(x) = \sum_{i=1}^{n} p_i \times u_i(x) \tag{2.2}$$

Décider de mettre en oeuvre l'alternative ayant la plus forte valeur d'utilité espérée est un critère de décision (voir [Kast, 2002, p64-80]). Celui-ci semble légitime et est d'ailleurs fort ancien. Von Neumann et Morgenstern dans [von Neumann and Morgenstern, 1947] ont montré qu'une telle forme de fonction d'utilité (à une fonction affine près) découle de quelques hypothèses (axiomes) sur le décideur et son comportement. Plus précisément, une telle fonction d'utilité représente une préférence qui peut se modéliser par une relation binaire ayant les propriétés de complétude, de transitivité, de continuité et d'indépendance.

- La propriété de *complétude* signifie que toutes les loteries sont comparables deux à deux.
- La propriété de *transitivité* quant à elle recouvre le fait que si l'alternative a_1 est préférée à l'alternative a_2 et si a_2 est préférée à a_3, alors a_1 est préférée à a_3. Ces deux propriétés impliquent que les préférences définissent un pré-ordre total sur les alternatives.
- La propriété de *continuité* (faible) signifie que si a_1 est préférée à a_2 et si a_2 est préférée à a_3, alors il existe un réel α compris entre 0 et 1 tel que l'alternative qui consiste à mettre en oeuvre l'alternative a_1 ou l'alternative a_3 (avec une probabilité respectivement de α et de $1 - \alpha$), soit indifférente à l'alternative a_2. Le principe d'indifférence entre ces deux loteries formulé ici est aussi appelé axiome de réduction des loteries composées : il signifie essentiellement que l'ordre de succession des évènements aléatoires ne compte pas et que la valeur des différences entre valeurs d'utilité a du sens (l'hypothèse de cardinalité).
- La propriété d'*indépendance* (aussi appelée en anglais "the sure-thing principle") signifie enfin que si a et b sont deux alternatives ayant une partie en commun (the sure thing) et si cette partie, et uniquement elle, est changée (mais reste la même dans les deux alternatives), alors le choix entre a et b n'est pas affecté.

L'utilité espérée est particulièrement importante dans la mesure où elle est calculable très simplement. En particulier, les utilités des conséquences et la distribution de probabilité attachée à chacune de ces conséquences sont clairement séparées.

2.2.2 L'utilité multi-critères

Dans la pratique, il est en général difficile de déterminer la fonction d'utilité d'un décideur et spécialement lorsque les alternatives sont complexes. Afin de simplifier cette phase, la théorie de l'utilité multi-attributs (MAUT) a été proposée dans [Keeney and Raiffa, 1976]. Cette théorie exploite le fait que, souvent, chaque alternative d'un problème de décision peut être décrite par un ensemble de variables (appelées attributs ou propriétés) vérifiant l'hypothèse de cardinalité. Sur la base de ce découpage des alternatives, les fonctions d'utilités peuvent souvent être décomposées en plusieurs sous-utilités. Chacune d'elles concerne un seul attribut et est plus simple à obtenir et à manipuler que la fonction d'utilité initiale. En général, spécifier une fonction d'utilité sur plusieurs variables est coûteux (exponentiel par rapport au nombre de variables). Néanmoins, sous certaines conditions (indépendance entre les variables), la spécification peut être plus rapide.

Cette section va nous permettre de présenter rapidement cette théorie. Pour plus de détails, le lecteur intéressé pourra se reporter aux travaux originaux de Keeney et Raiffa ([Keeney and Raiffa, 1976]) ou, pour un aperçu, au troisième chapitre de [Ha, 2001].

a) La notion d'indépendance

La notion d'*indépendance* est à la base de la théorie de l'utilité multi-attributs. Celle-ci présuppose qu'une alternative peut être décrite par une valuation d'un ensemble de variables $V = \{X_1; X_2; \ldots; X_n\}$. Par exemple, l'alternative x correspond à la valuation $\{X_1 = x_1; X_2 = x_2; \ldots; X_n = x_n\}$. Un ensemble de variables $Y \subset V$ est dit *utilement indépendant* (UI) (par rapport aux préférences du décideur) de son complément[2] Z si et seulement si les préférences entre les alternatives (dont les variables dans Z sont fixées) ne dépendent pas des valeurs des variables de Z.

Lorsque l'ensemble de variables Y est UI, les auteurs de [Keeney and Raiffa, 1976] ont montré que l'on peut écrire la fonction d'utilité du décideur $u(x)$ comme une expression fonction de deux fonctions sur Z et d'une fonction sur Y. En particulier, on a le théorème suivant : Si Y est UI, alors il existe trois fonctions (g, h et u_y) telles que g et h sont des fonctions qui dépendent uniquement des variables de Z et u_y est une fonction[3] qui dépend uniquement des variables de Y et telles que la fonction d'utilité u(x) a la forme suivante :

$$u(x) = g(x) + h(x) \times u_Y(x) \tag{2.3}$$

b) Utilisation de l'hypothèse MUI

L'indépendance des variables joue un rôle fondamental dans le calcul de l'utilité multi-attributs un peu comme l'indépendance des probabilités dans le calcul de l'utilité espérée. Elle permet de décomposer l'utilité et ainsi de réduire la dimension et donc la complexité de la fonction d'utilité du décideur. En particulier, si pour tout sous-ensemble Y de V, Y est *utilité indépendant* de son complémentaire Z, on dit que l'on est dans le cas d'*indépendance*

[2] $Z = \{V \setminus Y\}$

[3] La fonction u_Y est appelée la sous-utilité pour les variables de Y.

utilité mutuelle des variables (MUI) et on peut écrire la fonction d'utilité $u(x)$ sous la forme d'une somme (ou d'un produit[4]) des fonctions de sous-utilités :

$$u(x) = \sum_{i=1}^{n} k_i \times u_i(x_i) \tag{2.4}$$

Lorsque l'on se trouve dans le cas MUI, la fonction d'utilité peut être obtenue en appliquant les deux étapes suivantes :

1. Déterminer les sous-utilités u_i. Pour cela, le décideur doit, pour chaque sous-utilité en rapport avec l'attribut i, ranger les alternatives qui ont les attributs/variables autres que i fixés à la même valeur.
2. Déterminer les cœfficients k_i (le rapport relatif ou taux de substitution). Pour cela, le décideur doit ranger tous les couples d'alternatives se différenciant uniquement par deux critères.

À cause de la simplicité de la fonction d'utilité engendrée, l'hypothèse MUI est souvent faite même si malheureusement en pratique l'hypothèse MUI n'est pas réaliste. Certains chercheurs ont toutefois tenté de traiter pratiquement le cas où l'hypothèse MUI n'est pas vérifiée. En particulier, Ha et Haddaway ont étudié le cas où les variables sont seulement UI entre elles. Dans une telle situation, la fonction d'utilité a la forme multilinéaire suivante :

$$u(x) = \sum_{\emptyset \neq Y \subseteq X} k_y \prod_{X_i \in Y} u_i(x_i) \tag{2.5}$$

où les u_i sont les fonctions de sous-utilités et les k_y les coefficients. Comme dans ce cas il est nécessaire de déterminer $2^n - 1$ coefficients, dans la pratique si le problème comporte plus de trois variables, les coefficients ne sont pas déterminés et cela même si obtenir les fonctions de sous-utilité est assez simple. Aussi ces auteurs ont proposé des techniques pour réduire le nombre de coefficients à déterminer (voir [Ha, 2001]). Ces dernières se basent sur les trois hypothèses suivantes qui permettent d'obtenir des contraintes sur les coefficients :

1. L'ensemble des alternatives est connu et est fini.
2. La fonction d'utilité du décideur est multilinéaire. Les fonctions de sous-utilités sont connues. Les coefficients sont encore à déterminer.
3. Le décideur peut comparer des alternatives si on le lui demande.

2.2.3 Conclusion

Après avoir présenté dans la section 2.1 la théorie de la décision dans ses grandes lignes, nous avons présenté dans cette section les formalismes qui sont les plus souvent employés dans la littérature et en particulier dans les réalisations pratiques. Plus précisément, dans cette partie de chapitre, nous avons présenté le concept d'*utilité espérée* formalisé par von

[4]L'équation 2.4 est équivalente à $u(x) = \prod_{i=1}^{n} (1 + k_i \times u_i(x_i))$

Neuman et Morgenstern dans [von Neumann and Morgenstern, 1947] ainsi que la théorie de l'*utilité multi-attributs* (abrégée par « MAUT » en anglais) proposée dans [Keeney and Raiffa, 1976]. Ces derniers sont très attrayants car ils utilisent des représentations numériques des préférences et qu'il existe de nombreux outils mathématiques performants pour manipuler les représentations numériques ou quantitatives et en particulier pour maximiser les fonctions. Néanmoins, comme nous le verrons dans la section 2.4, ces formalismes reposent sur des hypothèses qui ne sont pas toujours vérifiées et en particulier dans le contexte de dialogue naturel.

2.3 Modélisations logiques des préférences

Comme nous l'avons remarqué dans la section 2.1, même si cela est moins fréquent, les informations sur lesquelles se base la décision peuvent aussi être des comparaisons. C'est pourquoi nous présentons dans cette partie de chapitre quatre modélisations qualitatives de la préférence. Plus précisément, après avoir exposé les hypothèses fréquemment effectuées dans ce domaine, nous décrirons le formalisme proposé par von Wright dans [von Wright, 1972], celui proposé par Doyle et Wellman dans [Doyle and Wellman, 1994], celui proposé par Boutilier dans [Boutilier, 1994] ainsi que les CP-nets proposés par Boutilier dans [Boutilier et al., 1999].

2.3.1 Hypothèses fréquentes

La connaissance dans de nombreux domaines fait intervenir des comparaisons. Souvent ce ne sont pas des comparaisons entre des alternatives à proprement parlé (« individuals » pour Doyle) mais entre des propriétés de ces alternatives. En effet, les comparaisons sont souvent effectuées afin de mettre en relation des groupes d'alternatives (« classe of individuals » pour Doyle) ayant chacun une propriété caractéristique. Par exemple, un importateur de vêtements utilise le fait que les siciliens sont plus petits que les danois pour envoyer plus de petites tailles en Sicile qu'au Danemark. C'est ce dernier type de comparaison qui pose le plus de problèmes. Aussi, depuis une cinquantaine d'années, de nombreux chercheurs en philosophie se sont proposés d'étudier la façon dont une personne pouvait exprimer ces comparaisons et en particulier les préférences. Même s'il n'existe pas de consensus sur la façon dont les préférences sont exprimées[5], il en ressort néanmoins un petit nombre de principes généraux fréquemment admis et en particulier les principes d'*Expansion*, de *Transitivité*, de *Dépendance au contexte* et *Ceteris Paribus*. Ceux-ci se basent sur l'hypothèse déjà faite par von Wright dans [von Wright, 1963] selon laquelle le sens des préférences se fonde sur les comparaisons entre alternatives (i.e. les objets du choix). C'est ce que Doyle et Wellman appellent le processus de « lifting ».

[5]Le fait que des questions comme « est-ce que deux propriétés sont forcément comparables ? » « si une propriété est préférée à une autre, est-ce que la négation de la seconde est préférée à la première ? » admettent plusieurs réponses contradictoires indique qu'il existe plusieurs conceptions différentes de la préférence

a) Expansion (ou Exclusion Mutuelle) :

Le principe le plus généralement admis est celui d'*Expansion*. Il indique que les préférences sont généralement exprimées par des comparaisons entre des objets mutuellement exclusifs. En particulier, si ce sont des propriétés qui sont comparées, alors celles-ci doivent être inconsistantes entre elles :

> « If we say that it will be better if p than if q, then we mean that it would be better if $p \wedge \neg q$ than if $\neg p \wedge q$. » [Halldén, 1957]

Cette définition indique qu'il faut comparer des alternatives vérifiant $p \wedge \neg q$ et $\neg p \wedge q$ plutôt que p et q. Si deux alternatives vérifient toutes les deux p, alors la seule chose qui peut permettre la comparaison est de savoir si elles vérifient aussi q. Mais ceci n'est alors plus une comparaison entre p et q mais entre q et $\neg q$. Dans le reste de ce document, toutes les préférences présentées respectent ce principe.

b) Transitivité :

Comme nous l'avons remarqué dans la partie (b) de la section 2.1.3, la notion d'ordre est pour beaucoup centrale aux préférences. C'est elle qui permet d'exprimer qu'un objet est meilleur qu'un autre, voire que tous les autres. Cette notion est communément modélisée avec des relations binaires transitives (voir l'annexe B.2.3). Normalement appliquée aux alternatives, elle est aussi parfois appliquée aux propriétés. Elle indique alors que si la propriété p_1 est préférée à p_2 et la propriété p_2 est préférée à p_3, cela signifie alors aussi que la propriété p_1 est préférée à p_3.

c) Dépendance au contexte :

Une préférence est relative à un contexte [Hansson, 1996]. En particulier, un agent peut préférer l'objet a à l'objet b dans un certain contexte et, dans un autre, préférer l'inverse sans pour autant être qualifié d'inconsistant. En d'autrement termes, cela signifie que les préférences exprimées et/ou manipulées sont généralement conditionnées par le contexte. C'est ce que von Wright appelle des préférences *holistiques* dans [von Wright, 1972]. En intelligence artificielle, elles sont plutôt qualifiées de *conditionnelles*.

d) Ceteris Paribus :

Le principe *Ceteris Paribus* indique pour sa part que les alternatives à considérer pour donner du sens aux comparaisons, doivent vérifier les mêmes autres propriétés :

$$\textit{Ceteris Paribus} \quad \equiv \quad \text{« toutes choses égales par ailleurs »} \tag{2.6}$$

En particulier, si la propriété p_1 est dite être préférée à la propriété p_2, cela signifie qu'une alternative vérifiant la propriété p_1 est préférée à une alternative vérifiant la propriété p_2 si elles sont « toutes choses égales par ailleurs » c'est-à-dire si elles ont les mêmes autres propriétés. Là encore, pour beaucoup de philosophes (en particulier von Wright et Hansson) et de chercheurs en Intelligence Artificielle (Doyle, Wellman, ...), nos préférences sont de ce type [Boutilier et al., 2004b, p.4]. Par exemple, dans [Hansson, 1996] nous pouvons lire que :

> « When discussing with my wife what table to buy for our living room, I said : "A round table is better than a square one" By this I did not mean that irrespectively of their other properties, any round table is better than any square-shape table. Rather, I meant that any round table is better (for our living room) than any square table that does not differ significantky in its other characteristics, such as height, sort of wood, finishing, price, etc. This is preference *ceteris paribus* or "everything else being equal". Most of the preferences that we express or act upon seem to be of this type ».

Ce principe est exprimé de façon très générale. Aussi comme nous le verrons dans la suite de cette section, de nombreux chercheurs ont proposé leur propre définition (voir entre autres [von Wright, 1972][Doyle and Wellman, 1994][Boutilier, 1994][Lang, 2003, p.121]). Néanmoins, lorsque les alternatives sont décrites par n variables logiquement indépendantes $V_1, \ldots, V_n$ (hypothèse usuelle), dire que la valeur x_i de la variable V_i est préférable *Ceteris Paribus* à la valeur x_j de la variable V_j signifie souvent simplement qu'une alternative ayant V_i à la valeur x_i est préférée à une alternative ayant V_j à la valeur x_j si les autres variables ont les mêmes valeurs. Notons ici que l'hypothèse *Ceteris paribus* est une forme qualitative de la notion d'indépendance entre attributs de la théorie classique de la décision [Keeney and Raiffa, 1976].

2.3.2 Formalisme proposé par von Wright dans [von Wright, 1972]

Bien que la notion de préférence soit depuis longtemps un sujet d'étude important en économie (voir la section 2.2 sur la théorie de la décision), son étude d'un point de vue logique a commencé seulement à partir des années cinquante et en particulier avec les travaux de von Wright (voir [von Wright, 1963] et [von Wright, 1972]). Nous présentons ici ces travaux car ils nous ont grandement influencés. Plus précisément, et sauf mention contraire, nous présentons ce qui dans [von Wright, 1972] permet de comprendre et situer notre travail.

a) La vision des préférences de von Wright :

Pour von Wright, il n'existe pas de consensus sur les lois (logiques) à la base de la préférence. Ceci est en partie dû au fait que le concept de préférence a plusieurs facettes. Par conséquent, comme le travail du logicien est de déterminer quelles règles sont compatibles entre elles (ainsi que les conséquences d'un ensemble de règles), il est dans l'ordre des choses que différentes modélisations logiques de la préférence soient proposées. Elles correspondent à chaque fois à une idée particulière que l'on se fait intuitivement des propriétés vérifiées par la préférence. C'est dans cette optique que von Wright propose une logique asymétrique basée sur des relations (de préférences) entre mondes possibles (au sens de Kripke) qui sont relatives à un décideur (ses goûts) et surtout à une situation donnée (les circonstances).

Plus précisément, pour von Wright, une préférence est souvent relative aux circonstances. Par exemple, la préférence envers les vins servis lors d'un repas, varie en fonction des plats servis : du vin rouge s'il y a de la viande, du vin blanc s'il y a du poisson. Lorsque c'est le cas, il qualifie la préférence d'*holistique*. Une préférence est dite *Ceteris Paribus* (toute chose égale par ailleurs) si elle est vérifiée quelles que soient les circonstances. De plus, pour lui le sens d'une préférence est relatif à un temps et à une personne. Pour des considérations de simplicité il ne le formalise pas mais insiste sur le fait d'avoir à l'esprit cette idée. Nous sommes d'accord avec lui. En particulier, dans notre formalisme, cet aspect relatif des préférences (par rapport à une situation dans le temps et à un décideur) est apporté/explicité en modulant les préférences par les croyances de l'agent (voir la section 8.1.3 de la quatrième partie).

Pour von Wright, on peut distinguer les préférences *intrinsèques* et les préférences *extrinsèques*. D'un côté, un objet est dit être préféré extrinsèquement à un autre s'il existe une raison (non circulaire) pour préférer le premier au second. D'un autre coté, un objet est dit être préféré intrinsèquement à un autre s'il est exprimé explicitement que le premier est préféré au second. Comme il est vraisemblable que toute préférence extrinsèque a pour raison *in fine* des références intrinsèques, il propose une modélisation logique qui se limite au traitement des préférences intrinsèques. Pour nous, cette distinction est une conséquence de l'aspect *argumentatif* des préférences. Celui-ci est surtout important pour l'utilisation pratique des préférences dans une phase de décision. Nous formaliserons cette idée en introduisant une phase d'extension à notre modèle (voir le chapitre 6). C'est pourquoi, les préférences intrinsèques de von Wright peuvent être rapprochées de nos préférences partielles primitives (voir le chapitre 5) et les préférences extrinsèques, de nos préférences partielles étendues (voir le chapitre 6). À notre avis, c'est de la formalisation du rapport entre les deux types de préférences que découle toute la force discriminante de notre modélisation.

Pour lui les préférences peuvent être exprimées soit en mettant en relation des propriétés (appelées en anglais *properties* ou *states*) ou des assemblages de propriétés (qu'il appelle en anglais *state of affairs*) soit en mettant en relation des états du monde ou des entités réelles (qu'il appelle en anglais *total states* ou *thing-like entities*). Pour lui, il faut avoir à l'esprit ces deux aspects. Il remarque néanmoins que, à chaque fois, il semble que les préférences peuvent être exprimées en mettant en relation des propriétés. Nous pouvons remarquer que pour von Wright, au vu des définitions qu'il propose (voir suite du document), même si les préférences sont exprimées explicitement en mettant en relation des propriétés, le sens de ces dernières concerne lui implicitement des relations entre des états du monde différents.

b) Formalisation :

von Wright reprend l'idée de préférences comparatives (i.e. « la propriété dénotée par la formule p est préférable à la propriété dénotée par la formule q si tout état qui vérifie p et $\neg q$ est préféré à celui qui ne diffère de celui-ci que parce qu'il vérifie q et $\neg p$. » [von Wright, 1963]) et ajoute que, d'une part, les informations spécifiées le sont sur des propriétés des objets que l'on veut *in fine* comparer (appelés *outcome* par von Wright) et d'autre part, indique que ces objets sont identiques par ailleurs.

Plus précisément, afin de formaliser sa vision de la préférence, von Wright étend une logique propositionnelle en ajoutant une relation binaire P. Les formules bien formées d'une telle logique sont ce qu'il appelle des *P-expressions*. Les *P-expressions atomiques* sont des formules formées de la lettre P, d'une variable à droite ainsi que d'une variable à gauche. La formule xPy est lue « x est préféré à y ». Les *P-expressions* sont pour leur part soit des *P-expressions atomiques* soit des formules formées à partir de *P-expressions atomiques* et des connecteurs logiques classiques. Même si à ce niveau les variables peuvent représenter indifféremment des propriétés ou des états possibles du monde, pour von Wright, les principes qu'il va imposer ne peuvent être appliqués que si les variables représentent des états possibles du monde. Dans un tel cas, puisque les variables x et y représentent des états possibles du monde (des *total state*), la notation xPy indique que l'état possible du monde représenté par x est préféré à l'état possible du monde représenté par y.

Afin de définir un système de préférence sur les états possibles du monde qui soit cohérent et transitif, il impose à la relation P d'être asymétrique ($xPy \Rightarrow \neg(yPx)$) et de vérifier ce qu'il appelle le principe de comparabilité ($xPy \Rightarrow xPz \vee zPy$). Il en déduit que la relation P est par conséquent transitive ($xPz \wedge zPy \Rightarrow xPy$) et irréflexive ($\neg(xPx)$). P définit donc un ordre total sur les configurations possibles du monde (total state). Il introduit explicitement la notion d'équivalence en définissant la relation I de la façon suivante : $xIy = \neg(xPy) \wedge \neg(yPx)$. Il en déduit que I est réflexive (xIx), symétrique ($xIy \Leftrightarrow yIx$), transitive ($xIy \wedge yIz \Rightarrow xIz$), et vérifie ce qu'il appelle le principe d'*équivalence* ($xPy \wedge xIz \Rightarrow zPy$ et $xIy \wedge yIz \Rightarrow xPz$). C'est une logique décidable car la logique propositionnelle est décidable.

Afin de formaliser sa notion de préférence entre propriétés (holistique), il considère un ensemble S de n propriétés élémentaires indépendantes (notées $p_1, \ldots, p_n$) et leurs 2^n combinaisons possibles qu'il apparente aux états possibles du monde (notés $w_1, \ldots, w_{2^n}$). Chaque propriété (élémentaire ou composée) est représentée par une forme normale qui est une disjonction de k ($0 \geq k \geq 2^n$) états possibles du monde. En particulier, une propriété est vérifiée par un monde possible si celui-ci fait partie de cette disjonction. Le cas particulier où k est égal à 0 dénote une propriété qui ne peut être obtenue dans aucun état possible du monde : une propriété qui porte en elle une contradiction. Si s est une propriété (élémentaire ou composée) alors un *s-word* est un état du monde qui vérifie la propriété s. Si s et t sont deux propriétés qui sont composées de m éléments de S, alors les 2^{n-m} combinaisons possibles des $n - m$ propriétés élémentaires de S restantes ($C_1, \ldots, C_{n-m}$) sont des propriétés de S au même titre que s et t ; il les appelle des *circonstances*. von Wright propose de définir le sens de « s est préférée à t dans la circonstance C_i » de deux manières différentes mais qui sont pour lui tout aussi acceptables :

- **Définition 1 :** s est préféré à t dans la circonstance C_i si et seulement si chaque C_i-word qui est aussi un s-word mais pas un t-word est préféré à chaque C_i-word qui est aussi un t-word mais pas un s-word.
- **Définition 2 :** s est préféré à t dans la circonstance C_i si et seulement si, d'une

part, il existe un C_i-word qui est aussi un s-word qui est préféré à un C_i-word qui est aussi un t-word, et d'autre part, il n'existe pas de C_i-word qui est aussi un t-word qui soit préféré à un C_i-word qui est aussi un s-word.

La formule $sP_{C_i}t$ dénote le fait que sous la circonstance C_i, la propriété s est préféré à t (au sens de la définition 1 ou de la définition 2). Comme la relation P, les relations P_{C_i} sont irréflexives, asymétriques et transitives. Néanmoins, elles ne vérifient pas nécessairement les principes de comparabilité et d'équivalence. En effet, considérons par exemple la situation où les quatre états du monde w_1, w_2, w_3 et w_4 vérifient la propriété C_i et tels que l'état du monde w_1 est préféré à l'état du monde w_2 (i.e. w_1Pw_2) qui est lui même préféré à l'état du monde w_3 (i.e. w_2Pw_3) qui est lui même préféré à l'état du monde w_4 (i.e. w_3Pw_4). Si s dénote la propriété qui est la disjonction des états w_1 et w_4 auxquels sont enlevés les éléments apparaissant dans C_i, si t dénote les propriétés de l'état w_2 auxquels sont enlevés les éléments apparaissant dans C_i, et si u dénote les propriétés de l'état w_3 auxquels sont enlevés les éléments apparaissant dans C_i, alors on n'a ni $sP_{C_i}u$ ni $tP_{C_i}s$ et pourtant $tP_{C_i}u$ (i.e. violation du principe de comparabilité). Le cas limite ou C_i est vide dénote le fait que la propriété s est préférée à la propriété t.

La formule sP_Ct dénote quant à elle le fait que sous toutes les circonstances possibles C_i, la propriété s est préférée à t (au sens de la définition 1 ou de la définition 2). La propriété s est dite alors être préférée *Ceteris Paribus* à t. Comme la relation P, la relation P_C est irréflexive, asymétrique et transitive. Néanmoins, elle ne vérifie pas nécessairement les principes de comparabilité et d'équivalence puisque les relations P_{C_i} ne vérifient pas nécessairement ces deux propriétés. Par conséquent, les préférences *holistiques* (i.e. les P_{C_i} et P_C) ne constituent pas des ordres totaux.

c) Intérêts et critiques de la formalisation proposée par von Wright :

Bien qu'il présente une formalisation basée sur un ensemble fini de propriétés indépendantes, cette hypothèse peut être relâchée sans modifier les deux définitions de la préférence. En effet, s'il existe une dépendance entre les propriétés alors le nombre d'états du monde (total space) sera inférieur au maximum. De plus, si le nombre de propriétés élémentaires est infini alors le nombre d'états du monde peut être infini. Il est seulement nécessaire de connaître les dépendances entre propriétés élémentaires pour utiliser pratiquement ce formalisme.

On peut reprocher à von Wright de mettre en évidence des problèmes (relativité des préférences à un décideur, préférence extrinsèque), mais ne pas proposer explicitement de solution. D'autre part son modèle se limite à l'expressivité de la logique propositionnelle Enfin, l'hypothèse *Ceteris Paribus* est vue comme le résultat d'une configuration spécifique des préférences et non comme un moyen de rendre l'expression des préférences plus concise.

2.3.3 Formalisme proposé par Doyle et Wellman dans [Doyle and Wellman, 1994]

Indépendamment du travail des philosophes dans ce domaine, l'hypothèse *Ceteris Paribus* a aussi attiré depuis le début des années 90 des chercheurs en Intelligence Artificielle. En particulier, Doyle, Shoham et Wellman ont proposé dans [Doyle et al., 1991] une *logique du désir relatif* afin de manipuler des préférences exprimées sous l'hypothèse *Ceteris Paribus*. Cette logique est une extension de leurs travaux parus dans [Wellman and Doyle, 1991]. Elle a par la suite été étendue avec entre autres [Doyle and Wellman, 1994] en précisant la notion d'équivalence. Il est à remarquer que même si ce sont des travaux indépendants, le travail de Doyle, Shoham et Wellman (surtout celui présenté dans [Doyle et al., 1991]) ressemble beaucoup à celui de von Wright. Il permet toutefois des inférences plus complexes justement grâce au concept d'équivalence (voir [Doyle and Wellman, 1994]). Dans cette section, nous présentons plus particulièrement les travaux publiés dans [Doyle and Wellman, 1994]. Ces derniers proposent une théorie de la comparaison basée sur l'hypothèse *Ceteris Paribus* qui permet de faire des comparaisons sur des classes d'individus en se basant sur des comparaisons entre individus.

a) La vision des préférences de Doyle et Wellman

Du point de vue de la théorie de la décision, un agent doit choisir la meilleure alternative parmi celles disponibles. Par conséquent, pour Doyle et Wellman, un agent doit ordonner les réactions à sa disposition. Malheureusement, cet ordonnancement ne peut être obtenu facilement. Il nécessite en effet un nombre exponentiel (par rapport au nombre d'alternatives) de comparaisons.

C'est pour cela qu'ils ont proposé de déduire cet ordonnancement en se basant sur des comparaisons entre les propriétés que peuvent vérifier les différentes alternatives mises en jeu. En effet, pour eux, même si les comparaisons entre alternatives sont utiles en soi dans certaines situations, dans de nombreux cas pratiques il est suffisant de comparer des ensembles d'alternatives entre eux au regard des propriétés qu'ils vérifient. Par exemple, un manufacturier de vêtements doit utiliser le fait que les danois sont plus grand que les siciliens pour importer plus de grandes tailles au Danemark qu'en Sicile. Ils remarquent à ce propos que les propriétés pertinentes varient d'une situation à une autre : elles dépendent du contexte.

Aussi, ils ont proposé une logique pour gérer des comparaisons[6] sur des propriétés que vérifient des ensembles d'alternatives et de situations. Cette logique base le sens d'une comparaison entre deux propriétés sur des comparaisons d'alternatives les vérifiant (processus dit de « Lifting ») tout en respectant le principe d'exclusion mutuelle et le fait que les préférences sur les attributs soient relatives à un contexte. Afin de permettre de distinguer plus de trois degrés dans la préférence, ils proposent de comparer les alternatives seulement lorsqu'elles sont *Ceteris Paribus* (i.e. « toutes choses égales par ailleurs »). Ils interprètent le fait que deux propositions soient *Ceteris Paribus* par le fait que tout ce qui

[6]Pour Doyle et Wellman, les préférences peuvent être exprimées par des comparaisons.

rentre dans le contexte en considération peut varier mais que le reste doit être équivalent. Dans les cas simples, le contexte n'est en relation qu'avec les propositions comparées. Cependant, souvent il peut aussi être en relation avec d'autres informations importantes pour la comparaison ; la condition d'une préférence conditionnelle par exemple.

b) Formalisation

L'ensemble des alternatives (noté Ω) est considéré comme ordonné par un pré-ordre total (noté $\succeq$) : une alternative ω est dite être *faiblement préférée* (i.e. au moins autant préférée) à une autre alternative ω' si la première est reliée à la seconde par la relation $\succeq$ (notée $\omega \succeq \omega'$). Une proposition p représente un ensemble d'alternatives. $P(\Omega)$ dénote l'ensemble des propositions construites sur l'ensemble Ω des alternatives. La formule $\overline{p}$ dénote le complément de la proposition p, c'est-à-dire l'ensemble des individus de Ω qui n'appartiennent pas à p (i.e. $\overline{p} = \Omega \setminus p$). La formule pq dénote les alternatives en commun à p et q (i.e. $pq = p \cap q$). Une alternative ω est dite satisfaire une proposition p si et seulement si ω appartient à la proposition p.

La notation $\mathcal{E}(\Omega)$ dénote l'ensemble de toutes les relations d'équivalence (i.e. les relations binaires réflexives, symétriques et transitives) sur Ω. Le contexte d'une préférence est identifié à un ensemble de propositions. Celui-ci, en plus de contraindre les préférences d'un agent, contraint aussi la façon dont l'agent juge des propriétés équivalentes. C'est ce que Doyle et Wellman appellent la notion d'*équivalence contextuelle*. Elle est définie formellement de la manière suivante :

> une équivalence contextuelle sur Ω est une fonction η de $P(P(\Omega))$ dans $\mathcal{E}(\Omega)$ qui assigne à chaque ensemble de propositions $\{p, q, \ldots\}$, une relation d'équivalence notée $\eta(p, q, \ldots)$.

Pour plus de lisibilité la formule $\omega\ \eta(p, q, \ldots)\ \omega'$ est écrite $\omega \sim \omega'\ \mathrm{mod}_\eta p, q, \ldots$. Cette notion permet de définir celle de « comparaison *Ceteris Paribus* », c'est-à-dire d'indiquer que les individus doivent être comparés si les autres propriétés qu'ils vérifient sont contextuellement équivalentes :

> La propriété p est dite *faiblement préférée* à la propriété q (noté $p \trianglerighteq q$) si et seulement si quelle que soit l'alternative ω appartenant à $p\overline{q}$ et quelle que soit l'alternative ω' appartenant à $\overline{p}q$ telles que ω est équivalent à ω' étant donné le contexte $\{p\overline{q}, \overline{p}q\}$, on a ω faiblement préféré à ω :

$$p \trianglerighteq q \quad \text{ssi} \quad \forall \omega, \omega' \in \Omega^2 \quad \left[\text{si} \left| \begin{array}{l} \omega \in p\overline{q} \\ \omega' \in \overline{p}q \\ \omega \sim \omega' \ \mathrm{mod}_\eta p\overline{q}, \overline{p}q \end{array} \right. \quad \text{Alors} \quad \omega \succeq \omega' \right]$$

La condition dite *Ceteris Paribus* a deux intérêts. Tout d'abord, la référence à un contexte permet de se prémunir d'assertions irréalistes car justement hors contexte. D'autre part, en quantifiant sur les contextes il est possible de contextualiser les comparaisons :

La propriété p est dite *faiblement préférée* à la propriété q lorsqu'elle est restreinte à un contexte r (noté $p \trianglerighteq^r q$) si et seulement si quelle que soit l'alternative ω appartenant à $p\overline{q}r$ et quelle que soit l'alternative ω' appartenant à $\overline{p}qr$ telles que ω est équivalent à ω' étant donné le contexte $\{p\overline{q}, \overline{p}q\}$ on a ω est faiblement préféré à ω :

$$p \trianglerighteq^r q \quad \text{ssi} \quad \forall \omega, \omega' \in \Omega^2 \quad \left[\text{ si } \left| \begin{array}{l} \omega \in p\overline{q}r \\ \omega' \in \overline{p}qr \\ \omega \sim \omega' \bmod_\eta p\overline{q}, \overline{p}q \end{array} \right. \quad \text{Alors} \quad \omega \succeq \omega' \right]$$

Cela permet de restreindre les comparaisons à des individus dans r au lieu de prendre des individus dans Ω. En d'autres termes, $p \trianglerighteq^\Omega q$ est équivalent à $p \trianglerighteq q$

c) Travaux précédents

Ce travail a été développé à l'origine pour relier la notion de préférence de la théorie de la décision à celle de buts de la planification. Plus précisément, dans [Wellman and Doyle, 1991] ils ont proposé une sémantique aux (propositions) buts basée sur des comparaisons d'individus. Celle-ci permet de gérer des situations où les buts sont satisfaits par différentes réactions suivant différents degrés où avec des garanties plus ou moins importantes.

Par la suite, ils ont généralisé ce travail afin de traiter des comparaisons entre propriétés [Doyle et al., 1991]. Les préférences sont « par nature » entre les alternatives. Elles sont étendues aux propositions par le processus dit de lifting.

La définition du *Ceteris Paribus* donnée par Doyle est très similaire à celle donnée par von Wright. Notons cependant que le formalisme de von Wright ne permet que de traiter des inférences simples à cause, d'une part, de la non utilisation de la notion d'équivalence et, d'autre part, à cause du traitement syntaxique du *Ceteris Paribus*.

2.3.4 Formalisme proposé par Boutilier dans [Boutilier, 1994]

Afin d'utiliser des préférences pour déterminer l'ensemble des buts d'un agent étant donné un contexte C, Boutilier a proposé dans [Boutilier, 1994] une logique basée sur une sémantique des mondes possibles permettant de représenter et raisonner sur des préférences et des probabilités avec des informations qualitatives. La sémantique des préférences est donnée grâce à un ordre sur les mondes possibles. Il en est de même pour la sémantique des probabilités. Cette logique permet, entre autres, de représenter des préférences conditionnelles (i.e. dépendantes du contexte) ainsi que des préférences défaisables (i.e. on peut en général préférer a à b mais, dans le cas particulier c donné, préférer b à a).

a) Logique proposée :

Pour ce faire, Boutilier a défini une logique propositionnelle quadri-modale appelée QDT (pour qualitative decision theory). Les formules bien formées sont construites sur la base d'un ensemble P de variables propositionnelles atomiques avec les opérateurs logiques classiques ainsi que les quatre opérateurs modaux $\Box_P$, $\overleftarrow{\Box}_P$, $\Box_N$ et $\overleftarrow{\Box}_N$.

La sémantique de ces formules est basée sur des modèles M qui sont des quadruplets de type $< W \ , \ \leq_P \ , \ \leq_N \ , \ \phi >$ où W est l'ensemble des mondes possibles, ϕ est une fonction de valuation, $\leq_P$ et $\leq_N$ sont des relations binaires complètes et transitives [7] sur W qui représentent respectivement les préférences de l'agent et la normalité (i.e. la probabilité d'occurrence) des différentes situations : $v \leq_P w$ indique que le monde v est au moins aussi préféré que w et $v \leq_N w$ indique que le monde v est au moins aussi normal que w. Plus précisément,

- La formule $\Box_P \alpha$ est vraie dans le monde w si et seulement si α est vraie dans tous les mondes au moins aussi préférés que w.
- La formule $\overleftarrow{\Box}_P \alpha$ est vraie dans le monde w si et seulement si α est vraie dans tous les mondes moins préférés que w.
- La formule $\overleftrightarrow{\Box}_P \alpha$ est vraie dans le monde w si et seulement si les formules $\overleftarrow{\Box}_P \alpha$ et $\Box_P \ \alpha$ sont vraies dans ce monde et signifie que α est vraie dans tous les mondes.
- La formule $\overleftrightarrow{\Diamond}_P \alpha$ est le dual de $\overleftrightarrow{\Box}_P \alpha$ (i.e. $\overleftrightarrow{\Diamond}_P \alpha \ = \ \neg \overleftrightarrow{\Box}_P \neg\alpha$) et signifie que α est vrai dans au moins un des mondes.
- La formule $\Box_N \alpha$ (respectivement $\overleftarrow{\Box}_N \alpha$) a le même sens que la formule $\Box_P \alpha$ (respectivement $\overleftarrow{\Box}_N \alpha$) excepté que l'on utilise la relation $\leq_N$ à la place de $\leq_P$.

L'axiomatique de cette logique est obtenue en ajoutant aux axiomes et aux règles d'inférence classique les règles d'inférence et les axiomes suivants :

NEC	Si A alors $\overleftrightarrow{\Box}_P A$	
	Si A alors $\overleftrightarrow{\Box}_N A$	
K	$\Box_P(A \Rightarrow B)$	$\Rightarrow \ (\Box_P A \Rightarrow \Box_P B)$
	$\Box_N(A \Rightarrow B)$	$\Rightarrow \ (\Box_N A \Rightarrow \Box_N B)$
K'	$\overleftarrow{\Box}_P(A \Rightarrow B)$	$\Rightarrow \ (\overleftarrow{\Box}_P A \Rightarrow \overleftarrow{\Box}_P B)$
	$\overleftarrow{\Box}_N(A \Rightarrow B)$	$\Rightarrow \ (\overleftarrow{\Box}_N A \Rightarrow \overleftarrow{\Box}_N B)$
H	$\overleftrightarrow{\Diamond}_P (\Box_P A \wedge \overleftarrow{\Box}_P B)$	$\Rightarrow \ \overleftrightarrow{\Box}_P(A \vee B)$
	$\overleftrightarrow{\Diamond}_N (\Box_N A \wedge \overleftarrow{\Box}_N B)$	$\Rightarrow \ \overleftrightarrow{\Box}_N(A \vee B)$
T	$\Box_P A$	$\Rightarrow \ A$
	$\Box_N A$	$\Rightarrow \ A$
4	$\Box_P A$	$\Rightarrow \ \Box_P \Box_P A$
	$\Box_N A$	$\Rightarrow \ \Box_N \Box_N A$
S	A	$\Rightarrow \ \overleftarrow{\Box}_P \Diamond_P A$
	A	$\Rightarrow \ \overleftarrow{\Box}_N \Diamond_N A$
PN	$\overleftrightarrow{\Box}_P A$	$\equiv \ \overleftrightarrow{\Box}_N A$

[7] $\leq_P$ et $\leq_N$ sont des pré-ordre totaux sur W

b) Le concept de préférence conditionnelle :

Pour Boutilier, une préférence est subjective et est en particulier relative à un contexte. Pour mettre en évidence ce phénomène il introduit une nouvelle notation : $I(\_|\_)$. La notation $I(B|A)$ dénote le fait que « dans les situations les plus préférées où A est vrai, B l'est aussi ». Elle est définie par la formule suivante :

$$I(B|A) \;=\; \overleftrightarrow{\Box}_P(\neg A) \vee \overleftrightarrow{\Diamond}_P(A \wedge \overleftrightarrow{\Box}_P(A \Rightarrow B))$$

Une préférence *absolue*, c'est-à-dire sans condition, est ainsi exprimée avec la formule $I(B|\top)$. Elle signifie que « Idéalement B est vrai » et est abrégée en $I(B)$. De plus, les comparaisons « classiques » (i.e. ce que Boutilier appelle des *préférences relatives*) peuvent être directement exprimées dans ce formalisme. Plus précisément, la propriété A est au moins aussi préférée que B si et seulement si les meilleurs A-worlds sont au moins aussi bons que les meilleurs B-worlds (i.e. $\overleftrightarrow{\Box}_P(B \Rightarrow \Diamond_P A)$).

La formule $I(B|A) \Rightarrow \neg I(\neg B|A))$ est un théorème de cette logique. Par conséquent les préférences absolues, tout comme les préférences dans un contexte fixé, sont consistantes entre elles.

c) Le concept de normalité :

Pour Boutilier, un agent ne doit pas seulement agir comme si seulement ses croyances (l'ensemble KB) étaient vraies mais aussi comme si tout ce qu'il pense le plus probable (l'ensemble $Cl(KB)$) l'est. Pour mettre en évidence ce phénomène, il introduit une nouvelle notation : $N(B|A)$ (notée $A \Rightarrow B$ dans [Boutilier, 1994]). La formule $N(B|A)$ signifie que B est vrai dans les mondes les plus normaux qui vérifient A. B peut donc être vue comme une conjecture par défaut étant donné A. Elle est définie de la même façon que $I(B|A)$ en utilisant les opérateurs $\overleftrightarrow{\Box}_N$ et $\Box_N$:

$$N(B|A) \;=\; \overleftrightarrow{\Box}_N(\neg A) \vee \overleftrightarrow{\Diamond}_N(A \wedge \overleftrightarrow{\Box}_N(A \Rightarrow B))$$

Étant donné un modèle de cette logique ainsi qu'un ensemble (fini) de faits KB, la clôture de KB est définie comme suit :

$$Cl(KB) = \{\alpha \text{ tel que } N(\alpha|KB)$$

d) Le concept de but idéal

C'est sur les notions de normalité et de préférences conditionnelles que Boutilier construit sa notion de but (idéal). Intuitivement, l'ensemble des buts (idéaux) est l'ensemble des propriétés vérifiées par les meilleures situations qui ont le plus de chance de se produire. Plus précisément, un but (idéal) est, en première approximation, une formule telle que dans les situations les plus préférées où l'ensemble des propriétés les plus probables sont vraies, la formule α est aussi vraie : étant donné un modèle M de QDT qui valide l'ensemble de fait KB,

$$\alpha \text{ est un but si et seulement si } M \models I(\alpha|Cl(KB))$$

e) Rapports avec l'approche de Doyle et Wellman :

Pour [Boutilier, 1994], les buts conditionnels de Doyle et Wellman sont très similaires à son approche. Cependant les buts conditionnels de [Boutilier, 1994] sont plus faibles et donc permettent plus de flexibilité : intuitivement le *toutes choses par ailleurs étant égales (équivalentes)* de [Doyle et al., 1991] est remplacé par un *toutes choses par ailleurs étant normales*. Pour [Doyle and Wellman, 1994] l'approche de Boutilier semble être plus proche de l'intuition même si elle limite la comparaison à tout ce qui est *normal*.

Ces deux approches nécessitent toutes les deux de définir, en plus de la relation binaire support des préférences, une autre relation : soit la relation indiquant la normalité des mondes, soit la relation indiquant l'équivalence entre mondes. Une voie intéressante serait donc de mixer ces deux approches mais ce n'est pas notre propos.

2.3.5 Les CP-nets : formalisme proposé dans [Boutilier et al., 1999]

Les CP-nets ont été introduits dans [Boutilier et al., 1999] comme un outil pour représenter de façon compacte et intuitive des préférences. C'est une représentation graphique qui permet de représenter des préférences conditionnelles de façon qualitative. Plus particulièrement c'est le premier formalisme graphique qui exploite l'indépendance de variables via l'hypothèse *Ceteris Paribus* afin de représenter des préférences de façon compacte et efficace. En fait, bien que l'hypothèse *Ceteris Paribus* ait été étudiée en Intelligence Artificielle bien avant les CP-nets (et en particulier dans [Wellman and Doyle, 1991]), cela n'a jamais été fait de façon graphique.

De nos jours, ce formalisme attire l'attention de beaucoup d'autres chercheurs (voir entre autres [Domshlak, 2002] et [Rossi et al., 2004]) car dans de nombreux domaines il est important de représenter les préférences d'un utilisateur de façon qualitative plutôt que de façon quantitative et surtout car son aspect graphique le rend attrayant et simple. C'est pour cela que nous présentons ce formalisme dans cette section. Pour plus de détails, le lecteur intéressé pourra se reporter au chapitre 4 de [Domshlak, 2002] et à [Boutilier et al., 2004a] (présentation plus complète) ainsi qu'à [Boutilier et al., 2004b] (un exemple d'utilisation des CP-net dans le cadre d'un problème de satisfaction de contraintes (CSP)).

a) Préliminaires

Ce modèle a été proposé afin de traiter les problèmes de décision sans ignorance qui se concentrent sur une seule décision. Il laisse donc de côté les problèmes liés aux décisions séquentielles ainsi que ceux dûs aux environnements incertains. Plus précisément, Boutilier et al. font l'hypothèse qu'il n'y a pas d'incertitude associée aux alternatives (les alternatives sont appelées « outcomes » par Boutilier) et que la situation est parfaitement connue. En effet, au moment de la décision, le monde est dans une situation donnée S. Celle-ci contraint l'ensemble A_S des actions qui peuvent être effectuées.

Les auteurs de [Boutilier et al., 1999] reprennent l'hypothèse (fréquente) selon laquelle chaque alternative peut être décrite par un ensemble fini de n variables logiquement indépendantes $X_1, \ldots, X_n$ ([von Wright, 1972], [Keeney and Raiffa, 1976], [Wellman and Doyle, 1991], ...). En particulier, lorsque ces dernières sont toutes instanciées, elles définissent une unique alternative. Une décision revient donc à choisir une assignation complète. Pour ce faire, le décideur doit spécifier, d'une manière ou d'une autre, un pré-ordre total sur les alternatives. Cela va être fait ici sur la base des n variables.

b) La notion d'indépendance

Spécifier directement un pré-ordre sur les alternatives est souvent impossible en raison de la taille exponentielle de cet ensemble. Heureusement, cela est plus facile (du moins partiellement) si les préférences ont des structures particulières. C'est pourquoi Boutilier et al. se sont concentrés sur les préférences qui sont exprimées sous l'hypothèse *Ceteris Paribus*. La notion d'indépendance joue donc ici un rôle important : Un sous-ensemble de variables X de V est *indépendant préférentiellement* (par rapport à la préférence $>$) de son complément Y (i.e. $Y = V \setminus X$) si et seulement si, pour toutes valuations x_1 et x_2 des variables de X et pour toutes valuations y_1 et y_2 des variables de Y, on a :

$$x_1 y_1 > x_2 y_1 \quad \text{si et seulement si} \quad x_1 y_2 > x_2 y_2$$

Si cette relation est vérifiée, on dit alors que x_1 est préféré *Ceteris Paribus* à x_2. Cette notion est une variante qualitative de la notion d'*utilité indépendance* de la théorie de l'utilité multi-attributs (voir la section 2.2.2).

Le fait que les préférences soient relatives à un contexte (au sens de Hansson) est pris en compte en conditionnant les préférences sur les valeurs d'une variable avec les valeurs de certaines autres variables. Par conséquent, et de façon analogue, X est dit *indépendant préférentiellement* (par rapport à la préférence $>$) de son complément Y étant donnée la valuation z des variables de Z si et seulement si, pour toutes valuations x_1 et x_2 des variables de X et pour toutes valuations y_1 et y_2 des variables de Y, on a :

$$x_1 y_1 z > x_2 y_1 z \quad \text{si et seulement si} \quad x_1 y_2 z > x_2 y_2 z$$

c) Formalisation

CP-net (Définition) : Un CP-net sur l'ensemble des variables[8] $V = \{X_1, \ldots, X_n\}$ est un graphe orienté G qui a pour nœuds les variables de V, ainsi qu'un ensemble de tables (une par nœud). Le graphe représente l'ensemble des dépendances entre les variables : les *parents* d'une variable sont les variables qui influencent les préférences sur les valeurs de cette variable. Chacune de ces tables ($CPT(X_i)$) décrit un ordre total sur les valeurs de la variable X_i pour chaque valuation des variables des parents de X_i ($Pa(X_i)$).

[8] Chaque variable a un domaine fini

Exemple : Afin d'illustrer cette définition, considérons l'exemple suivant qui exprime les préférences d'une personne au sujet des différents vêtements possibles à mettre pour sortir. Ces préférences portent sur les quatre variables (binaires pour plus de simplicité) V, P, C et S indiquant respectivement les couleurs de la veste, du pantalon, de la chemise et des sandales à mettre. Cette personne préfère quoi qu'il arrive (i.e. sans condition) que son pantalon soit noir plutôt que blanc. Il en est de même pour sa veste. Sa préférence pour la chemise est plus complexe : si le pantalon et la veste ont la même couleur alors elle préfère une chemise rouge ; si le pantalon et la veste ont des couleurs différentes alors elle préfère une chemise blanche. Enfin, si elle a une chemise rouge, elle préfère mettre des sandales rouges. Dans le cas contraire, elle préfère mettre des sandales noires. Toutes ces informations sont décrites par le CP-net de la figure 2.4.

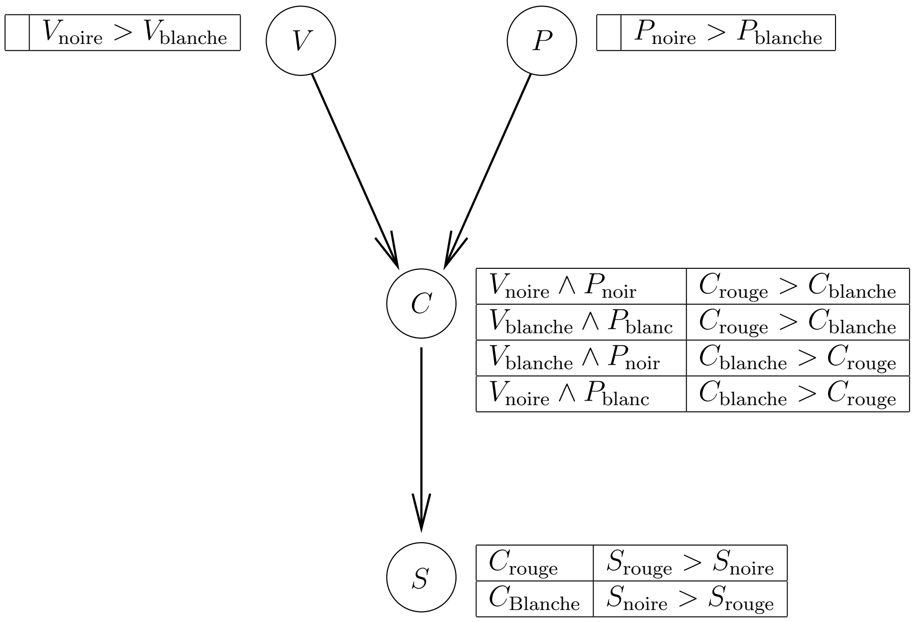


FIG. 2.4 – Exemple de CP-net

Conséquence de la définition : Un CP-net induit une relation binaire sur les assignations complètes des variables du problème (i.e. les alternatives). Cette dernière peut être illustrée par un graphe (voir le point (B.2.3 de l'annexe B.2) dont les nœuds représentent les alternatives possibles et dont les arcs représentent les préférences entre ces alternatives : il existe un arc (orienté) du nœud o au nœud o' si et seulement si les assignations en o et o' diffèrent seulement de la valeur de la variable X, et si, étant données les valeurs données aux parents de X (par o et o'), la valeur assignée par o' à X est préférée à la valeur assignée par o à X. L'intérêt du CP-net est de permettre de représenter très succinctement cette relation binaire sur l'ensemble des alternatives. Pour s'en convaincre, considérons le graphe de préférences sur les alternatives possibles défini par les informations de l'exemple précédent (voir la figure 2.5). Ce dernier est bien plus grand et complexe à appréhender.

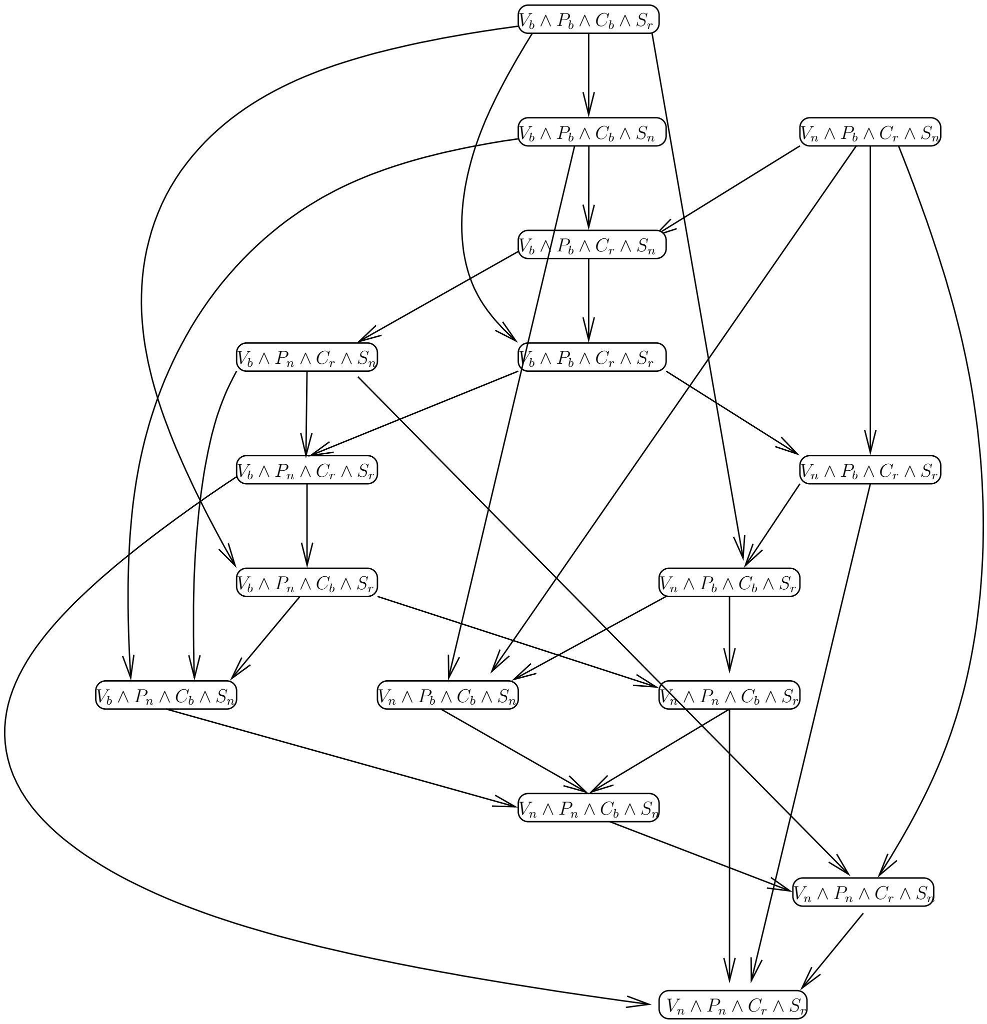


FIG. 2.5 – Graphe de préférences correspondant au CP-net précédant

d) Schéma général d'élicitation

Durant l'élicitation des préférences, le décideur est amené à indiquer, pour chaque variable X, les variables qui sont susceptibles d'affecter les préférences du décideur sur les valeurs de la variables X. L'ensemble de ces variables, appelé les parents de la variable X, est noté $Pa(X)$. En particulier, pour chaque valuation particulière des variables de $Pa(X)$, le décideur doit être capable de déterminer un ordre sur les valeurs de la variable X, toutes choses étant égales par ailleurs. Formellement, X et $V \setminus \{X, Pa(X)\}$ sont « conditionally preferentially independent » étant donné $Pa(X)$. Ces informations sont utilisées pour créer le graphe (voir la figure 2.4) du CP-net dans lequel chaque nœud est une variable (X) et a pour parents les variables appartenant à $Pa(X)$. Étant données ces informations structurelles, le décideur est invité à spécifier explicitement ses préférences sur les valeurs possibles de chaque variable X pour chaque valuation des variables de $Pa(X)$. Ces informations sont inscrites dans la *conditional preference table* (CPT) de chaque variable. Ainsi, pour toute valuation des variables appartenant à $Pa(X)$, la table CPT(X) définit un ordre total sur les valuations de la variable X.

e) Raisonner avec les CP-nets

L'objectif de toute modélisation des préférences est de permettre de comparer des alternatives entre elles et de déterminer la (ou une des) meilleure(s).

Afin de déterminer la meilleure alternative il suffit de parcourir le CP-net de haut en bas en positionnant la valeur des variables à la valeur préférée étant donnée la valeur des parents. Si le CP-net contient des cycles alors il peut exister plusieurs meilleures alternatives.

Afin de comparer des alternatives entre elles, la notion de changement élémentaire (dénommé *Flip*) est introduite. Il existe un changement élémentaire (*Flip*) entre deux alternatives si elles ne diffèrent que par la valeur d'une des variables (i.e si on passe de l'une à l'autre simplement en changeant la valeur d'une variable). Ce changement est dit *aggravant* (*worsening flip*) si l'alternative ainsi obtenue est moins préférée que l'alternative initiale. Une alternative A_1 est dite préférée à une alternative A_2 s'il existe une suite de changements élémentaires aggravants permettant de transformer A_1 en A_2. En d'autres termes A est préférée à B s'il existe une suite d'alternatives $\{A_i\}$ telle que $A_1 = A$, $A_n = B$ et telle que, quel que soit i, l'alternative A_i est obtenue de l'alternative A_{i-1} via un unique changement élémentaire et A_{i-1} est préférée à A_i.

f) Critique

La notion de changement élémentaire (*Flip*) utilisée dans CP-nets permet de comparer deux alternatives sans passer par le calcul d'une fonction d'utilité. Malheureusement, dès que les critères sont contradictoires, plus rien ne peut être déduit. C'est le cas par exemple lorsqu'il existe une suite de changements élémentaires entre les deux alternatives constituées par deux suites de changements élémentaires « aggravants » reliées par un changement « améliorant ». De plus l'existence d'une suite de changements élémentaires aggravants n'implique pas, a priori, l'absence de suite de changements élémentaires améliorante. Que faire dans de tels cas ? Enfin, comme les comparaisons se font entre les alternatives à départager via toutes les alternatives imaginables, il ne nous semble pas que cette approche soit adaptée pour traiter les situations réelles où les caractéristiques de chaque alternative sont potentiellement infinies.

2.3.6 Autres travaux sur les préférences qualitatives

Nous venons de présenter quatre formalisations qualitatives de la préférence. Notre propos ne se prétend pas être exhaustif ; il en existe d'autres (voir en particulier [Lang, 2004] pour une étude des représentations logique de la préférence d'un point de vue computationnel). Nous avons choisi de ne pas approfondir leur étude car soit elles semblent moins intéressantes d'un points de vue cognitif soit elles considèrent souvent les alternatives dans leur globalité au lieu de les appréhender par des descriptions partielles.

Tsoukias et Vincke ont proposé dans [Tsoukias and Vincke, 1995] d'utiliser une logique à quatre valuations (true, false, none, both) pour représenter les préférences (voir [Tsoukias, 2002] pour un exposé de cette logique). Leur travail est motivé par le fait que dans certaines situations il est aussi important de modéliser le manque, l'incertitude, l'ambiguïté, ou le conflit des informations déterminant les préférences d'un décideur.

Pour Kaci et van der Torre, il est intéressant de représenter simultanément plusieurs types de préférences afin de modéliser le comportement d'un agent. C'est pourquoi ils ont défini dans [Kaci and van der Torre, 2005] seize types de préférences différentes (basées pour certaines sur l'hypothèse *Ceteris Paribus* définie à la façon de [Doyle and Wellman, 1994]) et développé une logique non-monotone pour raisonner sur les interactions entre ces différents types de préférences.

Notons aussi l'utilisation possible pour modéliser des préférences, de la logique possibiliste (voir [Dubois et al., 1994] [Dubois et al., 2005]), de logiques non monotone (voir [Brewka et al., 2004]), ainsi que des relations de comparaisons *floues* (« fuzzy preference » en anglais) (voir [Chiclana et al., 1996] [Herrera et al., 2001]).

Notons enfin l'existence de nombreux travaux sur les préférence qualitatives dans d'autres domaines, en particulier dans les domaines relatifs à l'aide à la décision (voir [Vincke, 1989], [Druzdzel and Flynn, 2002] et [Mousseau, 2003]) et à l'agrégation des préférences (voir par exemple [Dubois et al., 2001], [Tsoukias et al., 2002], [Rossi et al., 2004]). Par exemple, dans [Tsoukias et al., 2002] Tsoukias a proposé de définir le sens d'une préférence par le fait qu'une alternative x est mieux qu'une alternative y s'il existe une majorité de « raisons » indiquant que x est mieux que y et qu'il n'existe pas d'opposition forte pour préférer x à y. L'idée sous-jacente est de considérer de façon séparée les arguments positifs et négatifs pour préférer une alternative à une autre. L'idée d'utiliser les principes de concordance / discordance dans une décision date de 1968 avec les travaux de Roy. C'est d'ailleurs sur cette idée que se base le système d'analyse de décision multicritères *ELECTRE* (voir [Vincke, 1989, p.87]).

2.4 Conclusion

Nous avons présenté dans les sections précédentes les représentations numériques et qualitatives de la notion de préférence. Dans cette partie de chapitre, nous allons exposer les limites de la théorie de la décision classique ainsi que les intérêts d'une représentation des préférences basée sur des comparaisons dans l'optique de les utiliser pour modéliser une phase de décision au sein d'un agent rationnel dialoguant.

2.4.1 Limites de la théorie de la décision « classique »

Dans la littérature et dans la pratique, les représentations numériques de la préférence sont les plus utilisées ; l'utilité de chaque alternative est calculée et celle ayant la valeur la plus forte est choisie comme réaction. Ceci est une conséquence du fait que les formalismes développés avec ces représentations ont atteint une grande maturité et surtout permettent de représenter et de manipuler de façon simple et puissante la désirabilité des différentes alternatives ainsi que le risque et l'incertitude des différentes situations mises en jeu. C'est pourquoi nous avons présenté rapidement dans la section 2.2 les deux formalismes les plus importants dans ce domaine (i.e. [von Neumann and Morgenstern, 1947] et [Keeney and Raiffa, 1976]). Malheureusement, la mise en œuvre de tels formalismes numériques n'est souvent pas une option réaliste car ils souffrent des trois inconvénients suivants.

Comme nous l'avons remarqué dans la partie (a) de la section 2.1.3, il n'est pas toujours possible de représenter des préférences de manière numérique. Cela suppose en particulier que tout peut être comparé.

De plus, le concept d'utilité espérée ne capture pas de façon réaliste la notion de préférence. En effet, l'expérience prouve que les préférences n'évoluent pas de façon linéaire avec la probabilité (voir le paradoxe d'Alais [Allais, 1953]) et que les préférences vérifient rarement l'hypothèse de cardinalité :

> « Expected utility does not capture a realistic notion of preference, either because expectations are nonlinear in the probabilities, or because utility is not cardinal. » [Doyle, 1994, p.16]

Enfin, les efforts nécessaires pour mettre en œuvre ces formalismes ne se justifient pas toujours et en particulier si « l'optimalité » du résultat n'est pas cruciale. En effet, et de façon générale, une phase d'élicitation (récupération des informations) qui permet de générer une fonction d'utilité adaptée requiert beaucoup de temps et d'efforts de la part de la personne qui doit exprimer ces préférences (le décideur de référence). Elle nécessite un effort de spécification très important puisqu'elle demande de décrire précisément chaque alternative en ayant à l'esprit le problème dans sa globalité (voir [Doyle and Wellman, 1994], [Ha and Haddawy, 1999] et [Faltings et al., 2004]). Dans certaines applications, ces efforts sont nécessaires et possibles. C'est le cas par exemple pour des décisions politiques ou médicales. Toutefois, dans de nombreuses situations courantes, éliciter une bonne fonction d'utilité n'est pas une option réaliste. En effet, comme le remarque Simon dans [Simon, 1960, p2], la phase de sélection est souvent très réduite chez tout un chacun. C'est le cas en particulier lorsque l'incertitude n'est pas un facteur déterminant ou lorsque le décideur de référence n'a pas le temps de le faire. Ainsi par exemple, un utilisateur d'un site commerçant sur internet comprendrait difficilement qu'un système d'aide lui impose de donner des notes à tous les différents produits mis en vente ou à toutes les caractéristiques qui les différencient. Dans de telles situations, il est préférable d'utiliser d'autres méthodes qui permettent d'obtenir les préférences du décideur (nécessaires pour le choix) plus facilement et de façon moins intrusive, moins indiscrète et moins importune [Brafman and Tennenholtz, 1997] [Boutilier et al., 2004b][Ha and Haddawy, 1999].

2.4.2 Intérêts d'une représentation des préférences basée sur des comparaisons

L'approche symbolique fournit une représentation des préférences plus proche de celle de l'être humain que ne l'est l'approche numérique. En effet, une personne voulant par exemple indiquer que, face à un choix entre un plat contenant de la viande et un plat contenant du poisson, il prendra celui contenant du poisson, spécifiera plus volontiers cette information par « prefere(viande,poisson) » que par « viande $= 0,12$ et poisson $= 0,567$ ». Une telle représentation est donc souvent naturelle ; elle est aussi compacte et adaptée aux situations de la vie courante et donc à de nombreuses situations dans lesquelles le traitement automatique de la langue naturelle (TALN) peut être utilisé. Pour s'en convaincre, il suffit de considérer les exemples de dialogue imaginés 2.6 et 2.7. Ceci est dû au fait que, (1) l'humain utilise des représentations compactes pour ne pas dépasser ses capacités limitées et que, (2) ses raisonnements se basent souvent sur les représentations qu'il utilise pour s'exprimer et que, (3) la cognition humaine a l'habitude d'aller directement à l'essentiel. Pour s'en convaincre, il suffit de remarquer l'utilisation fréquente chez tout un chacun de comparatifs et de superlatifs. Par conséquent, si l'on s'inspire du modèle humain pour modéliser les agents, il semble plus naturel de faire appel à des techniques de décision qualitatives que quantitatives [Brafman and Tennenholtz, 1997][Rossi et al., 2004].

Système	:	Sur une échelle de 1 à 10 telle que 1 indique que vous détestez et 10 que vous adorez, quelle valeur attribuez-vous à la couleur bleue ?
Utilisateur	:	Heu... , je sais pas trop, 6 ?
Système	:	Sur cette même échelle, quelle valeur attribuez-vous à la couleur rouge ?
Utilisateur	:	Heuu... laissez-moi réfléchir ... 8 ?

FIG. 2.6 – Dialogue élicitant des valeurs d'utilité

Système	:	Préférez vous la couleur bleue ou la couleur rouge ?
Utilisateur	:	La couleur rouge
Système	:	Et entre la couleur bleue et la couleur verte ?
Utilisateur	:	La bleue
Utilisateur	:	En fait, je préfère toutes les couleurs au vert.

FIG. 2.7 – Dialogue élicitant des comparaisons

De plus, comme le remarquent les auteurs de [Dubois et al., 2001] en introduction, afin d'avoir une méthode de décision plus homogène avec la description interne de l'agent via des attitudes mentales, utiliser une technique *qualitative* plutôt que *quantitative* semble préférable. En effet, dans les modèles cognitifs d'agents qui nous intéressent, les autres attitudes mentales sont décrites logiquement.

Enfin, dans la perspective d'une utilisation par le « grand public », l'utilisation de représentations logiques de la préférence est plus intéressante car elles sont plus intuitives. De façon générale, trouver un moyen de formaliser la connaissance qui permette aux humains de la spécifier facilement et aux machines de la manipuler efficacement est un point central dans le domaine de la représentation des connaissances. L'expérience montre que, pour que cela soit spécifié facilement par des humains, il est souvent nécessaire que ce soit proche de la façon dont un humain s'exprime. Ceci est en particulier vrai pour la préférence. Aussi, comme il est a priori plus naturel d'exprimer des préférences de façon qualitative que numériquement, certains chercheurs ont proposé d'utiliser des informations qualitatives afin de simplifier la spécification des fonctions d'utilité [McGeachie and Doyle, 2002] [Ha and Haddawy, 1999] [Ha, 2001]. Nous pensons pour notre part qu'il est intéressant de développer des méthodes entièrement qualitatives pour la raison suivante. Les êtres humains, qu'ils soient experts ou non, résolvent fréquemment et de manière naturelle des problèmes de décision que les formalismes et les méthodes numériques ne sont pas capables de résoudre. Or, d'après Simon un humain dispose de ressources psychologiques limitées, donc de capacités calculatoires et prédictives limitées et, par conséquent, ses efforts de raisonnement peuvent au mieux être une approximation grossière de la rationalité globale qui est par exemple appliquée dans la théorie des jeux ou dans la théorie de la décision « classique » [Simon, 1982]. Plus précisément, comme le laisse entrevoir le premier paragraphe de cette section, tout un chacun traite les préférences plutôt de manière logique que de manière numérique. Par conséquent, il est intéressant de considérer des méthodes différentes de celles utilisant des représentations numériques car ainsi le nombre de problèmes de décision traités pourra être augmenté. En particulier l'approche symbolique semble utile pour traiter les informations conflictuelles dans des raisonnements non monotones, pour raisonner sur les actions et le temps, éliciter les informations nécessaires au choix auprès d'un décideur et plus généralement pour la représentation des connaissances et le raisonnement.

Il en résulte que l'approche symbolique fournit une nouvelle perspective pour formaliser des informations qui sont essentielles pour de nombreux problèmes de décision comme par exemple l'aide à la décision, l'ordonnancement ou la planification en robotique. C'est pourquoi les approches symboliques/logiques visant à représenter les préférences attirent de plus en plus l'attention en Intelligence Artificielle. [Wellman and Doyle, 1991][Boutilier, 1994][Doyle and Wellman, 1994][Boutilier et al., 1999] [Doyle and Thomason, 1999] [Ha and Haddawy, 1999] [Dubois et al., 2001] [Ha, 2001] [Domshlak, 2002] [Lang, 2004] [Mousseau, 2003] [Benferhat et al., 2004] [Faltings et al., 2004] [Coste-Marquis et al., 2004] [Rossi et al., 2004] [Kaci and van der Torre, 2005].

2.4.3 Conclusion

Comme le décrit l'article de Doyle et Thomason [Doyle and Thomason, 1999], l'Intelligence Artificielle fournit des méthodes qualitatives pour traiter les préférences qui peuvent améliorer ou compléter les méthodes numériques de la théorie de la décision « classique » et ainsi en augmenter leur champ d'application (planification, apprentissage, collaboration). Pour Doyle et Thomason ceci est nécessaire car, malgré la remarquable maturité des outils développés dans le cadre de la théorie de la décision classique (quantitative), ceux-ci ne sont pas adaptés aux récents problèmes (comme le TALN ou l'introduction dans des applications de « sens commun ») que l'Intelligence Artificielle se propose de résoudre et en particulier ceux qui doivent mettre en œuvre des décisions automatiques. En effet, dans de nombreux domaines il est important de représenter les préférences d'un utilisateur de façon qualitative plutôt que de façon quantitative et de le faire le plus intuitivement possible. Ceci est particulièrement vrai pour formaliser des systèmes à base d'attitudes mentales devant être opérationnels dans des situations « courantes » comme celles qui peuvent se rencontrer dans le cadre du TALN.

Deuxième partie

Principes et outils retenus pour un nouveau formalisme logique des préférences adapté aux prises de décision d'un agent rationnel

Chapitre 3

Notre proposition dans les grandes lignes

Dans ce chapitre nous présentons notre proposition pour une nouvelle formalisation logique de la préférence adaptée aux agents rationnels dialoguants. Pour ce faire, dans un premier temps nous exposons les raisons qui nous ont poussés à proposer une formalisation logique de la préférence ainsi que les hypothèses sur lesquelles se base notre proposition. Dans un second temps, nous décrivons de façon globale le schéma de construction de notre modèle. Enfin, dans un troisième temps, nous terminons ce chapitre en comparant notre travail à quelques travaux connexes ainsi qu'en exposant les intérêts de notre proposition.

Ce chapitre ne vise qu'à donner un aperçu de notre proposition dans sa globalité. Le développement technique détaillé de notre proposition est présenté dans les chapitres 4, 5, 6 et 7.

3.1 Pourquoi proposer une nouvelle modélisation logique des préférences ?

Comme nous l'avons souligné dans la section 1.1, les modèles cognitifs d'agents sont attrayants pour modéliser des systèmes informatiques complexes. Ceci est en partie du au fait que les systèmes informatiques ainsi modélisés par un ou plusieurs agents rationnels, sont représentés de façon abstraite avec des notions intuitives. Ceci est particulièrement intéressant pour les systèmes dans lesquels le « bon sens » tient une place prépondérante, comme en traitement automatique de la langue naturelle.

Un agent rationnel est avant tout une entité qui perçoit son environnement et qui agit dessus selon le principe de rationalité : il met en oeuvre à chaque instant la réaction qu'il juge la meilleure. C'est pourquoi tout agent rationnel a besoin d'un mécanisme de « sélection » pour déterminer la réaction à mettre en œuvre. Comme un agent a des capacités limitées ,en particulier parce qu'un agent n'a généralement pas un accès complet à toutes les informations qui lui sont nécessaires, Bratman a défini dans [Bratman, 1987] le raisonnement pratique (et donc le mécanisme de sélection de la réaction de l'agent) en termes d'équilibre entre des connaissances (conflictuelles) pour ou contre les choix possibles en compétition (voir la section 1.1.2). Toutefois, comme nous l'avons remarqué dans la section 1.2, les mécanismes et informations utilisés par les théories basées sur cette approche sont souvent insuffisants pour spécifier complètement la réaction de l'agent. En particulier, dans les situations complexes où l'agent a plusieurs motivations et/ou plusieurs moyens de les satisfaire, ce dernier se retrouve souvent sans savoir choisir explicitement sa réaction alors que n'importe quel humain moyen le ferait sans difficulté. C'est le cas par exemple pour un agent dialoguant qui, dans le cadre d'une interaction multimodale et pour présenter de façon adéquate un itinéraire à un utilisateur, doit choisir entre les modalités vocale ou graphique selon le contexte. Dans les formalismes actuels, il se retrouve dans une situation où, ayant l'intention de fournir un itinéraire particulier[1] et bien que connaissant deux moyens de le faire, il reste « bloqué ».

C'est pourquoi il est nécessaire d'introduire une phase de décision explicite dans un agent rationnel et ainsi faire explicitement le lien entre l'état mental et la réaction de l'agent [Haddawy and Hanks, 1993][Louis, 2002]. Notons que cette phase se base sur des informations supplémentaires décrivant la désirabilité des différentes alternatives mises en jeu : les préférences.

Nous venons de voir, dans le chapitre 2, qu'il existe de nombreux formalismes pour mettre en oeuvre une phase de décision. Les plus aboutis manipulent des représentations numériques à travers la notion d'utilité. En ce qui nous concerne, nous pensons comme Domshlak que représenter les préférences « à la manière » de l'intuition d'utilisateurs naïfs, et en particulier en utilisant la façon dont ils expriment des préférences dans les activités de tous les jours (i.e. à base de comparaisons), peut permettre de spécifier des processus de décision plus adapté aux situations courantes [Domshlak, 2002, p17]. Plus précisément, nous pensons que pour être adaptée à une modélisation cognitive d'agent (et en particulier celle présentée dans la section 1.1.3), la problématique de décision doit se traiter logiquement et avec des informations qualitatives pour les deux raisons suivantes :

- Dans l'idéal, un agent rationnel doit posséder des aptitudes de raisonnement très générales et facilement adaptables à des domaines d'application variés. Il en résulte que d'ordinaire, on n'attend pas de lui la résolution détaillée de gros problèmes spécialisés en optimisant le temps de calcul. Il est plutôt destiné à réagir de façon sensée et appropriée dans des situations d'interaction avec d'autres entités « intelligentes » (humaines ou artificielles). Ceci nécessite plus de savoir appréhender le « sens commun » que d'assurer de bonnes performances calculatoires. À cet effet, l'utilisation de notions intuitives est un atout puisqu'il sera d'autant plus aisé de les acquérir au cours d'une interaction. Or, comme nous l'avons montré à la fin du chapitre précédent, dans cette optique, les représentations qualitatives sont plus adaptées que les représentations numériques. De façon générale, nous pensons que les représentations logiques de la préférence sont un bon compromis entre facilité de spécification et précision obtenue pour les informations nécessaires à la phase de décision.

[1]Dans le formalisme de [Sadek, 1991], si ce n'est pas explicitement demandé par l'utilisateur, il n'y a pas de sens qu'un agent ait des intentions assez précises pour indiquer explicitement s'il doit donner l'information en mode vocal ou graphique. En effet, dans ce formalisme, les intentions dirigent la planification (voir le point (e) de la section 1.1.3).

– De plus, comme les attitudes mentales de la théorie de l'interaction ([Sadek, 1991]) sont représentées logiquement, une modélisation logique de la préférence facilitera l'intégration de notre formalisme à ce modèle d'agent puisqu'ainsi toutes les notions auront une représentation homogène.

3.2 Hypothèses de notre travail

À ce niveau, il nous semple opportun de préciser les hypothèses que nous faisons pour modéliser les préférences afin de réduire notre champ d'investigation.

3.2.1 Caractéristiques des informations initiales

a) Les alternatives

Les alternatives correspondent aux objets parmi lesquels l'agent doit faire son choix. La définition de l'ensemble des alternatives (noté ALT^*) est une phase fondamentale et délicate de la modélisation du processus de décision. Une telle définition n'est ni neutre ni triviale :

> « L'ensemble ALT^* ne s'impose généralement pas comme une réalité objective facile à cerner. » [Vincke, 1989].

Pour notre part nous supposons que cet ensemble est connu et en particulier défini en extension. Il est de surcroît stable (n'évolue pas durant le choix) et fini. Plus précisément, dans notre vision, le résultat de la décision est le choix d'une alternative parmi un nombre fini d'alternatives quelconques résultant par exemple d'un processus de planification.

Néanmoins, comme nous pensons qu'un « bon » processus de décision doit permettre de comparer n'importe quel couple d'alternatives, nous considérons que l'ensemble des alternatives « imaginables » (noté ALT) est infini et tel que chacun de ses éléments est exclusif par rapport aux autres. Dans notre vision, l'ensemble des alternatives ALT^* est un sous ensemble fini de l'ensemble des alternatives « imaginables » ALT.

Afin de décrire ces alternatives (les éléments de ALT^*), nous supposons que l'agent possède une base de connaissances. Pour faire un parallèle avec les logiques de description, cette base peut être considérée en première approximation comme une K-Box. Plus précisément, chaque alternative est appréhendée par l'agent via une description partielle des propriétés qu'elle vérifie (A-box). De plus l'agent a des connaissances sur l'état du monde et sur sa dynamique (T-box). En particulier il est possible qu'il ait connaissance de certaines implications de ces alternatives. Par exemple, l'agent sait qu'un plat de poisson n'est pas une pizza.

b) Informations sur la « désirabilité » des alternatives

Comme nous l'avons indiqué dans le premier paragraphe de cette section, nous avons choisi de représenter les informations de « désirabilité » de manière logique car nous trouvons que c'est le mode de représentation qui correspond le mieux aux modélisations cognitives d'agents rationnels.

Nous avons indiqué dans la section 2.3.1 que ces informations sont données généralement de façon implicite dans le sens où il est rarement exprimé explicitement qu'une alternative est la meilleure ou qu'une alternative est mieux qu'une autre. Plus précisément, les informations données sont seulement des comparaisons entre certaines propriétés que les alternatives peuvent vérifier. Dans ce même point, nous avons aussi remarqué que ces comparaisons ont de surcroît souvent un sens implicite. Par exemple, lorsqu'un utilisateur indique simultanément qu'il préfère le « rouge » au « bleu » et le « bleu » au « vert », implicitement il indique aussi qu'il préfère le « rouge » au « vert ». Par conséquent, pour nous, tout formalisme destiné à spécifier des préférences de façon naturelle et intuitive doit prendre en compte ces deux aspects. Plus précisément, une « bonne » modélisation logique de la préférence doit ainsi, d'une part manipuler les propriétés des alternatives et, d'autre part, automatiquement expliciter toutes les conséquences intuitives des informations qui ont été spécifiées.

C'est pourquoi nous faisons l'hypothèse que les informations initiales sur la désirabilité des alternatives sont des comparaisons (1) entre des propriétés que peuvent vérifier les alternatives, (2) entre des propriétés mutuellement exclusives, (3) transitives, (4) *Ceteris Paribus*, (5) possiblement conditionnelles (voir la section 2.3.1).

D'autre part, afin de proposer une modélisation « robuste », nous faisons l'hypothèse que les informations initiales sur la désirabilité peuvent être contradictoires entre elles. Pour nous ces contradictions peuvent être de deux types. Tout d'abord, elles peuvent révéler que les propriétés mises en relation sont équivalentes (au regard de la préférence). D'autre part, elle peuvent aussi être le résultat de l'existence de différents aspects de la décision en contradiction entre eux. Dans le premier cas, nous décidons de considérer alors que les propriétés mises en relation sont équivalentes. C'est pourquoi nous supposons que les informations initiales sont des préférences « faibles » : chacune indique qu'une propriété p_1 est *préférable ou équivalente* à une propriété p_2. Dans le second cas, nous considérons que le résultat dépend du rapport relatif entre les points de vue qui indiquent que la première propriété est préférable à la seconde et ceux qui indiquent l'inverse. C'est pourquoi, nous considérons que les préférences sont spécifiées par points de vue (des groupes d'informations) indépendants les uns des autres.

Remarquons ici que les informations sur la désirabilité des alternatives ne sont pas de même nature que celles permettant d'appréhender et de raisonner sur les propriétés des alternatives. Plus précisément et en faisant toujours le parallèle avec les logiques de description, l'introduction des préférences dans un modèle agent nécessite l'ajout d'une « P-Box » en plus de la K-Box de l'agent (voir la figure 3.1).

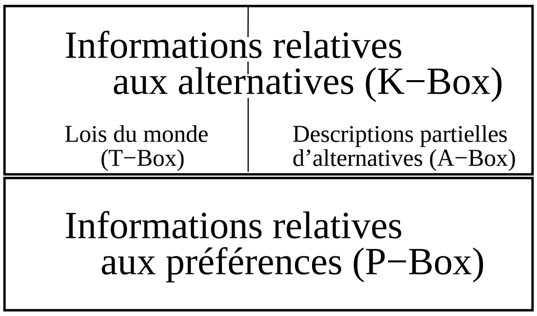


FIG. 3.1 – Schéma d'organisation des informations nécessaires à une décision

3.2.2 Limite de l'hypothèses *Ceteris Paribus*

L'hypothèse *Ceteris Paribus* a été largement étudiée et discutée en Philosophie car c'est une hypothèse relativement intuitive. Il en ressort que nombreux sont les chercheurs en Intelligence Artificielle et en Philosophie à penser que la spécification intuitive d'une préférence fait implicitement l'hypothèse *Ceteris Paribus* : lorsqu'une préférence est spécifiée via une comparaison, tous les aspects autres que ceux explicités dans l'une ou l'autre des parties mises en relation sont considérés comme étant égaux par ailleurs (voir le point (d) de la section 2.3.1). Par exemple, le fait d'exprimer que les plats à base de poisson sont préférés aux plats à base de viande signifie implicitement que les caractéristiques comme le prix et l'abondance sont les mêmes pour les deux plats comparés. De plus, les auteurs de [Coste-Marquis et al., 2004] ont montré que la spécification de préférences à l'aide de l'hypothèse *Ceteris Paribus* permet d'exprimer de façon succincte de nombreux pré-ordres partiels.

Néanmoins, rares sont les travaux qui proposent des modèles pratiques l'utilisant. Ceci provient du fait qu'une telle hypothèse ne permet pas de comparer tous les objets entre eux et en particulier ceux dont les caractéristiques (non spécifiées par la préférence) sont différentes : avec la préférence exprimée au paragraphe précédent, un poisson à la vapeur est-il préférable à une viande grillée ? En outre, cela devient encore plus délicat lorsque les préférences spécifiées induisent des contradictions (problème de gestion de contradictions) : qu'en est-il de l'exemple précédent lorsqu'il est spécifié de surcroît que les grillades sont préférées aux plats diététiques ? Plus précisément, la critique la plus fréquente de cette hypothèse, et en particulier de la formalisation qu'en a fait von Wright dans [von Wright, 1972], est qu'elle est trop forte pour être réaliste. Par exemple, comme le remarque Trapp dans [Trapp, 1985, p308], alors qu'il est sensé de dire qu'il est préférable d'avoir le choléra que le cancer, ce n'est pas une comparaison au sens de von Wight car ce ne serait pas le cas dans les mondes où le choléra serait incurable et le cancer curable. De la même façon, un client d'un restaurant qui préfère *Ceteris Paribus* les plats contenant du poisson à ceux contenant de la viande n'a, à la limite dans ces modèles, aucune raison de choisir un plat dénommé « dorade grillée » (contenant du poisson) plutôt qu'un plat dénommé « escalope de veau à la milanaise » (contenant de la viande). En effet puisqu'ils ont des noms différents, ils ne sont pas *Ceteris Paribus*. Il en résulte que ces modèles permettent seulement de comparer un petit nombre d'alternatives ce qui est bien insuffisant pour de nombreuses applications.

C'est pourquoi certains chercheurs ont proposé de modifier cette hypothèse. Ainsi par exemple, Doyle et Wellman ont proposé d'introduire le concept d'*équivalence* tandis que Boutilier a proposé d'introduire le concept de *normalité* (voir respectivement les sections 2.3.3 et 2.3.4). Pour Doyle et Wellman, ce problème provient du fait que de nombreuses propriétés ne sont pas pertinentes pour comparer les alternatives [Doyle and Wellman, 1994]. À l'inverse, pour Boutilier, ce problème provient du fait que le sens de nos préférences porte sur un nombre limité d'alternatives (i.e. pas sur toutes les alternatives « imaginables ») [Boutilier, 1994]. En effet, comme le remarque Domshlak dans [Domshlak, 2002], un humain ne raisonne pas en ayant à l'esprit les descriptions complètes des situations possibles. À cause de ses capacités cognitives limitées, il raisonne sur des modèles simplifiés de ces situations.

À notre avis, la limite de l'approche de von Wright se situe à un autre niveau et est plus profonde. Plus précisément, pour nous le problème provient du fait que cette hypothèse ne prend pas en considération le fait que chaque propriété (que ce soient celles comparées ou celles définissant le contexte d'une préférence) a des implications sur les autres propriétés. Par exemple, le nom d'une voiture (Testarossa, Cocinelle/Bettle, Mini) implique le nom de la marque de cette voiture (Ferrari, Wolkswagen, Austin). Or ces autres propriétés entrent aussi en considération afin de déterminer les alternatives qui vérifient cette propriété. Afin de dépasser cette limitation, il faut donc soit (1) déterminer les conséquences des formules entrant dans la comparaison et les écarter, soit (2) considérer ce qui n'est pas spécifié comme implicitement équivalent. Pour notre part, nous choisissons la seconde méthode. Toutefois, afin de ne pas interférer avec la propriété de transitivité, nous proposons de n'utiliser cette méthode qu'après avoir calculé la clôture transitive de l'ensemble des comparaisons initiales. Autrement dit, nous pensons qu'il faut différencier la phase d'élicitation des préférences de celle de son utilisation dans une phase de décision. Dans cette optique, nous pensons que les préférences exprimées sous l'hypothèse *Ceteris Paribus* doivent être étendues pour comparer les alternatives mises en jeu lors d'une phase de décision (pour plus de détails, voir la section suivante).

3.3 Le modèle proposé dans les grandes lignes

3.3.1 Idée directrice

L'objectif de cette thèse est de proposer un formalisme qui permette d'une part de spécifier de manière concise et naturelle les préférences (c'est à dire les informations de désirabilité pertinentes pour la décision), et d'autre part de faire des choix en se basant sur ces préférences.

Pour ce faire, l'idée directrice de notre travail est de bien différencier le moment de la spécification des informations qui servent à décrire la désirabilité des différentes alternatives mises en jeu et celui du choix entre ces alternatives à proprement parler. La spécification des informations permettant de départager les alternatives se fait en donnant des comparaisons entre des propriétés que les alternatives peuvent vérifier. Le choix, quant à lui, se fait sur les alternatives (c'est-à-dire les objets qui sont à départager) en utilisant les informations de désirabilité. Par exemple, lorsqu'il est indiqué que le poisson est préférable à la viande, c'est un ordre sur des propriétés que peuvent vérifier les alternatives qui est établi. Par contre lorsque, dans un restaurant, un client choisit un plat particulier (qui contient par exemple du poisson) c'est un objet qui est considéré : le plat que le client va manger.

Afin de mettre en œuvre cette idée et pour satisfaire nos objectifs tout en restant dans le cadre de nos hypothèses, nous pensons qu'il faut (1) supposer que les informations spécifiées initialement ont le sens qu'un utilisateur « commun » leur donnerait, (2) exploiter les informations spécifiées initialement afin de départager le plus d'alternatives possible et (3) gérer les contradictions entre les différents points de vue. Plus précisément,

1. Afin de permettre une spécification facile et intuitive des informations nécessaires aux choix, nous avons fait l'hypothèse que les préférences initiales vérifient les propriétés exposées dans la section 2.3.1 : ce sont des comparaisons (1) entre des propriétés que peuvent vérifier les alternatives, (2) entre des propriétés mutuellement exclusives, (3) transitives, (4) *Ceteris Paribus* et (5) possiblement conditionnelles (voir le point (b) de la section 3.2.1). En effet, comme nous l'avons noté au début de ce chapitre, le fait de permettre la spécification des informations de la manière dont tout un chacun les exprime, permet souvent de simplifier cette étape.
2. De plus, afin de comparer et de discriminer le plus d'alternatives sur la base de ces informations, il est nécessaire de dépasser la limitation de l'hypothèse *Ceteris Paribus* que nous avons mis en évidence précédemment dans la section 3.2.2. En effet, lorsqu'un choix est à effectuer, les préférences initiales ne sont souvent pas suffisantes pour départager les alternatives de ce choix. C'est pourquoi nous pensons qu'une phase dite « d'extension » est nécessaire. Plus précisément, nous considérons chacune des informations initiales comme un *argument* pour ou contre le choix d'une alternative plutôt qu'une autre.
3. Enfin, comme nous supposons que les informations sont spécifiées par points de vue indépendants, ces derniers peuvent être contradictoires entre eux. Aussi, afin de gérer ces contradictions, nous introduisons une phase d'élection (ou d'agrégation). Cette phase a pour but de générer un unique point de vue global à partir de l'ensemble des points de vue utilisés pour spécifier les préférences. À ce sujet, notons que déterminer une politique d'élection adaptée est difficile [Arrow, 1951]. Pour plus de détails voir [Bouyssou and Vincke, 2003, p18] et la section 7.6.2.

3.3.2 Schéma général

De façon schématique, nous proposons de scinder la phase de spécification de la préférence d'un agent en trois étapes (voir la figure 3.2).

Dans un premier temps, les données initiales de chaque point de vue sont interprétées selon les hypothèses du point (b) de la section 3.2.1 et en particulier selon l'hypothèse

Ceteris Paribus. Ceci génère ce que nous appelons une *préférence partielle primitive* (voir le chapitre 5). Cette phase dite de « génération » permet d'exprimer de façon naturelle et intuitive les données initiales.

Dans un deuxième temps, chaque préférence partielle primitive est étendue (au-delà de l'hypothèse *ceteris paribus*) en une *préférence partielle étendue* de telle sorte qu'une propriété p_1 est préférée à une propriété p_2 (via la préférence partielle étendue) si la propriété p_1 est préférée *strictement* à la propriété p_2 (via la préférence partielle primitive). Cette phase dite « d'extension » permet de discriminer de nombreuses alternatives qui ne l'étaient pas avec la préférence partielle primitive (voir le chapitre 6).

Dans un troisième temps, la *préférence globale* est construite, à partir de toutes les préférences partielles étendues spécifiées en parallèle pour chaque point de vue, via un mécanisme d'élection (voir le chapitre 7). Cette phase dite « d'agrégation » permet de gérer des contradictions entre les points de vue. C'est sur cette préférence globale que se base le choix final (voir la section 7.7).

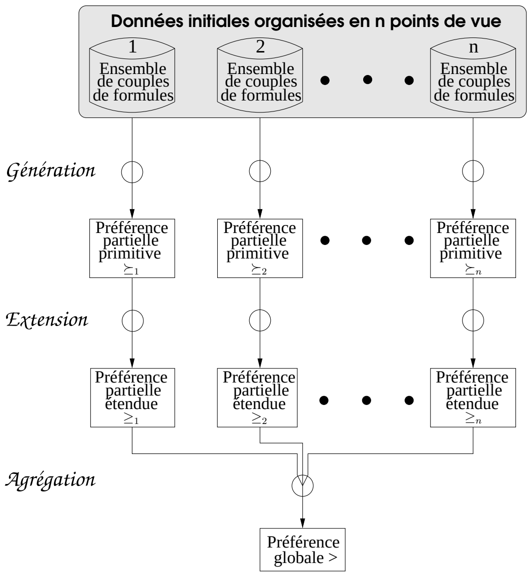


Fig. 3.2 – Construction des préférences.

Dans notre approche, chaque préférence est modélisée par une relation binaire entre les objets à comparer ou « alternatives ». Dans le contexte des agents rationnels auquel nous nous intéressons, ces objets correspondent aux états du monde (i.e. aux différents modèles de la logique que nous utilisons pour les formaliser). Or il est difficile voire impossible de représenter exhaustivement toutes les caractéristiques d'un état du monde (voir par exemple le problème du cadre [McCarthy and Hayes, 1969]). Notre modèle associe donc en outre à chaque préférence un opérateur logique portant sur des couples de formules décrivant partiellement les états à comparer. En résumé, chaque préférence partielle est associée à la fois à (1) une relation binaire sur l'ensemble des alternatives et à (2) un opérateur sur des couples de formules logiques représentant les alternatives. Dans ce sens, notre vision est proche de celle de Von Wright, Boutillier, Doyle et Wellman, selon laquelle les préférences sont spécifiées par des descriptions mais s'appliquent à des alternatives « concrètes » qui vérifient ces descriptions. Par ailleurs, la phase d'agrégation de notre modèle adapte le travail de Rossi et al. qui consiste à dériver une décision à partir des préférences de plusieurs agents [Rossi et al., 2004]. Cette adaptation s'appuie sur l'idée développée dans [Dubois et al., 2001] selon laquelle la décision multi-critères et la décision multi-agents sont deux approches différentes de la même problématique. Chacune de nos *préférences partielles* joue ainsi le rôle de la préférence d'un agent dans le modèle proposé par Rossi et al. : chaque point de vue est un « argument » en faveur du classement qu'il propose.

3.4 Travaux connexes et intérêts de l'approche

3.4.1 Rapports avec les travaux existants

Il existe dans la littérature de nombreux travaux en rapport avec le nôtre. Ceux-ci visent généralement soit (1) à utiliser le concept de préférence au sein d'un modèle cognitif agent [Horty and Pollack, 1998][Dastani et al., 2002b][Haddawy and Hanks, 1993][Wellman and Doyle, 1991][Brafman and Tennenholtz, 1997] soit (2) à formaliser ce concept logiquement à base de comparaisons [von Wright, 1972][Boutilier, 1994][Wellman and Doyle, 1991].

Travaux en rapport avec les modèles cognitifs d'agents : Dans le premier cas, ces travaux manipulent généralement le concept de préférence au travers de valeurs numériques (voir entre autres [Haddawy and Hanks, 1993] [Horty and Pollack, 1998] [van der Torre and Weydert, 1998] [Dastani et al., 2002b] [Brafman and Tennenholtz, 1997]). De plus, ils utilisent souvent ce concept afin de définir des attitudes mentales comme les buts ou les désirs de l'agent (voir entre autres [van der Torre and Weydert, 1998] [Wellman and Doyle, 1991] [Dastani et al., 2002b][Haddawy and Hanks, 1993]).

Ainsi par exemple, lorsque les auteurs de [Horty and Pollack, 1998] ont proposé un mécanisme pour permettre à l'agent d'évaluer la désirabilité des alternatives au cours de son évolution, ils ont employé des valeurs numériques.

De même, lorsque Brafman et Tennenholtz dans [Brafman and Tennenholtz, 1997] ont proposé de modéliser un agent en s'inspirant de la théorie de la décision, ils ont certes proposé de baser la modélisation d'un agent sur trois types de composantes clefs : les croyances, les préférences et un critère de décision. Cependant ils ont aussi utilisé une fonction d'utilité pour décrire les préférences de l'agent. Au sujet de cet article, ses auteurs sont à notre connaissance les premiers a avoir insisté, dans le cadre de modélisations cognitives d'agents, sur la nécessité d'un critère de décision lorsqu'une réaction peut avoir plusieurs résultats différents. Ce dernier indique la façon d'utiliser croyances et préférences afin de choisir la réaction à mettre effectivement en oeuvre. Ce critère peut consister, par exemple, à « choisir la réaction qui a en moyenne le meilleur résultat ». Ce critère peut se rapprocher des politiques d'agrégation et de décision que nous introduirons respectivement aux sections 7.4 et 7.7.

Travaux concernant les modélisations logiques de la préférence : En ce qui concerne les travaux relatifs à la modélisation logique de la préférence[2], ils nous semblent généralement mal adaptés à notre problématique. En effet, à notre connaissance, ils sont tous formalisés sur la base d'une logique propositionnelle (voir par exemple [von Wright, 1972][Wellman and Doyle, 1991][Doyle and Wellman, 1994] [Boutilier, 1994][Boutilier et al., 1999]). Or, cette dernière nous apparaît manquer d'expressivité dans l'optique d'une utilisation dans des applications réelles et en particulier pour des applications de TALN.

De plus, ils ne permettent généralement pas de manipuler des informations contradictoires sur la désirabilité des alternatives (voir par exemple les formalismes de [von Wright, 1972] et [Boutilier et al., 1999]). Or, faire cette hypothèse nécessite soit (1) de contraindre la spécification des données initiales, soit (2) de traiter préalablement ces informations pour résoudre les contradictions. Dans le premier cas, cela a pour conséquence de rendre la phase de spécification peu naturelle et nécessite en particulier une vue globale des préférences. Dans le second cas, cela implique un traitement a priori difficile puisqu'il s'agit d'obtenir une base de formules consistantes la plus « informative » possible (c'est un problème de révision des croyances).

Enfin, la mise en œuvre de l'hypothèse *Ceteris Paribus* ne nous semble pas satisfaisante. Nous avons vu dans la section 3.2.2 que la formalisation de cette hypothèse proposée par von Wright dans [von Wright, 1972] ne permet pas de discriminer des alternatives que le *bon sens* départagerait. De plus, à notre connaissance, les solutions proposées demandent de nombreux efforts pour être mises en œuvre puisqu'elles nécessitent la spécification d'une relation de normalité ou d'équivalence afin d'écarter les propriétés non pertinentes de la comparaison[3] : le problème est seulement déplacé.

3.4.2 Intérêts

L'architecture de construction des préférences d'un agent que nous avons proposée dans [Meyer et al., 2005a] et que nous exposons dans cette thèse avec plus de détails a de nombreux avantages. En particulier elle est adaptée aux modèles cognitifs d'agents, et permet une spécification aisée et intuitive des informations de désirabilité nécessaires à la mise en oeuvre d'une phase de décision. De plus, bien que mettant en œuvre l'hypothèse *Ceteris Paribus*, elle permet de départager un grand nombre d'alternatives. Enfin, elle offre un cadre formel pour gérer les contradictions qui peuvent exister entre les différents points de vue.

Un modèle adapté aux formalismes logiques d'agents rationnels : Notre modèle est adapté aux formalismes logiques d'agents rationnels car c'est un formalisme logique.

[2]Nous avons présenté dans la section 2.3 les modélisations logiques de la préférence qui nous semblent les plus pertinentes pour notre travail.

[3] Notons ici, que le lecteur intéressé pourra trouver une abstraction élégante en logique propositionnelle de certaines de ces solutions dans [Lang, 2003, p.120].

De plus, comme ce modèle est basé sur des notions intuitives (préférence, alternatives, politique de choix), il est aussi adapté aux modélisations cognitives d'agents (voir la section 1.1.2).

Un modèle permettant de spécifier des préférences facilement : Notre modèle permet de spécifier des préférences facilement puisque (1) il reprend une vision usuelle des préférences, (2) il ne nécessite pas une vue globale des préférences et (3) il est robuste.

Ce modèle reprend une vision usuelle des préférences car il présuppose que (a) les informations sur la désirabilité des alternatives sont données via des comparaisons entre les diverses propriétés que peuvent vérifier les alternatives et que (b) ces comparaisons peuvent avoir de surcroît un sens implicite. En particulier, ce modèle met en œuvre les principes d'expansion, de transitivité, *Ceteris Paribus* et de dépendance au contexte (voir la section 2.3.1).

Ce modèle ne nécessite pas d'avoir une vue globale de l'ensemble des préférences lors de la spécification de ces dernières car il présuppose que les comparaisons sont données par points de vue indépendants (donc possiblement contradictoires). Ceci permet entre autres, lors de la spécification des préférences, de se concentrer sur des sous-ensembles de ces préférences.

Enfin, ce modèle est robuste puisqu'il n'est pas gêné par la spécification d'informations contradictoires.

Un modèle permettant de discriminer un « grand » nombre d'alternatives : Notre modèle est une réponse à la critique faite à l'hypothèse *Ceteris Paribus* selon laquelle elle est trop forte pour permettre de comparer en pratique un grand nombre d'alternatives. En particulier, il utilise les informations initiales (spécifiées sous hypothèse *Ceteris Paribus*) comme des arguments pour comparer les diverses alternatives. De plus, ce formalisme fournit un cadre pour gérer les contradictions qui peuvent exister entre les informations sur la désirabilité des alternatives et ainsi permet de discriminer le plus d'alternatives possible.

Un modèle réaliste/prudent de la préférence : Le tour d'horizon de la littérature effectué au chapitre 2 indique qu'il n'existe pas de consensus sur la structure des préférences[4] : est-ce un ordre, un pré-ordre ou bien autre chose ? (voir en particulier le point (b) de la section 2.1.3). Pour notre part, nous ne faisons aucune hypothèse sur la forme de la relation de préférence globale d'un agent. En particulier, cette dernière n'est pas supposée être un pré-ordre total dans notre modèle.

[4]Ce manque de consensus est un argument pour développer des méthodes d'aide à la décision et en particulier des méthodes itératives (voir [Vincke, 1989]). En effet, ces méthodes cherchent plutôt à construire une relation de préférence qu'en révéler une [Mousseau, 2003, p.30].

Chapitre 4

Langages, notations et exemple utilisés

Avant d'aller plus loin dans l'exposition de notre proposition et de afin de clarifier notre propos, il nous apparaît important de préciser les outils que nous allons employer. C'est pourquoi, dans ce chapitre, nous rappelons, dans un premier temps, la définition d'une logique des prédicats du premier ordre. Dans un deuxième temps, nous introduisons la notation $\phi + \overline{\psi}$ qui nous permet de spécifier aisément le principe d'*expansion* dans notre modèle. Dans un troisième temps, nous définissons formellement les concepts importants de notre formalisme : les concepts d'*alternative*, de *préférence* et d'*indifférence*. Enfin nous terminons ce chapitre par l'introduction d'exemple qui nous permettra par la suite d'illustrer notre proposition au fur et à mesure de son développement.

4.1 Le langage des prédicats du premier ordre $\mathscr{L}$

Dans ce document, les langages $\mathscr{L}$, $\mathscr{L}^1$, $\mathscr{L}^2$ et $\mathscr{L}^3$ sont être utilisés. Le langage $\mathscr{L}$ permet de décrire les alternatives (voir la section 4.3.1) tandis que les langages $\mathscr{L}^1$, $\mathscr{L}^2$ et $\mathscr{L}^3$ permettent de décrire les préférences sur ces alternatives (voir la section 4.3.2). Ces derniers seront introduits un peu plus loin, respectivement dans les sections (5.3), (6.2) et (7.2), et définis à partir du langage $\mathscr{L}$. Le langage $\mathscr{L}$ est un langage des prédicats du premier ordre. Ceci nous permet d'utiliser l'expressivité de la Logique des Prédicats du Premier Ordre (dénommée par la suite LPPO) pour décrire les alternatives ainsi que de proposer un formalisme d'utilisation réaliste pour un grand nombre d'applications.

Cette section fixe les notations et rappelle la définition d'un tel langage. Elle se base sur les ouvrages [Herzig, 2005], [Thayse et al., 1989, pp36–41] et [Russell and Norvig, 2003, pp240–266].

4.1.1 Vocabulaire

Le langage $\mathscr{L}$ est un langage des prédicats du premier ordre construit classiquement sur un vocabulaire constitué d'un ensemble dénombrable de symboles logiques et d'un ensemble dénombrable de symboles non logiques (i.e. la signature). Les symboles logiques désignent des variables (notées $x_1, x_2, \ldots$), deux connecteurs (notés $\wedge$ et $\neg$) et un quantificateur (noté $\forall$). Les symboles non logiques désignent des prédicats à n places (notés $P, P_1, P_2, \ldots$) et des fonctions à m places (notées $f, f_1, f_2, \ldots$). En particulier, les symboles tautologie (noté $\top$) et contradiction (noté $\bot$) sont des prédicats à 0 place ; les constantes (notées $c_1, c_2, \ldots$) sont des fonctions à 0 place.

4.1.2 Syntaxe

La syntaxe est l'ensemble des lois qui permettent de distinguer les formules bien formées du langage parmi les assemblages quelconques de symboles et plus généralement de parler de la forme de ces formules. Pour plus de clarté, les notions de *terme*, de *formule atomique*, de *sous-formules*, de *variable libre et liée* sont classiquement introduites.

a) Terme, formule atomique et formule bien formée :

Les **termes** sont définis exhaustivement par les deux règles suivantes : (1) Si t est une constante ou une variable, alors t est un terme ; (2) Si $t_1, \ldots, t_n$ sont n termes et f le symbole d'une fonction à n places, alors $f(t_1, \ldots, t_n)$ est un terme. Les **formules atomiques** sont, quant à elles, définies exhaustivement par la règle suivante : Si $t_1, \ldots, t_n$ sont n termes et P est un prédicat à n places, alors $P(t_1, \ldots, t_n)$ est une formule atomique. Remarquons que $\top$ et $\bot$ sont toujours des formules atomiques de $\mathscr{L}$. Enfin, les **formules bien formées** sont définies exhaustivement par les quatre règles suivantes :

- Si F est une formule atomique, alors F est une formule bien formée.
- Si F est une formule bien formée, alors $(\neg F)$ est une formule bien formée.
- Si F et G sont des formules bien formées, alors $(F \wedge G)$ est une formule bien formée.
- Si F est une formule bien formée et x est une variable, alors $(\forall x F)$ est une formule bien formée.

Par la suite, toutes les formules seront considérées comme bien formées. Par souci de concision, nous omettrons donc de le spécifier. D'autre part, les parenthèses seront aussi omises suivant les règles de priorité usuelles.

b) Abréviations syntaxiques usuelles :

La notation $\phi \vee \psi$ est l'abréviation syntaxique (classique) de la notation $\neg(\neg\phi \wedge \neg\psi)$. De même, la notation $\phi \Rightarrow \psi$ est l'abréviation de la notation $\neg(\phi \wedge \neg\psi)$, la notation $\phi \Leftrightarrow \psi$ est l'abréviation de la notation $(\phi \wedge \psi) \vee (\neg\phi \wedge \neg\psi)$ et la notation $\exists\phi$ est l'abréviation de la notation $\neg(\forall x(\neg\phi))$.

c) Sous-formules et portée d'un quantificateur :

L'ensemble Subf(ϕ) des **sous-formules** d'une formule ϕ est défini récursivement par les quatre règles suivantes :

- Si ϕ est atomique, alors $\mathrm{Subf}(\phi) = \{\phi\}$
- Si ϕ est de la forme $\neg\psi$, alors $\mathrm{Subf}(\phi) = \mathrm{Subf}(\psi) \cup \{\neg\psi\}$

– Si ϕ est de la forme $\psi_1 \wedge \psi_2$, alors $\text{Subf}(\phi) = \text{Subf}(\psi_1) \cup \text{Subf}(\psi_2) \cup \{\psi_1 \wedge \psi_2\}$
– Si ϕ est de la forme $\forall x \psi$, alors $\text{Subf}(\phi) = \text{Subf}(\psi) \cup \{\forall x \psi\}$

Si la formule $\forall x \phi$ est une sous-formule de ψ, alors la formule ϕ est appelée « **la portée** » du quantificateur $\forall x$ dans la formule ψ.

d) Variable libre/liée et formule ouverte/close :

Une occurrence d'une variable x dans une formule ϕ est dite « **libre** » dans ϕ si et seulement si elle n'est pas dans la portée du quantificateur $\forall x$ présent dans ϕ ou s'il n'existe pas de quantificateur $\forall x$ dans ϕ. Inversement, si $\forall x \phi$ est une sous-formule de ψ, et si x est une occurrence de variable libre dans ϕ, alors x est dite « **liée** » par le quantificateur $\forall x$ dans ψ. Une variable dont les occurrences sont liées dans une formule peut être renommée dans cette formule (on dit qu'elle est muette). Par conséquent, et pour plus de clarté, les variables dont certaines occurrences sont liées et d'autres libres seront systématiquement renommées ; par exemple, la formule $\forall x[P_1(x) \wedge \forall x P_2(x)]$ deviendra $\forall x[P_1(x) \wedge \forall y P_2(y)]$.

Une formule **ouverte** est une formule qui comporte au moins une occurrence de variable libre. Inversement, une formule **close** (aussi appelée énoncé) est une formule qui ne comporte aucune occurrence de variable libre. La **clôture universelle** d'une formule ϕ, ayant uniquement les variables $x_1, \ldots, x_n$ comme variables libres, correspond à la formule close $\forall x_1 \ldots \forall x_n \; \phi$.

e) Substitutions :

Soient ϕ une formule, x une variable et t un terme. $\phi[x := t]$ dénote la formule obtenue en substituant (i.e. remplaçant) dans ϕ toutes les occurrences libres de x par t (après un renommage éventuel des occurrences de variables liées de ϕ qui apparaissent libres dans t). $\phi[x_1 := t_1, \ldots, x_n := t_n]$ désigne la formule obtenue par substitution simultanée des variables x_i par les termes t_i. Par exemple, $P(f_1(x, y))[x := f_2(z), y := t]$ dénote la formule $P(f_1(f_2(z), t))$.

Remarquons que de façon générale, la formule $\phi[x_1 := t_1, x_2 := t_2]$ est différente de la formule $\phi[x_1 := t_1][x_2 := t_2]$. Par exemple $P(x, y)[x := y, y := x]$ donne $P(y, x)$, mais $P(x, y)[x := y][y := x]$ donne $P(x, x)$.

4.1.3 Sémantique

La sémantique est l'ensemble des lois qui permettent d'attribuer une signification aux formules (i.e. de les interpréter comme vraies ($\top$) ou fausses ($\bot$)) et plus généralement de parler du sens des formules. L'interprétation est le mécanisme qui permet d'attribuer une valeur sémantique aux expressions du langage (i.e. variables, symboles de constantes prédicatives et de fonctions, termes, formules ouvertes et closes). L'interprétation des expressions est (1) relative à un modèle et à une assignation et (2) se base, de façon compositionnelle (suivant des règles logiques), sur la valeur sémantique des termes.

a) Modèle et assignation :

Un **modèle** (aussi appelé structure d'interprétation) pour une logique des prédicats est un couple (S,V) composé de l'ensemble non vide des « objets » S et d'une fonction de valuation V.

- L'ensemble S correspond à l'ensemble des valeurs que peuvent prendre les variables.
- La fonction de valuation V associe (1) à chaque symbole de fonction f à n places, une fonction de S^n dans S (notée V_f) et (2) à chaque symbole de prédicat P à n places, une fonction de S^n dans $\{\top, \bot\}$ (notée V_P).

Une **assignation** g est une fonction qui assigne un objet de S à chacune des variables du langage. L'assignation ne modifie donc pas la valeur des constantes.

b) Satisfaction des formules :

De façon générale, la valeur sémantique d'un terme t est notée $t^M[g]$. C'est un élément du domaine d'interprétation des variables S. Si un terme t est une variable x alors la valeur sémantique de ce terme relativement au modèle M et à l'assignation g est égale à $g(x)$; elle ne dépend pas du modèle M. Si un terme t est de la forme $f(t_1, \ldots, t_n)$ alors la valeur sémantique de ce terme relativement au modèle M et à l'assignation g est égale à $V_f(t_1^M[g], \ldots, t_n^M[g])$.

Une formule ϕ est satisfaite dans le modèle M par la fonction d'assignation g (notée $M \vDash \phi[g]$) si et seulement si :

- lorsque ϕ est de la forme $P(t_1, \ldots, t_n)$, on a $V_P(t_1^M[g], \ldots, t_n^M[g]) = \top$.
- lorsque ϕ est de la forme $\neg\psi$, on n'a pas $M \vDash \psi[g]$.
- lorsque ϕ est de la forme $\psi_1 \wedge \psi_2$, on a $M \vDash \psi_2[g]$ et $M \vDash \psi_1[g]$.
- lorsque ϕ est de la forme $\forall x \psi[g]$, pour tout objet d de S on a $M \vDash \psi[g[x/d]]$; $g[x/d]$ désignant la variante de g qui assigne l'objet d à la variable x.

c) Satisfiabilité d'une formule et formule vraie/fausse/valide :

Une formule ϕ est **satisfiable** (aussi dite *satisfaisable* ou *consistante*) si et seulement si il existe (au moins) un modèle M et une assignation g qui satisfont ϕ (notée $M \vDash \phi[g]$). Une formule ϕ est **vraie** dans un modèle M (notée $M \vDash \phi$) si et seulement si elle est satisfaite par toute assignation dans ce modèle. M est dit alors être un modèle de ϕ. Une formule ϕ est **fausse** dans M (noté $M \nvDash \phi$) si et seulement si $\neg\phi$ est vrai dans M (noté $M \vDash \neg\phi$). C'est le principe du tiers exclu. Une formule ϕ est (logiquement) **valide** (noté $\vDash \phi$) si elle est satisfaite par toute assignation dans tout modèle (i.e $\forall M, M \vDash \phi$). La notation $\nvDash \phi$ signifie que la formule ϕ n'est pas valide : elle n'est pas vraie dans tous les modèles possibles du langage (i.e $\exists M, M \nvDash \phi$).

Si ϕ est vraie dans un modèle alors ϕ est satisfiable. Dans le cas général l'inverse est faux. Cependant comme pour un modèle donné, un énoncé (i.e. une formule close) est soit vrai, soit faux, si ϕ est un énoncé et si ϕ est satisfaite par une assignation dans le modèle M, alors elle est satisfaite par toute assignation dans ce même modèle : elle est vraie dans ce modèle. Par la suite, les symboles $\phi, \phi_1, \ldots \psi, \psi_1 \ldots, \alpha, \beta, \gamma, \ldots$ dénoteront des formules bien formées closes de $\mathscr{L}$ (i.e. des énoncés). Ceci nous permettra de ne pas considérer la satisfiabilité des formules.

d) Formules équivalentes et formule consistante avec une autre

Deux formules ϕ et ψ sont dites **logiquement équivalentes** (noté $\phi \equiv \psi$) si et seulement si elles ont exactement les mêmes modèles (i.e. pour tout (M, g), $M \vDash \phi[g]$ si et seulement si $M \vDash \psi[g]$). La notation $\phi \not\equiv \psi$ indique pour sa part que ϕ et ψ ne sont pas logiquement équivalentes (i.e. il existe (M, g) tel que soit ($M \vDash \phi[g]$ et $M \vDash \neg\psi[g]$) soit ($M \vDash \psi[g]$ et $M \vDash \neg\phi[g]$)).

Une formule ϕ est dite **consistante avec une formule** ψ si et seulement si la formule $\phi \wedge \psi$ est *satisfiable* (i.e. il existe M et g tels que $M \vDash \phi \wedge \psi[g]$). Dans le cas contraire (i.e. si $\neg(\phi \wedge \psi)$ est une formule valide) on dit qu'elle est **inconsistante avec** ψ. En particulier, la formule ϕ sera dite logiquement inconsistante dans le cas où elle sera logiquement équivalent à $\bot$.

e) Subsomption, théorie et conséquence logique

Il résulte de ces définitions que la notation $\vDash \phi \Rightarrow \psi$ signifie que les modèles de ϕ sont aussi des modèles de ψ (i.e $\forall M$, si $M \vDash \phi$ alors $M \vDash \psi$). ψ est dite alors être **subsumée par** ϕ (ou ϕ subsume ψ). La notation $\vDash \phi \Leftrightarrow \psi$ signifie donc pour sa part que M est un modèle de ϕ si et seulement s'il est un modèle de ψ. On a donc $\vDash \phi \Leftrightarrow \psi$ si et seulement si $\phi \equiv \psi$. De même, on a $\nvDash \phi \Leftrightarrow \psi$ si et seulement si $\phi \not\equiv \psi$.

Par commodité, nous introduisons la notation $\phi \vDash \psi$. Cette dernière a le même sens que la notation $\vDash \phi \Rightarrow \psi$. Une **théorie** est un ensemble de formules qui est clos par conséquence logique. Aussi, par extension, si $\mathscr{E}$ dénote une théorie, la notation $\mathscr{E} \vDash \phi$ indique que les modèles de l'ensemble des formules de la théorie $\mathscr{E}$ sont aussi des modèles de la formule ϕ : (i.e $\forall M$, si $M \vDash \mathscr{E}$ alors $M \vDash \phi$). ϕ est dit alors être une **conséquence logique** de $\mathscr{E}$.

4.1.4 Axiomatique (à la Hilbert)

Afin de « coller » à la sémantique, des règles d'inférences et des axiomes sont imposés à la syntaxe. Ils permettent de déduire de manière mécanique et constructive de nouvelles formules : les théorèmes de la logique. Ceci est particulièrement intéressant car ces méthodes peuvent être automatisées et donc constituer la base de programmes conforme à la sémantique.

a) Système axiomatique

Un système axiomatique (ou axiomatisation) d'une logique est un ensemble d'axiomes ainsi qu'un ensemble de règles d'inférence. Il existe de nombreuses axiomatisations possibles. Pour notre part, nous choisissons d'utiliser l'axiomatique classique à la Hilbert, constituée de cinq schémas d'axiomes (A1, A2, A3, A4, A5) et des règles d'inférences de Modus Ponens (MP) et de généralisation (G) suivantes :

(A1) $A \Rightarrow (B \Rightarrow A)$
(A2) $[A \Rightarrow (B \Rightarrow C)] \Rightarrow [(A \Rightarrow B) \Rightarrow (A \Rightarrow C)]$
(A3) $[\neg A \Rightarrow \neg B] \Rightarrow [(\neg A \Rightarrow B) \Rightarrow A]$
(A4) $[\forall x(A \Rightarrow B)] \Rightarrow [A \Rightarrow \forall x B]$ (si la variable x ne figure pas dans A et est libre dans B)
(A5) $\forall x A \Rightarrow A[x := t]$ (si la variable x est libre dans A)

(MP) Si A et $A \Rightarrow B$ sont des théorèmes alors B est un théorème,
(G) Si A est un théorème et si la variable x est libre dans A alors $\forall x A$ est un théorème.

b) Notion de preuve et de théorème

Une **preuve** d'une formule F est une liste ordonnée finie de formules $(f_1, \ldots, f_n)$ de $\mathscr{L}$ telles que f_n est égale à F et quel que soit l'indice $i \in [1, n]$, (1) ou bien f_i est un axiome ou une hypothèse, ou bien il existe une suite d'indices $(i_1, \ldots, i_k)$ tous inférieurs à i, et une règle d'inférence qui permet de déduire la formules f_i lorsqu'elle est appliquée à $f_{i_1}, \ldots, f_{i_k}$.

S'il existe une preuve pour une formule, alors cette formule est dite **prouvable**. Si aucune hypothèse n'apparaît dans l'une des preuves d'une formule, alors cette formule est qualifiée de **théorème**. Un système axiomatique étant donné, le fait que la formule ϕ soit un théorème est noté $\vdash \phi$. Le fait que la formule ϕ soit prouvable à partir d'un ensemble de formules $\mathscr{E}$ (les hypothèses) est noté $\mathscr{E} \vdash \phi$.

c) Propriétés importantes de l'axiomatique et de la LPPO

Le système axiomatique que nous avons présenté est *complet* et *adéquat* : il permet de construire (1) toutes[1] les formules valides et (2) seulement[2] les formules valides. Par conséquent, il existe une correspondance parfaite entre la sémantique et l'axiomatique. En particulier, une formule est une conséquence logique d'une théorie si et seulement si cette formule est prouvable à partir de l'ensemble des formules de cette théorie. Par conséquent, le problème de déduction $\mathscr{E} \vdash B$ peut être réduit au problème de prouvabilité $\vdash \mathscr{E} \Rightarrow B$ et inversement (Herbrand, 1930).

Remarquons enfin que la logique des prédicats du premier ordre n'est pas **décidable** : il n'existe pas de procédé effectif pour décider si une expression donnée est, ou n'est pas, théorème du système. Elle est néanmoins **semi-décidable** : il existe une procédure effective telle que pour toute formule F en entrée,

– si F est un théorème alors la procédure s'arrête et retourne « oui »
– sinon (si F n'est pas un théorème) ou bien la procédure s'arrête et retourne « non », ou bien elle ne s'arrête pas.

[1] Propriété de complétude : si $\models A$ alors $\vdash A$.
[2] Propriété d'adéquation : si $\vdash A$ alors $\models A$.

4.2 La notation $\phi{+}\overline{\psi}$

Afin de spécifier aisément dans notre modèle le principe d'*expansion* (voir le point (2.3.1) de la section 2.3), nous introduisons et définissons dans cette section une nouvelle notation : $\phi{+}\overline{\psi}$. Celle-ci sera utilisée par la suite pour définir formellement les préférences. Elle désigne la formule ϕ « augmentée » si possible de la formule $\neg\psi$. Intuitivement, M est un modèle de la formule $\phi{+}\overline{\psi}$ si, d'une part, il est un modèle de la formule ϕ et si, d'autre part, il valide, autant que faire se peut, la négation de ψ. Dans un contexte de révision des croyances, la formule $\phi{+}\overline{\psi}$ peut être identifiée à la formule $\neg\psi$ révisée par la formule ϕ (voir la section 4.2.3).

4.2.1 Définition

Pour toutes formules closes ϕ et ψ, la notation $\phi{+}\overline{\psi}$ dénote la formule close ϕ de $\mathscr{L}$ si ϕ et $\neg\psi$ sont inconsistantes et la formule close $\phi\wedge\neg\psi$ sinon :

$$\text{Si}\quad \phi\wedge\neg\psi \equiv \bot \quad \text{alors}\quad \phi{+}\overline{\psi} \overset{\text{déf}}{=} \phi,\quad \text{sinon}\quad \phi{+}\overline{\psi} \overset{\text{déf}}{=} \phi\wedge\neg\psi \tag{4.1}$$

Remarquons ici que la formule $\phi\wedge\neg\psi \equiv \bot$ est équivalente à $\models \neg(\phi\wedge\neg\psi)$ c'est-à-dire à $\models \neg\phi\vee\psi$, soit encore à $\models \phi\Rightarrow\psi$.

4.2.2 Propriétés remarquables

Dans cette section nous listons quelques propriétés remarquables de la formule $\phi{+}\overline{\psi}$ induites par la définition (4.1). Afin de ne pas surcharger la lecture de ce document, nous ne listons ici que celles qui nous semblent les plus importantes. Le lecteur intéressé se référer à l'annexe (B.1) pour un complément.

(i). $\phi{+}\overline{\phi} \equiv \phi$ Preuve immédiate puisque $\phi\wedge\neg\phi \equiv \bot$.

(ii). $\phi{+}\overline{\neg\phi} \equiv \phi$ Preuve immédiate puisque $\phi\wedge\neg(\neg\phi) \equiv \phi\wedge\phi \equiv \phi$.

(iii). Si $[\ \phi \equiv \phi'$ et $\psi \equiv \psi'\ ]$ alors $[\phi{+}\overline{\psi} \equiv \phi'{+}\overline{\psi'}]$

Preuve immédiate puisque $\phi{+}\overline{\psi} \equiv \phi$ ou $\phi{+}\overline{\psi} \equiv \phi\wedge(\neg\psi)$

(iv). $\models \phi{+}\overline{\psi}\Rightarrow\phi$ En particulier, si $\models \phi \Rightarrow \phi'{+}\overline{\psi'}$ alors pour tout ψ, $\models \phi{+}\overline{\psi} \Rightarrow \phi'{+}\overline{\psi'}$.

Preuve immédiate puisque par définition $\phi{+}\overline{\psi} \equiv \phi$ ou $\phi{+}\overline{\psi} \equiv \phi\wedge(\neg\psi)$

(v). Si ϕ a un seul modèle alors, quel que soit ψ, $\phi{+}\overline{\psi} \equiv \phi$.

Preuve : Quelle que soit la formule close ψ et quel que soit le modèle M, soit $M \models \neg\psi$, soit $M \not\models \neg\psi$. En particulier, si M est le seul modèle de ϕ alors, soit M est un modèle pour $\neg\psi$, soit M n'est pas un modèle pour $\neg\psi$. Dans le premier cas cela implique que M est le seul modèle de $\phi\wedge\neg\psi$. Dans le second cas cela implique que $\phi\wedge\neg\psi$ n'a aucun modèle. En effet, pour qu'un modèle M soit un modèle de $\phi\wedge\neg\psi$, il faut que M soit un modèle de

ϕ et aussi de $\neg\psi$. Dans le premier cas cela implique que $\models \phi \Rightarrow \neg\psi$ et donc $\phi\wedge\neg\psi \equiv \phi$ et donc que $\phi+\overline{\psi} \equiv \phi$. Dans le second cas, cela implique que ϕ et $\neg\psi$ sont inconsitants (i.e. $\phi\wedge\neg\psi \equiv \bot$) et donc aussi que $\phi+\overline{\psi} \equiv \phi$.

(vi). Si $\left[\models \psi \Rightarrow \psi' \text{ et } \phi\wedge\neg\psi' \not\equiv \bot\right]$ alors $\left[\models \phi+\overline{\psi'} \Rightarrow \phi+\overline{\psi}\right]$.

Preuve : Si $\models \psi \Rightarrow \psi'$ et $\phi\wedge\neg\psi' \not\equiv \bot$, alors $\phi\wedge\neg\psi \not\equiv \bot$. De plus, et par définition, si $\phi\wedge\neg\psi \not\equiv \bot$ alors $\phi+\overline{\psi} \equiv \phi\wedge\neg\psi$. De même, par définition, si $\phi\wedge\neg\psi' \not\equiv \bot$ alors $\phi+\overline{\psi'} \equiv \phi\wedge\neg\psi'$.
Par conséquent, comme $\models \psi \Rightarrow \psi'$ équivaut à $\models \neg\psi\vee\psi'$, cela implique que $\models \psi \Rightarrow \psi'$ implique que $\models \neg\psi\vee\psi'\vee\neg\phi$, donc aussi que $\models (\neg\psi\wedge\phi)\vee\psi'\vee\neg\phi$, soit que $\models (\neg\psi\wedge\phi) \Leftarrow (\neg\psi'\wedge\phi)$. Par conséquent, si $\models \psi \Rightarrow \psi'$ et $\phi\wedge\neg\psi' \not\equiv \bot$ alors $\phi+\overline{\psi}$ est impliqué par $\phi+\overline{\psi'}$.

(vii). Si $\left[\models \phi \Rightarrow \phi' \text{ et } \models \phi' \Rightarrow (\phi\vee\psi)\right]$ alors $\left[\models \phi+\overline{\psi} \Rightarrow \phi'+\overline{\psi}\right]$.

Preuve en remarquant que si $\models \phi' \Rightarrow (\phi\vee\psi)$ alors de trois choses l'une :

- Soit $\phi\wedge\neg\psi \equiv \bot$ et $\phi'\wedge\neg\psi \equiv \bot$.
  Dans ce cas, et par définition $\phi+\overline{\psi} \equiv \phi$ et $\phi'+\overline{\psi} \equiv \phi'$. Or comme par hypothèse $\models \phi \Rightarrow \phi'$, on a aussi $\models \phi+\overline{\psi} \Rightarrow \phi'+\overline{\psi}$.
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\phi'\wedge\neg\psi \equiv \bot$.
  Dans ce cas, et par définition $\phi+\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\phi'+\overline{\psi} \equiv \phi'$. Or comme par hypothèse $\models \phi \Rightarrow \phi'$, on a aussi $\models \phi\wedge\neg\psi \Rightarrow \phi'$. Par conséquent on a $\models \phi+\overline{\psi} \Rightarrow \phi'+\overline{\psi}$.
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\phi'\wedge\neg\psi \not\equiv \bot$.
  Dans ce cas, et par définition $\phi+\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\phi'+\overline{\psi} \equiv \phi'\wedge\neg\psi$. Or comme par hypothèse $\models \phi \Rightarrow \phi'$, on a aussi $\models \phi\wedge\neg\psi \Rightarrow \phi'\wedge\neg\psi$. Par conséquent on a $\models \phi+\overline{\psi} \Rightarrow \phi'+\overline{\psi}$.

(viii). Si $\phi \not\equiv \psi$ alors $\left[\models \phi+\overline{\psi} \Rightarrow \neg(\psi+\overline{\phi})\right]$ (et si $\phi \equiv \psi$ alors $\phi+\overline{\psi} \equiv \psi+\overline{\phi}$).

Preuve en remarquant que de quatre choses l'une :

- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$
  Dans ce cas, $\phi+\overline{\psi} = \phi\wedge\neg\psi$ et $\neg(\psi+\overline{\phi}) = \neg(\psi\wedge\neg\phi) = \neg\psi\vee\phi$
  Par conséquent, comme $\phi\wedge\neg\psi$ implique $\neg\psi\vee\phi$, on a $\phi+\overline{\psi} \Rightarrow \neg(\psi+\overline{\phi})$.
- Soit $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$
  Dans ce cas, $\phi+\overline{\psi} = \phi$ et $\neg(\psi+\overline{\phi}) = \neg(\psi\wedge\neg\phi) = \neg\psi\vee\phi$
  Par conséquent, comme ϕ implique $\neg\psi\vee\phi$, on a $\phi+\overline{\psi} \Rightarrow \neg(\psi+\overline{\phi})$.
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$
  Dans ce cas, $\phi+\overline{\psi} = \phi\wedge\neg\psi$ et $\neg(\psi+\overline{\phi}) = \neg\psi$
  Par conséquent, comme $\phi\wedge\neg\psi$ implique $\neg\psi$, on a $\phi+\overline{\psi} \Rightarrow \neg(\psi+\overline{\phi})$.
- Soit $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$
  Dans ce cas, on a $\models \phi \Rightarrow \psi$ et $\models \psi \Rightarrow \phi$ et donc $\phi \equiv \psi$. De plus, dans ce cas on a aussi, par définition que $\phi+\overline{\psi} \equiv \phi$ et $\psi+\overline{\phi} \equiv \psi$.
  Par conséquent, on a $\phi+\overline{\psi} \equiv \psi+\overline{\phi}$.

(ix). $(\phi+\overline{\psi})+\overline{(\psi+\overline{\phi})} \equiv \phi+\overline{\psi}$.

Preuve en remarquant que de quatre choses l'une :

- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$
  Dans ce cas, on a $\phi{+}\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\psi{+}\overline{\phi} \equiv \psi\wedge\neg\phi$. et donc, par définition, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv (\phi\wedge\neg\psi){+}(\overline{\psi\wedge\neg\phi})$. De plus, comme $(\phi\wedge\neg\psi) \wedge \neg(\psi\wedge\neg\phi) \equiv (\phi\wedge\neg\psi) \wedge (\neg\psi \vee \phi) \equiv (\phi\wedge\neg\psi)$, on a $(\phi\wedge\neg\psi){+}(\overline{\psi\wedge\neg\phi}) \equiv (\phi{+}(\overline{\psi})$ Par conséquent, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv \phi{+}\overline{\psi}$
- Soit $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$
  Dans ce cas, on a $\phi{+}\overline{\psi} \equiv \phi$ et $\psi{+}\overline{\phi} \equiv \psi\wedge\neg\phi$. et donc, par définition, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv (\phi){+}(\overline{\psi\wedge\neg\phi})$. De plus, comme $\phi \wedge \neg(\psi\wedge\neg\phi) \equiv \phi \wedge (\neg\psi \vee \phi) \equiv \phi$, on a $(\phi\wedge\neg\psi){+}(\overline{\psi\wedge\neg\phi}) \equiv (\phi{+}(\overline{\psi})$ Par conséquent, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv \phi{+}\overline{\psi}$
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$
  Dans ce cas, on a $\phi{+}\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\psi{+}\overline{\phi} \equiv \psi$. et donc, par définition, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv (\phi\wedge\neg\psi){+}(\overline{\psi})$. De plus, comme $(\phi\wedge\neg\psi)\wedge\neg\psi \equiv (\phi\wedge\neg\psi)$, on a $(\phi\wedge\neg\psi){+}(\overline{\psi}) \equiv (\phi{+}(\overline{\psi})$ Par conséquent, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv \phi{+}\overline{\psi}$
- Soit $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$
  Dans ce cas, on a $\phi{+}\overline{\psi} \equiv \phi$ et $\psi{+}\overline{\phi} \equiv \psi$. et donc, par définition, $(\phi{+}\overline{\psi}){+}(\overline{\psi{+}\overline{\phi}}) \equiv \phi{+}\overline{\psi}$.

(x). Si $\models (\phi{+}\overline{\psi}) \wedge \gamma$ alors $\models (\phi\wedge\gamma){+}\overline{\psi}$ et donc aussi $\models (\phi\wedge\gamma){+}\overline{(\psi\wedge\gamma)}$ (Distributivité)

Preuve en remarquant que de trois choses l'une :

- Soit $\phi\wedge\neg\psi \equiv \bot$. Dans ce cas, $\phi{+}\overline{\psi} = \phi$. De plus comme alors $\phi\wedge\gamma\wedge\neg\psi \equiv \bot$, on a aussi que $(\phi\wedge\gamma){+}\overline{\psi} = \phi\wedge\gamma$. Par conséquent, $(\phi{+}\overline{\psi}) \wedge \gamma = \phi\wedge\gamma = (\phi\wedge\gamma){+}\overline{\psi}$
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\phi\wedge\gamma\wedge\neg\psi \not\equiv \bot$. Dans ce cas, $\phi{+}\overline{\psi} = \phi\wedge\neg\psi$ et $(\phi\wedge\gamma){+}\overline{\psi} = \phi\wedge\gamma\wedge\neg\psi$. Par conséquent, $(\phi{+}\overline{\psi}) \wedge \gamma = \phi\wedge\neg\psi\wedge\gamma = (\phi\wedge\gamma){+}\overline{\psi}$
- Soit $\phi\wedge\neg\psi \not\equiv \bot$ et $\phi\wedge\gamma\wedge\neg\psi \equiv \bot$. Dans ce cas, $\phi{+}\overline{\psi} = \phi\wedge\neg\psi$ et $(\phi\wedge\gamma){+}\overline{\psi} = \phi\wedge\gamma$. Par conséquent, $(\phi{+}\overline{\psi}) \wedge \gamma$ (qui équivaut à $\phi\wedge\neg\psi\wedge\gamma$) implique $(\phi\wedge\gamma){+}\overline{\psi}$ (qui équivaut à $\phi\wedge\gamma$).

(xi). $\phi{+}\overline{\psi} \equiv \psi{+}\overline{\phi}$ si et seulement si $\phi \equiv \psi$.

Preuve en remarquant que deux quatre choses l'une :

- $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$ Dans ce cas $\phi{+}\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\psi{+}\overline{\phi} \equiv \psi\wedge\neg\phi$ et donc $\phi{+}\overline{\psi} \equiv \psi{+}\overline{\phi}$ si et seulement si $\phi\wedge\neg\psi \equiv \psi\wedge\neg\phi$. Or ceci n'est possible que si $\phi\wedge\neg\psi \equiv \phi\wedge\neg\psi \equiv \bot$, et ceci est en contradiction avec les hypothèses.
- $\phi\wedge\neg\psi \not\equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$ Dans ce cas $\phi{+}\overline{\psi} \equiv \phi\wedge\neg\psi$ et $\psi{+}\overline{\phi} \equiv \psi$ et donc $\phi{+}\overline{\psi} \equiv \psi{+}\overline{\phi}$ si et seulement si $\phi\wedge\neg\psi \equiv \psi$. Or ceci n'est possible que si $\phi\wedge\neg\psi \equiv \psi \equiv \bot$, et ceci est en contradiction avec les hypothèses.
- $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \not\equiv \bot$ Dans ce cas $\phi{+}\overline{\psi} \equiv \phi$ et $\psi{+}\overline{\phi} \equiv \psi\wedge\neg\phi$ et donc $\phi{+}\overline{\psi} \equiv \psi{+}\overline{\phi}$ si et seulement si $\phi \equiv \psi\wedge\neg\phi$. Or ceci n'est possible que si $\phi \equiv \psi\wedge\neg\phi \equiv \bot$, et ceci est en contradiction avec les hypothèses.
- $\phi\wedge\neg\psi \equiv \bot$ et $\psi\wedge\neg\phi \equiv \bot$ Dans ce cas $\phi{+}\overline{\psi} \equiv \phi$ et $\psi{+}\overline{\phi} \equiv \phi$ et donc $\phi{+}\overline{\psi} \equiv \psi{+}\overline{\phi}$ si et seulement si $\phi \equiv \psi$

(xii). $\phi+\overline{\psi} \equiv (\neg\psi)+\overline{\neg\phi}$

Preuve : d'après la propriété ((xxx)) de l'annexe B.1, on a $\phi+\overline{\psi} \Rightarrow (\neg\psi)+\overline{\neg\phi}$ et en particulier (en renommant les variables et prenant leur négation) $\neg(\psi)+\overline{\neg\phi} \Rightarrow (\neg\neg\phi)+\overline{\neg\neg\psi}$ soit $\neg(\psi)+\overline{\neg\phi} \Rightarrow (\phi)+\overline{\psi}$

4.2.3 Notation $\phi + \overline{\psi}$ et opérateur de révision des croyances

Dans un contexte de révision des croyances, la formule $\phi+\overline{\psi}$ peut être identifiée à la formule $\neg\psi$ révisée par la formule ϕ. Plus précisément, si $\circ$ est un opérateur de révision des croyances au sens de Katsuno et Mendelzon [Katsuno and Mendelzon, 1991], l'égalité suivante est vérifiée :

$$\phi+\overline{\psi} =_{\text{déf}} (\neg\psi) \circ \phi \tag{4.2}$$

En effet, l'opérateur $\circ$ ainsi défini vérifie alors les six propriétés suivantes :

1. $\phi \circ \mu$ implique μ
   Preuve :

   D'après la définition (4.2), on a $\phi \circ \mu \equiv \mu+\overline{\neg\phi}$. De plus, d'après la propriété (iv), on a aussi $\mu+\overline{\neg\phi}$ implique μ. Par conséquent, on a donc aussi que $\phi \circ \mu$ implique μ.
2. si $\phi \wedge \mu$ est satisfiable alors $\phi \circ \mu \equiv \phi \wedge \mu$
   Preuve :

   D'après la définition (4.2), on a $\phi \circ \mu \equiv \mu+\overline{\neg\phi}$. De plus, d'après la définition (4.1), si $\phi \wedge \mu$ est satisfiable alors $\mu+\overline{\neg\phi} \equiv \phi \wedge \mu$. Par conséquent $\phi \circ \mu \equiv \phi \wedge \mu$.
3. si μ est satisfiable alors $\phi \circ \mu$ est aussi satisfiable.
   Preuve :

   D'après la définition (4.2), on a $\phi \circ \mu \equiv \mu+\overline{\neg\phi}$. De plus, d'après la propriété (xxi), si μ est satisfiable, alors $\mu+\overline{\neg\phi}$ est satisfiable. Par conséquent, si μ est satisfiable, alors $\phi \circ \mu$ est satisfiable.
4. si $\phi_1 \equiv \phi_2$ et $\mu_1 \equiv \mu_2$ alors $\phi_1 \circ \mu_1 \equiv \phi_2 \circ \mu_2$
   Preuve :

   D'après la définition (4.2), on a $\phi_1 \circ \mu_1 \equiv \mu_1+\overline{\neg\phi_1}$ et $\phi_2 \circ \mu_2 \equiv \mu_2+\overline{\neg\phi_2}$. D'autre part, d'après la propriété (iii), si $\phi_1 \equiv \phi_2$ et $\mu_1 \equiv \mu_2$ alors $\mu_1+\overline{\neg\phi_1} \equiv \mu_2+\overline{\neg\phi_2}$. Par conséquent, si $\phi_1 \equiv \phi_2$ et $\mu_1 \equiv \mu_2$ alors $\phi_1 \circ \mu_1 \equiv \phi_2 \circ \mu_2$.
5. $(\phi \circ \mu) \wedge \psi$ implique $\phi \circ (\mu \wedge \psi)$
   Preuve :

   D'après la définition (4.2), on a $(\phi \circ \mu) \wedge \psi \equiv (\mu+\overline{\neg\phi}) \wedge \psi$ et $\phi \circ (\mu \wedge \psi) \equiv (\mu \wedge \psi)+\overline{\neg\phi}$. De plus, d'après la propriété (x), $(\mu+\overline{\neg\phi}) \wedge \psi$ implique $(\mu \wedge \psi)+\overline{\neg\phi}$. Par conséquent, $(\phi \circ \mu) \wedge \psi$ implique $\phi \circ (\mu \wedge \psi)$.
6. si $(\phi \circ \mu) \wedge \psi$ est satisfiable alors $\phi \circ (\mu \wedge \psi)$ implique $(\phi \circ \mu) \wedge \psi$
   Preuve :

Du fait de la définition (4.2), on a $\phi \circ (\mu \wedge \psi) \equiv (\mu \wedge \psi) + \overline{\neg\phi}$. Par conséquent, du fait de la définition (4.1), si $\mu \wedge \psi \wedge \phi$ est satisfiable alors $\phi \circ (\mu \wedge \psi) \equiv \mu \wedge \psi \wedge \phi$ sinon $\phi \circ (\mu \wedge \psi) \equiv \mu \wedge \psi$.
De même, du fait de la définition (4.2), on a $(\phi \circ \mu) \circ \psi \equiv (\mu + \overline{\neg\phi}) \circ \psi$ soit $(\phi \circ \mu) \circ \psi \equiv \psi + \overline{\neg(\mu + \overline{\neg\phi})}$. Par conséquent, là aussi, du fait de la définition (4.1), si $\mu \wedge \psi \wedge \phi$ est satisfiable alors $(\phi \circ \mu) \circ \psi \equiv \mu \wedge \psi \wedge \phi$, sinon, si $\mu \wedge \phi$ est satisfiable alors $(\phi \circ \mu) \circ \psi \equiv \mu \wedge \psi$, sinon $(\phi \circ \mu) \circ \psi \equiv \psi$.
En considérant tous les cas possibles, il apparaît donc que $\phi \circ (\mu \wedge \psi)$ implique $(\phi \circ \mu) \circ \psi$.
D'un autre côté, d'après la seconde propriété de l'opérateur $\circ$, si $(\phi \circ \mu) \wedge \psi$ est satisfiable alors $(\phi \circ \mu) \circ \psi \equiv (\phi \circ \mu) \wedge \psi$. Par conséquent, il en résulte que si $(\phi \circ \mu) \wedge \psi$ est satisfiable alors $(\phi \circ \mu) \circ \psi$ implique $(\phi \circ \mu) \wedge \psi$ et donc que si $(\phi \circ \mu) \wedge \psi$ est satisfiable alors $\phi \circ (\mu \wedge \psi)$ implique $(\phi \circ \mu) \wedge \psi$.

4.3 Les concepts importants

Dans cette section nous allons définir les concepts clefs que nous utilisons par la suite pour formaliser notre proposition. Plus précisément, cette section précise les concepts d'*alternative*, de *préférence* et d'*indifférence*.

4.3.1 Alternatives

Intuitivement, chaque alternative désigne un objet qui peut être comparé. Le terme *alternative* est employé ici pour bien mettre en évidence la notion de choix qui y est associée. Nous les formalisons en identifiant chaque alternative à un modèle de $\mathscr{L}$ (i.e. une structure d'interprétation). Nous les notons $a, a_1, \ldots, b, \ldots$ et ALT dénote l'ensemble des alternatives imaginables. Une alternative a valide un ensemble de formules (cohérentes) : les formules dont a est un modèle. Ces formules désignent les « propriétés » ou les « caractéristiques » de l'alternative a : ϕ est une propriété de l'alternative a si a est un modèle de ϕ (i.e. $a \vDash \phi$). Remarquons que, comme énoncé dans le point (c) de la section 4.1.3, pour toute alternative a de ALT, et pour tout formule bien formée close ϕ de $\mathscr{L}$, soit $a \vDash \phi$ soit $a \vDash \neg\phi$: soit ϕ est une propriété de l'alternative a, soit $\neg\phi$ est une propriété de l'alternative a.

Une formule ϕ de $\mathscr{L}$ désigne en général un ensemble de modèles de $\mathscr{L}$ et donc un ensemble d'alternatives : l'ensemble des alternatives qui valident ϕ (i.e. les modèles de ϕ). En particulier la formule $\top$ désigne l'ensemble de toutes les alternatives et $\perp$ ne désigne aucune alternative. Toutefois, lorsqu'une formule du langage $\mathscr{L}$ subsume chaque formule close de $\mathscr{L}$ ou sa négation, cette formule représente une unique alternative. Cette alternative est alors son unique modèle. Nous notons f_a la formule canonique qui représente uniquement l'alternative a. C'est ce que nous appelons la description « totale » de l'alternative a. Par conséquent, pour toute alternative a et pour toute formule bien formée close ϕ de $\mathscr{L}$, de façon exclusive soit $\vDash f_a \Rightarrow \phi$ soit $\vDash f_a \Rightarrow \neg\phi$. Formellement, la formule f_a est telle que : $\forall \alpha \in \mathscr{L},\ a \vDash \alpha$ si et seulement si $\vDash f_a \Rightarrow \alpha$. Notons ici que la formule f_a existe pour toute alternative a dans une logique propositionnelle puisque qu'il est possible d'énumérer tous les modèles d'une telle logique et de construire des formules qui sont vraies pour chacun des modèles et seulement pour celui-là. Il en est de même pour une logique des prédicats sans symboles de fonction autre que les symbole de constantes et avec des symboles de prédicats et de constantes en nombre fini. Par la suite, nous supposons que la formule f_a existe pour toute alternative a. Dans le cas général, cette formule ne peut pas être explicitée car le nombre de modèles de la logique considérée est généralement infini. La formule f_a sera donc appréhendée partiellement à travers les formules qu'elle vérifie.

4.3.2 Préférence

De façon générale, nous modélisons une *préférence*[3] par une relation binaire définie sur l'ensemble des alternatives. En particulier, une *préférence partielle* est une relation de préférence circonscrite à un point de vue. Nous noterons par la suite de telles relations avec les symboles $G, G_1, G_2, \ldots, G_i, G'_1, \ldots$ où i désigne le numéro du point de vue associé. Plus précisément, la notation $a_1 G_i a_2$ signifie que, du point de vue numéro i, l'alternative a_1 est préférée à l'alternative a_2.

Une *préférence* étant une relation binaire, elle peut-être vue comme un ensemble de couples d'alternatives. Dans notre formalisme, cet ensemble n'a pas de propriété a priori (transitivité, symétrie, complétude etc.). Il peut par exemple contenir des couples « contradictoires » entre eux : à la fois aG_ib et bG_ia.

Lorsque des préférences sont spécifiées, elles le sont généralement au travers de généralités, c'est-à-dire au travers de comparaisons de propriétés (voir le point (2.3.1) de la section 2.3). C'est pour cela que pour nous associons à chaque préférence un opérateur (logique modal) sur les formules de $\mathscr{L}$. De façon générale, nous notons respectivement $\succeq_i$, $\geq_i$, $\geq$ les opérateurs logiques modal relatifs respectivement aux relations de préférence G_i, G'_i et G. C'est l'ajout de ces opérateurs au langage $\mathscr{L}$ qui définit respectivement les langages $\mathscr{L}^1$, $\mathscr{L}^2$ et $\mathscr{L}^3$ (voir les trois chapitres suivants pour plus de détails). Par extension, la propriété ϕ sera dite préférée à la propriété ψ par rapport à la préférence G_i (respectivement à G'_i et G) si et seulement si la formule $\phi \succeq_i \psi$ (respectivement $\phi \geq_i \psi$ et $\phi \geq \psi$) est vraie.

4.3.3 Indifférence

Deux alternatives sont dites *indifférentes* suivant une préférence donnée (et donc selon un point de vue) si et seulement si chacune d'elles est préférée à l'autre suivant la préférence considérée ou bien si aucune d'elles n'est préférée à l'autre suivant la préférence considérée. Plus précisément, l'indifférence est définie comme la contraposée de la *préférence stricte*. En particulier, pour nous, l'indifférence n'est pas forcément transitive.

Dans [Roubens and Vincke, 1985], un individu (*a decision maker*) confronté à deux alternatives distinctes est supposé, de façon exclusive, soit (1) en préférer une à l'autre, soit

[3] Remarquons ici que ce que nous appelons *préférence* est quelquefois appelé dans la littérature *préférence large* (voir [Roubens and Vincke, 1985]).

(2) les trouver indifférentes, soit (3) les trouver incomparables. Dans le contexte de notre travail, étant donné que nous voulons faire un choix, les deux derniers cas sont regroupés dans ce que nous appelons l'*indifférence*. En particulier, deux alternatives incomparables seront considérées comme indifférentes au regard de la décision à prendre. Par conséquent si nous devons choisir entre deux alternatives que nous jugeons incomparables, nous en prendrons une au hasard.

Ceci pourrait être considéré comme gênant au regard de la critique faite dans la section 1.2 : « l'agent doit pouvoir déterminer sans ambiguïté sa réaction ». Il n'en est rien. En effet, dans cette section nous avons seulement soutenu qu'il est nécessaire pour un agent ,à chaque instant, (1) de mettre en œuvre une unique alternative, (2) d'avoir une description précise de ce qu'il met en œuvre et (3) d'avoir les moyens de décider ce qu'il doit mettre en œuvre. Plus précisément, il est tout a fait possible (à cause d'un manque d'information) qu'il n'ait aucun argument pour départager deux alternatives. Dans une telle situation, il choisira au hasard une des alternatives qui lui reste.

4.4 Exemple de scénario

Afin d'illustrer le formalisme que nous développons dans les prochaines sections, nous considérons une application visant à proposer à un utilisateur dénommé Pierre, un plat dans un restaurant afin que celui-ci puisse aller se restaurer. Cette application peut être modélisée par un agent dont l'objectif est de choisir, pour cet utilisateur, le plat qu'il va manger ce soir.

Afin de traiter ce scénario, nous supposons que le langage $\mathscr{L}$ permettant de décrire les alternatives contient entre autre les quatre constantes "Pierre", "le Goéland", "la Légende", "le Capuccino" ; ainsi que les six prédicats listés ci-dessous avec leur signification :

- Poisson(y) : y contient du poisson,
- Viande(y) : y contient de la viande,
- Pizza(y) : y est une pizza,
- Classique(y) : y est un plat classique,
- Plat(x, y) : y est sur la carte du restaurant x,
- Manger(x, y) : la personne dénommée x mange le plat y.

Ce langage permet de spécifier que nous appelons une théorie des restaurants (notée $\mathscr{R}$). Plus précisément, $\mathscr{R}$ est un ensemble d'énoncés applicatifs qui décrivent les relations qu'entretiennent les propriétés (des restaurants) entre elles. Cette théorie admet des modèles : ce sont les alternatives imaginables (i.e. si a appartient à ALT alors $a \vDash \mathscr{R}$). Parmi ces dernières, l'agent en considère explicitement une partie : les alternatives de l'ensemble ALT^*. Considérons que l'agent a par exemple réussi à déterminer (d'une manière non précisée ici) que les trois alternatives a, b et c sont à considérer. Celles-ci correspondent (informellement) respectivement à « aller manger du poisson au restaurant le Goéland » ; « aller manger du poisson au restaurant la Légende » ; « aller manger une pizza au restaurant le Capuccino ». Dans notre formalisme cela signifie que les modèles a, b et c sont tels que :

$$f_a \vdash \exists y,\ \text{Manger(“Pierre”}, y) \wedge \text{Plat}(y, \text{“le Goéland”}) \wedge \text{Poisson}(y) \quad \text{(ex.1)}$$

$$f_b \vdash \exists y,\ \text{Manger}(\text{“Pierre”}, y) \wedge \text{Plat}(y, \text{“la Légende”}) \wedge \text{Poisson}(y) \tag{ex.2}$$

$$f_c \vdash \exists y,\ \text{Manger}(\text{“Pierre”}, y) \wedge \text{Plat}(y, \text{“le Capuccino”}) \wedge \text{Pizza}(y) \tag{ex.3}$$

Afin de choisir l'une de ces trois alternatives, l'agent dispose de certaines *connaissances* sur l'univers. Ces connaissances sont modélisées en ajoutant des énoncés (du domaine) à la théorie $\mathscr{R}$. Techniquement, ceux-ci sont décrits par des schémas. En particulier, dans la situation qui nous intéresse, l'agent sait que

– Une pizza n'est pas un plat Classique :

$$\mathscr{R} \vdash \text{Pizza}(Y) \Rightarrow \neg\text{Classique}(Y) \tag{ex.4}$$

– Le Capuccino sert des Pizzas :

$$\mathscr{R} \vdash \exists x,\ \text{Plat}(x, \text{“le Capuccino”}) \wedge \text{Pizza}(x) \tag{ex.5}$$

– Le Goéland et la Légende ne servent que des plats classiques :

$$\mathscr{R} \vdash \forall x,\ \text{Plat}(x, \text{“le Goéland”}) \Rightarrow \text{Classique}(x) \tag{ex.6}$$

$$\mathscr{R} \vdash \forall x,\ \text{Plat}(x, \text{“la Légende”}) \Rightarrow \text{Classique}(x) \tag{ex.7}$$

– Il n'y a que du poisson au Goéland :

$$\mathscr{R} \vdash \forall x,\ (\text{Plat}(x, \text{“le Goéland”})) \Rightarrow (\text{Poisson}(x) \wedge \neg\text{Viande}(x)) \tag{ex.8}$$

– Il y a au moins un plat à la Légende contenant de la viande et sans poisson :

$$\mathscr{R} \vdash \exists x,\ \text{Plat}(x, \text{“la Légende”}) \wedge \text{Viande}(x) \wedge \neg\text{Poisson}(x) \tag{ex.9}$$

Ces connaissances indiquent, entre autres, que les propriétés de l'univers ont des relations entres elles et que par conséquent le choix de l'une des alternatives, même s'il peut apparaître comme le choix d'une valuation d'un petit nombre de propriétés (ici uniquement le nom du restaurant), porte en réalités sur de nombreuses propriétés. Par conséquent, ces schémas permettent à l'agent de déterminer plus précisément les alternatives parmi lesquelles il doit faire son choix. En particulier, dans la situation qui nous intéresse, cela lui permet de déduire que :

– Pierre peut aller manger un plat (classique) de poisson (sans viande) au restaurant le Goéland, un restaurant qui ne fait que des plats de poisson.

$$\begin{aligned} f_a, \mathscr{R} \vdash\ & \exists y,\ \text{Manger}(\text{“Pierre”}, y) \wedge \text{Plat}(y, \text{“le Goéland”}) \wedge \text{Poisson}(y) \\ \wedge\ & \text{Classique}(y) \wedge \neg\text{Viande}(y) \\ \wedge\ & \forall x,\ \text{Plat}(x, \text{“le Goéland”}) \Rightarrow [\text{Poisson}(x) \wedge \neg\text{Viande}(x)] \end{aligned} \tag{ex.10}$$

– Pierre peut aller manger un plat (classique) de poisson (sans viande) au restaurant la Légende, un restaurant qui fait aussi des plats de viande (sans poisson).

$$\begin{aligned} f_b, \mathscr{R} \vdash\ & \exists y,\ \text{Manger}(\text{“Pierre”}, y) \wedge \text{Plat}(y, \text{“la Légende”}) \wedge \text{Poisson}(y) \\ \wedge\ & \text{Classique}(x) \\ \wedge\ & \exists x,\ \text{Plat}(x, \text{“la Légende”}) \wedge \text{Viande}(x) \wedge \neg\text{Poisson}(x) \end{aligned} \tag{ex.11}$$

– Pierre peut aller manger une pizza (un plat non classique) au restaurant le Capuccino.

$$\begin{aligned} f_c, \mathscr{R} \vdash\ & \exists y,\ \text{Manger}(\text{“Pierre”}, y) \wedge \text{Plat}(y, \text{“le Capuccino”}) \wedge \text{Pizza}(y) \\ \wedge\ & \neg\text{Classique}(y) \end{aligned} \tag{ex.12}$$

Ces informations peuvent être interprétées comme un moyen pour l'agent de connaître les « conséquences » du choix de chacune des alternatives. Elles seront reprises tout comme cet exemple au fil des chapitres 5, 6 et 7 afin d'expliciter notre formalisme de la préférence. En effet, à ce point, aucune information sur la désirabilité des alternatives ou de leurs propriétés n'a encore été donnée. Pour reprendre le parallélisme fait avec les logiques de description, jusqu'ici seule la K-box de l'agent a été décrite.

Troisième partie

Formalisation des préférences en logique des prédicats du premier ordre

Chapitre 5

Préférences partielles primitives

Afin de permettre à l'agent de faire un choix motivé, les informations qui peuvent être spécifiées dans un langage classique de logique des prédicats du premier ordre sont insuffisantes. En effet, comme nous l'avons remarqué au début de la section 2.1.3, il est indispensable de posséder de surcroît des informations indiquant la « désirabilité » des diverses alternatives possibles. Ainsi, dans l'exemple présenté précédemment dans la section 4.4 rien ne permet de départager les trois alternatives. Dans notre formalisme, la « désirabilité » de chaque alternative est en premier lieu basée sur ce que nous appelons les *préférences partielles primitives*. Ce chapitre présente la partie de notre formalisme relative à cette notion.

Pour nous, les informations décrivant les préférences ont, en plus de leur sens « premier », un sens implicite qui permet de spécifier les préférences de façon synthétique. Par conséquent, afin de raisonner correctement sur ces informations pour en particulier prendre une décision, il est nécessaire d'expliciter ce sens. Dans notre formalisme c'est le rôle des *préférences partielles primitives*. Plus précisément, l'objectif des *préférences partielles primitives* est de fournir un cadre logique afin de d'expliciter le sens des informations fournies lors de la spécification des préférences (voir le schéma 3.2). En particulier, nous pensons que des contradictions entre les informations exprimées, plutôt que de mener à des inconsistances dans la théorie, doivent simplement indiquer une indifférence. De plus, les informations spécifiées sont implicitement transitives. Par exemple, lorsqu'un utilisateur indique simultanément qu'il préfère le « rouge » au « bleu » et le « bleu » au « vert », il indique implicitement qu'il préfère aussi le « rouge » au « vert ». Enfin, pour nous, un aspect important de tout formalisme permettant de spécifier des préférences de façon naturelle et intuitive est de permettre, lors de la spécification des préférences, d'indiquer facilement les informations qui peuvent avoir des dépendances entre elles et les informations qui a priori n'en ont pas.

Pour ce faire, le formalisme que nous proposons se base sur la répartition des informations nécessaires aux choix en différents points de vue indépendants. Plus précisément nous proposons l'introduction d'une relation de préférence G_i pour chaque point de vue i : c'est précisément ce que nous appelons une *préférence partielle primitive*. La relation de *préférence partielle primitive* est transitive afin d'exprimer le fait que si un agent indique qu'il préfère une alternative a_1 à une alternative a_2 ainsi que l'alternative a_2 à une alternative a_3 alors il indique aussi implicitement qu'il préfère l'alternative a_1 à l'alternative a_3. Elle est de plus réflexive afin d'exprimer le fait que deux alternatives identiques sont jugées indifférentes.

Nous avons vu au point (2.3.1) de la section 2.3 que les préférences sont généralement spécifiées par le biais de comparaisons entre les propriétés. Pour intégrer ce fait nous introduisons de surcroît n opérateurs logiques modaux ; un associé à chaque préférence partielle primitive. Le sens de ces comparaisons se base sur les *préférences partielles primitives* via l'hypothèse *Ceteris Paribus*. C'est pourquoi, dans un premier temps nous définissons plus précisément la propriété *Ceteris Paribus*. Dans un second temps nous présentons la syntaxe, la sémantique et l'axiomatique des opérateurs associés aux préférences partielles primitives. Par la suite, nous dégagerons les propriétés remarquables de tels opérateurs. Enfin, nous reprenons l'exemple introduit dans la section 4.4 pour concrétiser notre proposition.

5.1 L'hypothèse *Ceteris Paribus*

5.1.1 Introduction :

Afin de coller à l'intuition, nous faisons l'hypothèse *Ceteris Paribus*, c'est-à-dire que les préférences spécifiées le sont « toutes choses égales par ailleurs ». En effet, comme nous l'avons remarqué dans le point (2.3.1) de la section 2.3, pour beaucoup de chercheurs, la majorité de nos préférences sont de ce type.

Deux alternatives a et b sont « égales » si elles vérifient les mêmes propriétés. Par extension, étant donné la formule ϕ, nous disons que deux alternatives a et b sont « égales par ailleurs » si elles vérifient les mêmes propriétés pourvu que ces dernières n'aient pas de rapport avec la formule ϕ. De même, étant donné les formules ϕ et ψ, deux alternatives a et b sont dites « égales par ailleurs » si elles vérifient les mêmes propriétés pourvu que ces dernières n'aient pas de rapport avec les formules ϕ et ψ. C'est pour cela que nous interprétons dans ce document le sens de « toute choses égales par ailleurs », comme le fait que « toute propriété sans rapport avec les propriétés considérées, soit est vérifiée par les deux alternatives considérées, soit ne l'est par aucune ». Si c'est le cas, les deux alternatives a et b seront dites alors *Ceteris Paribus* vis-à-vis des deux formules ϕ et ψ.

Intuitivement, une formule a un rapport avec une autre si l'une de ses parties (ou sa négation) est impliquée par une partie de l'autre (ou sa négation). Similairement, une formule a un rapport avec deux formules si l'une de ses parties (ou sa négation) est impliquée par une partie (ou la négation d'une partie) de l'une ou de l'autre.

Pour formaliser ceci, nous introduisons pour toutes formules closes ϕ et ψ l'ensemble $\mathscr{L}^{\phi\psi}$. Ce dernier désigne la partie du langage $\mathscr{L}$ construite seulement (1) avec les opérateurs $\wedge$ et $\neg$, (2) sur une partie des sous-formules des formules équivalentes à ϕ ou à ψ (voir le prochain paragraphe), et (3) ne contenant aucune formule équivalente ni à $\top$ ni à $\bot$.

Formellement, $\mathscr{L}^{\phi\psi}$ est tel que :

$$\begin{array}{l} \bullet\ \text{Si } \models \phi' \Leftrightarrow \phi \text{ et } \alpha \in \text{Subf*}(\phi') \text{ alors } \alpha \in \mathscr{L}^{\phi\psi} \\ \bullet\ \text{Si } \models \psi' \Leftrightarrow \psi \text{ et } \alpha \in \text{Subf*}(\psi') \text{ alors } \alpha \in \mathscr{L}^{\phi\psi} \\ \bullet\ \text{Si } \alpha \in \mathscr{L}^{\phi\psi} \text{ alors } \neg\alpha \in \mathscr{L}^{\phi\psi} \\ \bullet\ \text{Si } \big[(\alpha,\beta) \in \mathscr{L}^{\phi\psi} \times \mathscr{L}^{\phi\psi} \text{ et } \not\models \neg(\alpha \wedge \beta)\big] \text{ alors } \alpha \wedge \beta \in \mathscr{L}^{\phi\psi} \end{array} \tag{5.1}$$

Il est important de ne pas considérer dans la définition (5.1) toutes les sous-formules des formules équivalentes à ϕ ou à ψ. Plus précisément, il est nécessaire d'écarter les sous-formules qui sont, dans les formules équivalentes à ϕ ou à ψ, des sous-formules de la constante $\top$. En effet, leur valeur de vérité n'est pas pertinente pour la valeur de vérité des formules ϕ et ψ. C'est pourquoi nous n'utilisons pas dans la définition (5.1) les ensembles des sous-formules de ϕ' et ψ' (i.e. Subf(ϕ') et Subf(ψ')) mais les ensembles Subf*(ϕ') et Subf*(ψ') : respectivement l'ensemble des *sous-formules pertinentes* pour ϕ' et l'ensemble des *sous-formules pertinentes* pour ψ'. L'ensemble Subf*(ϕ) des **sous-formules pertinentes** d'une formule ϕ est défini récursivement par les six règles suivantes (dans l'ordre) :

$$\begin{array}{l} \bullet\ \text{Si } \phi \text{ est équivalent à la formule } \top, \text{ alors Subf*}(\phi) = \emptyset \\ \bullet\ \text{Si } \phi \text{ est équivalent à la formule } \bot, \text{ alors Subf*}(\phi) = \emptyset \\ \bullet\ \text{Si } \phi \text{ est atomique, alors Subf*}(\phi) = \{\phi\} \\ \bullet\ \text{Si } \phi \text{ est de la forme } \neg\psi, \text{ alors Subf*}(\phi) = \text{Subf*}(\psi) \cup \{\neg\psi\} \\ \bullet\ \text{Si } \phi \text{ est de la forme } \alpha \wedge \beta \text{ alors,} \\ \quad - \text{ si seule } \alpha \text{ est équivalente à la constante } \top, \text{ alors Subf*}(\phi) = \text{Subf*}(\beta) \\ \quad - \text{ si seule } \beta \text{ est équivalente à la constante } \top, \text{ alors Subf*}(\phi) = \text{Subf*}(\alpha) \\ \quad - \text{ si } \alpha \text{ et } \beta \text{ ne sont pas équivalentes à la constante } \top, \\ \quad\ \ \text{alors Subf*}(\phi) = \text{Subf*}(\alpha) \cup \text{Subf*}(\beta) \cup \{\alpha \wedge \beta\} \\ \bullet\ \text{Si } \phi \text{ est de la forme } \forall x\psi, \text{ alors Subf*}(\phi) = \text{Subf*}(\psi) \cup \{\forall x\psi\} \end{array} \tag{5.2}$$

Étant donnée une formule ϕ quelconque, il résulte de cette définition que, pour toute formule γ consistante avec ϕ, l'ensemble Subf*(ϕ) est un sous-ensemble de Subf*($\phi \wedge \gamma$). En particulier, Subf*(ϕ) est un sous-ensemble de Subf*($\phi + \overline{\psi}$). Plus précisément, si les formules ϕ et $\neg\psi$) sont inconsistantes entre elles, alors Subf*($\phi + \overline{\psi}$) = Subf*(ϕ) ; sinon, si $\neg\psi$ est valide alors Subf*($\phi + \overline{\psi}$) = Subf*(ϕ) ; sinon, si ϕ est valide (i.e équivalente à la constante $\top$) alors Subf*($\phi + \overline{\psi}$) = Subf*($\neg\psi$) ; sinon, Subf*($\phi + \overline{\psi}$) = Subf*(ϕ) $\cup$ Subf*($\neg\psi$) $\cup$ $\{\phi \wedge \neg\psi\}$.

Remarquons qu'avec de telles définitions, les constantes $\top$ et $\bot$ n'appartiennent jamais à l'ensemble $\mathscr{L}^{\phi\psi}$. En particulier, on a $\mathscr{L}^{\top\top} = \mathscr{L}^{\bot\top} = \mathscr{L}^{\top\bot} = \mathscr{L}^{\bot\bot} = \emptyset$. De plus, si ϕ_2 et ψ_2 sont des formules équivalentes respectivement aux formules ϕ_1 et ψ_1 alors $\mathscr{L}^{\phi_1\psi_1}$ et $\mathscr{L}^{\phi_2\psi_2}$ désignent le même ensemble. Enfin, par construction, les ensembles $\mathscr{L}^{\phi\psi}$ et $\mathscr{L}^{\phi+\overline{\psi}\phi+\overline{\psi}}$ sont les mêmes. Afin de raccourcir les notations dans la suite du document, nous abrégeons la notation $\mathscr{L}^{\phi\phi}$ par la notation $\mathscr{L}^{\phi}$.

5.1.2 Ceteris Paribus (Définition) :

Formellement, les alternatives a et b sont *Ceteris Paribus* vis-à-vis des formules closes ϕ et ψ (par commodité noté $CP(a, b, \phi, \psi)$) si et seulement si les alternatives a et b sont des modèles pour les mêmes formules closes si les *sous-formules* de ces dernières ainsi que leurs négations sont consistantes avec l'ensemble des formules construites sur l'ensemble des *sous-formules pertinentes* (voir la définition (5.2)) des formules équivalentes à ϕ ou à ψ (voir la définition (5.1)). En d'autre termes, a et b sont *Ceteris Paribus* vis-à-vis des formules closes ϕ et ψ si et seulement si, toutes les propriétés dont (i) les *sous-formules* sont consistantes avec l'ensemble des formules construites sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ et dont (ii) la négation de ces mêmes *sous-formules* sont aussi consistantes avec l'ensemble des formules construites sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ, sont (iii) des propriétés de a si et seulement si elles sont des propriétés de b :

$$CP(a,b,\phi,\psi) \;\; \text{ssi} \;\; \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{Si, } \forall\alpha' \in \mathrm{Subf}(\alpha) \text{ et } \forall\gamma \in \mathscr{L}^{\phi\psi}, & \\ \quad \left| \; \exists M,\;\; M \vDash \gamma\wedge\alpha' \right. & (\text{i}) \\ \quad \left| \; \exists M',\;\; M' \vDash \gamma\wedge\neg\alpha' \right. & (\text{ii}) \\ \text{alors} \;\; \left[a\vDash\alpha \;\;\text{ssi}\;\; b\vDash\alpha\right] & (\text{iii}) \end{array}\right] \tag{5.3}$$

En appliquant les règles de logique classique, on obtient de manière équivalente[1] que deux alternatives a et b sont *Ceteris Paribus* vis-à-vis des deux formules closes ϕ et ψ de $\mathscr{L}$ si et seulement si toute formule close α de $\mathscr{L}$ est telle que soit (i) une sous-formule de α ou une formule construite sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ est vraie, soit (ii) la négation d'une sous-formule de α ou une formule construite sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ est vraie, soit (iii) la formule α a pour modèle a si et seulement si elle a pour modèle b :

$$CP(a,b,\phi,\psi) \;\; \text{ssi} \;\; \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \;\text{soit} \;\; \exists\alpha' \in \mathrm{Subf}(\alpha) \text{ et } \exists\gamma \in \mathscr{L}^{\phi\psi} \text{ tels que } \vDash \gamma\vee\alpha' & (\text{i}) \\ \;\text{soit} \;\; \exists\alpha' \in \mathrm{Subf}(\alpha) \text{ et } \exists\gamma \in \mathscr{L}^{\phi\psi} \text{ tels que } \vDash \gamma\vee\neg\alpha' & (\text{ii}) \\ \;\text{soit} \;\; \left[a\vDash\alpha \;\;\text{ssi}\;\; b\vDash\alpha\right] & (\text{iii}) \end{array}\right] \tag{5.4}$$

De manière analogue, deux formules closes β et γ sont dites *Ceteris Paribus* vis-à-vis des deux formules closes ϕ et ψ de $\mathscr{L}$ (noté $CP(\beta,\gamma,\phi,\psi)$), si et seulement si, toute formule close α de $\mathscr{L}$ est telle que soit (i) une sous-formule de α ou une formule construite sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ est vraie, soit (ii) la négation d'une sous-formule de α ou une formule construite sur les *sous-formules pertinentes* des formules équivalentes à ϕ ou à ψ est vraie, soit (iii-a) la formule α est subsumée par β si et seulement si elle l'est aussi par γ, soit (iii-b) la négation de la formule α est subsumée

[1] Voir l'annexe A.1.3 pour une preuve.

par β si et seulement si elle l'est aussi par γ :

$$CP(\beta,\gamma,\phi,\psi) \ \text{ssi} \ \begin{bmatrix} \forall\alpha \in \mathscr{L}, & \\ \text{soit} \ \ \exists\alpha' \in \mathrm{Subf}(\alpha) \text{ et } \exists\gamma \in \mathscr{L}^{\phi\psi} \text{ tels que } \models \gamma \vee \alpha' & (\ \text{i}\ ) \\ \text{soit} \ \ \exists\alpha' \in \mathrm{Subf}(\alpha) \text{ et } \exists\gamma \in \mathscr{L}^{\phi\psi} \text{ tels que } \models \gamma \vee \neg\alpha' & (\ \text{ii}\ ) \\ \text{soit} \ \ \left[\models \beta \Rightarrow \alpha \quad \text{ssi} \ \models \gamma \Rightarrow \alpha\right] & (\text{iii-a}) \\ \text{soit} \ \ \left[\models \beta \Rightarrow \neg\alpha \quad \text{ssi} \ \models \gamma \Rightarrow \neg\alpha\right] & (\text{iii-b}) \end{bmatrix} \tag{5.5}$$

5.1.3 Exemples

Pour fixer cette notion, reprenons l'exemple présenté à la section 4.4. Considérons plus particulièrement les alternatives a et b (respectivement « manger du poisson au Goéland » et « manger du poisson à la Légende »). Manifestement, ces deux alternatives ne sont pas « Ceteris Paribus » vis-à-vis des formules $\exists y\mathrm{Poisson}(y)$ et $\exists y\mathrm{Manger}(\text{“Pierre”}, y)$. En effet, la formule $\alpha = \exists y\mathrm{Plat}(y, \text{“la Légende”})$ ne vérifie aucune des conditions (i), (ii) et (iii) de la définition (5.4) Pour s'en convaincre, il suffit de remplacer dans la définition (5.4) respectivement ϕ par $\exists y\mathrm{Poisson}(y)$, ψ par $\exists y\mathrm{Manger}(\text{“Pierre”}, y)$, α par $\exists y\mathrm{Plat}(y, \text{“la Légende”})$ et de remarquer que :

- $\mathrm{Subf}(\exists y\mathrm{Plat}(y, \text{“la Légende”})) = \left\{ \begin{array}{c} \exists y\mathrm{Plat}(y, \text{“la Légende”}) \\ \mathrm{Plat}(y, \text{“la Légende”}) \end{array} \right\}$,
- $\mathrm{Subf}^*(\exists y\mathrm{Manger}(\text{“Pierre”}, y)) = \left\{ \begin{array}{c} \neg(\forall x \neg(\mathrm{Manger}(\text{“Pierre”}, x))) \\ \forall x \neg(\mathrm{Manger}(\text{“Pierre”}, x)) \\ \neg(\mathrm{Manger}(\text{“Pierre”}, x)) \\ \mathrm{Manger}(\text{“Pierre”}, x) \end{array} \right\}$,
- $\mathrm{Subf}^*(\exists y\mathrm{Poisson}(y)) = \left\{ \begin{array}{c} \neg(\forall z \neg\mathrm{Poisson}(z)) \\ \forall z \neg\mathrm{Poisson}(z) \\ \neg\mathrm{Poisson}(z) \\ \mathrm{Poisson}(v) \end{array} \right\}$,
- $\mathscr{L}^{\exists y\mathrm{Poisson}(y)\ \exists y\mathrm{Manger}(\text{“Pierre”}, y)}$ correspond à l'ensemble des formules contingentes (i.e. non valides et non inconsistantes) construites sur la base des formules des ensembles $\mathrm{Subf}^*(\exists y\mathrm{Manger}(\text{“Pierre”}, y))$ et $\mathrm{Subf}^*(\exists y\mathrm{Poisson}(y))$. Pour plus de simplicité, nous supposons qu'il n'existe aucune formule équivalente à ϕ ou à ψ.
- La formule $\exists y\mathrm{Plat}(y, \text{“la Légende”})$ est une conséquence de la seconde alternative mais pas de la première.

5.1.4 Propriétés remarquables :

a) Comportement avec l'équivalence logique

La forme des formules ϕ et ψ considérées n'a pas d'importance puisque la notation $CP(a, b, \phi, \psi)$ est définie en considérant l'ensemble des formules équivalentes à ϕ et à ψ.

En particulier la propriété suivante est vérifiée :

$$\text{Si } \left[\ \vDash \phi_1 \Leftrightarrow \phi_2 \quad \text{et} \quad \vDash \psi_1 \Leftrightarrow \psi_2\ \right] \text{ alors } \left[\ CP(a,b,\phi_1,\psi_1) \quad \text{ssi} \quad CP(a,b,\phi_2,\psi_2)\ \right] \tag{5.6}$$

Pour une preuve voir l'annexe A.1.1.

b) Comportement avec la propriété d'Expansion

Par construction, les ensembles $\mathscr{L}^{\phi\psi}$ et $\mathscr{L}^{\phi+\overline{\psi}}$ sont les mêmes. Par conséquent, l'utilisation dans les définitions (5.3), (5.4) et (5.5) des formules ϕ et ψ ou $\phi+\overline{\psi}$ et $\psi+\overline{\phi}$ n'a pas d'importance. Il en résulte que la propriété suivante est vérifiée :

$$CP(a,b,\phi,\psi) \quad \text{ssi} \quad CP(a,b,\phi+\overline{\psi},\psi+\overline{\phi}) \tag{5.7}$$

c) Comportement avec la propriété de Spécialisation

Intuitivement, étant données deux formules ϕ et ψ, si deux alternatives sont « égales par ailleurs » alors ces deux alternatives sont aussi « égales par ailleurs » si les formules ϕ et ψ sont « spécialisées ». C'est le cas dans notre formalisme. En effet, si la formule γ n'est ni inconsistante avec ϕ ni inconsistante avec ψ alors l'ensemble $\mathscr{L}^{\phi\psi}$ est un sous-ensemble de l'ensemble $\mathscr{L}^{\phi\wedge\gamma\psi\wedge\gamma}$.

Il en résulte que pour toute formule close γ qui n'est ni inconsistante avec ϕ ni inconsistante avec ψ, si les alternatives a et b sont Ceteris Paribus vis-à-vis des formules closes ϕ et ψ alors les alternatives a et b sont aussi Ceteris Paribus vis-à-vis des formules closes $\phi\wedge\gamma$ et $\psi\wedge\gamma$:

$$\text{Si } \left[\begin{array}{l} CP(a,b,\phi,\psi) \\ \nvDash \neg(\phi\wedge\gamma) \\ \nvDash \neg(\psi\wedge\gamma) \end{array}\right] \text{ alors } \quad CP(a,b,\phi\wedge\gamma,\psi\wedge\gamma) \tag{5.8}$$

d) Comportement avec les descriptions totales d'alternatives

Si ϕ ou ψ représente une unique alternative alors quelles que soient les alternatives a et b, a sera égale b toutes choses égales par ailleurs. En particulier, quelles que soient les alternatives a et b, $CP(a,b,f_a,f_b)$ est toujours vérifiée. Pour une preuve, reportez-vous à l'annexe A.1.2.

e) Comportement avec les constantes $\top$ et $\bot$

La notation $CP(a,b,\phi,\psi)$ indique intuitivement que a et b ont les mêmes propriétés si on ne regarde pas les propriétés relatives à ϕ et à ψ. Par conséquent plus ϕ et ψ désignent un ensemble petit de propriétés, plus a et b vérifient les mêmes propriétés. En particulier, si ϕ et ψ désignent un ensemble de propriétés vide (i.e. $\top$) alors a et b représentent la même alternative :

$$CP(a,b,\top,\top) \quad \text{si et seulement si} \quad a=b \tag{5.9}$$

Dans notre formalisme, ceci est vérifié puisque l'ensemble Subf*($\top$) et donc l'ensemble $\mathscr{L}^{\top\top}$ est vide. De même, $CP(a, b, \perp, \perp)$ est vérifié si et seulement si a et b représentent la même alternative. Pour la même raison, $CP(a, b, \phi, \top)$, $CP(a, b, \phi, \perp)$, $CP(a, b, \top, \phi)$) et $CP(a, b, \perp, \phi)$) sont vérifiées si et seulement si $CP(a, b, \phi, \phi)$ l'est.

f) Propriété *Ceteris Paribus* et égalité « Classique »

Intuitivement, la propriété *Ceteris Paribus* est une propriété plus faible que l'égalité classique. Par conséquent, comme une alternative est toujours « égale » à elle même, étant données n'importes quelles formules ϕ et ψ, une alternative est aussi toujours « égale par ailleurs » à elle même. Plus précisément, si a et b désignent la même alternative alors, quelles que soient les formules ϕ et ψ, la propriété $CP(a, b, \phi, \psi)$ est vérifiée :

$$\text{Si } a = b \quad \text{alors } CP(a, b, \phi, \psi) \tag{5.10}$$

Preuve immédiate en remarquant que si a et b représente la même alternative alors pour toute formule close α, soit a et b implique α, soit a et b implique $\neg\alpha$. Par conséquent le cas (iii) de la définition (5.4) (et donc CP) est toujours vrai si a et b désignent la même alternative.

5.2 La relation de *préférence partielle primitive* G_i

Afin de formaliser notre vision, nous proposons l'introduction d'une relation binaire G_i pour chaque point de vue i : c'est précisément ce que nous appelons une *préférence partielle primitive*. Chaque relation de *préférence partielle primitive* G_i (circonscrite au point de vue i) est une relation binaire définie sur l'ensemble des alternatives ALT qui est transitive et réflexive. Chaque relation de *préférence partielle primitive* G_i est donc un pré-ordre partiel[2] (aussi appelé *quasi order* en anglais). La relation de *préférence partielle primitive* est transitive afin d'exprimer le fait que si un agent indique qu'il préfère une alternative a_1 à une alternative a_2 ainsi que l'alternative a_2 à une alternative a_3 alors il indique aussi implicitement qu'il préfère l'alternative a_1 à l'alternative a_3. Elle est de plus réflexive afin d'exprimer le fait que deux alternatives identiques sont jugées indifférentes.

5.3 Syntaxe de l'opérateur $\succeq_i$

À chaque relation binaire de *préférence partielle primitive* G_i est associé un opérateur logique modale défini sur l'ensemble des formules de $\mathscr{L}$. Nous le noterons avec le symbole $\succeq_i$ lorsque la relation correspondante est associée au point de vue i. L'ensemble des opérateurs $\succeq_i$ définissent le langage $\mathscr{L}^1$: si ϕ et ψ sont des formules bien formées de $\mathscr{L}$ (et non de $\mathscr{L}^1$) et si $\succeq_i$ dénote un opérateur logique associé à une relation binaire de *préférence partielle*

[2]Cette relation de préférence n'est pas nécessairement un ordre car deux alternatives qui sont chacune préférées à l'autre (et qui sont donc indifférentes) ne sont pas nécessairement les mêmes.

primitive, alors les formules $\phi \succeq_i \psi$ et $\neg(\phi \succeq_i \psi)$ sont des formules bien formées de $\mathscr{L}^1$. De plus, si F_1 et F_2 sont deux formules bien formées de $\mathscr{L}^1$ alors $F_1 \wedge F_2$ est une formule bien formée de $\mathscr{L}^1$.

Afin d'alléger les notations, nous introduisons l'abréviation syntaxique $\phi \succ_i \psi$. Elle indique que la propriété ϕ est préférée à la propriété ψ (par rapport à la préférence partielle primitive associée au point de vue i) mais que l'inverse n'est pas vrai (i.e. $\phi \succ_i \psi$ si et seulement si $\phi \succeq_i \psi \wedge \neg(\psi \succeq_i \phi)$). On dira alors que la propriété ϕ sera *préférée strictement*, par rapport à la préférence partielle primitive associée au point de vue i, à la propriété ψ.

5.4 Sémantique de l'opérateur $\succeq_i$

5.4.1 Les modèles du langage $\mathscr{L}^1$

Afin d'interpréter ces nouvelles formules, nous définissons les modèles du langage $\mathscr{L}^1$ comme des triplets $(ALT, \models, \mathscr{G})$ où

- ALT est l'ensemble des alternatives : l'ensemble des modèles (S, V) de $\mathscr{L}$.
- $\models$ est la validité sémantique du langage servant à décrire les alternatives ($\mathscr{L}$).
- $\mathscr{G}$ est un ensemble des relations binaires sur ALT. Il contient toutes les relations de préférence nécessaires à notre formalisme et en particulier toutes les relations de préférence partielle primitive G_i.

Dans l'optique de clarifier les notations, nous introduisons les symboles $\models_{\mathscr{L}^1}$, $\vdash_{\mathscr{L}^1}$, $\not\models_{\mathscr{L}^1}$ et $\nvdash_{\mathscr{L}^1}$. Ils représentent respectivement la validité sémantique, la conséquence logique, la non validité sémantique et la non conséquence logique pour le langage $\mathscr{L}^1$.

5.4.2 Définition

Dire que la formule ϕ est préférée à la formule ψ via la préférence partielle primitive i (noté avec la formule $\phi \succeq_i \psi$ de $\mathscr{L}^1$) signifie que les alternatives qui satisfont la formule ϕ (et, si cela est cohérent, la négation de la formule ψ) sont préférées, suivant le point de vue numéro i, aux alternatives qui satisfont la formule ψ (et, si cela est cohérent, la négation de la formule ϕ) si elles sont *Ceteris Paribus* vis-à-vis de ϕ et ψ (i.e. identiques par ailleurs). La sémantique de cet opérateur est formalisée comme suit :

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}^1} \phi \succeq_i \psi \quad \text{ssi} \quad \begin{array}{l} \bullet \quad \not\models \neg\phi \quad \text{et} \quad \not\models \neg\psi \\ \bullet \left[\begin{array}{l} \forall (a,b) \in \mathrm{ALT}^2; \quad \text{tel que} \left| \begin{array}{l} a \models \phi + \overline{\psi} \\ b \models \psi + \overline{\phi} \end{array} \right. \\ \text{si} \quad CP(a,b,\phi+\overline{\psi},\psi+\overline{\phi}) \quad \text{alors} \quad aG_ib \end{array} \right] \end{array} \tag{5.11}$$

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}^1} \neg(\phi \succeq_i \psi) \quad \text{ssi} \quad (\mathrm{ALT}, \models, \mathscr{G}) \not\models_{\mathscr{L}^1} \phi \succeq_i \psi \tag{5.12}$$

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} F_1 \wedge F_2 \quad \text{ssi} \quad \Big[(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} F_1 \ \text{ et } \ (\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} F_2\Big] \tag{5.13}$$

Il pourrait sembler suffisant d'utiliser directement dans la définition (5.11) la formule ϕ à la place de $\phi+\overline{\psi}$. Nous ne le faisons pas car comme nous supposons, d'une part, que les préférences sont décrites via des comparaisons de formules mutuellement exclusives et que, d'autre part, les informations explicitement spécifiées ne vérifient pas forcément cette propriété, il est nécessaire d'expliciter cet aspect.

5.4.3 Corollaire 1

D'après la définition (5.11), si ϕ est préférée à la formule ψ via la préférence partielle primitive i (i.e. $(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} \phi \succeq_i \psi$) alors $\neg\phi$ et $\neg\psi$ ne sont pas des tautologies (i.e. $\nvDash \neg\phi$ et $\nvDash \neg\psi$). De plus, pour tout couple d'alternatives a et b, s'il existe un couple de formules (ϕ, ψ) de $\mathscr{L}$ tel que ϕ est préférée à ψ, a est un modèle de $\phi+\overline{\psi}$, b est un modèle de $\psi+\overline{\phi}$ et a et b sont égales toutes choses par ailleurs par rapport à ϕ etψ, alors a est préférée à b. Formellement, il en résulte que, pour tout couple d'alternatives a et b,

$$\text{Si} \quad \left[\begin{array}{ll} \exists(\phi,\psi) \in \mathscr{L}^2 \quad \text{tels que :} & \\ \bullet\ (\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} \phi \succeq_i \psi & \bullet\ a \vDash \phi+\overline{\psi} \\ \bullet\ CP(a,b,\phi+\overline{\psi},\psi+\overline{\phi}) & \bullet\ b \vDash \psi+\overline{\phi} \end{array}\right] \quad \text{alors} \quad aG_ib \tag{5.14}$$

5.4.4 Corollaire 2

D'après la définition (5.11), les formules $\phi \succeq_i \psi$ et $\phi+\overline{\psi} \succeq_i \psi+\overline{\phi}$ ont le même sens :

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} \phi \succeq_i \psi \quad \text{si et seulement si} \quad (\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} \phi+\overline{\psi} \succeq_i \psi+\overline{\phi} \tag{5.15}$$

On prouve que les formules $\phi \succeq_i \psi$ et $\phi+\overline{\psi} \succeq_i \psi+\overline{\phi}$ ont le même sens en remplaçant dans la définition sémantique de l'opérateur de préférence partielle (5.11) la formule ϕ par $\phi+\overline{\psi}$ et la formule ψ par $\psi+\overline{\phi}$, et en remarquant que $[\phi+\overline{\psi}]+\overline{[\psi+\overline{\phi}]}$ est équivalent à $\phi+\overline{\psi}$ et que $[\psi+\overline{\phi}]+\overline{[\phi+\overline{\psi}]}$ est équivalent à $\psi+\overline{\phi}$ (voir la propriété (ix) de la section 4.2.2), et que de surcroît $CP(a, b, \psi+\overline{\phi}, \phi+\overline{\psi})$ est vérifiée si et seulement si $CP(a, b, \phi, \psi)$ l'est (propriété (5.7)).

5.4.5 Lien entre l'opérateur $\succeq_i$ et la relation G_i

Comme énoncé dans la section 4.3.1, à chaque alternative a de ALT est associée une formule f_a de $\mathscr{L}$ telle que pour chaque formule close α de $\mathscr{L}$, de façon exclusive, soit f_a subsume α, soit f_a subsume $\neg\alpha$ et telle que seule l'alternative a est un modèle de cette formule (i.e. $\forall b \in ALT,\ b \vDash f_a$ ssi $a = b$). Par conséquent, si un couple d'alternatives appartient à une relation binaire de préférence partielle primitive alors les formules associées sont reliées par l'opérateur logique correspondant. De plus si deux formules représentant chacune une unique alternative sont mises en relation par un opérateur logique $\succeq_i$ alors les deux alternatives correspondantes appartiennent à la relation binaire de préférence partielle primitive associée. Formellement,

$$\forall (a,b) \in \text{ ALT}^2_? \qquad aG_ib \quad \text{ssi} \quad (\text{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} f_a \succeq_i f_b \tag{5.16}$$

Preuve immédiate en identifiant dans la définition (5.11) les formules ϕ et ψ avec les formules f_a et f_b, et en remarquant de surcroît que f_a et f_b ne sont pas équivalent à $\bot$, que $f_a + \overline{f_b}$ équivaut à f_a et $f_b + \overline{f_a}$ équivaut à f_b (voir la propriété (v) de la section 4.2.2), et surtout que la propriété $CP(a, b, f_a, f_b)$ est toujours vérifiée (voir le point (d) de la section 5.1.4).

5.5 Axiomatisation de l'opérateur $\succeq_i$

Dans la pratique, chaque préférence partielle primitive G_i est générée à partir d'un ensemble initial de formules de la forme $\phi \succeq_i \psi$, où ϕ et ψ sont des formules de $\mathscr{L}$: ce sont les données initiales représentées dans la figure (3.2). Ces formules n'induisent pas forcément une relation de préférence antisymétrique ni même complète dans le sens où, par exemple, elles peuvent spécifier à la fois que $\phi \succeq_i \psi$ et que $\psi \succeq_i \phi$ ou à la fois ni l'une ni l'autre. Par contre, ces formules sont spécifiées de telle sorte qu'implicitement leur clôture transitive, Ceteris Paribus, et réflexive l'est aussi. C'est pourquoi les règles d'inférence syntaxique suivantes sont imposées à chaque opérateur (sur les formules de $\mathscr{L}$) associé à chacune des relations binaires (sur les alternatives) de préférence partielle primitive.

L'axiomatisation que nous proposons ici vise à « coller » au plus prés de la sémantique de cet opérateur. Elle est correcte mais sûrement pas complète.

5.5.1 Défaut (D)

Afin d'être en accord avec l'intuition, par défaut deux propriétés quelconques sont jugées indifférentes. Par conséquent Nous imposons que par défaut, s'il ne contraire n'est pas vrai alors une formule ϕ de $\mathscr{L}$ n'est pas préférée à une formule ψ de $\mathscr{L}$. Formellement le défaut normal sans prérequis (à la Reiter) suivant est valide dans le langage $\mathscr{L}^1$:

$$\top : \neg(\phi \succeq_i \psi) \rightarrow \neg(\phi \succeq_i \psi) \tag{5.17}$$

En reprenant la terminologie utilisée dans [Roubens and Vincke, 1985], Cela signifie que deux propriétés sont qualifiées par défaut « d'incomparables ».

5.5.2 Comportement de l'opérateur $\succeq_i$ avec l'équivalence logique

Afin d'être en accord avec l'intuition, des propriétés logiquement équivalentes peuvent être substituées les unes aux autres dans des préférences. Par conséquent Nous imposons que si deux formules sont logiquement équivalentes alors toute occurrence de l'une dans une préférence partielle primitive peut-être substituée par une occurrence de l'autre et inversement. Formellement,

$$\text{Si} \quad \left[\begin{array}{l} \vdash \phi_1 \Leftrightarrow \phi_2 \\ \vdash \psi_1 \Leftrightarrow \psi_2 \end{array}\right] \quad \text{alors} \quad \left[\ \vdash_{\mathscr{L}^1} \phi_2 \succeq_i \psi_2 \quad \text{si et seulement si} \quad \vdash_{\mathscr{L}^1} \phi_1 \succeq_i \psi_1\ \right] \tag{5.18}$$

L'adéquation de cette règle par rapport à la sémantique se prouve en remplaçant, dans la définition sémantique de l'opérateur de préférence partielle (5.11), les formule ϕ et ψ par des formules équivalentes (voir l'annexe A.2.2).

5.5.3 Règle d'inférence imposant la réflexivité (R)

Pour « forcer » l'indifférence entre deux propriétés il est plus sûr d'imposer que la première est préférée à la seconde et inversement. En effet, ainsi aucune déduction logique ne peut changer cette appréciation. C'est pourquoi, comme une indifférence existe toujours entre la même propriété on impose que toute formule de $\mathscr{L}$ non inconsistante est préférée à elle-même :

$$\text{si } \nvdash \neg\phi \quad \text{alors} \quad \vdash_{\mathscr{L}^1} \phi \succeq_i \phi \tag{5.19}$$

Cette règle est correcte par rapport à la sémantique de l'opérateur de *préférence partielle primitive* car la relation associée G_i est réflexive. Pour une preuve, se reporter à l'annexe A.2.3.

5.5.4 Règle d'inférence imposant la transitivité (T)

Comme indiqué en introduction, l'intuition indique que les préférences vérifient une certaine forme de transitivité. C'est pourquoi on impose que si une formule ϕ_1 est préférée à une formule ϕ_2 et si de même la formule ϕ_2 est préférée à une formule ϕ_3, alors la formule ϕ_1 est préférée à la formule ϕ_3. C'est ce que formalise la règle d'inférence suivante :

$$\text{si } \vdash_{\mathscr{L}^1} \phi_1 \succeq_i \phi_2 \ \text{ et } \vdash_{\mathscr{L}^1} \phi_2 \succeq_i \phi_3 \quad \text{alors } \vdash_{\mathscr{L}^1} \phi_1 \succeq_i \phi_3 \tag{5.20}$$

Cette règle est correcte par rapport à la sémantique de l'opérateur de *préférence partielle primitive* car la relation associée G_i est transitive. Pour une preuve, se reporter à l'annexe A.2.4.

5.5.5 Règles d'expansion (E)

Afin de rester en accord avec le corollaire (5.15) de la section 5.4 relatif à la sémantique de la préférence partielle primitive, nous imposons qu'une formule du type $\phi \succeq_i \psi$ est équivalente logiquement à la formule $\phi + \overline{\psi} \succeq_i \psi + \overline{\phi}$. Formellement les deux règles suivantes sont imposées :

$$\text{Si} \ \vdash_{\mathscr{L}^1} \phi \succeq_i \psi \quad \text{alors} \quad \vdash_{\mathscr{L}^1} \phi + \overline{\psi} \succeq_i \psi + \overline{\phi} \tag{5.21}$$

$$\text{Si} \ \vdash_{\mathscr{L}^1} \phi + \overline{\psi} \succeq_i \psi + \overline{\phi} \quad \text{alors} \quad \vdash_{\mathscr{L}^1} \phi \succeq_i \psi \tag{5.22}$$

Ces règles permettent entre autres d'exprimer le fait que lorsqu'il est spécifié que la propriété ϕ est préférable à ψ, cela indique que la propriété décrite par la formule ϕ (et si cela est cohérent, la négation de ψ) est préférée, suivant le point de vue numéro i, à la propriété décrite par la formule ψ (et si cela est cohérent, la négation de ϕ). Remarquons ici qu'il nous semble que la première règle sera dans la pratique la plus souvent utilisée afin d'indiquer que lorsqu'une formule du type $\phi \succeq_i \psi$ est spécifiée alors la formule $\phi{+}\overline{\psi} \succeq_i \psi{+}\overline{\phi}$ est aussi implicitement spécifiée.

Ces règles sont correctes par rapport à la sémantique de l'opérateur de *préférence partielle primitive* puisqu'elles axiomatisent le corollaire (5.15).

5.5.6 Règle de spécialisation (S)

Afin de définir pratiquement une relation de *préférence partielle primitive* sur les alternatives G_i, il est nécessaire de pouvoir « spécialiser » les formules apparaissant de part et d'autre de l'opérateur associé $\succeq_i$. Ainsi, en partant d'une comparaison entre formules « simples », on peut obtenir, petit à petit, des comparaisons entre des descriptions totales d'alternatives. Néanmoins, afin que ce mécanisme soit en accord par rapport à la sémantique de l'opérateur $\succeq_i$, il est nécessaire que cette spécialisation se fasse de la même façon des deux côtés de l'opérateur et qu'elle n'entraîne pas d'incohérences (voir l'annexe A.2.1).

C'est pourquoi nous imposons à notre modèle que si une propriété ϕ est dite être préférée à une propriété ψ alors la spécialisation de ϕ suivant un aspect, non déjà spécifié de façon contraire, sera aussi dite être préférée à la même spécialisation de la propriété ψ. C'est ce que formalise la règle d'inférence suivante :

$$\text{si} \begin{bmatrix} \nvdash \neg(\phi{+}\overline{\psi}\wedge\alpha) \\ \nvdash \neg(\psi{+}\overline{\phi}\wedge\alpha) \\ \vdash_{\mathscr{L}^1} \phi \succeq_i \psi \end{bmatrix} \quad \text{alors} \quad \vdash_{\mathscr{L}^1} (\phi\wedge\alpha)\succeq_i(\psi\wedge\alpha) \tag{5.23}$$

Ainsi par exemple, en identifiant la formule α avec la $\forall y\ r(y)$, cette règle implique en particulier que si la formule $\exists x\ p(x) \succeq_i \exists x\ q(x)$ est vérifiée alors la formule $\exists x\ p(x) \wedge \forall y\ r(y) \succeq_i \exists x\ q(x) \wedge \forall y\ r(y)$ est aussi vérifiée.

Nous avons nommé cette règle « *Ceteris paribus* (CP) » dans nos précédents travaux afin d'insister sur le fait qu'elle intègre l'hypothèse *Ceteris Paribus* dans l'axiomatique (voir [Meyer et al., 2005a], [Meyer et al., 2005b] et [Meyer et al., 2005c]). Nous préférons ici changer d'appellation afin de limiter les confusions et pour insister sur la façon de l'utiliser.

5.5.7 Opérateur de préférence et constantes $\top$ et $\bot$

On prouve aisément, avec la règle d'inférence imposant la réflexivité, que la fait que $\top$ soit préféré à $\top$ est un théorème (i.e. $\vdash_{\mathscr{L}^1} \top \succeq_i \top$). On prouve aussi, avec la définition sémantique de l'opérateur $\succeq_i$, que c'est aussi une tautologie (i.e. $\models_{\mathscr{L}^1} \top \succeq_i \top$).

Par contre, même si la définition sémantique de l'opérateur $\succeq_i$ indique que cela n'a pas de sens d'avoir $\bot$ préféré à $\bot$ (i.e. $\models_{\mathscr{L}^1} \neg(\bot \succeq_i \bot)$) rien ne permet de le prouver dans le cas général (i.e. $\vdash_{\mathscr{L}^1} \neg(\bot \succeq_i \bot)$). On peut juste prouver compte tenu des axiomes existants que si la formule $\bot$ n'apparaît pas dans une préférence alors on a $\vdash_{\mathscr{L}^1} \neg(\bot \succeq_i \bot)$. Pour se prémunir de ce possible effet de bord, nous introduisons deux nouveaux axiomes :

$$\vdash_{\mathscr{L}^1} \neg(\phi \succeq_i \bot) \tag{5.24}$$

$$\vdash_{\mathscr{L}^1} \neg(\bot \succeq_i \phi) \tag{5.25}$$

5.6 Propriétés remarquables

La définition d'une telle axiomatique induit des propriétés intéressantes pour chaque opérateur logique de préférence partielle primitive. Nous listons ici celles qui nous sembles les plus remarquables.

5.6.1 Contraposition

Si une propriété est préférée via la préférence partielle primitive i à une autre propriété, alors la négation de cette dernière est préférée (via la préférence partielle primitive i) à la négation de la première :

$$\vdash_{\mathscr{L}^1} \phi \succeq_i \psi \quad \text{si et seulement si} \quad \vdash_{\mathscr{L}^1} \neg\psi \succeq_i \neg\phi \tag{5.26}$$

Preuve : d'après les règles d'expansion (5.21) et (5.22), on a $\phi \succeq \psi$ si et seulement si on a aussi $\phi{+}\overline{\psi} \succeq \psi{+}\overline{\phi}$. Or, d'après la propriété (xii) de la notation $\phi{+}\overline{\psi}$ on a donc de manière équivalent $(\neg\psi){+}\overline{\neg\phi} \succeq (\neg\phi){+}\overline{(\neg\psi)}$ ce qui, d'après les règles d'expansion, est équivalent à $\neg\psi \succeq \neg\phi$.

5.6.2 Représentation des préférences conditionnelles

Une propriété γ est dite être *commune* aux propriétés α et β si et seulement si γ est subsumée à la fois par α et par γ (i.e. $\vDash \alpha \Rightarrow \gamma$ et $\vDash \beta \Rightarrow \gamma$). Dit autrement[3], cela signifie que la formule γ est subsumée par la formule $\alpha \vee \beta$. Remarquons que :

- La propriété $\top$ est toujours une propriété commune à deux propriétés.
- Si $\top$ est la seule propriété commune aux propriétés α et β alors α est la négation de β et vice et versa ($\alpha \equiv \neg\beta$).
- Si α implique α' et α est une propriété commune, alors α' est une propriété commune.

Une *préférence conditionnelle* est une préférence entre deux propriétés ayant une propriété commune. On appelle une telle propriété commune la *condition* ou le *contexte* de la préférence conditionnelle. On dira que ϕ est préférée à ψ sous la condition C, si et seulement

[3] $\left[\vDash \alpha \Rightarrow \gamma \text{ et } \vDash \beta \Rightarrow \gamma\right]$ si et seulement si $\left[\vDash \neg\alpha \vee \gamma \text{ et } \vDash \neg\beta \vee \gamma\right]$ si et seulement si $\left[\vDash (\neg\alpha \vee \gamma) \wedge (\neg\beta \vee \gamma)\right]$ si et seulement si $\left[\vDash (\neg\alpha \wedge \neg\beta) \vee \gamma\right]$ si et seulement si $\left[\vDash \neg(\alpha \vee \beta) \vee \gamma\right]$ si et seulement si $\left[\vDash (\alpha \vee \beta) \Rightarrow \gamma\right]$.

si la propriété $C\wedge\phi$ est préférée à $C\wedge\psi$. Pour plus de clarté et de concision nous reprendrons la notation usuelle des préférences conditionnelles. Plus précisément, cette dernière est définie dans notre formalisme par l'équivalence suivante :

$$C : \phi \succeq_i \psi \quad \text{si et seulement si} \quad C\wedge\phi \succeq_i C\wedge\psi \tag{5.27}$$

À la lueur de cette définition il apparaît que l'application de la règle de spécialisation (voir 5.5.6) permet de « remplir » la condition d'une préférence conditionnelle. Dit autrement, cela signifie qu'une préférence peut être contextualisée par n'importe quel critère non encore spécifié.

5.6.3 Addition des préférences

Si la formule ϕ est préférée à la formule ϕ via la préférence partielle primitive i, et si, de la même façon la formule γ est préférée à la formule ω, alors, de la même façon, la formule $\phi\wedge\gamma$ est préférée à la formule $\phi\wedge\omega$:

$$\text{si} \begin{bmatrix} \vdash_{\mathscr{L}^1} \phi \succeq_i \psi & \text{et} & \nvdash \neg((\phi+\overline{\psi})\wedge\gamma) & \text{et} & \nvdash \neg((\gamma+\overline{\varpi})\wedge\psi) \\ \vdash_{\mathscr{L}^1} \gamma \succeq_i \omega & & \nvdash \neg((\psi+\overline{\phi})\wedge\gamma) & & \nvdash \neg((\omega+\overline{\gamma})\wedge\psi) \end{bmatrix} \quad \text{alors} \quad \vdash_{\mathscr{L}^1} (\phi\wedge\gamma) \succeq_i (\psi\wedge\omega) \tag{5.28}$$

Preuve en utilisant les règles d'inférence de spécialisation et de transitivité (voir l'annexe A.2.5).

5.6.4 Classes d'équivalences

Si ϕ et ψ sont préférées chacune l'une à l'autre alors toute occurrence de l'une dans une préférence peut être remplacée par une une occurrence de l'autre. Formellement :

$$\text{si} \begin{bmatrix} \vdash_{\mathscr{L}^1} \phi \succeq_i \psi \\ \vdash_{\mathscr{L}^1} \psi \succeq_i \phi \end{bmatrix} \quad \text{alors} \quad \begin{bmatrix} \vdash_{\mathscr{L}^1} \phi \succeq_i \alpha & \text{implique} & \vdash_{\mathscr{L}^1} \psi \succeq_i \alpha \\ \vdash_{\mathscr{L}^1} \alpha \succeq_i \phi & \text{implique} & \vdash_{\mathscr{L}^1} \alpha \succeq_i \psi \end{bmatrix} \tag{5.29}$$

Cela permet de considérer des *classes d'équivalence* sur l'ensemble des propriétés. En effet, dans une conjonction intervenant dans une préférence, une propriété peut substituer une propriété tout aussi préférée lorsque cette substitution ne provoque pas une contradiction :

$$\text{si} \begin{bmatrix} \vdash_{\mathscr{L}^1} \phi \succeq_i \psi & \text{et} & \nvdash \neg(\phi+\overline{\psi}\wedge\alpha) \\ \vdash_{\mathscr{L}^1} \psi \succeq_i \phi & & \nvdash \neg(\psi+\overline{\phi}\wedge\alpha) \end{bmatrix} \quad \text{alors} \quad \begin{bmatrix} \vdash_{\mathscr{L}^1} \phi\wedge\gamma \succeq_i \alpha & \text{implique} & \vdash_{\mathscr{L}^1} \psi\wedge\gamma \succeq_i \alpha \\ \vdash_{\mathscr{L}^1} \alpha \succeq_i \phi\wedge\gamma & \text{implique} & \vdash_{\mathscr{L}^1} \alpha \succeq_i \psi\wedge\gamma \end{bmatrix} \tag{5.30}$$

Preuve de la propriété (5.29) en utilisant la règle d'inférence de transitivité et de la propriété (5.30) en utilisant les règles d'inférence de spécialisation et de transitivité (voir l'annexe A.2.6).

5.7 Exemple (suite 1)

Afin de rendre un peu plus concret le formalisme que nous venons de présenter nous montrons dans cette section ce qu'il peut apporter à l'exemple introduit dans la section 4.4.

Pour faire un choix, les informations que nous avons présentées précédemment ne sont pas suffisantes. En effet, en plus de connaître les « conséquences » des réaction, il faut connaître ce qui est « préférable ». À cet effet, nous avons introduit dans ce chapitre le langage $\mathscr{L}^1$. Celui-ci se base sur le langage $\mathscr{L}$ et permet de construire que nous appelons une théorie des préférences sur les restaurants que nous notons $\mathscr{P}$. Plus précisément, $\mathscr{P}$ est un ensemble d'hypothèses applicatives qui décrivent les préférences de l'agent. Celui-ci est, par construction, contraint par les théories du langage $\mathscr{L}$ et ici en particulier par la théorie $\mathscr{R}$. Cet ensemble d'hypothèses peut se diviser en deux parties : d'un côté **les préférences initiales** et de l'autre **les préférences déduites**.

Préférences initiales : Supposons ici qu'il ait été spécifié, au sujet du point de vue « culinaire », uniquement les trois informations conditionnelles suivantes :

- Lorsque Pierre mange un plat, il est préférable que celui-ci contienne du poisson plutôt qu'il ne contienne de la viande,
- Lorsque Pierre mange un plat, il est préférable que celui-ci soit un plat « classique » contenant de la viande plutôt que ce soit une pizza (préférence conditionnelle),
- Lorsque Pierre mange un plat contenant du poisson, il est préférable que ce soit dans un restaurant qui ne fait que des plats contenant du poisson plutôt que soit dans un restaurant faisant aussi des plats contenant de la viande.

Dans notre formalisme celles-ci correspondent respectivement aux schémas d'hypothèses suivants :

$$\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \land \text{Plat}(Y, R) \ : \ \text{Poisson}(Y) \ \succeq \ \text{Viande}(Y) \tag{ex.13}$$

$$\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \land \text{Plat}(Y, R) \ : \ \text{Classique}(Y) \land \text{Viande}(Y) \ \succeq \ \text{Pizza}(Y) \tag{ex.14}$$

$$\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \land \text{Plat}(Y, R) \land \text{Poisson}(Y) \ : \ \forall x,\ (\text{Plat}(x, R) \Rightarrow \text{Poisson}(x)) \ \succeq \ \exists x,\ \text{Plat}(x, R) \land \text{Viande}(x) \tag{ex.15}$$

Préférences déduites par PPP : Nous avons postulé dans la section 3.2 que les préférences fournies initialement sont des données qui « cachent » implicitement de nombreuses autres préférences. Grâce au formalisme des Préférences Partielles Primitives d'une part et d'autre part à la théorie des restaurants $\mathscr{R}$, ces préférences implicites peuvent être explicitées. En particulier,[4] les informations suivantes sont déduites :

[4] Pour des raisons de lisibilité, seules les préférences déduites pertinentes pour la décision sont indiquées.

De la préférence (ex.13), on dérive en utilisant la règle de spécialisation avec la formule Classique(Y) que, lorsque Pierre mange un plat classique, il est préférable que celui-ci contienne du poisson plutôt qu'il ne contienne de la viande (préférence conditionnelle) :

$$\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \wedge \text{Plat}(Y, R) \wedge \text{Classique}(Y) \ : \ \text{Poisson}(Y) \succeq \text{Viande}(Y) \tag{ex.16}$$

Des préférences (ex.16) et (ex.14), on dérive en utilisant la règle de transitivité que, lorsque Pierre mange un plat, il est préférable que celui-ci soit un plat « classique » contenant du poisson plutôt que ce soit une pizza :

$$\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \wedge \text{Plat}(Y, R) \ : \ \text{Classique}(Y) \wedge \text{Poisson}(Y) \succeq \text{Pizza}(Y) \tag{ex.17}$$

De la préférence (ex.15), on dérive en utilisant la règle d'expansion que, lorsque Pierre mange un plat contenant du poisson, il est préférable que ce soit dans un restaurant qui ne fait que des plats contenant du poisson mais pas de la viande plutôt que soit dans un restaurant faisant des plats contenant de la viande et des plats contenant du poisson (préférence conditionnelle) :

$$\begin{aligned} &\mathscr{P} \vdash_{\mathscr{L}^1} \text{Manger}(\text{“Pierre”}, Y) \wedge \text{Plat}(Y, R) \wedge \text{Poisson}(Y) \ : \\ &\left[\forall x,\ (\text{Plat}(x, R) \Rightarrow \left| \begin{array}{c} \text{Poisson}(x) \\ \wedge \\ \neg\text{Viande}(x) \end{array} \right. \right] \succeq \left[\begin{array}{c} \exists x,\ \text{Plat}(x, R) \wedge \text{Viande}(x) \\ \wedge \\ \exists y,\ \text{Plat}(y, R) \wedge \neg\text{Poisson}(y) \end{array} \right] \end{aligned} \tag{ex.18}$$

Par défaut, Pierre n'a pas de préférence. En particulier, lorsque Pierre mange un plat contenant du poisson, il ne préfère pas que ce soit dans un restaurant faisant des plats contenant de la viande et des plats contenant du poisson plutôt que ce ne soit dans un restaurant faisant uniquement des plats contenant du poisson et sans viande (inverse de la préférence précédente). C'est ce qu'on dérive grâce au défaut (5.17) :

$$\mathscr{P} \vdash_{\mathscr{L}^1} \neg \left(\begin{array}{l} \text{Manger}(\text{“Pierre”}, Y) \wedge \text{Plat}(Y, R) \wedge \text{Poisson}(Y) \ : \\ \quad \left[\begin{array}{c} \exists x,\ \text{Plat}(x, R) \wedge \text{Viande}(x) \\ \wedge \\ \exists y,\ \text{Plat}(y, R) \wedge \neg\text{Poisson}(y) \end{array} \right] \succeq \left[\forall x,\ (\text{Plat}(x, R) \Rightarrow \left| \begin{array}{c} \text{Poisson}(x) \\ \wedge \\ \neg\text{Viande}(x) \end{array} \right. \right] \end{array} \right) \tag{ex.19}$$

De même, par défaut, lorsque Pierre mange un plat, il ne préfère pas que celui-ci soit une pizza plutôt que ce ne soit un plat « classique » contenant du poisson. C'est ce qu'on dérive grâce au défaut (5.17) :

$$\mathscr{P} \vdash_{\mathscr{L}^1} \neg\big(\text{Manger}(\text{“Pierre”}, Y) \wedge \text{Plat}(Y, R) \ : \ \text{Pizza}(Y) \succeq \text{Classique}(Y) \wedge \text{Poisson}(Y)\big) \tag{ex.20}$$

Chapitre 6

Préférences partielles étendues

Le formalisme que nous avons présenté dans la section précédente permet de construire pratiquement des opérateurs de préférence partielle primitive (sur des couples de formules), donc de définir commodément une relation binaire de préférence partielle primitive (entre les différentes alternatives possibles) et donc de décrire de façon précise et fidèle (si nos hypothèses sont correctes) les préférences d'un décideur à partir des informations données.

Néanmoins, dans la pratique, il est fréquent d'avoir à choisir parmi des alternatives qui ne sont pas explicitement ordonnées par une préférence partielle primitive. En effet, les propriétés sont généralement interdépendantes entre elles si bien qu'une différence entre alternatives causée par des propriétés explicitement ordonnées par une préférence partielle primitive implique souvent des différences sur des propriétés qui ne sont pas ordonnées par cette préférence. Par exemple, le point exact de départ d'un voyage peut dépendre du moyen de transport utilisé : un voyage depuis Paris peut avoir comme point de départ l'aéroport de Roissy ou la gare Montparnasse. D'un autre côté, les préférences sur les moyens de transport ont peu de chance d'exprimer explicitement le fait qu'un départ de l'aéroport de Roissy est tout aussi préférable que de la gare Montparnasse. Par conséquent, l'ordonnancement que nous avons proposé dans les sections précédentes est souvent insuffisant tout seul pour choisir une réaction. Pourtant, même dans de telles situations, il semble naturel de choisir une réaction en se basant sur l'ordonnancement défini précédemment. Plus précisément, il semble naturel d'ordonner des couples d'alternatives tels que la première alternative vérifie la formule $\phi \wedge \gamma$ et la seconde la formule $\psi \wedge \omega$ et tels que seule la formule $(\phi \succeq_i \psi)$ de $\mathscr{L}^1$ a été spécifiée d'une manière ou d'une autre lors de la définition des préférences.

Aussi, afin d'ordonner le plus d'alternatives possibles sur la base des données initiales, nous « étendons » chaque *préférence partielle primitive*. Pour se faire, nous définissons dans ce chapitre la notion de *préférence partielle étendue*. À cet effet nous introduisons pour chaque point de vue i, une relation de préférence G'_i ainsi qu'un opérateur logique associé. Ce dernier se base, d'une part, sur un opérateur logique de *préférence partielle primitive* et, d'autre part, sur l'hypothèse *Ceteris Imparibus* (les préférences sont des arguments pour départager les alternatives). Notons dès à présent qu'une relation de *préférence partielle étendue* G'_i est une relation binaire (définie sur l'ensemble des alternatives ALT) qui contient la relation G_i et qui par conséquent est, a priori, seulement réflexive.

Dans un premier temps nous formalisons ce que nous appelons l'hypothèse *Ceteris Imparibus*. Par la suite, nous présentons la syntaxe, la sémantique et la façon d'obtenir les opérateurs logiques modaux relatifs aux préférences partielles primitives. Enfin nous terminons ce chapitre en pointant sur quelques propriétés intéressantes de ce formalisme puis en l'instanciant sur l'exemple introduit à la section 4.4.

6.1 L'hypothèse *Ceteris Imparibus*

6.1.1 Introduction

Afin d'ordonner le plus d'alternatives possibles sur la base des données initiales, nous émettons l'hypothèse que chaque *préférence partielle primitive* doit être appréhendée comme un « argument » pour choisir une alternative plutôt qu'une autre. C'est ce que nous appelons l'hypothèse *Ceteris Imparibus*. Intuitivement, l'hypothèse *Ceteris Imparibus* indique qu'une condition suffisante pour qu'une alternative soit préférée à une autre est qu'il existe un argument qui l'indique, c'est-à-dire qu'il existe une *préférence partielle primitive* dans ce sens. Cette hypothèse a pour fondement le fait que, comme nous l'avons remarqué en section (5.5), par défaut deux propriétés sont tout aussi préférées. Par conséquent, une différence entre deux alternatives au sujet d'une propriété ne doit pas, jusqu'à preuve du contraire, interdire la comparaison.

6.1.2 La relation de *préférence partielle étendue* G'_i

Afin de formaliser cette vision, nous proposons l'introduction d'une relation binaire G'_i pour chaque point de vue i : c'est précisément ce que nous appelons une *préférence partielle étendue*. Chaque relation de *préférence partielle étendue* G'_i (circonscrite au point de vue i) est une relation binaire définie sur l'ensemble des alternatives ALT qui n'est a priori que réflexive (contrairement aux relations de *préférence partielle primitive*). Elle est néanmoins construite à partir de la relation de *préférence partielle primitive* circonscrite au même point de vue i (G_i).

Plus précisément, pour nous, une alternative est préférée à une autre alternative via la préférence partielle étendue G'_i si et seulement si (1) c'est la cas via la préférence partielle primitive associé au même point de vue (i.e. G_i) **ou** si (2) il existe une régularité dans le réseau de de la préférence partielle primitive G_i indiquant que les alternatives vérifiant une propriété ϕ sont préférées strictement à celles vérifiant une propriété ψ si elles sont toutes choses égales par ailleurs et que de plus la première alternative vérifie la propriété ϕ et la seconde alternative vérifie la propriété ψ :

$$aG'_ib \quad \text{ssi} \quad \left[\; aG_ib \quad \text{ou} \quad \begin{pmatrix} \exists(\phi,\psi) \in \mathscr{L}\times\mathscr{L} \text{ tel que} \\ a \vDash \phi{+}\overline{\psi} \quad \text{et} \quad b \vDash \psi{+}\overline{\phi} \quad \text{et} \quad Arg_i(\phi,\psi) \end{pmatrix} \;\right] \tag{6.1}$$

où $Arg_i(\phi,\psi)$ est l'abréviation définie de la façon suivante :

$$Arg_i(\phi,\psi) \quad \text{ssi} \quad \left[\begin{array}{llllll} \forall(a,b)\in ALT^2; & & & & & \\ \text{si} & a\vDash \phi{+}\overline{\psi} & \text{et} & b\vDash \psi{+}\overline{\phi} & \text{et} & CP(a,b,\phi,\psi) \\ \text{alors} & aG_ib & \text{et} & \neg(bG_ia) & & \end{array}\right]$$

L'abréviation $Arg_i(\phi,\psi)$ est très proche de la définition sémantique (5.11) de l'opérateur $\succeq_i$. Elle en diffère néanmoins puisqu'elle et fait appelle au fait que le couple (b,a) ne fait pas partie de la relation G_i. Il en résulte que seuls les couples appartenant à la relation de préférence (partielle primitive) stricte sont des arguments. Ainsi, dans la situation où il existe seulement trois alternatives a, b, c telles que $a\Rightarrow\psi$, $b\Rightarrow\psi$, $c\Rightarrow\phi$, $\neg CP(c,a,\phi,\psi)$, $CP(c,b,\phi,\psi)$, si aG_ib, bG_ia, cG_ib, alors, du fait de la première partie de l'équation (6.1), on a aG'_ib, bG'_ia, cG'_ib, et comme on a $Arg_i(\phi,\psi)$, du fait de la seconde partie de l'équation (6.1), on a aussi cG'_ia.

Il résulte de cette définition que pour tout point de vue i, la relation de préférence partielle primitive G_i est un sous-ensemble de la relation de préférence partielle étendue G'_i.

6.1.3 Ceteris Imparibus (Définition) :

On dira qu'il existe un argument pour préférer la propriété ϕ à la propriété ψ selon le point de vue i (noté par commodité $CI_i(\phi,\psi)$) si et seulement si il existe un couple de formules (ϕ',ψ') tel qu'il est prouvé que ϕ' est préférée à ψ' selon la préférence partielle primitive numéro i mais pas inversement[1], et que les formules $\phi'{+}\overline{\psi'}$ et $\psi'{+}\overline{\phi'}$ sont subsumées respectivement par les formules ϕ et ψ. Formellement :

$$\mathrm{CI}_i(\phi,\psi) \quad \text{ssi} \quad \left[\begin{array}{ccccc} \nvdash \neg\phi & \text{et} & \vdash_{\mathscr{L}^1} \phi'\succeq_i\psi' & \text{et} & \vdash \phi\Rightarrow\phi'{+}\overline{\psi'} \\ \nvdash \neg\psi & & \vdash_{\mathscr{L}^1} \neg(\psi'\succeq_i\phi') & & \vdash \psi\Rightarrow\psi'{+}\overline{\phi'} \end{array}\right] \tag{6.2}$$

6.1.4 Propriétés remarquables :

- La notation $\mathrm{CI}_i(\phi,\psi)$ est stable par équivalence logique dans $\mathscr{L}$:

$$\text{Si} \quad \vdash \phi_1\Leftrightarrow\phi_2 \text{ et } \vdash \psi_1\Leftrightarrow\psi_2 \qquad \text{alors} \qquad \mathrm{CI}_i(\phi_1,\psi_1) \quad \text{ssi} \quad \mathrm{CI}_i(\phi_2,\psi_2) \tag{6.3}$$

  Preuve avec l'axiome imposant la stabilité de l'opérateur $\succeq_i$ par rapport à l'équivalence logique (5.18) et en remarquant que si la formule ϕ_1 est équivalente à ϕ_2 alors on a, pour tout γ, ϕ_1 implique γ si et seulement si on a ϕ_2 implique γ.
- Quel que soit le point de vue i et les formules ϕ et ψ, s'il existe un « argument » pour préférer ϕ à ψ suivant le point de vue i, alors il existe aussi un « argument » pour préférer $\phi{+}\overline{\psi}$ à $\psi{+}\overline{\phi}$ suivant le point de vue i :

$$\text{Si} \quad \mathrm{CI}_i(\phi,\psi) \qquad \text{alors} \qquad \mathrm{CI}_i(\phi{+}\overline{\psi},\psi{+}\overline{\phi}) \tag{6.4}$$

  Preuve avec les propriétés d'expansion (5.21) et (5.22) de l'opérateur de préférence partielle primitive et avec la propriété (iv) de la notation $\phi{+}\overline{\psi}$.

[1] ϕ' est donc préférée *strictement* à ψ' selon la préférence partielle primitive numéro i.

– Quel que soit le point de vue i et les formules ϕ et ψ, s'il existe un « argument » pour préférer ϕ à ψ suivant le point de vue i alors pour toutes formules ϕ' et ψ' dont la négation n'est pas un thèorème et telles que ϕ' subsume ϕ et ψ' subsume ψ, il existe aussi un « argument » pour préférer ϕ' à ψ' suivant le point de vue i :

$$\text{Si} \quad \begin{bmatrix} \mathrm{CI}_i(\phi,\psi) \\ \vdash \phi' \Rightarrow \phi \\ \vdash \psi' \Rightarrow \psi \\ \nvdash \neg\phi' \\ \nvdash \neg\psi' \end{bmatrix} \qquad \text{alors} \qquad \mathrm{CI}_i(\phi',\psi') \tag{6.5}$$

En particulier, pour toutes formules ϕ et ψ, et pour toute formule α qui n'est pas inconsistante avec ϕ ou ψ, s'il existe un « argument » pour préférer ϕ à ψ suivant le point de vue i, alors il existe un « argument » pour préférer $\phi\wedge\alpha$ à $\psi\wedge\alpha$ suivant le point de vue i,

$$\text{Si} \quad \begin{bmatrix} \mathrm{CI}_i(\phi,\psi) \\ \nvdash \neg(\phi\wedge\alpha) \\ \nvdash \neg(\psi\wedge\alpha) \end{bmatrix} \qquad \text{alors} \qquad \mathrm{CI}_i(\phi\wedge\alpha,\psi\wedge\alpha) \tag{6.6}$$

Remarquons que cet « argument » est le même. Preuve immédiate en considérant la définition (6.2).

6.2 Syntaxe de l'opérateur $\geq_i$

À chaque relation binaire de *préférence partielle étendue* G_i' est associé un opérateur logique modal défini sur l'ensemble des formules de $\mathscr{L}$. Nous noterons un tel opérateur avec le symbole $\geq_i$ lorsque la relation correspondante est associée au point de vue i. L'ensemble des opérateurs $\geq_i$ définissent le langage $\mathscr{L}^2$: si ϕ et ψ sont des formules bien formées de $\mathscr{L}$ (et non de $\mathscr{L}^2$) et si $\geq_i$ dénote un opérateur logique associé à une relation binaire de *préférence partielle étendue*, alors les formules $\phi \geq_i \psi$ et $\neg(\phi \geq_i \psi)$ sont des formules bien formées de $\mathscr{L}^2$. De plus, si F_1 et F_2 sont deux formules bien formées de $\mathscr{L}^2$ alors $F_1 \wedge F_2$ est une formule bien formée de $\mathscr{L}^2$.

Afin d'alléger les notations, nous introduisons l'abréviation syntaxique $\phi >_i \psi$. Elle indique que la propriété ϕ est préférée à la propriété ψ (par rapport à la préférence partielle étendue associée au point de vue i) mais que l'inverse n'est pas vrai (i.e. $\phi >_i \psi$ si et seulement si $\phi \geq_i \psi \wedge \neg(\psi \geq_i \phi)$). On dira alors que la propriété ϕ sera *préférée strictement*, par rapport à la préférence partielle étendue associée au point de vue i, à la propriété ψ.

6.3 Sémantique de l'opérateur $\geq_i$

6.3.1 Les modèles du langage $\mathscr{L}^2$

Afin d'interpréter ces nouvelles formules, nous définissons les modèles du langage $\mathscr{L}^2$ comme des triplets $(ALT, \vDash, \mathscr{G})$ où

- ALT est l'ensemble des alternatives : l'ensemble des modèles (S, V) de $\mathscr{L}$.
- $\vDash$ est la validité sémantique du langage servant à décrire les alternatives ($\mathscr{L}$).
- $\mathscr{G}$ est un ensemble des relations binaires sur ALT. Il contient toutes les relations de préférence nécessaires à notre formalisme et en particulier toutes les relations de préférence partielle étendue G'_i.

Dans l'optique de clarifier les notations, nous introduisons les symboles $\vDash_{\mathscr{L}^2}$, $\vdash_{\mathscr{L}^2}$, $\nvDash_{\mathscr{L}^2}$ et $\nvdash_{\mathscr{L}^2}$. Ils représentent respectivement la validité sémantique, la conséquence logique, la non validité sémantique et la non conséquence logique pour le langage $\mathscr{L}^2$.

6.3.2 Définition

Dire que la formule ϕ est préférée à la formule ϕ via la préférence partielle étendue numéro i (noté avec la formule $\phi \geq_i \phi$) signifie que les alternatives qui vérifient les propriétés décrites par la formule ϕ et, si cela est cohérent, le contraire de celles décrites par la formule ψ, sont préférées (sans la restriction Ceteris Paribus), suivant le point de vue numéro i, aux alternatives qui vérifient les propriétés décrites par la formule ψ et, si cela est cohérent, le contraire de celles décrites par la formule ϕ. La sémantique de cet opérateur est formalisée comme suit :

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} \phi \geq_i \psi \quad \text{ssi} \quad \begin{array}{l} \bullet \quad \nvDash \neg\phi \quad \text{et} \quad \nvDash \neg\psi \\ \bullet \quad \left[\begin{array}{l} \forall (a,b) \in \mathrm{ALT}^2, \\ \text{si} \left| \begin{array}{l} a \vDash \phi + \overline{\psi} \\ b \vDash \psi + \overline{\phi} \end{array} \right. \quad \text{alors} \quad aG'_i b \end{array} \right] \end{array} \tag{6.7}$$

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} \neg(\phi \geq_i \psi) \quad \text{ssi} \quad (\mathrm{ALT}, \vDash, \mathscr{G}) \nvDash_{\mathscr{L}^2} \phi \geq_i \psi \tag{6.8}$$

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} F_1 \wedge F_2 \quad \text{ssi} \quad \Big[(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} F_1 \;\; \text{et} \;\; (\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} F_2 \Big] \tag{6.9}$$

6.3.3 Corollaire 1

D'après la définition (6.7), si ϕ est préférée à la formule ψ via la préférence partielle étendue i (i.e. $\vDash \phi \geq_i \psi$) alors $\neg\phi$ et $\neg\psi$ ne sont pas des tautologies (i.e. $\nvDash \neg\phi$ et $\nvDash \neg\psi$). De plus, pour tout couple d'alternatives a et b, s'il existe un couple de formules (ϕ, ψ) de $\mathscr{L}$

tel que ϕ est préférée à ψ et tel que a est un modèle de $\phi{+}\overline{\psi}$ et b est un modèle de $\psi{+}\overline{\phi}$, alors a est préférée à b. Formellement pour tout couple d'alternative a et b,

$$\text{Si} \quad \left[\begin{array}{ll} \exists(\phi,\psi) \in \mathscr{L}^2 & \text{tels que :} \\ \bullet & (\text{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} \phi \geq_i \psi \\ \bullet & a \vDash \phi{+}\overline{\psi} \\ \bullet & b \vDash \psi{+}\overline{\phi} \end{array} \right] \quad \text{alors} \quad aG'_i b \tag{6.10}$$

6.3.4 Corollaire 2

D'après la définition (6.7), les formules $\phi \geq_i \psi$ et $\phi{+}\overline{\psi} \geq_i \psi{+}\overline{\phi}$ ont le même sens :

$$(\text{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} \phi \geq_i \psi \quad \text{si et seulement} \quad (\text{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} \phi{+}\overline{\psi} \geq_i \psi{+}\overline{\phi} \tag{6.11}$$

On prouve que les formules $\phi \geq_i \psi$ et $\phi{+}\overline{\psi} \geq_i \psi{+}\overline{\phi}$ ont le même sens en remplaçant dans la définition sémantique (6.7) de l'opérateur de préférence partielle étendue la formule ϕ par $\phi{+}\overline{\psi}$ et la formule ψ par $\psi{+}\overline{\phi}$, et en remarquant que $[\phi{+}\overline{\psi}]{+}\overline{[\psi{+}\overline{\phi}]}$ est équivalent à $\phi{+}\overline{\psi}$, et que $[\psi{+}\overline{\phi}]{+}\overline{[\phi{+}\overline{\psi}]}$ est équivalent à $\psi{+}\overline{\phi}$ et que de surcroît $\vDash \psi{+}\overline{\phi} \Rightarrow \psi$ et $\vDash \phi{+}\overline{\psi} \Rightarrow \phi$.

6.3.5 Lien entre l'opérateur $\geq_i$ et la relation G'_i

Comme énoncé dans la section (4.3.1), à chaque alternative a de ALT est associée une formule f_a de $\mathscr{L}$ telle que pour chaque formule close α de $\mathscr{L}$, de façon exclusive, soit f_a subsume α, soit f_a subsume $\neg\alpha$ et telle que seule l'alternative a est un modèle de cette formule (i.e. $\forall b \in ALT,\ b \vDash f_a$ ssi $a = b$). Par conséquent, si un couple d'alternatives appartient à une relation de préférence partielle étendue alors les formules associées sont reliées par l'opérateur logique correspondant. De plus si deux formules représentant chacune une unique alternative sont mises en relation par l'opérateur logique $\geq_i$ alors les deux alternatives correspondantes appartiennent à la relation de préférence partielle étendue associée. Formellement,

$$\forall (a,b) \in \text{ALT}^2_, \qquad aG'_i b \quad \text{ssi} \quad (\text{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^2} f_a \geq_i f_b \tag{6.12}$$

Preuve immédiate en identifiant dans la définition (6.7) les formules ϕ et ψ avec les formules f_a et f_b, et en remarquant de surcroît que f_a et f_b ne sont pas équivalents à $\perp$, que $f_a{+}\overline{f_b}$ équivaut à f_a et $f_b{+}\overline{f_a}$ équivaut à f_b (voir la propriété (v) de la section 4.2.2).

6.4 Axiomatisation de l'opérateur $\geq_i$

Afin de « coller » au plus prés de la sémantique de cet opérateur, nous proposons l'axiomatisation suivante. Comme pour l'axiomatisation de l'opérateur $\succeq_i$, celle-ci n'est sûrement pas complète.

6.4.1 Défaut (D)

Afin d'être en accord avec l'intuition, pour tout couple de formules $(\phi, \psi$ de $\mathscr{L} \times \mathscr{L}$, par défaut, s'il n'est pas déjà exprimée la formule ϕ est préférée à ψ alors la formule ϕ n'est pas préférée à une formule ψ. Formellement le défaut normal sans prérequis (à la Reiter) suivant est valide dans le langage $\mathscr{L}^2$:

$$\top : \neg(\phi \geq_i \psi) \to \neg(\phi \geq_i \psi) \tag{6.13}$$

6.4.2 Règles de construction de l'opérateur $\geq_i$

Dans la pratique, chaque préférence partielle étendue G_i' est générée à partir d'un ensemble des formules[2] de $\mathscr{L}^2$ de la forme $\phi \geq_i \psi$, où ϕ et ψ sont des formules de $\mathscr{L}$ et i est le point de vue considéré. Ces formules sont déduites des formules de $\mathscr{L}'$ relatives à l'opérateur de la préférence partielle primitive associée de telle sorte que si on a $\psi \succeq_i \phi$ alors on a $\psi \geq_i \phi$ et que si on a $\psi \succeq_i \phi$ mais pas $\phi \succeq_i \psi$ et si $\phi \Rightarrow \phi'$ et $\psi \Rightarrow \psi'$ alors $\psi' \succeq_i \phi'$. Formellement, les deux règles suivantes sont valides :

$$\text{si } \vdash_{\mathscr{L}^1} \phi \succeq_i \psi \quad \text{alors } \vdash_{\mathscr{L}^2} \phi \geq_i \psi \tag{6.14}$$

$$\text{si } \mathrm{CI}_i(\phi, \psi) \quad \text{alors } \vdash_{\mathscr{L}^2} \phi \geq_i \psi \tag{6.15}$$

Chaque opérateur de *préférence partielle étendue* est donc construit comme l'extension *Ceteris Imparibus* de l'opérateur de *préférence partielle primitive* associé au même point de vue.

6.4.3 Remarques

La condition $\neg(\psi \succeq_i \phi)$, spécifiée dans la définition (6.2) de la notation CI_i, permet de « bloquer » la règle de construction lorsque les formules ϕ et ψ sont tout aussi préférables par rapport à la préférence partielle primitive associée. Sans cette condition, l'existence d'une telle règle entraînerait l'indifférence entre toutes les formules et donc entre toutes les alternatives. En effet, dans cette configuration il existe un « contre-argument » trivial : $\top \succeq_i \top$. Preuve en considérant la version alors « simplifiée » de l'hypothèse *Ceteris Imparibus* et en remarquant que la formule $\top$ est toujours préférée à elle-même du fait de la réflexivité de l'opérateur $\succeq_i$ (voir l'annexe A.3.1).

L'opérateur de préférence partielle étendue ne doit pas être transitif. En effet, si tel était le cas, n'importe quelle alternative vérifiant une formule ϕ serait préférée (pour la relation de préférence associée à cet opérateur) à une autre alternative vérifiant ψ dès lors qu'il existerait deux formules ϕ_1 et ψ_1 telles que $\vDash \phi \geq_i \phi_1$ et $\vDash \psi_1 \geq_i \psi$. (voir l'annexe A.3.2 pour une preuve).

Les relations binaires induites par les opérateurs $\succeq_i$ et $\geq_i$ sont différentes. Alors que la première (G_i) est transitive et réflexive, la seconde (G_i') n'est que réflexive. Cela peut

[2] Ces formules n'induisent pas forcément des ensembles de préférences cohérents : elles peuvent spécifier à la fois que $\phi \geq_i \psi$ et que $\psi \geq_i \phi$.

se justifier par le fait que G_i permet d'interpréter les données d'entrée et que G'_i permet d'exploiter ces informations pour la décision.

6.4.4 Expansion (E)

L'axiomatique spécifiée jusqu'ici implique que l'opérateur de préférence partielle étendue vérifie partiellement la propriété d'expansion (E). Plus précisément, avec une telle axiomatique, les formules du type $\phi \geq_i \psi$ impliquent les formules du type $\phi + \overline{\psi} \geq_i \psi + \overline{\phi}$:

$$\text{Si} \quad \vdash_{\mathscr{L}^2} \phi \geq_i \psi \quad \text{alors} \quad \vdash_{\mathscr{L}^2} \phi + \overline{\psi} \geq_i \psi + \overline{\phi} \tag{6.16}$$

Preuve en remarquant que, comme c'est le cas pour l'opérateur de préférence partielle primitive, c'est aussi le cas pour l'opérateur de préférence partielle étendue (voir l'annexe A.3.4).

La réciproque n'est toutefois pas vérifiée. Aussi, afin de rester en accord avec le corollaire (6.11) de la section (6.3) relative à la sémantique de la préférence partielle étendue, nous imposons qu'une formule du type $\phi \geq_i \psi$ est équivalente logiquement à la formule $\phi + \overline{\psi} \geq_i \psi + \overline{\phi}$ en ajoutant à l'axiomatique le règle suivante :

$$\text{Si} \quad \vdash_{\mathscr{L}^2} \phi + \overline{\psi} \geq_i \psi + \overline{\phi} \quad \text{alors} \quad \vdash_{\mathscr{L}^2} \phi \geq_i \psi \tag{6.17}$$

6.5 Propriétés remarquables

La définition axiomatique préférence induit des propriétés intéressantes pour chaque opérateur logique de préférence partielle étendue. Nous listons ici celles qui nous semblent les plus remarquables.

6.5.1 Comportement de l'opérateur $\geq_i$ avec l'équivalence logique

Si deux formules sont logiquement équivalentes alors toute occurrence de l'une dans une préférence partielle étendue peut-être substituée par une occurrence de l'autre et inversement. Formellement,

$$\text{Si} \quad \left[\vdash \phi_1 \Leftrightarrow \phi_2 \quad \text{et} \quad \vdash \psi_1 \Leftrightarrow \psi_2\right] \quad \text{alors} \quad \left[\vdash_{\mathscr{L}^2} \phi_2 \geq_i \psi_2 \quad \text{ssi} \quad \vdash_{\mathscr{L}^2} \phi_1 \geq_i \psi_1\right] \tag{6.18}$$

Preuve en remarquant que c'est le cas pour l'opérateur de préférence partielle primitive et que si $\vdash \phi_1 \Leftrightarrow \phi_2$ et $\vdash \psi_1 \Leftrightarrow \psi_2$ alors $CI_i(\phi_1, \psi_1)$ est équivalent à $CI_i(\phi_2, \psi_2)$.

6.5.2 Réflexive (R)

Toute formule de $\mathscr{L}$ non inconsistante est préférée à elle même :

$$\text{si} \quad \nvdash \neg\phi \quad \text{alors} \quad \vdash_{\mathscr{L}^2} \phi \geq_i \phi \tag{6.19}$$

C'est pourquoi une indifférence existe toujours entre une propriété et elle-même. Preuve en remarquant que c'est le cas pour l'opérateur de préférence partielle primitive. Ceci entraîne que chaque relation binaire de *préférence partielle primitive* associée G_i est réflexive.

6.5.3 Spécialisation (S)

Si une propriété ϕ est préférée à une propriété ψ alors la spécialisation de ϕ suivant un aspect, non déjà spécifié de façon contraire, sera aussi préférée à la même spécialisation de la propriété ψ :

$$\text{si } \begin{bmatrix} \nvdash \neg(\phi + \overline{\psi} \wedge \alpha) \\ \nvdash \neg(\psi + \overline{\phi} \wedge \alpha) \\ \vdash_{\mathscr{L}^2} \phi \geq_i \psi \end{bmatrix} \quad \text{alors} \quad \vdash_{\mathscr{L}^2} (\phi \wedge \alpha) \geq_i (\psi \wedge \alpha) \tag{6.20}$$

Preuve en remarquant que, comme c'est le cas pour l'opérateur de préférence partielle primitive, c'est aussi le cas pour l'opérateur de préférence partielle étendue (voir l'annexe A.3.3).

6.6 Exemple (suite 2)

Comme nous l'avons fait à la fin du chapitre précédent, nous reprenons ici l'exemple introduit dans la section 4.4 afin de rendre notre approche plus concrète.

Dans cette situation, les formules mises en relation par les Préférences Partielles Primitives sont malheureusement encore trop différentes des formules décrivant les alternatives mises en jeu pour pouvoir dire qu'une de ces alternatives est meilleure que les autres. En effet, les alternatives sont trop différentes entre elles pour que leurs descriptions « totales » puissent se retrouver conjointement de part et d'autre d'un opérateur de Préférence Partielle Primitive. Pourtant, afin de choisir une alternative, l'intuition nous indique que les informations spécifiés précédemment sont suffisantes pour déterminer si l'une d'elles est plus « intéressante » que les autres. C'est pourquoi, le formalisme des Préférences Partielles Étendues permet d'étendre les informations disponibles afin de comparer les descriptions totales des différentes alternatives.

Par définition, les formules f_a et f_b ne sont pas inconsistantes (i.e. $\nvdash \neg f_a$ et $\nvdash \neg f_b$). De plus en considérant que

$$\phi' = \begin{bmatrix} \text{Manger(“Pierre”}, Y) \\ \wedge \\ \text{Plat}(Y, R) \wedge \text{Poisson}(Y) \end{bmatrix} \wedge \begin{bmatrix} \forall x,\ (\text{Plat}(x, R) \Rightarrow \left| \begin{matrix} \text{Poisson}(x) \\ \wedge \\ \neg\text{Viande}(x) \end{matrix} \right. \end{bmatrix}$$

$$\psi' = \begin{bmatrix} \text{Manger(“Pierre”}, Y) \\ \wedge \\ \text{Plat}(Y, R) \wedge \text{Poisson}(Y) \end{bmatrix} \wedge \begin{bmatrix} \exists x,\ \text{Plat}(x, R) \wedge \text{Viande}(x) \\ \wedge \\ \exists y,\ \text{Plat}(y, R) \wedge \neg\text{Poisson}(y) \end{bmatrix}$$

on remarque que les équations (ex.18) et (ex.19) sont respectivement de la forme, $\models_{\mathscr{L}^1} \phi' \succeq_i \psi'$ et $\models_{\mathscr{L}^1} \neg(\psi' \succeq_i \phi')$. Enfin, des équations (ex.10) et (ex.11) on obtient respectivement que $f_a \Rightarrow \phi' + \overline{\psi'}$ et $f_a \Rightarrow \psi' + \overline{\phi'}$. Tout ceci permet de déduire en en utilisant la règle de construction (6.15), qu'il est préférable pour Pierre d'aller manger du poisson au Goéland plutôt que d'aller manger du poisson à la Légende :

$$\mathscr{P} \vdash_{\mathscr{L}2} f_a \geq f_b \tag{ex.21}$$

De la même manière, des formules (ex.17), (ex.20), (ex.10) et (ex.12) on déduit, en utilisant la règle de construction (6.15), qu'il est préférable pour Pierre d'aller manger du poisson au Goéland plutôt que d'aller manger une Pizza au Capuccino :

$$\mathscr{P} \vdash_{\mathscr{L}2} f_a \geq f_c \tag{ex.22}$$

De la même manière, des formules (ex.17), (ex.20), (ex.11) et (ex.12) on déduit, en utilisant la règle de construction (6.15), qu'il est préférable pour Pierre d'aller manger du poisson à la Légende plutôt que d'aller manger une Pizza au Capuccino :

$$\mathscr{P} \vdash_{\mathscr{L}2} f_b \geq f_c \tag{ex.23}$$

Puisque que rien n'indique le contraire, grâce au défaut (6.13), les préférences spécifiées et les hypothèses énoncées sur le domaine ne portent pas à croire qu'il est préférable pour Pierre d'aller manger une Pizza au Capuccino plutôt que d'aller manger du poisson au Goéland :

$$\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq f_a) \tag{ex.24}$$

Puisque que rien n'indique le contraire, grâce au défaut (6.13), les préférences spécifiées et les hypothèses énoncées sur le domaine ne portent pas à croire qu'il est préférable pour Pierre d'aller manger une Pizza au Capuccino plutôt que d'aller manger du poisson à la Légende :

$$\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq f_b) \tag{ex.25}$$

Puisque que rien n'indique le contraire, grâce au défaut (6.13), les préférences spécifiées et les hypothèses énoncées sur le domaine ne portent pas à croire qu'il est préférable pour Pierre d'aller manger du poisson à la Légende plutôt que d'aller manger du poisson au Goéland :

$$\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_b \geq f_a) \tag{ex.26}$$

Il est à remarquer que sans la Préférence Partielle Étendue, les alternatives a, b et c sont incomparables : rien ne permet de dire que l'une est préférable à l'autre avec la préférence partielle primitive spécifiée à la section 5.7. En effet les descriptions totales des alternatives a, b et c ne peuvent former un couple de la relation de cette préférence partielle primitive étant donné que les alternatives considérées sont trop différentes les une des autres. Pour nous, elles sont trop différentes car le degré de description des alternatives est « trop précis ». En particulier, les descriptions des alternatives explicitent des caractéristiques (ici en particulier le nom du restaurant) qui ne sont pas pertinentes pour la comparaison. Le calcul des Préférences Partielles Étendues permet donc de se passer d'une définition de pertinence (comme c'est le cas dans les formalismes de Doyle et Wellman [Doyle and Wellman, 1994] et Boutilier [Boutilier, 1994]) des propriétés décrites pour la comparaison.

Chapitre 7

Préférence globale et choix de la réaction

Les deux chapitres précédents nous ont permis de présenter les *préférences partielles primitives* (PPP) et les *préférences partielles étendues* (PPE) ainsi que la manière de les obtenir. À ce stade, le formalisme que nous proposons permet d'exprimer, de spécifier et de déterminer, pour chaque point de vue et chaque couple d'alternatives (a_1, a_2), le résultat du choix de l'agent[1] dans la configuration où il a à choisir entre les alternatives a_1 et a_2 et que seul le point de vue en question importe. Plus précisément, pour chaque point de vue i et chaque couple d'alternatives (a_1, a_2), ce formalisme permet de dire de façon exclusive,

– si a_1 est strictement préférée[2] à a_2 (i.e. $f_{a_1} >_i f_{a_2}$),
– si a_2 est strictement préférée[2] à a_1 (i.e. $f_{a_2} >_i f_{a_1}$),
– si a_1 est équivalente[3] ou incomparable[4] à a_2 (i.e. $f_{a_1} \sim_i f_{a_2}$).

De plus, par construction, chaque préférence partielle étendue permet de départager fréquemment les alternatives qui sont à comparer puisqu'il est uniquement nécessaire d'avoir un « argument » non contredit pour « faire pencher la balance » et donc pour préférer strictement une des deux alternatives.

Néanmoins, comme nous l'avons indiqué dans la section 3.2.1, la décision que doit prendre l'agent a plusieurs « facettes » a priori indépendantes les unes des autres. Ainsi, dans l'exemple introduit dans la section 4.4, en plus de la « facette » relative à la composition des plats des restaurant que nous avons développée, la décision a d'autres « facettes » comme celles relatives aux prix des plats ou à la localisation des restaurants. Ces dernières sont décrites dans notre formalisme grâce aux d'autres points de vue. Aussi, si l'agent veut prendre en considération toutes ces « facettes », il est nécessaire qu'il regroupe tous ces points de vue (partiels) en un point de vue global. De plus, comme ces « facettes » sont a priori indépendantes les unes des autres, il est possible que les points de vue expriment des informations contradictoires. L'agent doit donc gérer ces contradictions lors du regroupement de ces points de vue.

[1] Ce résultat est soit a_1 soit a_2 soit l'une des deux au hasard.
[2] ϕ est strictement préférée à ψ si et seulement si ϕ est préférée à ψ et ψ n'est pas préférée à ϕ
[3] ϕ est équivalente à ψ si et seulement si ϕ est préférée à ψ et ψ est préférée à ϕ
[4] ϕ est incomparable à ψ si et seulement si ϕ n'est pas préférée à ψ et ψ n'est pas préférée à ϕ

Afin de permettre à l'agent d'agréger ses différents points de vue et ainsi de n'en retenir qu'un seul, nous définissons dans ce chapitre la notion de *Préférence Globale*. À cet effet nous introduisons *la* relation de préférence G ainsi qu'un opérateur logique modal associé. Ce dernier se base, d'une part, sur l'ensemble des opérateurs de *Préférence Partielle Étendue* et, d'autre part, sur une politique d'agrégation donnée.

À ce niveau, l'objectif est d'obtenir une unique relation de préférence : La préférence globale G. Cependant en pratique cette dernière est définie, comme pour les autres relations de préférence de notre modèle, par des comparaisons sur les propriétés que peuvent vérifier les différentes alternatives c'est-à-dire grâce à l'opérateur logique qui lui est associé. C'est donc en premier lieu ce dernier qu'il faut déterminer. Plus précisément, pour chaque couple de formules à comparer, il faut pouvoir déterminer de manière globale si les alternatives vérifiant la première formule sont préférées strictement à celles vérifiant la seconde ou si c'est l'inverse ou si elles sont indifférentes.

Pour chaque couple de propriétés à comparer, chaque opérateur relatif à une Préférence Partielle Étendue indique si elles sont indifférentes (soit qu'aucune n'est préférée à l'autre, soit que chacune des deux est préférée à l'autre) ou si l'une est préférée strictement à l'autre et si oui laquelle. À chaque fois, ceci va être considéré comme une voix lors du « vote » pour choisir l'une plutôt que l'autre dans la phase d'agrégation de notre modèle. Cette phase d'agrégation s'inspire de celle présente dans le travail de Rossi et al. [Rossi et al., 2004]. Dans ce dernier, les préférences de plusieurs agents (représentées chaque fois par des CP-nets) sont agrégées via un mécanisme d'élection. Chacune de nos préférences partielles étendues joue ainsi le rôle de la préférence d'un agent dans le modèle proposé par Rossi et al. Cette adaptation s'appuie sur l'idée selon laquelle la décision multi-critères et la décision multi-agents sont très similaires dès que les critères de la première approche sont assimilés aux agents de la seconde (voir l'introduction de [Dubois et al., 2001]).

Dans un premier temps, nous présentons la syntaxe et la sémantique de l'opérateur de préférence globale. Par la suite, nous exposons différentes politiques d'agrégation envisageables et examinons comment elles influent sur la sémantique de l'opérateur de préférence globale. En conséquence de quoi, nous donnons l'axiomatique de cet opérateur pour différentes politiques. Nous poursuivrons alors notre présentation par une petite discussion sur les différentes politiques d'agrégation envisageables ainsi que sur la façon d'utiliser cette préférence pour déterminer explicitement le résultat de la décision de l'agent. Enfin, nous terminons ce chapitre en instanciant ce formalisme sur l'exemple introduit à la section 4.4.

7.1 La relation de *préférence globale* G

Afin de formaliser notre approche, nous proposons l'introduction de la relation binaire G : c'est précisément ce que nous appelons une *préférence globale*. La relation de *préférence globale* G est une relation binaire réflexive définie sur l'ensemble des alternatives ALT qui n'est, a priori, ni transitive ni complète. De façon similaire à l'opérateur logique qui lui est associé (voir plus loin), cette relation est obtenue via l'agrégation des relations de Préférences Partielles Étendues G'_i suivant une politique donnée. Pour différentes politiques d'agrégation, la préférence globale est a priori différente.

7.2 Syntaxe de l'opérateur $\geq$

À la relation binaire de *préférence totale* G est associé un opérateur logique modal défini sur l'ensemble des formules de $\mathscr{L}$. Nous notons un tel opérateur avec le symbole $\geq$. Ce dernier définit le langage $\mathscr{L}^3$: si ϕ et ψ sont des formules bien formées de $\mathscr{L}$ (et non de $\mathscr{L}^3$), alors les formules $\phi \geq \psi$ et $\neg(\phi \geq \psi)$ sont des formules bien formées de $\mathscr{L}^3$. De plus, si F_1 et F_2 sont deux formules bien formées de $\mathscr{L}^3$ alors $F_1 \wedge F_2$ est une formule bien formée de $\mathscr{L}^3$.

Afin d'alléger les notations, nous introduisons l'abréviation syntaxique $\phi > \psi$. Elle indique que la propriété ϕ est préférée à la propriété ψ (par rapport à la préférence globale) mais que l'inverse n'est pas vrai (i.e. $\phi > \psi$ si et seulement si $\phi \geq \psi \wedge \neg(\psi \geq \phi)$). On dira alors que la propriété ϕ est *préférée strictement*, à la propriété ψ.

7.3 Sémantique de l'opérateur $\geq$

7.3.1 Les modèles du langage $\mathscr{L}^3$

Afin d'interpréter ces nouvelles formules, nous définissons les modèles du langage $\mathscr{L}^3$ comme des triplets $(ALT, \models, \mathscr{G})$ où

- ALT est l'ensemble des alternatives : l'ensemble des modèles (S, V) de $\mathscr{L}$.
- $\models$ est la validité sémantique du langage servant à décrire les alternatives ($\mathscr{L}$).
- $\mathscr{G}$ est un ensemble des relations binaires sur ALT. Il contient toutes les relations de préférence nécessaires à notre formalisme et en particulier la relation de préférence globale G.

Dans l'optique de clarifier les notations, nous introduisons les symboles $\models_{\mathscr{L}^3}$, $\vdash_{\mathscr{L}^3}$, $\not\models_{\mathscr{L}^3}$ et $\nvdash_{\mathscr{L}^3}$. Ils représentent respectivement la validité sémantique, la conséquence logique, la non validité sémantique et la non conséquence logique pour le langage $\mathscr{L}^3$.

7.3.2 Définition

Pour une politique d'agrégation donnée, dire que la formule ϕ est préférée à la formule ψ via la préférence globale (noté avec la formule $\phi \geq \psi$) signifie que les alternatives qui vérifient les propriétés décrites par la formule ϕ et, si cela est cohérent, le contraire de celles décrites par la formule ψ, sont préférées suivant la préférence globale (sans la restriction Ceteris Paribus), aux alternatives qui vérifient les propriétés décrites par la formule ψ et, si cela est cohérent, le contraire de celles décrites par la formule ϕ.

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} \phi \geq \psi \quad \text{ssi} \quad \begin{array}{l} \bullet \quad \not\models \neg\phi \quad \text{et} \quad \not\models \neg\psi \\ \bullet \quad \left[\begin{array}{l} \forall(a,b) \in \mathrm{ALT}^2, \\ \text{si} \left| \begin{array}{l} a \models \phi + \overline{\psi} \\ b \models \psi + \overline{\phi} \end{array} \right. \\ \text{alors} \quad aGb \end{array} \right] \end{array} \tag{7.1}$$

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} \neg(\phi \geq \psi) \quad \text{ssi} \quad (\mathrm{ALT}, \models, \mathscr{G}) \not\models_{\mathscr{L}3} \phi \geq \psi \tag{7.2}$$

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} F_1 \wedge F_2 \quad \text{ssi} \quad \Big[(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} F_1 \;\; \text{et} \;\; (\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} F_2 \Big] \tag{7.3}$$

7.3.3 Corollaire 1

D'après la définition, si ϕ est préférée à la formule ψ via la préférence globale (i.e. $\models_{\mathscr{L}3} \phi \geq \psi$) alors $\neg\phi$ et $\neg\psi$ ne sont pas des tautologies (i.e. $\not\models \neg\phi$ et $\not\models \neg\psi$). De plus, pour tout couple d'alternatives a et b, s'il existe un couple de formules (ϕ, ψ) de $\mathscr{L}$ telles que ϕ est préférée à ψ et telles que a est un modèle de $\phi + \overline{\psi}$ et b est un modèle de $\psi + \overline{\phi}$, alors a est préférée à b. Formellement pour tout couple d'alternative a et b,

$$\text{Si} \quad \left[\begin{array}{l} \exists(\phi,\psi) \in \mathscr{L}^2 \quad \text{tels que :} \\ \bullet \quad (\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} \phi \geq \psi \\ \bullet \quad a \models \phi + \overline{\psi} \\ \bullet \quad b \models \psi + \overline{\phi} \end{array} \right] \quad \text{alors} \quad aGb \tag{7.4}$$

7.3.4 Corollaire 2

D'après la définition (7.1), les formules $\phi \geq \psi$ et $\phi + \overline{\psi} \geq \psi + \overline{\phi}$ ont le même sens :

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} \phi \geq \psi \quad \text{si et seulement} \quad (\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}3} \phi + \overline{\psi} \geq \psi + \overline{\phi} \tag{7.5}$$

On prouve que les formules $\phi \geq \psi$ et $\phi + \overline{\psi} \geq \psi + \overline{\phi}$ ont le même sens en remplaçant dans la définition sémantique (7.1) de l'opérateur de préférence partielle étendue la formule ϕ par $\phi + \overline{\psi}$ et la formule ψ par $\psi + \overline{\phi}$, et en remarquant que $[\phi + \overline{\psi}] + \overline{[\psi + \overline{\phi}]}$ est équivalent à $\phi + \overline{\psi}$, et que $[\psi + \overline{\phi}] + \overline{[\phi + \overline{\psi}]}$ est équivalent à $\psi + \overline{\phi}$ et que de surcroît $\models \psi + \overline{\phi} \Rightarrow \psi$ et $\models \phi + \overline{\psi} \Rightarrow \phi$.

7.3.5 Lien entre l'opérateur $\geq$ et la relation G

Comme énoncé dans la section (4.3.1), à chaque alternative a de ALT est associée une formule f_a de $\mathscr{L}$ telle que pour chaque formule close α de $\mathscr{L}$, de façon exclusive, soit f_a subsume α, soit f_a subsume $\neg\alpha$ et telle que seule l'alternative a est un modèle de cette formule (i.e. $\forall b \in ALT,\ b \vDash f_a$ ssi $a = b$). Par conséquent, si un couple d'alternatives appartient à une relation de préférence globale alors les formules associées sont reliées par l'opérateur logique correspondant. De plus si deux formules représentant chacune une unique alternative sont mises en relation par l'opérateur logique $\geq$ alors les deux alternatives correspondantes appartiennent à la relation de préférence globale associée. Formellement,

$$\forall (a,b) \in \ \mathrm{ALT}^2_, \qquad aGb \quad \text{ssi} \quad (\mathrm{ALT}, \vDash, \mathscr{G}) \vDash f_a \geq f_b \tag{7.6}$$

Preuve immédiate en identifiant dans la définition (7.1) les formules ϕ et ψ avec les formules f_a et f_b, et en remarquant de surcroît que f_a et f_b ne sont pas équivalentes à $\perp$, que $f_a + \overline{f_b}$ équivaut à f_a et $f_b + \overline{f_a}$ équivaut à f_b (voir la propriété (v)de la section 4.2.2).

7.4 Politiques d'agrégation

7.4.1 Objectif

L'objectif dans ce chapitre est de proposer un moyen adapté au formalisme que nous avons présenté jusqu'ici afin de synthétiser une relation binaire permettant de comparer n'importe quel couple d'alternatives (i.e. pouvoir dire, pour tout couple d'alternative (a,b), s'il est préférable de choisir a à b, s'il est préférable de choisir b à a ou s'il est indifférent de choisir a ou b). Il n'existe pas a priori de propriétés particulière que cette relation doit vérifier. En particulier elle n'est pas forcément transitive ni complète et n'est donc pas un pré-ordre. Toutefois, il nous semble logique que cette relation soit synthétisée à partir des préférences partielles étendues car intuitivement elle correspond à l'agrégation des points de vue de l'agent dans l'optique de la décision. Plus précisément, le moyen que nous cherchons à proposer doit permettre de passer d'un ensemble de préférences (les préférences partielles étendues) à une unique préférence (la préférence globale). Informellement ceci peut se justifier car les premières permettent d'exprimer des préférences (i.e. des informations) tandis que la dernière permet d'effectuer un choix qui est par nature unique.

7.4.2 Méthode envisageable

Dans notre approche, chaque point de vue est considéré comme indépendant. Il peut donc être appréhendé, dans une première approximation, comme un critère de la théorie de l'utilité multi-critères (voir la section 2.2.2). Cependant, comme d'après nos hypothèses les préférences ne vérifient pas l'hypothèse de cardinalité ni celle de complétude il n'est pas possible d'utiliser directement ce formalisme. Rappelons ici que le cadre d'application que nous envisageons est tel que (1) la relation de préférence globale est généralement ni transitive, ni complète et que (2) les critères à agréger ne sont pas forcément des pré-ordres (voir les remarques de la section 6.4.3).

7.4.3 Méthode utilisé

La préférence globale G est obtenue par l'agrégation de l'ensemble des préférences partielles étendues. À cet effet, les *points de vue* sont ici identifiés aux agents d'un problème de décision multi-agents (Multi-Agent Decision Making en anglais). Par conséquent, la sémantique de l'opérateur logique $\geq$ est relative à une politique d'agrégation donnée ; c'est-à-dire à une façon de regrouper l'ensemble des points de vue. En un sens, il y a autant de préférences globales que de politiques d'agrégation. Par exemple des politiques d'agrégation possibles (et simples !) sont l'*unanimité*, la *dominance* et la *dictature* du point de vue j. Ces dernières peuvent être respectivement appréhendées intuitivement comme le vote à l'unanimité, le vote reflétant la présence d'un « véritable argument » en même temps que l'absence de « véritable contre-argument » ainsi que la situation où le point de vue j détermine seul le résultat de l'agrégation.

La politique *unanimité* : Une alternative a est préférée (globalement) à une alternative b (noté $aG_{All}b$) suivant la politique *unanimité* si toutes les préférences partielles étendues indiquent que l'alternative a est préférée (de leurs points de vue respectif) à b :

$$\forall a, b \in ALT^2; \quad aG_{All}b \quad \text{ssi} \quad \forall i,\ aG'_i b \tag{7.7}$$

La politique *dominance* : Une alternative a est préférée (globalement) à une alternative b (noté $aG_{Dom}b$) suivant la politique *dominance* si elle est strictement préférée pour au moins une préférence partielle primitive et qu'aucune autre ne trouve b strictement préférée à a :

$$\forall a, b \in ALT^2; \quad aG_{Dom}b \quad \text{ssi} \quad \left[\begin{array}{lll} \exists j, & \text{tel que} & \left[aG'_j b \wedge \neg(bG'_j a)\right] \\ \forall i, & \text{si } i \neq j \text{ alors} & \neg\left[bG'_i a \wedge \neg(aG'_i b)\right] \end{array}\right] \tag{7.8}$$

La dominance traduit aussi une forme d'unanimité sur tous les points de vue. En effet, tous les points de vue indiquent entre autre que la seconde alternative n'est pas strictement préférée à la première. Elle est néanmoins un peu moins « stricte » que ne l'est la précédente puisque qu'elle nécessite l'absence et non la présence de préférences.

La politique *dictature* du point de vue j : Une alternative a est préférée (globalement) à une alternative b (noté $aG_{dict_j}b$) suivant la politique *dictature*, si elle est préférée pour la préférence partielle primitive relative au point de vu j :

$$\forall a, b \in ALT^2; \quad aG_{dict_j}b \quad \text{ssi} \quad aG'_j b \tag{7.9}$$

Afin de réaliser la phase d'agrégation, il existe bien entendu d'autres politiques d'élections envisageable que les trois que nous venons de présenter. Par exemple des politiques basées sur la majorité (Condorcet) ou sur un ordonnancement des critères (Bordas). Chacune d'elles donne a priori un résultat différent des autres car elles ont peu de contraintes à vérifier. Nous avons proposé d'utiliser des méthodes d'agrégation élémentaires car ce sont celles qui viennent à l'esprit immédiatement lorsque l'on est confronté à un problème de vote. Il existe néanmoins une littérature abondante sur le sujet que ce soit dans les domaines de la fusion d'informations (voir par exemple [Everaere et al., 2005]), de la théorie du choix social (décision multi-agent) et de la décision multi-critères (voir en particulier [Vincke, 1989] pour une introduction). À ce sujet le lecteur intéressé pourra regarder [Dubois and Prade, 2004] pour une discussion pointant sur les champs potentiels d'application des techniques d'agrégation. En ce qui concerne les techniques utilisées dans le champs la décision multi-critères, les méthodes de « surclassement » pourrait être utilisées mais nécessitent une adaptation importante car ces dernières nécessitent de savoir comparer les différents ensembles possibles de points de vue (ceci est réalisé généralement par une pondération des différents points de vue) et surtout supposent que les préférences à agréger vérifient l'hypothèse de cardinalité ce qui n'est pas le cas ici. Elles suivent toutefois le même esprit : comparer les alternatives avec chaque préférence partielle puis agréger les résultats. Enfin, les méthodes « itératives », très prisées dans le domaine de l'aide à la décision (voir [Vincke, 1989] et [Bouyssou et al., 2002]), ne correspondent pas du tout à notre approche car elles nécessitent une participation active d'un décideur de référence durant la décision.

En tout état de cause, il en résulte à chaque fois que la sémantique de l'opérateur $\geq$ peut-être exprimée en fonction des relations de préférences partielles primitives. Pour expliciter notre propos, nous présentons ci-dessous la sémantique de l'opérateur $\geq$ pour deux politiques particulière : les politiques *All* et *Dictature j*.

Pour la politique *All*, dire que la formule ϕ est préférée à la formule ψ via la préférence globale signifie que les alternatives qui vérifient les propriétés décrites par la formule ϕ et, si cela est cohérent, le contraire de celles décrites par la formule ψ, sont préférées (sans la restriction Ceteris Paribus), suivant **tous les points de vue**, aux alternatives qui vérifient les propriétés décrites par la formule ψ et, si cela est cohérent, le contraire de celles décrites par la formule ϕ :

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}3} \phi \geq \psi \quad \text{ssi} \quad \begin{array}{l} \bullet \quad \nvDash \neg\phi \quad \text{et} \quad \nvDash \neg\psi \\ \bullet \quad \left[\begin{array}{l} \forall (a,b) \in \mathrm{ALT}^2, \\ \text{si} \left| \begin{array}{l} a \vDash \phi + \overline{\psi} \\ b \vDash \psi + \overline{\phi} \end{array} \right. \\ \text{alors, } \textbf{pour tout } i,\ aG'_i b \end{array} \right] \end{array} \tag{7.10}$$

Pour la politique *Dictature j*, dire que la formule ϕ est préférée à la formule ψ via la préférence globale signifie que les alternatives qui vérifient les propriétés décrites par la formule ϕ et, si cela est cohérent, le contraire de celles décrites par la formule ψ, sont préférées (sans la restriction Ceteris Paribus), suivant **le point de vue** j, aux alternatives qui vérifient les propriétés décrites par la formule ψ et, si cela est cohérent, le contraire de celles décrites par la formule ϕ.

$$(\mathrm{ALT}, \models, \mathscr{G}) \models_{\mathscr{L}^3} \phi \geq \psi \quad \text{ssi} \quad \bullet \quad \not\models \neg\phi \quad \text{et} \quad \not\models \neg\psi$$
$$\bullet \quad \left[\begin{array}{l} \forall (a,b) \in \mathrm{ALT}^2, \\ \text{si} \left| \begin{array}{l} a \models \phi + \overline{\psi} \\ b \models \psi + \overline{\phi} \end{array} \right. \\ \text{alors} \quad a G'_{\mathbf{j}} b \end{array}\right] \tag{7.11}$$

7.5 Axiomatisation des politiques

Dans la pratique la relation de préférence globale d'un agent est déterminée par l'opérateur $\geq$. C'est donc ce dernier qui résulte en premier lieu de l'agrégation des préférences partielles étendues. Plus précisément, dans notre approche ce sont les résultat des comparaisons des préférences partielles étendues qui vont être agrégées via un mécanisme *d'élection* afin de donner le résultat des comparaisons de la préférence globale. Notre but est donc ici de déterminer par extension l'opérateur de préférence globale d'un agent en se basant sur les opérateurs de préférence partielle étendue. Dans cette approche, le résultat de la comparaison de deux formules par une préférence partielle étendue est considéré comme un vote en faveur du résultat qu'il indique.

Comme nous l'avons remarqué à la section 7.4, chaque politique d'agrégation donne a priori un résultat différent des autres. Par conséquent, chaque politique d'agrégation est mise en oeuvre par une axiomatique particulière. À titre d'exemple, nous proposons dans ce qui suit une axiomatisation pour quatre mécanismes différents : les politiques *All*, *Dictature* j, *Lex* et *Maj*. Afin de raccourcir les formules, nous introduisons une nouvelle notation : la notation $\mathrm{incomp}_i(\phi, \psi)$ est l'abréviation syntaxique de la formule $\neg(\phi \geq_i \psi) \wedge \neg(\psi \geq_i \phi)$. Elle indique intuitivement qu'il n'existe aucune information selon le point de vue numéro i pour comparer ϕ et ψ.

7.5.1 Axiomatique de la politique *All*

La formule ϕ est préférée globalement à ψ au sens *All* (noté $\phi \geq_{All} \psi$) si et seulement si toutes les préférences partielles étendues de l'agent indiquent que, de leurs points de vue respectifs, la formule ϕ est préférée à ψ.

Ceci est équivalent au fait qu'il n'existe aucune préférence partielle étendue de l'agent qui indique que, de son point de vue, une alternative vérifiant la formule ψ est préférée strictement à une vérifiant la formule ϕ (i.e. $\psi >_i \phi$) ou qu'elles sont incomparables (i.e. $\neg(\phi \geq_i \psi) \wedge \neg(\psi \geq_i \phi)$). Par conséquent, le défaut et les règles de déduction suivants sont ajoutés à la théorie :

$$\top : \phi \geq_{All} \psi \rightarrow \phi \geq_{All} \psi \tag{7.12}$$

$$\text{si} \quad \vdash_{\mathscr{L}^2} \psi >_i \phi \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \neg(\phi \geq_{All} \psi) \tag{7.13}$$

$$\text{si} \quad \vdash_{\mathscr{L}2} \text{incomp}_i(\phi, \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}3} \neg(\phi \geq_{All} \psi) \tag{7.14}$$

7.5.2 Axiomatique de la politique *Dictature* j

La formule ϕ est préférée globalement à ψ au sens *Dictature* j (noté $\phi >_{\text{Dict}_j} \psi$), si et seulement si la préférence partielle étendue relative au point de vue j indique que ϕ est préférée à ψ. Formellement la règle suivante est ajoutée à la théorie :

$$\text{si} \quad \vdash_{\mathscr{L}2} \phi \geq_j \psi \quad \text{alors} \quad \vdash_{\mathscr{L}3} \phi \geq_{\text{Dict}_j} \psi \tag{7.15}$$

De la même façon, les alternatives vérifiant respectivement les formules ϕ ne sont pas préférées au sens *Dictature* j aux alternatives ψ (noté $\neg(\phi \geq_{\text{Dict}_j} \psi)$), si et seulement si la préférence partielle étendue jugée relative au point de vue j indique que les alternatives vérifiant respectivement les formules ϕ ne sont pas préférées au sens *Dictature* j aux alternatives ψ (noté $\neg(\phi \geq_j \psi)$). Formellement la règle suivante est ajouté à la théorie :

$$\text{si} \quad \vdash_{\mathscr{L}2} \neg(\phi \geq_j \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}3} \neg(\phi \geq_{\text{Dict}_j} \psi) \tag{7.16}$$

Par conséquent, du fait de la définition de l'indifférence, il résulte de cette axiomatisation que les alternatives vérifiant respectivement les formules ϕ et ψ sont indifférentes au sens *Dictature* j (noté $\phi \sim_{\text{Dict}_j} \psi$), si et seulement si la préférence partielle étendue jugée relative au point de vue j indique que ϕ et ψ sont indifférents :

$$\text{si} \quad \vdash_{\mathscr{L}2} \phi \sim_j \psi \quad \text{alors} \quad \vdash_{\mathscr{L}3} \phi \sim_{\text{Dict}_j} \psi$$

$$\text{si} \quad \vdash_{\mathscr{L}2} \neg(\phi \sim_j \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}3} \neg(\phi \sim_{\text{Dict}_j} \psi)$$

7.5.3 Axiomatique de la politique *Lex*

Dans cette partie de document, les préférences partielles étendues sont supposées être ordonnées par niveau d'importance. Par souci de simplicité, nous supposons ici que si $i < j$ alors le point de vue i est plus important que le point de vue j. Cela ne réduit néanmoins pas la généralité de notre propos.

Par défaut, si ce n'est pas contradictoire avec le reste de la préférence partielle étendue, une formule ϕ de $\mathscr{L}$ n'est pas préférée à une formule ψ de $\mathscr{L}$. Formellement le défaut normal sans prérequis (à la Reiter) suivant est imposé :

$$\top : \neg(\phi \geq_{Lex} \psi) \rightarrow \neg(\phi \geq_{Lex} \psi) \tag{7.17}$$

De plus, la formule ϕ est préférée globalement à ψ au sens *Lex* (noté $\phi \geq_{Lex} \psi$), si et seulement si la préférence partielle étendue pertinente jugée la plus importante (i.e le plus petit i tel que $\neg\text{incomp}_i(\phi, \psi)$) indique que ϕ est préférée à ψ (i.e. $\phi \geq_i \psi$). Plus formellement, s'il existe n points de vue alors les n (schémas d') axiomes suivant sont imposés :

$$\text{si } \vdash_{\mathscr{L}^2} (\phi \geq_1 \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{Lex} \psi \tag{7.18}$$
$$\text{si } \vdash_{\mathscr{L}^2} \text{incomp}_1(\phi,\psi) \wedge (\phi \geq_2 \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{Lex} \psi \tag{7.19}$$
$$\text{si } \vdash_{\mathscr{L}^2} \text{incomp}_1(\phi,\psi) \wedge \text{incomp}_2(\phi,\psi) \wedge (\phi \geq_3 \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{Lex} \psi \tag{7.20}$$
$$\dots$$
$$\text{si } \vdash_{\mathscr{L}^2} \text{incomp}_1(\phi,\psi) \wedge \ldots \wedge \text{incomp}_{n-1}(\phi,\psi) \wedge (\phi \geq_n \psi) \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{Lex} \psi \tag{7.21}$$

Il résulte de cette axiomatisation que les formules ϕ et ψ sont indifférentes au sens *Lex* (noté $\phi \sim_{Lex} \psi$), si et seulement si toutes les préférences partielles étendues indiquent que les formules ϕ et ψ sont incomparables ou si la préférence partielle étendue pertinente jugée la plus importante (i.e le plus petit i tel que $\neg\text{incomp}_i(\phi,\psi)$) indique à la fois que ϕ est préférée à ψ et que ψ est préférée à ϕ (i.e. $\phi \geq_i \psi \wedge \psi \geq_i \phi$).

7.5.4 Axiomatique de la politique *Maj*

Par défaut, si ce n'est pas contradictoire avec le reste de la préférence partielle étendue, une formule ϕ de $\mathscr{L}$ n'est pas préférée à une formule ψ de $\mathscr{L}$. Formellement le défaut normal sans prérequis (à la Reiter) suivant est valide :

$$\top : \neg(\phi \geq_{Maj} \psi) \rightarrow \neg(\phi \geq_{Maj} \psi) \tag{7.22}$$

La formule ϕ est préférée globalement à ψ au sens *Maj* (noté $\phi >_{Maj} \psi$) si et seulement si la majorité absolue des préférences partielles étendues indique que les alternatives vérifiant la propriété ϕ sont préférées à celles vérifiant ψ. S'il existe n points de vue, la majorité absolue est obtenue avec $E(\frac{n}{2}) + 1$ points de vue distincts. Il existe alors $k = C_n^{E(\frac{n}{2})+1}$ façons différentes de choisir un ensemble de points de vue ayant la majorité absolue[5]. Par conséquent, Nous imposons que les k (schémas d') axiomes suivant sont valides. Chacun d'eux est relatif à une des différentes façons distinctes de choisir un ensemble de $l = E(\frac{n}{2})+1$ points de vue.

$$\text{si} \quad \vdash_{\mathscr{L}^2} \phi \geq_{\sigma_{1,1}} \psi \wedge \phi \geq_{\sigma_{1,2}} \psi \wedge \phi \geq_{\sigma_{1,3}} \psi \wedge \ldots . \wedge \phi \geq_{\sigma_{1,l}} \psi \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{\text{Maj}} \psi \tag{7.23}$$
$$\text{si} \quad \vdash_{\mathscr{L}^2} \phi \geq_{\sigma_{2,1}} \psi \wedge \phi \geq_{\sigma_{2,2}} \psi \wedge \phi \geq_{\sigma_{2,3}} \psi \wedge \ldots . \wedge \phi \geq_{\sigma_{2,l}} \psi \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{\text{Maj}} \psi \tag{7.24}$$
$$\ldots$$
$$\text{si} \quad \vdash_{\mathscr{L}^2} \phi \geq_{\sigma_{k,1}} \psi \wedge \phi \geq_{\sigma_{k,2}} \psi \wedge \phi \geq_{\sigma_{k,3}} \psi \wedge \ldots . \wedge \phi \geq_{\sigma_{k,l}} \psi \quad \text{alors} \quad \vdash_{\mathscr{L}^3} \phi \geq_{\text{Maj}} \psi \tag{7.25}$$

où la notation $\sigma_{i,j}$ désigne le $j^{\text{ème}}$ élément de la $i^{\text{ème}}$ combinaison de l éléments de l'ensemble $\{1, \ldots, n\}$.

[5] $E(x)$ représente la partie entière de x et C_n^p représente le nombre de combinaison de p éléments pris parmi n.

7.6 Discussion sur les politiques agrégation

Comme nous l'avons remarqué dans la section précédente, le choix de la politique d'agrégation est très important pour le résultat du choix final. En effet, comme les points de vue sont spécifiés dans notre modèle de façon indépendante les uns des autres, il est à prévoir qu'ils indiqueront des informations contradictoires. Par conséquent, dans la plupart des situations il n'existera pas de meilleure alternative objective parmi toutes celles imaginables c'est-à-dire d'alternative optimale pour chacun des points de vue. Néanmoins, comme il faut choisir une alternative, il faudra faire des compromis. Dans notre modèle, c'est justement la politique d'agrégation choisie qui déterminera la manière d'effectuer ces compromis. Par conséquent, comme chaque façon de faire des compromis peut aboutir à un résultat différent, le choix d'une politique plutôt que d'une autre influencera grandement le résultat du choix final. Nous ne voulons par rentrer dans le débat pour savoir quelle politique d'agrégation utiliser. Cependant, même si nous laissons cet aspect pour des travaux futurs, il nous semble important de préciser quelques points afin d'éclairer ces prochains travaux.

7.6.1 Rappel des caractéristiques des situations à traiter

Tout d'abord, il faut remarquer que la formalisation que nous présentons ici se fait dans un cadre bien précis. En particulier les alternatives qu'il faudra départager seront en nombre très limité. En effet, dans le cadre dans lequel nous nous plaçons, les alternatives à départager résultent d'un filtrage de l'ensemble des alternatives imaginable ALT via l'étape de planification basée (entre autre) sur les intentions de l'agent.

De plus, il est tout a fait sensé que dans certaines circonstances l'agent ne trouve aucune raison pour départager des alternatives. Après tout, pourquoi un agent ne pourrait pas trouver des alternatives « équivalentes » ? Dans un tel cas, c'est un processus aléatoire qui prendra le relais si besoin est.

7.6.2 Impossibilité d'exhiber une politique « optimale »

Comme les différents points de vue à prendre en compte sont souvent conflictuels, généralement aucune alternative n'est meilleure simultanément vis-à-vis de chaque point de vue ; d'où la nécessité de compromis. À ce propos, Arrow a montré dans [Arrow, 1951] qu'il n'existe pas de technique d'agrégation « équitable » qui permette d'obtenir à coup sûr une préférence (globale) qui soit un pré-ordre total et donc qui définisse sans ambiguïté les meilleures alternatives (voir l'annexe B.3 pour plus de détails).

À la vue de nos hypothèses, les préférences partielles étendues ne sont pas forcément des pré-ordres totaux. En particulier, elles ne sont pas forcément transitives. Par conséquent, le résultat du théorème d'Arrow n'est pas directement applicable. Néanmoins, comme une façon de s'accommoder de ce théorème d'impossibilité est d'imposer plus de contraintes aux relations à agréger, nous pouvons conjecturer que dans notre cas, puisque que nous faisons encore moins d'hypothèses sur les contraintes des relations à agréger, la relation obtenue ne sera pas en général un pré-ordre total. Il faut donc accepter l'agrégation par une structure plus faible qu'un pré-ordre. En particulier la relation que nous obtiendrons ne sera sûrement ni totale ni transitive.

7.6.3 Importance (relative) d'un point de vue

Lors de la phase d'agrégation, il peut être intéressant que certains points de vue soient plus importants que d'autres. Dans notre formalisme, ceci peut-être obtenu de deux manières différentes. Il est possible soit (1) d'utiliser une politique comme *Lex* en ayant auparavant ordonné les points de vue par ordre d'importance, soit (2) d'utiliser une politique comme *Maj* basée sur le nombre de voix « pour » et « contre » tout en ayant auparavant dupliqué les points de vue importants. En effet, si un aspect de la décision est décrit (de façon intentionnelle ou pas) avec plusieurs points de vue alors dans de nombreuses politiques d'agrégation, cet aspect aura mécaniquement un poids plus important par rapport aux autres. Cela revient à pondérer les différents points de vue.

Une gestion inter-points de vue plus fine nécessite des informations supplémentaires comme par exemple la valeur des cœfficients de substitution de chaque point de vue. Pour intégrer ces derniers à notre formalisme, les méthodes de *surclassement* pourraient servir d'inspiration. À ce sujet, l'axiomatisation des règles d'*accordances* proposée dans [Dubois et al., 2001] pourrait être reprise.

7.6.4 Rapport entre les différentes politiques

Certaines politiques d'agrégation sont plus « restrictives » que d'autres dans le sens où les premières permettent d'ordonner, en plus de celles ordonnées par les secondes, des alternatives supplémentaires. Plus précisément, une politique pol_1 est plus « restrictive » qu'une politique pol_2 si pour tout couple de formules ϕ et ψ, si ϕ est préférée à la formule ψ via la politique pol_1 (i.e. $\phi >_{\mathrm{pol}_1} \psi$) alors la formule ϕ est aussi préférée à la formule ψ via la politique pol_2 (i.e. $\phi >_{\mathrm{pol}_2} \psi$). Ainsi par exemple, la politique *All* est plus restrictive que la politique *Maj*. De même la politique *Dictature* j est plus restrictive que la politique *Lex* si l'ordonnancement des points de vue indique que le point de vue j est le plus important.

Intuitivement, une politique est plus « restrictive » qu'une autre si elle permet d'ordonner les mêmes alternatives mais aussi si elle permet d'en ordonner d'autres. Dans ce mouvement visant à ordonner de plus en plus d'alternatives, il est à craindre que les « raisons » permettant cet ordonnancement soient de moins en moins évidentes. Il n'est donc pas a priori plus intéressant d'utiliser une politique la moins restrictive possible et en particulier si l'indifférence a un sens important. Par exemple, dans la situation où un agent recherche pour un utilisateur des voyages dans une grande base de donnée, le fait de ne pas départager toutes les alternatives permet de présenter à l'utilisateur des alternatives qui peuvent l'intéresser.

7.6.5 Politique et « personnalité » de l'agent

Comme nous l'avons remarqué en début de cette section, le choix d'une politique d'agrégation influence grandement le choix final. Dans le contexte d'agents rationnels de cette thèse, ceci peut être interprété par le fait que la politique d'agrégation choisie représente une partie de la « personnalité » de l'agent.

Via l'assignation de poids importants à des points de vue, ces derniers peuvent devenir plus important que les autres lors de l'agrégation avec des politique de type *Maj*. Cette situation peut s'interpréter comme le fait que l'agent privilégie les aspects correspondant à ces points de vue. Ainsi par exemple, si dans notre exemple le(s) point(s) de vue relatif(s) au prix des plats est (sont) privilégié(s) dans une optique de minimisation, alors cela indiquera que l'agent à un comportement « économe ». Le même phénomène est obtenu avec des politiques de type *Lex*. Il est néanmoins alors plus marqué : c'est une certaine forme de « dictature ».

D'un autre côté, un agent qui utilise une politique comme l'*unanimité* aura un comportement qualifiable de « tout ou rien ». En effet, ce dernier choisira soit l'alternative qui est meilleure que toutes celles qu'il peut choisir via chaque point de vue, soit une alternative au hasard.

7.7 Choix de la réaction

Le formalisme que nous venons de présenter dans ce chapitre permet de déterminer une relation binaire qui agrège l'ensemble points de vue selon une politique donnée : la préférence globale.

Cependant, comme nous l'avons remarqué dans la section 7.6.2, cette relation n'est a priori ni complète ni transitive. Par conséquent il n'existe pas a priori d'alternative (ou d'ensemble d'alternatives) qui domine les autres, c'est-à-dire qui est préférée à toutes les autres. Il en résulte qu'il faut utiliser les informations décrites via la préférence globale de façon plus ou moins arbitraire pour décider l'alternative à mettre œuvre. Plus précisément, en se basant sur l'ensemble des couples d'alternatives constituant la relation de préférence globale de l'agent[6], la « meilleure » alternative va être déterminée grâce à un critère de décision (i.e. une deuxième technique de décision qui n'a a priori aucun rapport avec la technique d'agrégation choisie précédemment). Là aussi, il existe de nombreuses techniques utilisables. Chacune d'elles déterminera un choix différent.

Étant donnée que le but de cette thèse est de proposer un cadre permettant de spécifier des informations afin de mettre en oeuvre une phase de décision, nous ne rentrons pas dans le débat visant à déterminer le critère de décision à employer. Toutefois, afin d'expliciter notre propos, nous exposons ci-dessous rapidement quelques politiques envisageables ainsi que leur application dans la situation où les alternative a, b, c, d et e sont à considérer et que :

[6] Il y a autant de couples à déterminer que de combinaisons possibles d'alternatives. Par conséquent, s'il y a n alternatives il y aura $(n^2 - n)/2$ couples.

a est préférée à b,	d est préférée à b,	c est préférée à b,	e est préférée à a.
c est préférée à d,	a est préférée à d,	a est préférée à d,	

Min- :

L'alternative à choisir est une de celles qui ont le moins d'alternatives qui leur sont préférées.
Dans l'exemple, l'emploi de la politique *Min-* implique que l'alternative à choisir est c ou e.

Max+ :

L'alternative à choisir est une de celles qui sont préférées au plus grand nombre d'alternatives.
Dans l'exemple, l'emploi de la politique *Max+* implique que l'alternative à choisir est a.

Mix-/+ :

L'alternative à choisir est une de celles qui, en moyenne, sont préférées au plus grand nombre d'alternatives et qui ont le moins d'alternatives qui leur sont préférées.
Dans l'exemple, l'emploi de la politique *Mix-/+* implique que l'alternative à choisir est a ou c.

Maj :

L'alternative à choisir est une de celles qui sont plus souvent préférées que non préférées.
Dans l'exemple, l'emploi de la politique *Maj* implique que l'alternative à choisir est a ou c ou e.

NotDom :

L'alternative à choisir est une de celles qui est telle qu'aucune autre alternative ne lui est préférée et à défaut l'alternative vide.
C'est la politique de choix formalisée et utilisée par Louis dans [Louis, 2002] (voir l'équation (1.1)). Dans l'exemple, l'emploi de la politique *NotDom* implique que l'alternative à choisir est c ou e.

Comme pour les politiques d'agrégation, il nous semble opportun de faire les remarques suivantes puisque celles-ci traduisent une certaine forme de « bon sens ». Tout d'abord, l'alternative à choisir est un élément de ALT. Plus précisément, elle fait partie du sous-ensemble d'alternatives de ALT parmi lesquelles le choix est explicitement à faire. Dans notre approche cet ensemble est supposé connu. Par conséquent, étant donné que les alternatives qui ne sont pas explicitement parmi les alternatives à comparer (i.e. parmi les données initiales) peuvent apparaître comme non pertinentes pour le choix final, un raffinement des critères cités ci-dessus peut être obtenu en ne considérant que les alternatives parmi lesquelles l'agent doit explicitement choisir. C'est ce que nous appelons les versions « étoilées » des critères précédents. D'autre part, il est fortement souhaitable que l'alternative à choisir vérifie les propriétés que nous nommons d'*unanimité* et de *dominance*. La propriété d'*unanimité* indique que si une alternative est préférée à toutes toutes les autres via la préférence globale, alors cette alternative est à choisir. La propriété de *dominance* indique quant à elle que s'il existe un ensemble non vide d'alternatives qui sont préférées (globalement) à toutes les autres alors l'alternative à choisir se trouve parmi cet ensemble. Plus précisément, si l'ensemble des alternatives à comparer peut être partitionné en deux ensembles distincts A et B non vides tels que toute alternative de A est préférée à toute alternative de B, alors l'alternative à choisir se trouve dans l'ensemble A.

7.8 Exemple (Fin)

Afin de rendre un peu plus concret le formalisme que nous venons de présenter nous allons voir dans cette section ce qu'il peut apporter à l'exemple introduit dans la section 4.4 et développé aux sections 5.7 et 6.6.

Les informations déduites précédemment avec les préférences partielles étendues permettent de dire dans la situation qui nous intéresse (i.e. l'exemple introduit dans la section 4.4) que l'alternative a est préférée strictement à b, que l'alternative b est préférée strictement à c et que l'alternative a est préférée strictement à c. Cependant, ces informations sont relatives à un point de vue particulier : le point de vue des préférences « culinaires » de Pierre. Aussi il est important de regarder d'autres dimensions. Par exemple, il est intéressant de considérer en outre les dimensions « économique » et « pratique » des alternatives[7]. C'est pour cela que nous introduisons de nouvelles préférences partielles et que nous indiçons nos notations (avec respectivement « C », « E » et « P ») pour les différencier les unes des autres.

Considérons donc la situation précédente mais où il a été déduit[8] en plus que, pour des raisons pratiques, Pierre préfère manger au Capuccino ou à la Légende étant donné que ces deux restaurants se trouvent à Lannion et que le Goéland se trouve à Trébeurden. De plus, il a été aussi déduit[8] que, pour des raisons financières, il est préférable de ne pas manger au Goéland ni au Capuccino du fait de leur mauvais rapport qualité-prix (voir la figure 7.1). Plus précisément, considérons la situation où les formules suivantes sont vérifiées :

[7] Il existe beaucoup d'autres points de vue intéressants à considérer. Citons par exemple l'appréciation des restaurants par des guides ou les habitudes de Pierre (Pierre habite dans la région et va quelque fois au restaurant. Une étude statistique de ses fréquentations permet de connaître ses habitudes). Dans un souci de clarté et de concision cet exemple n'en traite que trois.

[8] Les préférences partielles relatives aux points de vue « Pratique » et « Économique » se traitent de la même façon que celles relatives au point de vue « Culinaire ». En particulier, les nouvelles connaissances et préférences nécessaires n'interfèrent pas avec les déductions effectuées précédemment du fait du formalisme utilisé.

$$\begin{array}{lll}
\mathscr{P} \vdash_{\mathscr{L}2} \quad f_a \geq_{\mathrm{C}} f_b & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_b \geq_{\mathrm{P}} f_a & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_b \geq_{\mathrm{E}} f_a \\
\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_b \geq_{\mathrm{C}} f_a) & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_a \geq_{\mathrm{P}} f_b) & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_a \geq_{\mathrm{E}} f_b) \\
\mathscr{P} \vdash_{\mathscr{L}2} \quad f_b \geq_{\mathrm{C}} f_c & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_b \geq_{\mathrm{P}} f_c & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_b \geq_{\mathrm{E}} f_c \\
\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq_{\mathrm{C}} f_b) & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_c \geq_{\mathrm{P}} f_b & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq_{\mathrm{E}} f_b) \\
\mathscr{P} \vdash_{\mathscr{L}2} \quad f_a \geq_{\mathrm{C}} f_c & \mathscr{P} \vdash_{\mathscr{L}2} \quad f_c \geq_{\mathrm{P}} f_a & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_a \geq_{\mathrm{E}} f_c) \\
\mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq_{\mathrm{C}} f_a) & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_a \geq_{\mathrm{P}} f_c) & \mathscr{P} \vdash_{\mathscr{L}2} \neg(f_c \geq_{\mathrm{E}} f_a)
\end{array}$$

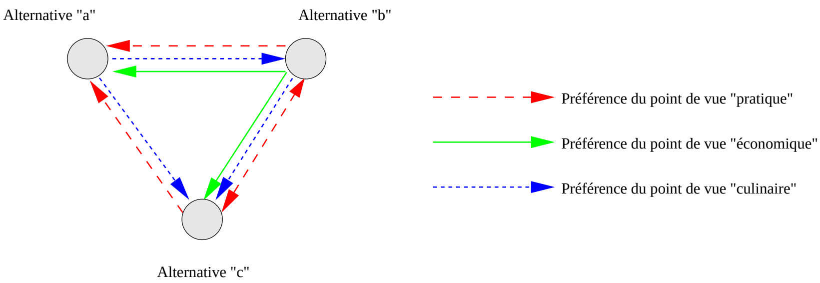


FIG. 7.1 – Schéma regroupant les trois types de préférences partielles primitives

Afin d'agréger ces préférences et donc de déterminer l'alternative à choisir, nous supposons ici connue[9] la politique à employer. En particulier, nous supposons que dans une telle situation, la politique « Maj » est utilisée. Comme nous considérons seulement trois points de vue dans cet exemple, la majorité est obtenue avec deux. C'est pourquoi, puisque $\mathscr{P} \vdash_{\mathscr{L}^2} f_b \geq_{\mathrm{E}} f_a$ et $\mathscr{P} \vdash_{\mathscr{L}^2} f_b \geq_{\mathrm{P}} f_a$, d'après l'une des règles (7.23), (7.24), …, (7.25), on obtient que $\mathscr{P} \vdash_{\mathscr{L}^3} f_b \geq_{\mathrm{Maj}} f_a$. De plus, puisqu'il n'existe pas plus d'un point de vue qui indique que f_a est préféré à f_b, le défault (7.22) permet de déduire que $\mathscr{P} \vdash_{\mathscr{L}3} \neg(f_a \geq_{\mathrm{Maj}} f_b)$. De la même façon, on déduit que la préférence globale de l'agent est telle que les propriétés suivantes (qui sont résumées sur la figure 7.2) sont vérifiées :

$$\begin{array}{lll}
\mathscr{P} \vdash_{\mathscr{L}3} \neg(f_a \geq_{\mathrm{Maj}} f_b) & \mathscr{P} \vdash_{\mathscr{L}3} \quad f_b \geq_{\mathrm{Maj}} f_c & \mathscr{P} \vdash_{\mathscr{L}3} \neg(f_a \geq_{\mathrm{Maj}} f_c) \\
\mathscr{P} \vdash_{\mathscr{L}3} \quad f_b \geq_{\mathrm{Maj}} f_a & \mathscr{P} \vdash_{\mathscr{L}3} \neg(f_c \geq_{\mathrm{Maj}} f_b) & \mathscr{P} \vdash_{\mathscr{L}3} \neg(f_c \geq_{\mathrm{Maj}} f_a)
\end{array}$$

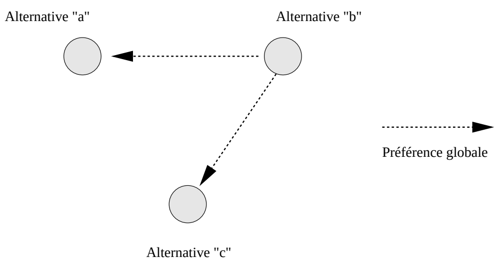


FIG. 7.2 – Schéma de la préférence globale avec la politique « Majorité »

[9]Nous ne cherchons pas ici à déterminer comment choisir la politique d'agrégation à utiliser. Pour nous c'est une donnée initiale.

Maintenant que la préférence globale de l'agent est déterminée (voir la figure 7.2), il nous reste à l'utiliser pour choisir. Si nous supposons que pour se faire l'agent utilise le critère *Max+**, l'alternative à choisir est l'une de celles (parmi a, b et c) qui est la plus préférée (via la préférence globale) aux autres alternatives. Dans cette situation, c'est donc l'alternative b qui est à choisir. Par conséquent l'agent propose à Pierre d'aller manger du poisson au restaurant la Légende.

Afin de montrer l'importance de la politique d'agrégation utilisée, considérons ici, qu'à la place d'utiliser la politique « *Maj* », l'agent utilise la politique « *Lex* » avec comme ordre d'importance des points de vue : (« Culinaire »,« Pratique »,« Économique »). Dans une telle situation, puisque le point de vue « Culinaire » est le plus important et qu'il indique que f_a est préféré à f_b (i.e. $\mathscr{P} \vdash_{\mathscr{L}^2} f_a \geq_C f_b$, on a aussi, d'après la règle (7.18), que $\mathscr{P} \vdash_{\mathscr{L}^3} f_a \geq_{\text{Lex}} f_b$. De plus, puisqu'aucune des règles (7.18) ...(7.21) ne permet d'inférer le contraire, on a, d'après le défaut (7.17), $\mathscr{P} \vdash_{\mathscr{L}^3} \neg(f_b \geq_{\text{Lex}} f_a)$. De la même manière, les formules suivantes sont aussi déduites :

$$\begin{array}{lll} \mathscr{P} \vdash_{\mathscr{L}^3} \quad f_a \geq_{\text{Lex}} f_b & \mathscr{P} \vdash_{\mathscr{L}^3} \quad f_b \geq_{\text{Lex}} f_c) & \mathscr{P} \vdash_{\mathscr{L}^3} \quad f_a \geq_{\text{Lex}} f_c) \\ \mathscr{P} \vdash_{\mathscr{L}^3} \neg(f_b \geq_{\text{Lex}} f_a) & \mathscr{P} \vdash_{\mathscr{L}^3} \neg(f_c \geq_{\text{Lex}} f_b) & \mathscr{P} \vdash_{\mathscr{L}^3} \neg(f_c \geq_{\text{Lex}} f_a) \end{array}$$

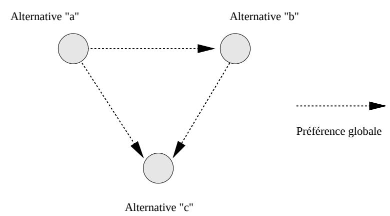


FIG. 7.3 – Schéma de la préférence globale avec la politique « Lex » et l'ordonnancement des points de vue (« Culinaire »,« Pratique »,« Économique »)

La relations de préférence globale ainsi obtenue (voir la figure 7.3) est différente de celle obtenue précédemment (voir la figure 7.2). Il en résulte que si l'agent utilise toujours le critère de décision *Max+**, alors l'alternative à choisir sera l'alternative a.

Quatrième partie

Extensions et conclusion générale

Chapitre 8

Extensions possibles de notre formalisme

Dans les chapitres précédents, nous avons présenté un formalisme précis permettant (1) de spécifier des préférences de manière intuitive et (2) de mettre en œuvre une phase de décision dans le cadre d'un formalisme logique tels que ceux utilisés pour définir des agents rationnels. Afin d'utiliser ce formalisme de façon pratique, nous proposons dans ce chapitre deux pistes d'extension de notre travail. Dans un premier temps, nous présentons une voie pour intégrer effectivement notre proposition à un formalisme d'agents basée sur une logique multimodale. Dans un second temps, nous présentons l'utilisation de notre proposition dans un schéma plus général de prise de décision faisant intervenir des informations de natures différentes en plus des préférences que nous avons traités jusqu'ici.

8.1 Vers l'intégration de notre proposition à un formalisme d'agents

Par construction, notre modèle est adapté aux modèles cognitifs d'agents puisqu'il est formalisé logiquement sur la base d'une notion intuitive : la préférence (voir le chapitre 3). Toutefois, jusqu'à présent, l'intégration de notre proposition à un formalisme d'agents n'a pas été explicitement abordée. C'est donc ce que nous discutons dans cette section. Plus précisément, nous présentons ici deux façons possible d'intégrer notre proposition au formalisme de la théorie de l'interaction rationnelle [Sadek, 1991]. Afin de simplifier notre discours, dans cette partie de document, seule la notion de préférence *globale* est examinée. Par conséquent, les langages $\mathscr{L}^1$ et $\mathscr{L}^2$ (introduits respectivement aux chapitres 5 et 6) ne sont pas considérés ici. C'est pourquoi, les indices, lorsqu'ils apparaissent, permettent simplement de différencier les agents mis en jeu et, plus généralement, les différents opérateurs modaux.

De façon générale, l'utilisation dans un formalisme logique d'agent cognitif de la notion de préférence (afin de mettre en œuvre une phase de décision) nécessite de pouvoir décrire (1) les alternatives, (2) les préférences et (3) les rapports des préférences entre elles et surtout avec les autres attitudes mentales (voir par exemple la façon dont Kaci et van der Torre définissent le langage qui leur permet de définir et manipuler les six types de préférences qu'ils comparent dans [Kaci and van der Torre, 2005]). C'est pourquoi, après avoir proposé des modifications aux langages $\mathscr{L}$ et $\mathscr{L}^3$ définis respectivement aux sections 4.1 et 7.2, nous introduisons ici un cinquième langage pour satisfaire le troisième point : le langage $\mathscr{L}^4$.

8.1.1 Modification à apporter au langage $\mathscr{L}$

Avant d'aller plus loin, il est nécessaire de modifier le langage servant à décrire les alternatives : le langage $\mathscr{L}$. En effet, même si le langage $\mathscr{L}$ introduit dans la section 4.1 a une grande expressivité, cette dernière est insuffisante pour décrire l'ensemble des alternatives qui s'offre à un agent formalisé sur la base de la théorie de l'interaction [Sadek, 1991]. Plus précisément, puisque l'environnement dans lequel évolue l'agent y est décrit par une logique multimodale quantifiée des attitudes mentales et de l'action (voir la section 1.1.3), les alternatives que l'agent peut considérer doivent être décrites de la même manière.

Les logiques (multi-)modales sont développées depuis le début du XX$^{\text{ieme}}$ siècle afin d'étendre l'expressivité des logiques classiques sans toutefois passer à l'ordre supérieur. Ainsi par exemple, une logique du premier ordre multimodale est plus expressive qu'une logique d'ordre zéro et moins expressive qu'une logique du second ordre. Cette extension est obtenue en introduisant (dans la logique à étendre) un ou plusieurs opérateurs modaux (notés ici M_i). La syntaxe du nouveau langage est obtenue en ajoutant à la syntaxe du langage à étendre des règles de la forme suivante : « si ϕ est une formule bien formée de ce langage, alors $M_i(\phi)$ est aussi une formule bien formée de ce langage ». Généralement, la sémantique des nouvelles formules ainsi construites se base quant à elle, en plus des structures utilisées pour donner un sens aux formules du langage à étendre, sur un ensemble de *mondes possibles* (noté W) ainsi que sur un ensemble de relations binaires R sur ces mondes [Kripke, 1963]. Plus précisément, dans ces logiques, la sémantique des formules (toujours compositionnelle) est relativisée à chaque monde possible (noté usuellement w) : la notation $\vDash_w \phi$ indique que dans le monde w la formule ϕ est valide. Pour déterminer si c'est le cas, lorsque la formule ϕ est construite avec au moins un opérateur modal, il est nécessaire d'examiner les mondes possibles en relation avec le monde w. Si ce n'est pas le cas, le sens de la formule ϕ dans le monde w est obtenu de la même manière que dans la logique à étendre. Enfin, l'axiomatique est obtenue en complétant l'axiomatique de la logique à étendre par des règles d'inférence ainsi que des axiomes appropriés afin de refléter le sens des opérateurs introduits.

Dans ce document, nous ne parlerons pas plus en détails de logiques modales. Le lecteur intéressé pourra se référer entre autres à [Thayse et al., 1989] pour un cours plus complet. Notons toutefois que dans le cas qui nous intéresse (i.e. la théorie de l'interaction) les modèles de la logique utilisée par Sadek sont des quadruplets de la forme (W, R, S, V) où

- W est un ensemble non vide de mondes possibles. Chaque monde possible fournit la description, dans le langage $\mathscr{L}$, d'un univers qui peut servir de support aux croyances

et aux choix[1] des agents ainsi qu'à la représentation des instants temporels.
- R est une famille de relations binaires sur W qui interprètent les opérateurs modaux normaux du langage $\mathscr{L}$ permettant de modéliser les croyances, les choix[1] ainsi que l'occurrence passée et future des actions (voir la section 1.1.3).
- S correspond à l'ensemble des valeurs que peuvent prendre les variables.
- V est une fonction de valuation qui, pour tout monde possible w de W, associe (1) à chaque symbole de fonction f à n places, une fonction de S^n dans S (notée $V_{f,w}$) et (2) à chaque symbole de prédicat P à n places, une fonction de S^n dans $\{\top, \bot\}$ (notée $V_{P,w}$).

Ceci est une vision simplifiée et schématisée de la structure des modèles de la logique utilisée dans la théorie de l'interaction. Le lecteur intéressé par un exposé plus complet pourra se reporter à [Louis, 2002, pp33-40].

Dans notre proposition, chaque alternative est identifiée à un modèle de la logique servant à décrire les propriétés des alternatives. Par conséquent, dans une adaptation de notre proposition au formalisme d'agents de Sadek, nous pensons que les alternatives seront des modèles de ce langage logique multimodale quantifié des attitudes mentales et de l'action. Plus précisément ce seront des quadruplets (W, R, S, V).

8.1.2 Modification à apporter au langage $\mathscr{L}^3$

Afin de décrire des préférences sur des formules d'une logique multimodale nous proposons de réutiliser le langage $\mathscr{L}^3$ introduit dans le chapitre 7. Cependant, ce langage ne permet pas originellement de « traiter » des telles formules. C'est pourquoi, nous proposons dans cette section une voie pour modifier le langage $\mathscr{L}^3$ afin qu'il puisse prendre en compte les diverses modalités.

Nous avons défini dans la section 4.3.2 une préférence comme une relation binaire sur l'ensemble des alternatives. Dans notre formalisme, cette relation est appréhendée via un opérateur logique sur les formules du langage $\mathscr{L}$ servant à décrire les alternatives. La syntaxe du langage ainsi obtenu est assez simple et peut se schématiser comme suit : si ϕ et ψ sont des formules du langage $\mathscr{L}$ alors $\phi > \psi$ est une formule de $\mathscr{L}^3$ (voir la section 7.2). Par conséquent, le fait d'utiliser comme langage $\mathscr{L}$ le langage multimodal permettant de modéliser les agents dans la théorie de l'interaction nécessite de modifier quelque peu le langage $\mathscr{L}^3$.

Au niveau syntaxique, le fait d'utiliser un langage $\mathscr{L}$ qui soit multimodal permet, sans changer la syntaxe, de comparer aussi des formules contenant des opérateurs modaux. Par exemple la formule $Agit(a) \wedge B(\phi) \geq Fait(b) \wedge B(\psi)$ fait alors partie du langage $\mathscr{L}^3$. Cette dernière indique « grosso modo » qu'« il est préférable que l'agent mette en oeuvre le plan a et qu'il croit que ϕ est vérifié plutôt que l'agent mette en oeuvre le plan b et qu'il croit que ψ est vérifié ».

Nous pensons que la sémantique de ces formules se baserait quant à elle toujours sur

[1]Le terme « choix » correspond ici à l'opérateur C_i de la théorie de l'interaction défini au paragraphe (b) de la section 1.1.3 et non au résultat de la décision

des modèles de la forme $(ALT, \models, \mathscr{G})$, mais où cette fois

- ALT est l'ensemble des alternatives et donc l'ensemble des modèles de $\mathscr{L}$: des quadruplets (W, R, S, V) et non plus des couples (S, V).
- $\models$ est la validité sémantique du langage servant à décrire les alternatives ($\mathscr{L}$).
- $\mathscr{G}$ est un ensemble des relations binaires sur ALT : les préférences.

Plus précisément, nous pensons que la sémantique que nous avons donnée dans la section 7.3.1 peut être reprise telle quelle pour interpréter ces formules avec ces nouveaux modèles. Néanmoins il nous semble nécessaire d'adapter la construction des ensembles $\mathscr{L}^{\phi\psi}$ et Subf*(ϕ) introduits dans la section 5.1.1. En effet c'est sur ces ensembles que se base la construction des préférences partielles primitives (voir l'équation (5.4)), donc des préférences partielles étendues et donc aussi de la préférence globale. Cette adaptation ne nous semble pas triviale a priori. Elle nécessite de décider si, par exemple, les formules $B_1(\phi)$ (L'agent 1 croit que la propriété ϕ est vérifiée) et $B_2(\phi)$ (L'agent 2 croit que la propriété ϕ est vérifiée) ont des rapports entre elles. En effet, d'un côté, la formule ϕ apparaît dans les deux formules ; d'un autre côté, il est intuitif de supposer que les croyances d'un agent sont indépendantes de celles d'un autre agent.

L'axiomatisation que nous avons proposée à travers les sections précédentes ne nous semble pas devoir être modifiée. Notons ici que ces axiomes dépendent de la politique d'agrégation choisie.

8.1.3 Le langage $\mathscr{L}^4$

Afin de raisonner sur les rapports des préférences entre elles et avec les autres attitudes mentales, un troisième langage est nécessaire. Nous le notons $\mathscr{L}^4$. Dans cette section nous présentons quelques idées pour définir la syntaxe, la sémantique et l'axiomatique d'un tel langage. Par la suite, nous insistons sur la nécessité de contraindre ce formalisme afin que la réaction de l'agent soit effectivement en rapport avec ses préférences et proposons quelques pistes pour le faire.

a) Syntaxe :

Le langage $\mathscr{L}^4$ doit permettre, entre autres, de considérer simultanément plusieurs préférences ainsi que des préférences et d'autres attitudes mentales conjointement. C'est pourquoi nous pensons que les formules de ce langage sont (1) les formules permettant de décrire les alternatives (les formules de $\mathscr{L}$), (2) les formules permettant de décrire les préférences (les formules de $\mathscr{L}^3$) et (3) les formules construites au dessus des précédentes avec les mêmes règles syntaxique que celles du langage $\mathscr{L}$ (les formules de $\mathscr{L}^4$ à proprement parlé). Formellement, pour nous, une formule bien formée de $\mathscr{L}^4$ est soit :

- une formule bien formée de $\mathscr{L}$ ou de $\mathscr{L}^3$;
- une formule de la forme de $\phi \wedge \psi$ avec ϕ et ψ des formules bien formées de $\mathscr{L}^4$;
- une formule de la forme de $\neg\phi$ avec ϕ une formule bien formée de $\mathscr{L}^4$;
- une formule de la forme de $\forall x\phi$ avec ϕ une formule bien formée de $\mathscr{L}^4$;
- une formule de la forme de $M_i(\phi)$ avec ϕ une formule bien formée de $\mathscr{L}^4$ et M_i un opérateur modal de $\mathscr{L}$.

Le fait d'utiliser les formules de $\mathscr{L}^3$ comme formules atomiques de $\mathscr{L}^4$ ainsi que le fait que le langage $\mathscr{L}^4$ ait une syntaxe similaire à celle du langage formalisant la théorie de l'interaction, permettent ensemble, entre autres, de relativiser une préférence à un instant temporel et à un état mental d'un agent (voir le point (a) de la section 2.3.2 sur la vision des préférences de von Wright). En particulier, la formule $B(\phi > \psi)$ appartient au langage $\mathscr{L}^4$. Celle-ci peut être interprétée comme le fait que l'agent croit que ϕ est préférée à ψ. De même la formule $Faisable(a, \phi > \psi)$ appartient aussi à $\mathscr{L}^4$ et indique que, après la réalisation de l'action a, ϕ est préférée à ψ. Remarquons que ces formules n'appartiennent ni à $\mathscr{L}^3$ ni à $\mathscr{L}$.

b) Sémantique :

Afin d'interpréter les formules du langage $\mathscr{L}^4$, nous proposons l'utilisation de modèles qui soient des quintuplets de la forme $(W, R, S, V, \mathscr{P})$ où W, R, S et V représentent les mêmes objets que précédemment et où $\mathscr{P}$ est une fonction qui associe à chaque monde possible w de W un ensemble d'alternatives (noté ALT_w) et un ensemble de relations binaires sur ces alternatives noté $\mathscr{G}_w$: les préférences dans le monde w. Quel que soit le monde w, l'ensemble ALT_w désigne le même ensemble : il correspond à l'ensemble des modèles de $\mathscr{L}$. Par contre, les ensembles $\mathscr{G}_w$ n'ont a priori aucune raison d'être identiques d'un monde possible à un autre : dans différentes situations, les préférences sont différentes.

La valeur sémantique d'une formule ϕ de $\mathscr{L}^4$ se base sur de tels modèles et est relative à une assignation g ainsi qu'à un monde w. La notation $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ indique que la formule ϕ de $\mathscr{L}^4$ est vrai dans le modèle $(W, R, S, V, \mathscr{P})$, dans le monde w de W et pour l'assignation g. Nous proposons que ceci soit déterminé de la façon suivante :

- Si ϕ est une formule de $\mathscr{L}$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si ϕ est satisfaite dans le modèle (W, R, S, V), dans le monde w et pour l'assignation g (i.e. $(W, R, S, V) \vDash_w \phi[g]$).
- Si ϕ est une formule de $\mathscr{L}^3$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si ϕ est satisfaite dans le monde w et pour l'assignation g suivant la sémantique présentée dans la section 7.3.1 (et adaptée suivant les remarques de la section 8.1.2) avec pour ensemble ALT, l'ensemble ALT_w et l'ensemble $\mathscr{G}$, l'ensemble $\mathscr{G}_w$ (i.e. $(W, R, S, V, ALT_w, \mathscr{G}_w) \vDash_w \phi[g]$).
- Si ϕ est une formule de la forme $\neg\psi$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si la formule ψ n'est pas telle que $(W, R, S, V, \mathscr{P}) \vDash_w \psi[g]$.
- Si ϕ est une formule de la forme $\psi_1 \wedge \psi_2$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si on a à le fois $(W, R, S, V, \mathscr{P}) \vDash_w \psi_2[g]$ et $(W, R, S, V, \mathscr{P}) \vDash_w \psi_1[g]$.
- Si ϕ est une formule de la forme $\forall x \psi[g]$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si pour tout objet d de S on a $(W, R, S, V, \mathscr{P}) \vDash_w \psi[g[x/d]]$. La notation $g[x/d]$ désignant la variante de la fonction g qui assigne l'objet d à la variable x.
- Si ϕ est une formule de la forme $M_i\psi[g]$, alors on a $(W, R, S, V, \mathscr{P}) \vDash_w \phi[g]$ si et seulement si on a $(W, R, S, V, \mathscr{P}) \vDash_{w'} \psi[g]$ pour l'ensemble des mondes w' en relation avec le monde w via la relation de R associée à l'opérateur M_i.

c) Axiomatique

Tout comme la syntaxe et la sémantique, nous pensons que l'axiomatique de ce nouveau langage peut être obtenue en combinant l'axiomatique du langage $\mathscr{L}$ avec l'axiomatique du langage $\mathscr{L}^3$. Plus précisément, nous proposons que les axiomes et règles d'inférences du langage $\mathscr{L}^4$ soient (1) ceux du langage à la base de la théorie de l'interaction (voir par exemple [Louis, 2002, pp34-37]) et (2) ceux que nous avons proposés dans la section 7.5.

A l'image de l'axiomatisation de l'extension du langage $\mathscr{L}^3$ que nous avons ébauché dans la section 8.1.2, nous ne pensons pas que l'axiomatisation du langage $\mathscr{L}^3$ (proposée dans la section 7.5) doit être modifiée pour être intégrée à l'axiomatique du langage $\mathscr{L}^4$. Toutefois, du fait que la règle d'inférence de *nécessitation* est applicable à chacun des opérateurs modaux du langage, il nous semble nécessaire d'ajouter, pour chaque défaut de la forme $\top : \neg(\phi \geq \psi) \rightarrow \neg(\phi \geq \psi)$, les défauts de la forme $\top : \mathscr{M}(\neg(\phi \geq \psi)) \rightarrow \mathscr{M}(\neg(\phi \geq \psi))$ où $\mathscr{M}$ est une suite d'opérateurs modaux normaux. Ceci permet par exemple d'indiquer que les agents pensent que, par défaut, une propriété n'est pas préférable à une autre. Plus généralement, nous pensons qu'il faut faire de même pour chacune des règles d'inférence introduites aux cours des chapitres 5, 6 et 7.

Remarquons ici, que ceci n'est qu'une proposition résultant d'une étude limité du problème. En conséquence, elle doit plus être appréhendée comme une voie de recherche que comme un résultat définitif.

d) Extension de la théorie de l'interaction

Nous pensons que l'axiomatique qui résultera des la mise en oeuvre des quelques pistes que nous venons d'exposées permettra de coller à la sémantique exposée précédemment. Néanmoins, afin de faire en sorte que, par exemple, la réaction de l'agent soit en rapport avec ses préférences, il est nécessaire de contraindre un peu plus les formules du système. Pour se faire, il est nécessaire d'intégrer à la théorie des « hypothèses applicatives ». Ces dernières sont celles déjà imposées par Sadek, Bretier et Louis pour contraindre la théorie de l'interaction (voir [Louis, 2002, pp37-40]). Ce sont aussi des schémas de formules permettant de mettre en relation les préférences avec les autres attitudes mentales et, en particulier, des schémas permettant de faire le lien entre les préférences et la réaction. Plus précisément, il est nécessaire d'intégrer les schémas qui permettent de mettre en oeuvre le critère de décision retenue (voir la section 7.7). À ce sujet, si le critère de décision est le critère *NotDom*, il est nécessaire de modifier l'axiome (1.1) proposé par Louis. Plus précisément, comme les objets mis en relation par l'opérateur $>$ sont des formules du langage $\mathscr{L}$ et non des « descripteurs de plans », il est nécessaire de remplacer $b > a$ par $Agit(a) > Agit(b)$. C'est pourquoi l'axiome (1.1) est remplacé par

$$B\left[\begin{pmatrix} & CPossible(\phi) \wedge Faisable(a,\phi) \\ \wedge & CPossible(\psi) \wedge Faisable(b,\psi) \\ \wedge & (\exists e, e \neq <> \wedge Debute(e,a)) \\ \wedge & (\exists e, e \neq <> \wedge Debute(e,b)) \\ \wedge & (Agit(a) > Agit(b)) \end{pmatrix} \Rightarrow \neg Agit(b)\right] \tag{8.1}$$

Si une autre politique de choix est utilisée, il est nécessaire de remplacer ce schéma par d'autres. Par exemple s'il l'on veut mettre en oeuvre la politique de choix *Min*-, et si le prédicat $Min(\phi)$ indique que toutes les alternatives qui vérifient ϕ sont celles qui ont le moins d'alternatives qui leur sont préférées, alors il faut remplacer le schéma précédent par le suivant :

$$B\left[\left(\begin{array}{ll} & CPossible(\phi) \wedge Faisable(a,\phi) \\ \wedge & (\exists e, e \neq <> \wedge Debute(e,a)) \\ \wedge & (\neg Min(Agit(a)) \end{array}\right) \Rightarrow \neg Agit(a)\right] \tag{8.2}$$

Remarquons que la formule $Min(Agit(a))$ est aussi vérifiée lorsque pour toute réaction a on a $(Agit(a) < Agit(b))$.

8.1.4 Vers une autre extension possible

L'extension de notre formalisme que nous venons d'ébaucher permet d'exprimer par exemple qu'un agent i préfère qu'un agent j pense que la propriété ϕ est vérifiée plutôt que si l'action a est réaliser il pense que la propriété ψ est vérifiée : $B_i(B_j(\phi) \geq Faisable(a, B_j(\psi)))$. Ceci peut être utile si l'on veut par exemple exprimer le fait qu'un agent i préfère que l'agent j connaisse la vérité *maintenant*, plutôt qu'il ne l'apprenne plus tard.

Néanmoins, cette extension ne permet pas d'exprimer des préférences sur des préférences. Par exemple, il est impossible d'exprimer avec ce formalisme qu'un agent i préfère qu'un agent j préfère la propriété ϕ à la propriété ψ. plutôt que l'agent k préfère la propriété ϕ à la propriété ψ : $B_i(B_j(\phi \geq \psi) \geq B_k(\phi \geq \psi))$.

Afin d'obtenir plus d'expressivité, il serait intéressant de considérer l'extension de notre formalisme qui consisterait à prendre comme langage de base sur lequel est construit le langage $\mathscr{L}^3$, en plus du langage $\mathscr{L}$, le langage $\mathscr{L}^3$ lui même. Cette récursivité dans la définition permettrait d'exprimer des préférences sur des préférences. Néanmoins, il ne nous semble pas qu'une telle généralisation de notre formalisme puisse être effectuée aussi simplement que celle que nous avons proposée précédemment. En particulier, il nous semble que la définition de l'ensemble ALT n'est pas alors trivial. En effet, les modèles d'une telle logique seront alors sûrement composés, pour partie, des modèles de cette logique. Nous laissons donc cette voie ouverte pour de futurs travaux.

8.2 Utilisation conjointe de préférences et d'autres types de désirabilités

Dans cette section nous exposons une voie de recherche qui fut à l'origine de nos travaux. Celle-ci se base sur le constat que les données initiales permettant de décrire la *désirabilité* des différentes alternatives sont souvent de nature très différentes. En particulier, ces données ne se limitent pas uniquement à des préférences « classiques » telles que nous les avons décrites jusqu'ici. Aussi, au lieu de rechercher un formalisme pivot pour les traiter ensemble de la même manière, une voie intéressante est de les traiter séparément selon leurs diversités respectives, puis, de regrouper les différents résultats.

8.2.1 Origine du travail

Pour tout un chacun, les informations entrant en jeu lors d'une décision sont souvent potentiellement très nombreuses et de nature hétérogène. Pour s'en convaincre, il suffit de remarquer la diversité du vocabulaire employé dans de telles situations. Un rapide examen permet d'identifier entre autres « les coûts », « les intérêts », « les préférences » et « les obligations ».

Bien que cet aspect hétérogène des données initiales puisse apparaître comme une limitation dans une optique de développement (il implique en effet la gestion de cette hétérogénéité), nous pensons au contraire qu'il peut être un avantage. En effet, comme nous l'avons souligné dans la section 2.4.2, utiliser les mêmes moyens de représentation que ceux utilisés par l'être humain permet de spécifier plus intuitivement les informations pertinentes pour une décision. De plus, comme l'humain utilise des représentations compactes et adaptées à son raisonnement limité, le fait d'organiser naturellement les informations utiles pour une décision en différentes catégories tend à montrer que ces catégories ont des intérêts et des inconvénients différents au regard de la décision à prendre. Ainsi par exemple, l'utilisation d'*obligations* et d'*interdictions* permet de restreindre rapidement l'ensemble des alternatives. D'un autre coté, l'utilisation de *coûts* et d'*intérêts* associés à chaque alternative permet d'extraire de chacune d'elles un aspect qui peut être ordonné facilement. Enfin, l'utilisation de *comparaisons* permet de se focaliser sur un sous-ensemble des alternatives afin de les départager. Il en résulte que l'utilisation, via des traitements adaptés, de ces différents types d'informations permet de mettre en place une « rationalité limitée » [Simon, 1982] [Doyle, 1994].

De façon plus générale, une interaction avec un système est d'autant plus naturelle que le système est capable de traiter de façon adéquate les informations de désirabilité exprimées dans toute leur diversité par un utilisateur « moyen ». Cette vision des choses est dans la même lignée que celles des auteurs de [Domshlak, 2002] ou de [Ha, 2001] dans le sens où nous pensons aussi que développer une technique hybride peut permettre d'avancer dans la résolution de l'automatisation de la décision dans des situations de la vie courante et en particulier dans des applications de TALN.

Cette vision « complémentaire » des modes de représentation des informations pertinentes pour une décision est très répandue dans la littérature. Ainsi par exemple, l'introduction de nouvelles représentations pour baser un processus de décision est très souvent proposée comme un complément de ce qui se fait usuellement (en général la manipulation de valeurs numériques). C'est ainsi que, par exemple, Boutilier dans [Boutilier, 1994] justifie l'introduction de son formalisme en expliquant que les techniques qualitatives et quantitatives sont complémentaires.

8.2.2 Proposition : l'utilisation séparée des données hétérogènes

Afin d'utiliser chaque mode de représentation de la désirabilité pour profiter de leurs avantages et, dans un souci de modularité, nous proposons de traiter chacun de ces modes séparément (mais de façon adaptée) par des techniques différentes. Par la suite, le résultat de chaque technique permettant de traiter de façon adaptée le mode de représentation correspondant sera combiné aux autres afin de déterminer la réaction de l'agent. Nous proposons donc une approche de type « Compare and Agregate » et non « Agregate and Compare » (voir [Dubois et al., 2001] pour une formalisation des deux approches). Remarquons ici que cette approche est généralement utilisée dans les problèmes de décision sociale/multi-agents (voir par exemple [Lang, 2004]).

8.2.3 1^er^ problème : comment traiter chaque type d'information ?

Pour mettre en oeuvre notre vision, chaque type d'information doit être traité d'une manière adaptée. Par conséquent, l'un des premiers problèmes à résoudre est d'associer, à chaque type d'information pertinente pour la décision, un traitement adéquat. Par exemple, les *comparaisons* pourront être traitées avec le formalisme que nous avons proposé au travers des chapitres 5, 6 et 7. D'un autre côté, les *coûts* pourront être traités avec le formalisme de l'*utilité espérée* présenté dans la section 2.2.1. Enfin, [Demolombe et al., 2005] et [Stratulat, 2002] pourrons être par exemple regardés afin de traiter les obligations et les interdictions.

Nous ne voulons pas rentrer dans les détails d'une telle association. Notons toutefois que cette association devra être effectuée dans l'objectif d'une intégration à un formalisme d'agent rationnel modélisé cognitivement. Par conséquent, chacun de ces traitements devra permettre de (1) récupérer/spécifier intuitivement les informations nécessaires à son fonctionnement et de (2) donner un résultat pertinent pour la décision, c'est-à-dire de départager certaines alternatives mises en jeu explicitement (i.e. celles appartenant à l'ensemble ALT^*).

8.2.4 2^ème^ problème : comment agréger les différents résultats ?

Pour mettre en oeuvre notre vision, chaque type d'information pertinente pour la décision doit être traité séparément. Par conséquent, l'un des problèmes à résoudre est de déterminer la façon de regrouper les résultats du traitement de chacun des types d'information. Afin de clarifier notre propos nous considérons successivement les deux cas suivant :

- soit chaque résultat est un sous-ensemble des alternatives ALT,
- soit chaque résultat est une relation binaire sur l'ensemble des alternatives ALT.

a) Chaque résultat est un sous-ensemble des alternatives *ALT*

Lorsque le résultat est un sous-ensemble des alternatives *ALT*, ce dernier peut être interprété comme le fait que les alternatives retournées sont meilleures[2] (du point de vue des informations utilisées dans la technique considérée) que celles qui ne sont pas retournées. Afin de simplifier la combinaison des différentes techniques, nous proposons de les abstraire de la façon suivante :

> Chaque technique de choix est considérée comme un filtre F prenant en entrée un ensemble d'alternatives ALT et retournant un sous-ensemble noté $F(ALT)$.

Cette abstraction est en accord avec le fait mis en évidence par Simon selon lequel l'alternative choisie par un humain n'est pas forcément la meilleure, mais en est une qui est « suffisamment bonne » [Simon, 1960]. De plus, dans cette vision, l'emploi des informations sur la *désirabilité* des alternatives est similaire (dans l'objectif) à l'emploi préconisé des *intentions* dans les systèmes BDI [Bratman, 1987] : elles servent à élaguer l'ensemble des alternatives (voir le point (a) de la section 1.1.2).

Dans un telle situation, l'agrégation des différentes méthodes correspond à une combinaison des différents filtres. À titre d'exemple, nous présentons rapidement ici une tentative de mécanisme permettant de déterminer la réaction du système qui utilise trois types d'informations différents : des obligations sur les états résultant des réactions, les coûts de la mise en oeuvre des réactions et des préférences sur ces réactions (voir la figure 8.1). Ce mécanisme utilise ces informations en séquence de la façon suivante :

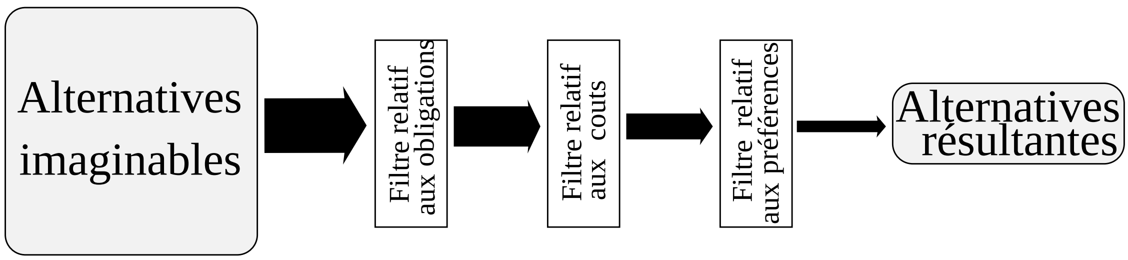


Fig. 8.1 – Un arrangement de trois mécanismes permettant de déterminer la réaction

1. La première étape a pour objectif d'écarter toutes les réactions qui sont « non acceptables ». Ceci est réalisé en filtrant l'ensemble des réactions du système afin d'exclure toutes celles dont l'état résultant est incompatible avec les obligations du système. Cette étape utilise uniquement les informations de désirabilité de type obligation.

[2]Notons qu'il peut y avoir plusieurs interprétations possibles de cette situation. Par exemple, soit chaque alternative de l'ensemble retourné est mieux que chaque alternative de l'ensemble qui n'est pas retourné, soit chaque alternative de l'ensemble retourné est mieux qu'au moins une alternative de l'ensemble qui n'est pas retourné, soit il existe une alternative de l'ensemble retourné qui est mieux que chaque alternative de l'ensemble qui n'est pas retourné, soit il existe une alternative de l'ensemble retourné qui est mieux qu'une alternative de l'ensemble qui n'est pas retourné. Il faudra choisir un interprétation !

2. La seconde étape a pour objectif de calculer le coût de chaque réaction puis de ne retenir que les alternatives « acceptables » qui ont le plus bas coût. Au lieu de prendre simplement une des réactions « acceptables » ayant le plus petit coût, nous sélectionnons l'ensemble des alternatives qui ont les coûts les plus bas. En pratique, cela est effectué en ayant au préalable fixé un coefficient γ : si C_{min} représente le coût minimum d'une l'alternative « acceptables » alors la réaction mise en oeuvre sera parmi toutes celles qui ont un coût inférieur à $(1+\gamma) \times C_{min}$.

   En effet, le bon sens nous indique que souvent, lorsque les coûts de deux réactions sont proches, les différences ne sont pas pertinentes pour le choix. Ceci est à rapprocher de la remarque faite sur Simon précédemment. Cette étape utilise uniquement les informations de désirabilité de type coût.

3. La troisième étape consiste, plus spécifiquement, à choisir la « meilleur » alternative. Cette dernière est une alternative « acceptable » qui a l'un des plus bas coûts et qui est préférée aux autres alternatives « acceptables » qui ont l'un des plus bas coûts. Cette étape utilise uniquement les informations de désirabilité de type préférence.

A priori, il y a une infinité de combinaisons possibles : les filtres peuvent être mis non seulement en séquence mais aussi en parallèle et être appelés plusieurs fois. Aussi, afin d'éclairer le choix de la façon de combiner les différentes méthodes, nous pensons qu'il est intéressant de répondre aux questions suivantes :

- existe t-il des filtres plus généraux que d'autre (i.e. tels que $F_1(ALT) \subset F_2(ALT)$) ?
- existe t-il des filtres contradictoires (i.e. tels que $F_1(ALT) \cap F_2(ALT) = \emptyset$) ?
- l'ensemble des filtres est-il contradictoire (i.e. tels que $F_1(ALT) \cap F_2(ALT) \cap \ldots \cap F_n(ALT) = \emptyset$) ?
- existe t-il des filtres contradictoires par composition (i.e. tels que $F_1(F_2(ALT) = \emptyset)$?
- existe t-il des filtres respectant la propriété d'indépendance vis-à-vis des autres alternatives ? (i.e. tels que si $ALT_1 \subset ALT$, alors $F(ALT_1) \subset F(ALT)$)
- existe t-il des filtres stables par composition ? (i.e. tels que $F(ALT) = F(F(ALT))$) (est-il intéressant d'utiliser un filtre plus d'une fois ?)
- quelle est la complexité des différents filtres par rapport au nombre d'alternatives à traiter ?

Si tous les filtres respectent la propriété d'indépendance vis-à-vis des autres alternatives, alors le problème s'apparente à un problème de satisfaction de contraintes[3] (abrégé par CSP en anglais) à une seule variable A (l'alternative à choisir), qui a pour domaine l'ensemble ALT et dont les contraintes sont de la forme « $A \in D_i(ALT)$ ». Ces contraintes indiquent que pour le filtre i, la variable A doit prendre une valeur parmi le résultat du filtre i. Nous pensons néanmoins que la majorité des filtres que nous pouvons imaginer intuitivement ne vérifient ni la propriété d'indépendance vis-à-vis des autres alternatives ni celle de stabilité par composition.

C'est pourquoi nous pensons qu'il serait aussi intéressant d'effectuer en complément une étude sociologique et psychologique afin d'identifier les pré-requis de chaque technique, leur

[3] voir par exemple [Kumar, 1992] ou plus récemment la première partie de [Batnini, 2005] pour un aperçu des problèmes de satisfaction de contraintes

résultats, leurs avantages et leurs inconvénients ainsi que leurs agencements. En particulier, elle pourrait permettre de déterminer la valeur des paramètres de chaque technique suivant les situations ainsi que leurs valeurs par défaut. En effet, dans notre vision l'utilisation des diverses informations est apparentée à des filtres. Par conséquent, ceux-ci pourront être configurés afin de retourner un sous-ensemble plus ou moins important. Par exemple, dans le mécanisme que nous venons de proposer, plus le coefficient γ est faible, plus le système accorde de l'importance à ses préférences et moins il en accorde aux coûts.

b) Chaque résultat est une relation binaire

Chaque technique permettant de traiter chacun des types d'informations initiales peut aussi être considérée comme un moyen de générer une relation binaire sur l'ensemble des alternatives « imaginables » ALT. Dans cette approche, chaque fois qu'un couple d'alternatives appartient à cette relation, la première alternative est meilleure (du point de vue des informations utilisées par la technique considérée) que la seconde.

Afin de regrouper ces diverses relations, il est nécessaire de déterminer une nouvelle phase d'agrégation. Pour se faire, les idées présentées dans le chapitre 7 pourront être reprises. En particulier, dans cette nouvelle phase chaque technique pourra être considérée comme un ensemble d'arguments pour ou contre le choix d'une alternative plutôt que d'une autre (voir le schéma 8.2).

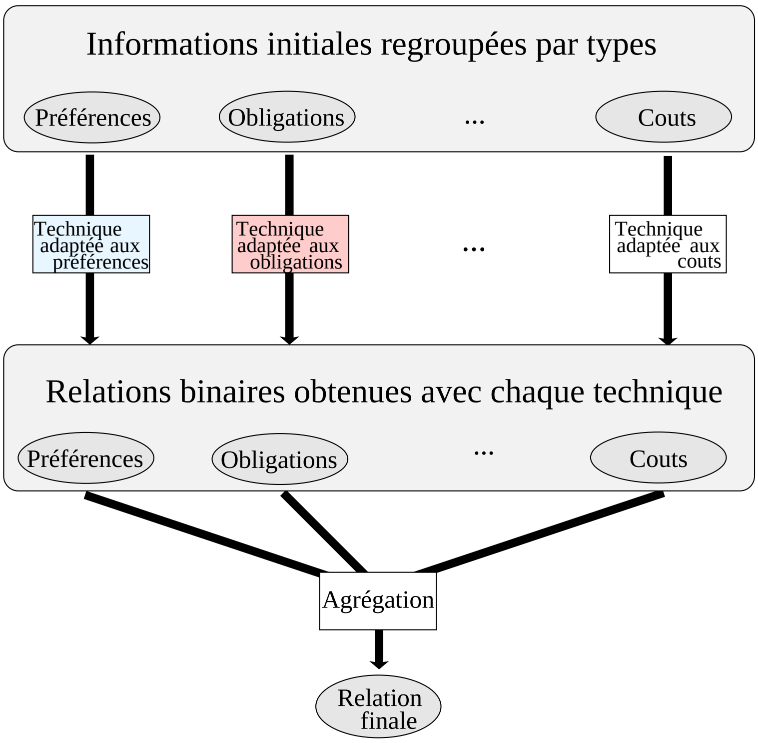


FIG. 8.2 – 1er schéma d'agrégation pour différents types d'informations de désirabilité

Le fait de considérer chaque technique comme un ensemble d'arguments permet d'envisager de reprendre la phase d'agrégation ébauchée dans le chapitre 7 mais aussi de la modifier afin d'intégrer les autres types informations. Plus précisément, il nous semble intéressant de considérer chaque technique comme un point de vue au même titre que les préférences partielles étendues. En effet, dans cette vision il n'y a plus qu'une seule phase d'agrégation (voir le schéma 8.3) puisque les préférences partielles étendues sont considérées comme toutes les autres techniques de traitement. De plus il devient alors possible et naturel de considérer plusieurs points de vue pour chacune des différents types de désirabilité (obligations, préférences, etc.). Enfin, cette vision généralise celle suivant laquelle chaque information sur la désirabilité est un argument (parmi d'autres) pour le résultat de la décision.

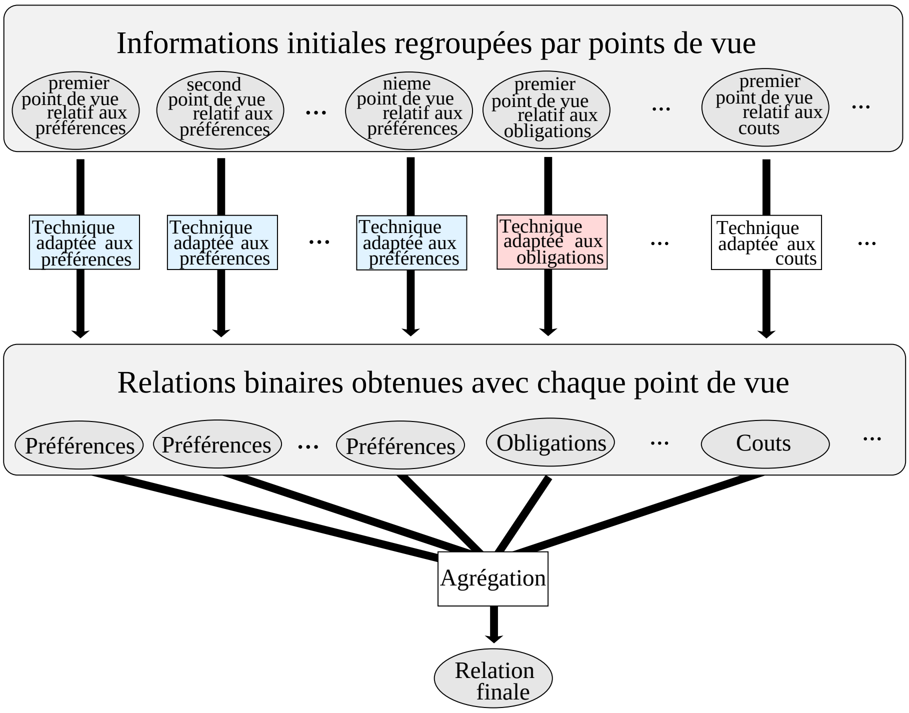


FIG. 8.3 – 2nd schéma d'agrégation pour différents types d'informations de désirabilité

8.2.5 Conclusion

Dans la littérature il existe de nombreux travaux cherchant à faire un lien entre les différents modes de représentation pouvant permettre de décrire la désirabilité des alternatives (voir la section 3.4.1). Il en résulte fréquemment que seul un mode de représentation est utilisé pour calculer la la réaction du système.

Cependant, comme chacun de ces modes a ses propres avantages, il est intéressant de les combiner. Cela peut permettre en particulier de manipuler dans chaque situation les informations nécessaires à la décision sous une forme intuitive. Ceci est particulièrement intéressant pour des systèmes de TALN puisque l'utilisation de formes intuitives (et pas seulement une seule) est un atout. En effet, ceci permet, par exemple, d'obtenir plus facilement les informations au cours d'une interaction.

Dans cette section, nous avons donc proposé pour notre part de considérer chacune de ces informations dans sa diversité et de la traiter comme telle. Plus précisément, nous avons proposé dans cette section deux voies pour le faire. La première voie consiste à utiliser le formalisme que nous avons exposée à travers les chapitres 5, 6 et 123 tel quel, et, par conséquent, nécessite de définir une deuxième phase d'agrégation pour prendre en compte les autres types d'informations sur la désirabilité. La seconde voie consiste à généraliser la phase d'agrégation proposée dans le chapitre 7 de ce document afin qu'elle prenne en compte, en plus des différentes préférences partielles, d'autres types d'informations sur la désirabilité des alternatives comme des obligations ou de coûts.

l

Chapitre 9

Conclusion générale du document

Dans ce chapitre nous présentons les conclusions générales de notre travail. Nous commençons par rappeler rapidement le contexte et la problématique générale. Par la suite, après avoir résumé notre proposition, nous exposons nos contributions. Enfin, nous dressons les perspectives à court et à long terme.

9.1 Contexte et problématique de la thèse

Le paradigme agent permet de représenter les systèmes informatiques complexes de façon abstraite avec des notions intuitives. Dans le cas qui nous intéresse, les agents constituant de tels systèmes sont formalisés à partir de notions primitives appelées *attitudes mentales* telles que la croyance, le désir, l'intention, etc. qui évoluent suivant des règles logiques [Rao and Georgeff, 1991b][Sadek, 1991]. Ces agents (rationnels) sont avant tout des entités qui perçoivent leurs environnements et agissent dessus suivant le principe de rationalité (c'est-à-dire visant à satisfaire ses intentions, en accord avec ses connaissances). Malheureusement, la mise ne œuvre de ce principe ne spécifie pas en général complètement leurs réactions. C'est pourquoi Louis a affirmé la nécessité de faire explicitement le lien entre l'état mental et la réaction de l'agent et entre autres proposé dans [Louis, 2002] d'introduire explicitement une phase de décision [Doyle and Thomason, 1999].

Comme nous l'avons remarqué dans la section 1.2.5, la phase de décision requiert la modélisation de la *désirabilité* de chaque alternative : les préférences de l'agent. Dans de nombreux domaines, une fonction d'utilité est souvent utilisée pour appréhender cette notion. C'est un outil très précis pour représenter et manipuler les connaissances sur la désirabilité des alternatives et permet, en plus, de prendre en compte la nature stochastique de l'environnement. Malheureusement, cet outil se base sur des représentations numériques qui souffrent d'inconvénients (difficulté pratique pour obtenir les informations nécessaires, précision souvent inutile, représentation non homogène avec la description interne d'un agent, représentation peu intuitive) qui le rendent peu adapté à une modélisation cognitive d'agents.

9.2 Notre proposition

Dans ce document, et afin de contourner ces limitations, nous avons défendu la thèse selon laquelle il est intéressant de formaliser qualitativement les informations décrivant la « désirabilité » des alternatives pour mettre en œuvre une phase de décision dans un agent rationnel modélisé cognitivement. Ce faisant nous avons proposé une formalisation logique de préférences qualitatives basée sur une logique des prédicats du premier ordre.

Afin de se rapprocher le plus possible du sens commun, nous avons supposé que les informations sur la désirabilité des alternatives sont des comparaisons qui vérifient les propriétés d'expansion, de transitivité, de *Ceteris Paribus*, qui sont spécifiées par points de vue éventuellement contradictoires, et telles que, par défaut, deux propriétés quelconques sont jugées indifférentes. Il en résulte que l'idée directrice de notre approche consiste à scinder la phase de spécification de la préférence en trois étapes. Dans un premier temps, les données initiales de chaque point de vue sont interprétées selon l'hypothèse *Ceteris Paribus*, ce qui génère, pour chaque point de vue i, une *préférence partielle primitive* (notée $\succeq_i$). Cette phase dite de génération permet d'exprimer de façon naturelle et intuitive les données initiales. Dans un deuxième temps, chaque préférence partielle primitive est étendue (au-delà de l'hypothèse *Ceteris Paribus*) en une *préférence partielle étendue* (notée $\geq_i$ lorsque le point de vue en question est le point de vue numéro i). Cette phase dite « d'extension » permet de départager un plus grand nombre d'alternatives. Dans un troisième temps, la *préférence globale* (notée $\geq_i$) est construite, à partir de toutes les préférences partielles étendues spécifiées en parallèle, via un mécanisme d'élection. Cette phase dite « d'agrégation » permet de gérer des contradictions entre points de vue.

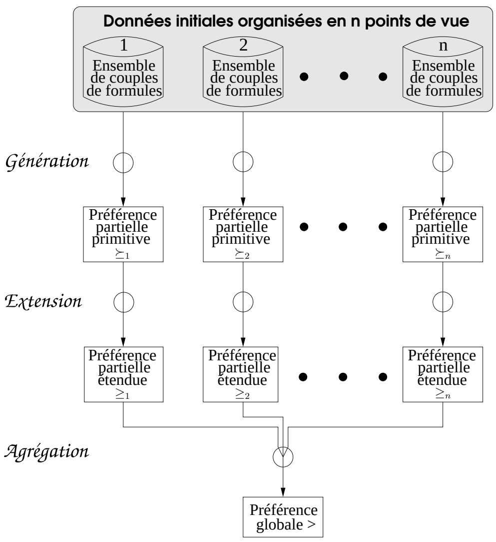


Schéma de construction des préférences

Dans notre approche, chaque préférence est modélisée par une relation binaire sur l'ensemble des alternatives. Dans le contexte des agents rationnels auquel nous nous intéressons, les alternatives correspondent aux états du monde (i.e. aux différents modèles de la logique que nous utilisons pour les formaliser). Or il est difficile voire impossible de représenter exhaustivement toutes les caractéristiques d'un état du monde (voir par exemple le problème du cadre [McCarthy and Hayes, 1969]). C'est pourquoi, notre modèle associe en outre à chaque préférence un opérateur logique modal portant sur des couples de formules décrivant partiellement les alternatives à comparer. En résumé, chaque préférence partielle est associée à la fois à (1) une relation binaire sur l'ensemble des alternatives et à (2) un opérateur modal sur des couples de formules logiques représentant les alternatives. Dans ce sens, notre vision est proche de celle de Von Wright, Boutillier, Doyle et Wellman, selon laquelle les préférences sont spécifiées par des descriptions mais s'appliquent à des alternatives « concrètes » qui vérifient ces descriptions. Par ailleurs, la phase d'agrégation de notre modèle adapte le travail de Rossi et al. qui consiste à dériver une décision à partir des préférences de plusieurs agents [Rossi et al., 2004]. Cette adaptation s'appuie sur l'idée développée dans [Dubois et al., 2001] selon laquelle la décision multi-critères et la décision multi-agents sont deux approches différentes de la même problématique. Chacune de nos *préférences partielles* joue ainsi le rôle de la préférence d'un agent dans le modèle proposé par Rossi et al. : chaque point de vue est un « argument » en faveur du classement qu'il propose.

9.3 Contributions

Dans ce document nous avons proposé une formalisation de la préférence adaptée aux modèles d'agents rationnels basés sur des attitudes mentales. Ce faisant, nous avons proposé, d'une part, des solutions originales pour contourner un inconvénient de l'hypothèse *Ceteris Paribus*[1] et pour mettre en œuvre le principe d'expansion[2] et d'autre part, une architecture de construction des préférences d'un agent en trois étapes qui a de nombreuses caractéristiques intéressantes. De plus, le formalisme que nous avons proposé peut a priori être adapté à de nombreuses logiques différentes.

Les mises en œuvre de l'hypothèse *Ceteris Paribus* sont souvent trop restrictives pour permettre de départager les alternatives parmi lesquelles la décision est explicitement à faire dans de nombreuses situations. La solution que nous proposons se base sur l'idée proposé dans [Meyer et al., 2005c] qui consiste à appréhender les informations sur la désirabilité des alternatives comme des arguments[3] pour comparer les alternatives entre elles. Plus précisément, dans ce document nous avons proposé de différencier la spécification des informations de désirabilité de leur utilisation. Un des intérêts de cette solution est que sa mise en œuvre ne nécessite pas la spécification d'informations supplémentaires comme une relation de normalité ou d'équivalence.

[1]Les préférences concernent des alternatives « toutes choses égale par ailleurs » (voir la section 2.3.1).

[2]Les préférences concernent des propriétés mutuellement exclusives (voir la section 2.3.1).

[3]Dans notre formalisation, le terme « argument » est utilisé pour son sens intuitif. Il ne réfère en aucun cas à une définition formelle telle que celles que l'on trouve dans les théories de l'argumentation.

Dans ce document, nous avons aussi proposé d'utiliser un opérateur de révision des croyances (via la notation $\phi+\overline{\psi}$) pour formaliser le principe d'expansion.

L'architecture de construction des préférences d'un agent en trois étapes que nous avons proposée dans [Meyer et al., 2005a] et exposée avec plus de détails dans ce document a entre autre les caractéristiques suivantes :

- Elle est adaptée aux modèles logiques d'agents cognitifs puisqu'elle leur est homogène : elle est aussi formalisée logiquement.
- Elle permet une spécification aisée et intuitive des informations de désirabilité nécessaires à la mise en oeuvre d'une phase de décision puisqu'elle (1) reprend une vision usuelle des préférences, (2) qu'elle ne nécessite pas une vue globale des préférences à chaque instant et (3) qu'elle est « robuste ».
- Elle offre un cadre formel pour gérer les contradictions qui peuvent exister entre les différents points de vue.
- Elle fait peu d'hypothèses sur la préférence globale d'un agent. En particulier, cette dernière n'est pas supposée être un pré-ordre total.

Le fait d'utiliser une logique expressive (logique des prédicats du premier ordre puis logique modale) permet de prendre en compte pour la décision ne nombreuses informations. En particulier, ces informations peuvent concerner la façon de mettre en œuvre les réactions disponibles, le résultat de leur mise en œuvre, ainsi que les informations elles même (par exemple la qualité des prédictions). Nous pensons de plus que la formalisation que nous avons proposée peut être adaptée à de nombreuses autres logiques puisque qu'elle base le sens des préférences entre alternatives sur les modèles de la logique permettant de décrire ces alternatives.

9.4 Perspectives

Les perspectives de nos travaux sont de différentes natures. À court terme tout d'abord, il nous semble important de poursuivre les deux extensions que nous avons proposées dans le chapitre 8. En particulier, en ce qui concerne l'intégration effective de notre formalisme à une logique modale, il faudra adapter à une telle logique les ensembles subf*(ϕ) et $\mathscr{L}^{\phi\psi}$ introduits dans la section 5.1.1. En ce qui concerne l'extension envisagée dans la section 8.2 visant à utiliser d'autres types d'informations de désirabilité, il faudra bien entendu formaliser de façon homogène les différentes techniques permettant de manipuler ces divers types de désirabilité.

Il serait aussi sûrement intéressant de modifier quelque peu l'axiomatique que nous avons proposée tout au long de ce document afin qu'elle ai un aspect « plus classique ». Plus précisément, il faudrait transformer les règles (5.18) (5.19) (5.20) (5.21) (5.22) (5.23) (5.24) (5.25) (6.14) (6.15) (6.17) (7.13) (7.14) (7.16) (7.16)(7.18) (7.19) (7.20)...(7.21) (7.23) (7.24)...(7.25) que nous avons introduites par un ensemble d'axiomes et un ensemble limité de règles de déduction (style *modus ponens*). Au sujet de cette axiomatique, il faudrait aussi montrer si elle est complète et, si ce n'est pas le cas, chercher à l'améliorer pour que ce le soit.

Enfin, dans la perspective d'utiliser ce formalisme dans un programme, il faudra proposer des algorithmes adaptés et étudier leurs complexités.

À plus long terme, plusieurs pistes de recherche sont aussi envisageables pour prolonger nos travaux. Il s'agira tout d'abord de définir des politiques ainsi que des critères de choix qui soient adaptés à des situations données et en particulier pour des agents dialoguants (comme par exemple le choix d'une modalité de communication).

Une autre voie qui nous semble aujourd'hui très prometteuse serait d'étudier la modélisation de l'hypothèse *Ceteris Paribus* ainsi que la gestion de points de vue avec des techniques issues de théorie de l'argumentation (voir à ce sujet [Morge, 2005, chapitre 2] pour un tour d'horizon). Dans ce domaine, Amgoud et Cayrol ont proposé dans [Amgoud and Cayrol, 2002] de structurer l'ensemble des arguments par une relation d'ordre stricte (appelée préférence). Pourquoi donc ne pas étendre leurs travaux avec une relation moins contraignante que celle qui y est utilisée et en particulier avec la préférence globale de notre formalisme ?

Ce travail a été réalisé dans l'optique de fournir à un système de dialogue les outils permettant la mise en oeuvre d'une phase de décision afin de favoriser l'interaction entre le système et un utilisateur. Ce faisant, nous nous somme focalisés sur le traitement de la notion de préférence de façon autonome c'est-à-dire sans interaction avec la personne spécifiant des préférences. Pourtant, si le système est utilisé comme un expert, ce dernier a tout intérêt à utiliser ces interactions pour impliquer l'utilisateur dans une prise de décision. C'est pourquoi, il serait intéressant d'inverser la problématique et ainsi d'étudier la façon d'utiliser ces interactions afin, entre autres, d'éliciter les préférences de l'utilsateur. Pour ce faire, les méthodes « itératives », qui sont très prisées dans le domaine de l'aide à la décision (voir [Vincke, 1989] et [Bouyssou et al., 2002]), pourraient servir d'inspiration puisqu'elles nécessitent une participation active d'un décideur de référence durant la décision.

Références

[Albouy, 2004] Albouy, D. (2004). Preference relations, social decision rules, single-peakedness, and social welfare functions. http://emlab.berkeley.edu/users/webfac/saez/e131_s04/prefer.pdf. Part of the "Economics 131" course at the university of California, USA.
(cité à la page 201)

[Allais, 1953] Allais, M. (1953). Le comportement de l'homme rationnnel devant le risque : Critique des postulats et axiomes de l'école americaine. *Econometrica*, 21 :503–546.
(cité à la page 59)

[Allen et al., 1991] Allen, J. F., Kautz, H. A., Pelavin, R. N., and Tenenberg, J. D. (1991). *Reasoning about plan.* Morgan Kaufmann.
(cité aux pages ii, 25)

[Amgoud and Cayrol, 2002] Amgoud, L. and Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Maths and AI*, 34(1-3) :197–215.
(cité à la page 161)

[Amgoud et al., 2003] Amgoud, L., Herzig, A., and Mercier, D. (2003). Calcul des intentions d'agent à partir de ses désirs. In Andréas Herzig, Brahim chaib-draa, p. M., editor, *proceedings of MFI'03, Modèles formels de l'interaction*, pages 3–10, Lille, France. Cepadues edition.
(cité aux pages 5, 8)

[Arrow, 1951] Arrow, K. (1951). *Social Choice and Individual Values.* Wiley, New York, USA.
(cité aux pages 71, 133, 201)

[Austin, 1962] Austin, J. L. (1962). *How to Do Things with Words.* Harvard University Press, Cambridge, MA, USA.
(cité à la page 12)

[Batnini, 2005] Batnini, H. (2005). *Contraintes Globales et Techniques de Résolution pour les CSPs Continus.* PhD thesis, Université de Nice Sophia-antipolis, France.
(cité à la page 153)

[Benferhat et al., 2004] Benferhat, S., Brewka, G., and Berre, D. L. (2004). On the relation between qualitative choice logic and possibilistic logic. In *proceedings of IPMU 04, Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Perugia, Italia.
(cité à la page 61)

[Black, 1948] Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, 56 :23–34.
(cité à la page 201)

[Boutilier, 1994] Boutilier, C. (1994). Toward a logic for qualitative decision theory. In Doyle, J., Sandewall, E., and Torasso, P., editors, *Principles of Knowledge Representation and Reasoning : proceedings of the 14th International Conference (KR'94)*, pages 75–86. Kaufmann, San Francisco, CA.
(cité aux pages iii, 42, 44, 50, 52, 53, 61, 70, 73, 74, 121, 150, 182)

[Boutilier et al., 2004a] Boutilier, C., Brafman, R., Domshlak, C., Hoos, H., and Poole, D. (2004a). Cp-nets : A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research (JAIR)*, 21 :135–191.
(cité à la page 53)

[Boutilier et al., 2004b] Boutilier, C., Brafman, R., Domshlak, C., Poole, D., and Hoos, H. H. (2004b). Preference-based constraint optimization with CP-nets. *Computational Intelligence*, 20(2) :137–157.
(cité aux pages 44, 53, 59)

[Boutilier et al., 1999] Boutilier, C., Brafman, R., Hoos, H. H., and Poole, D. (1999). Reasoning With Conditional Ceteris Paribus Preference Statements. In *proceedings of UAI'99 :the 15th Annual Conference on Uncertainty in Artificial Intelligence*, pages 71–80, Stockholm, Swedeen.
(cité aux pages 42, 53, 54, 61, 74, 182)

[Bouyssou et al., 2002] Bouyssou, D., Jacquet-Lagrèze, E., Perny, P., Slowinski, R., Vanderpooten, D., and Vincke, P. (2002). *Aiding Decisions with Multiple Criteria : Essays in Honour of Bernard Roy.* Kluwer Academic, Dordrecht.
(cité aux pages 129, 161)

[Bouyssou and Vincke, 2003] Bouyssou, D. and Vincke, P. (2003). Relations binaires et modélisation des préférences. Technical Report 208, Laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision (LAMSADE). Cet article est le premier chapitre d'un ouvrage collectif à paraître dans la série IC2 publiée par Hermès : Concepts et Méthodes pour l'Aide à la Décision, D. Bouyssou, D. Dubois, M. Pirlot et H. Prade.
(cité aux pages 32, 36, 71, 200)

[Brafman and Tennenholtz, 1997] Brafman, R. and Tennenholtz, M. (1997). Modeling agents as qualitative decision makers. *Artificial Intelligence*, 94(1) :217–268.
(cité aux pages 24, 59, 60, 73)

[Bratman, 1987] Bratman, M. (1987). *Intentions, Plans, and Practical Reason.* Harvard University Press, Cambridge, MA, USA.
(cité aux pages 5, 7, 8, 20, 23, 65, 152)

[Bratman et al., 1988] Bratman, M., Israel, D., and Pollack, M. (1988). Plans and resource-bounded practical reasoning. In Cummins, R. and Pollock, J. L., editors, *Philosophy and AI : Essays at the Interface*, pages 1–22. The MIT Press, Cambridge, Massachusetts.
(cité aux pages 5, 9, 24)

[Brazier et al., 1996] Brazier, F., Dunin-Keplicz, B., Treur, J., and Verbrugge, R. (1996). Beliefs, intentions and desire. In *proceedings of KAW'96 : the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Alberta, Canada.
(cité à la page 5)

[Bretier, 1995] Bretier, P. (1995). *La communication orale coopérative : Contribution à la modélisation logique et à la mise en oeuvre d'un agent rationnel dialoguant.* PhD thesis, Université de Paris Nord, France.
(cité aux pages ii, 3, 11, 12, 15, 16, 18)

[Brewka et al., 2004] Brewka, G., Benferhat, S., and Le Berre, D. (2004). Qualitative choice logic. *Artificial Intelligence*, 157 :203–237.
(cité à la page 58)

[Briot and Gasser, 1998] Briot, J. and Gasser, L. (1998). Actors & agents : Agents and concurrent objects. *IEEE Concurrency*, 6(4) :74–77, 81.
(cité à la page 4)

[Burkhard et al., 1998] Burkhard, H.-D., Hannebauer, M., and Wendler, J. (1998). Belief-desir-intention deliberation in artificial soccer. *AI magazine*.
(cité à la page 5)

[Chiclana et al., 1996] Chiclana, F., Herrera-viedma, E., and Herrera, F. (1996). Preference relations as the information representation base in multi-person decision making. In *proceedings of IPMU'96*, volume 1, pages 459–464. 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems.
(cité à la page 58)

[Cohen and Levesque, 1990a] Cohen, P. R. and Levesque, H. J. (1990a). Intention is choice with commitment. *Artificial Intelligence, AI*, 42(2-3) :213–261.
(cité aux pages 12, 16)

[Cohen and Levesque, 1990b] Cohen, P. R. and Levesque, H. J. (1990b). Persistence, intention, and commitment. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 33–69. MIT Press, Cambridge, MA, USA.
(cité aux pages 5, 10)

[Coste-Marquis et al., 2004] Coste-Marquis, S., Lang, J., Liberatore, P., and Marquis, P. (2004). Expressive power and succinctness of propositional languages for preference representation. In *proceedings of KR'04*, pages 203–212, Whistler, Canada.
(cité aux pages iii, 61, 69)

[Dastani et al., 2002a] Dastani, M., Boer, F., Dignum, F., van der Hoek, W., Meindert, K., and Meyer, J.-J. (2002a). Programming the deliberation cycle of cognitive robots. In *proceedings of the 3rd International Cognitive Robotics Workshop*, Edmonton, Canada. AAAI'2002.
(cité à la page 5)

[Dastani et al., 2002b] Dastani, M., Huang, Z., and van der Torre, L. (2002b). *Dynamic desires*, volume 5 of *Multiagent systems, artificial societies and simulated organizations*, chapter 1. Kluwer.
(cité aux pages 29, 73)

[Demolombe et al., 2005] Demolombe, R., Bretier, P., and Louis, V. (2005). Formalisation de l'obligation de faire avec delais. In Herzig, A., Lespérance, Y., and Mouaddib, A.-I., editors, *proceedings of MFI'05 : Modèles formels de l'interaction, Actes des troisièmes journées francophones*, Caen, France. Cépaduès.
(cité à la page 151)

[D'Inverno et al., 1997] D'Inverno, M., Kinny, D., Luck, M., and Wooldridge, M. (1997). A formal specification of dmars. In Springer-Verlag, editor, *proceedings of ATAL'97 :*

the 4th International Workshop on Intelligent Agents IV, Agent Theories, Architectures, and Languages, volume 1365 of *Lecture Notes In Computer Science*, pages 155 – 176, London, UK. Michael Wooldridge.
(cité à la page 8)

[D'Inverno et al., 2004] D'Inverno, M., Luck, M., Georgeff, M. P., Kinny, D., and Wooldridge, M. (2004). The dmars architecture : A specification of the distributed multi-agent reasoning system. *Autonomous Agents and Multi-Agent Systems*, 9(1-2) :5–53.
(cité à la page 8)

[Domshlak, 2002] Domshlak, C. (2002). *Modeling and Reasoning about Preferences with CP-nets*. PhD thesis, Ben-Gurion University, Israel.
(cité aux pages 31, 53, 61, 66, 70, 150)

[Doyle, 1994] Doyle, J. (1994). Rationality and its roles in reasoning. *Computational Intelligence*, 8(2) :376–409.
(cité aux pages 22, 28, 29, 31, 59, 150)

[Doyle et al., 1991] Doyle, J., Shoham, Y., and Wellman, M. (1991). A logic of relative desire. In Ras, Z. W. and Zemankova, M., editors, *proceedings of ISMIS'91 : the 6th International Synmposium on Methodologies for Intelligent Systems*, pages 16–31, Berlin, Germany. Springer-Verlag.
(cité aux pages 48, 50, 53)

[Doyle and Thomason, 1999] Doyle, J. and Thomason, R. H. (1999). Background to qualitative decision theory. *AI Magazine*, 20(2) :55–68.
(cité aux pages ii, 25, 61, 157)

[Doyle and Wellman, 1994] Doyle, J. and Wellman, M. (1994). Representing preferences as ceteris paribus comparatives. In *proceedings of the AAAI Symposium on Decision-Theoretic Planning*, Stanford, CA, USA. AAAI press.
(cité aux pages iii, 42, 44, 48, 53, 58, 59, 61, 70, 74, 121, 182)

[Druzdzel and Flynn, 2002] Druzdzel, M. and Flynn, R. (2002). Decision support systems. In Kent, A., editor, *Encyclopedia of Library and Information Science*. Marcel Dekker, New York, USA, second edition.
(cité à la page 58)

[Dubois et al., 2001] Dubois, D., Fargier, H., Perny, P., and Prade, H. (2001). Towards a qualitative multicriteria decision theory. In *proceedings of the Eurofuse Workshop on Preference Modeling and Applications*, pages 121–129, Granada, Spain.
(cité aux pages 58, 60, 61, 73, 124, 134, 151, 159)

[Dubois et al., 2005] Dubois, D., Kaci, S., and Prade, H. (2005). Cp-nets and possibilistic logic : Two approaches to preference modeling , steps towards a comparison. In Brafman, R. and Junker, U., editors, *proceedings of PREF'05 : the IJCAI'05 Multidisciplinary Workshop on Advances in Preference Handling*, Edinburgh, Scotland.
(cité à la page 58)

[Dubois et al., 1994] Dubois, D., Lang, J., and Prade, H. (1994). *Handbook of Logic in Artificial Intelligence and Logic Programming*, chapter Possibilistic logic, pages 439–513.

Oxford University Press, 3 edition.
(cité à la page 58)

[Dubois and Prade, 2004] Dubois, D. and Prade, H. (2004). On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142 :143–161.
(cité à la page 129)

[Eberhard, 1972] Eberhard, W. (1972). Field research on complex decision-making processes - the phase theorem. *International Studies of Management and Organization*, pages 156–182.
(cité à la page 32)

[Emerson and Srinivasan, 1989] Emerson, E. and Srinivasan, J. (1989). Branching time temporal logic. In de Bakker, J. W., de Roever, W. P., and Rozenberg, G., editors, *proceedings of the REX School/Workshop on Linear Time, Branching Time, and Partial Order in Logics and Models for Concurrency Noordwijkerhout, The Netherlands, May 30 - June 3, 1988*, volume 354 of *Lecture Notes in Computer Science*, pages 123–172. Springer.
(cité à la page 9)

[Everaere et al., 2005] Everaere, P., Konieczny, S., and Marquis, P. (2005). Quota and gmin merging operators. In *proceedings of IJCAI'05*.
(cité à la page 129)

[Faltings et al., 2004] Faltings, B., Torrens, M., and Pu, P. (2004). Solution generation with qualitative models of preferences. *Computational Intelligence*, 20(2) :246–263.
(cité aux pages 59, 61)

[Fishburn, 1970] Fishburn, P. (1970). *Utility theory for decision making.* John Wiley & Sons, New-York, USA.
(cité à la page 32)

[Fishburn, 1999] Fishburn, P. (1999). Preference structures and their numerical representations. *Theoretical Computer Science*, 217(2) :359–383.
(cité aux pages 32, 36)

[Franklin and Graesser, 1996] Franklin, S. and Graesser, A. (1996). Is it an agent, or just a program ? : A taxonomy for autonomous agents. In *Intelligent Agents III, proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL'96)*, volume 1193, Berlin, Germany. Springer Verlag.
(cité à la page 4)

[Georgeff et al., 1999] Georgeff, M., Pell, B., Pollack, M., Tambe, M., and Wooldridge, M. (1999). The belief-desire-intention model of agency. In Müller, J., Singh, M. P., and Rao, A. S., editors, *proceedings of ATAL'98*, volume 1555 of *LNAI*, pages 1–10, Berlin, Germany. Springer.
(cité aux pages 5, 6, 24)

[Georgeff and Ingrand, 1989] Georgeff, M. P. and Ingrand, F. F. (1989). Decision-making in an embedded reasoning system. In *proceedings of IJCAI'89*, pages 972–978, Detroit,

Michigan, USA.
(cité à la page 8)

[Ha, 2001] Ha, V. (2001). *Reasoning with Partial Preference Models*. PhD thesis, University of Wisconsin-Milwaukee, USA.
(cité aux pages 31, 40, 41, 61, 150)

[Ha and Haddawy, 1999] Ha, V. and Haddawy, P. (1999). A hybrid approach to reasoning with partial preference models. In *proceedings of UAI'99*, pages 263–270, Stockholm, Sweden.
(cité aux pages 59, 61)

[Haddawy and Hanks, 1993] Haddawy, P. and Hanks, S. (1993). Utility models for goal-directed decision theoretic planners. Technical Report TR-93-06-04, University of Washington - Department of Computer Science & Engineering, USA.
(cité aux pages ii, 29, 66, 73)

[Halldén, 1957] Halldén, S. (1957). *On the Logic of Better*. Library of Theoria, Lund.
(cité à la page 43)

[Hansson, 1994] Hansson, S. (1994). Decision theory : a brief introduction. http://www.infra.kth.se/~soh/decisiontheory.pdf. This text is a non-technical overview of modern decision theory. It is intended for university students with no previous acquaintance with the subject, and was primarily written for the participants of a course on risk analysis at Uppsala University in 1994.
(cité à la page 31)

[Hansson, 1996] Hansson, S. (1996). What is ceteris paribus preference. *Journal of Philosophical Logic*, 25(3) :307–332.
(cité aux pages 43, 44)

[Harel, 1984] Harel, D. (1984). Dynamic logic. In Gabbay, D. and Guenther, F., editors, *Handbook of Philosophical Logic Volume II — Extensions of Classical Logic*, pages 497–604. D. Reidel Publishing Company : Dordrecht, The Netherlands.
(cité à la page 14)

[Harsanyi, 1955] Harsanyi, J. C. (1955). Cardinal welfare, individual ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63 :309–321.
(cité à la page 201)

[Herrera et al., 2001] Herrera, F., Herrera-Viedma, E., and Chiclana, F. (2001). Multiperson decision-making based on multiplicative preference relations. *European journal of operational research*, 129 :372–385.
(cité à la page 58)

[Herzig, 2005] Herzig, A. (2005). Introduction à la logique, un hyper-cours de répétition. http://www.irit.fr/ACTIVITES/LILaC/Pers/Herzig/C/index.html. En collaboration avec Gabriella Crocco, Olivier Gasquet et Bruno Gaume.
(cité à la page 77)

[Horty and Pollack, 1998] Horty, J. and Pollack, M. (1998). Option evaluation in context. In Gilboa, I., editor, *proceedings of TARK-VII : the 7th conference on Theoretical Aspects of Rationality and Knowledge*, pages 249–262, Evanston, Illinois, USA.
(cité aux pages 24, 73)

[Horvitz et al., 1988] Horvitz, E., Breese, J., and Henrion, M. (1988). Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2(special issue on Uncertainty in Artificial Intelligence) :247–308.
(cité à la page 24)

[Howden et al., 2001] Howden, N., Rönnquist, R., Hodgson, A., and Lucas, A. (2001). JACK - summary of an agent infrastructure. In *proceedings of the 5th International Conference on Autonomous Agents*.
(cité à la page 8)

[Huber, 1999] Huber, M. J. (1999). JAM : a BDI-theoretic mobile agent architecture. In *proceedings of AGENTS'99 : the 3rd annual conference on Autonomous Agents*, pages 236–243, New York, NY, USA. ACM Press.
(cité à la page 8)

[Ingrand et al., 1992] Ingrand, F. F., Georgeff, M. P., and Rao., A. S. (1992). An architecture for real-time reasoning and system control. *IEEE Expert*, 7(6) :34–44.
(cité à la page 8)

[Jennings and Wooldridge, 1998] Jennings, N. and Wooldridge, M. (1998). Applications of intelligent agents. In Jennings, N. R. and Wooldridge, M. J., editors, *Agent Technology : Foundations, Applications, and Markets*, pages 3–28. Springer-Verlag : Heidelberg, Germany.
(cité à la page 4)

[Kaci and van der Torre, 2005] Kaci, S. and van der Torre, L. (2005). Non-monotonic reasoning with various kinds of preferences. In Brafman, R. and Junker, U., editors, *proceedings of PREF'05 : the IJCAI'05 Multidisciplinary Workshop on Advances in Preference Handling*, pages 111–117, Edinburgh, Scotland.
(cité aux pages 58, 61, 144)

[Kast, 2002] Kast, R. (2002). *THEORIE DE LA DECISION*. La Découverte, Paris, France, 2nd edition.
(cité aux pages 25, 31, 39)

[Katoen, 2004] Katoen, J.-P. (2004). Principles of model checking. /www.iro.umontreal.ca/~vachon/Ift6222/Documentation/katoenLectureNotesl.ps. Lecture Notes of the System validation course (214012), the University of Twente, The Netherlands.
(cité à la page 9)

[Katsuno and Mendelzon, 1991] Katsuno, H. and Mendelzon, A. (1991). On the difference between updating a knowledge base and revising it. In Allen, J. F., Fikes, R., and Sandewall, E., editors, *proceedings of KR'91 : Principles of Knowledge Representation and Reasoning*, pages 387–394, Boston, MA, USA. Morgan Kaufmann.
(cité à la page 86)

[Keeney and Raiffa, 1976] Keeney, R. and Raiffa, H. (1976). *Decisions with multiple objectives : Preferences and value tradeoffs*. J. Wiley, New York.
(cité aux pages 36, 40, 42, 44, 54, 59)

[Konolige and Pollack, 1993] Konolige, K. and Pollack, M. (1993). A representationalist theory of intention. In *proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chambéry, France.
(cité à la page 5)

[Kripke, 1963] Kripke, S. A. (1963). *Semantical considerations on modal logic*, volume 16. Acta philosophica Fennica.
(cité aux pages 9, 144)

[Kumar, 1992] Kumar, V. (1992). Algorithms for constraint-satisfaction problems : A survey. *AI Magazine*, 13(1) :32–44.
(cité à la page 153)

[Lang, 2003] Lang, J. (2003). Contributions à l'étude de modèles, de langages et d'algorithmes pour le raisonnement et la prise de décision en intelligence artificielle. Mémoire présenté en vue de l'obtention de l'habiltation à diriger des recherches (HDR).
(cité aux pages 44, 74)

[Lang, 2004] Lang, J. (2004). Logical preference representation and combinatorial vote. *Annals of Mathematics and Artificial Intelligence*, 42 :37–71.
(cité aux pages iii, 57, 61, 151)

[Lang et al., 2002] Lang, J., van der Torre, L., and Weydert, E. (2002). Utilitarian desires. *International Journal of Autonomous Agents and Multi-Agent Systems*, 5 :329–363.
(cité à la page 29)

[Louis, 2002] Louis, V. (2002). *Conception et mise en oeuvre de modèles formels de calcul de plans d'action complexes par un agent rationnel dialoguant*. PhD thesis, Université de Caen/Basse-Normandie, France.
(cité aux pages ii, 3, 8, 11, 12, 13, 15, 16, 17, 18, 20, 24, 25, 28, 31, 66, 136, 145, 148, 157, 181)

[May, 1952] May, K. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20 :680–684.
(cité à la page 201)

[McCarthy, 1979] McCarthy, J. (1979). Ascribing mental qualities to machines. In Ringle, M., editor, *Philosophical Perspectives in Artificial Intelligence*, pages 161–195. Humanities Press, Atlantic Highlands, New Jersey, USA.
(cité aux pages 3, 5)

[McCarthy and Hayes, 1969] McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press.
(cité aux pages 3, 5, 72, 159)

[McGeachie and Doyle, 2002] McGeachie, M. and Doyle, J. (2002). Efficient utility functions for ceteris paribus preferences. In *proceedings of AAAI'02*, Edmonton, Alberta,

Canada.
(cité à la page 61)

[Meyer et al., 2005a] Meyer, G., Louis, V., and Sansonnet, J.-P. (2005a). Un opérateur de préférence adapté aux agents dialoguant. In Herzig, A., Lespérance, Y., and Mouaddib, A.-I., editors, *proceedings of MFI'05 : Modèles formels de l'interaction, Actes des troisièmes journées francophones*, pages 163–172, Caen, France. Cépaduès.
(cité aux pages 74, 106, 160)

[Meyer et al., 2005b] Meyer, G., Louis, V., Sansonnet, J.-P., and Larvor, Y. (2005b). Deux opérateurs logiques adaptés à la représentation et à la manipulation des préférences. In Guéré, E., editor, *proceedings of RJCIA'05 : Septièmes Rencontres nationales des Jeunes Chercheurs en Intelligence Artificielle*, pages 281–294. PUG.
(cité à la page 106)

[Meyer et al., 2005c] Meyer, G., Louis, V., Sansonnet, J.-P., and Larvor, Y. (2005c). Two logical operators for representing and handling preferences. In Ronen I.Brafman and Junker, U., editors, *proceedings of PREF'05 : the IJCAI'05 Multidisciplinary Workshop on Advances in Preference Handling*, pages 140–145, Edinburgh, Scotland.
(cité aux pages 106, 159)

[Mintzberg et al., 1976] Mintzberg, H., Raisinghani, D., and Théoret, A. (1976). The structure of "unstructured" decision processes. *Administrative Science Quarterly - empirical study of 25 decision processes*, 21 :246–275.
(cité aux pages 32, 33)

[Morge, 2005] Morge, M. (2005). *Systéme dialectique multi-agents pour l'aide à la concertation*. PhD thesis, Ecole Nationale Supéerieure des Mines de Saint Etienne, France.
(cité à la page 161)

[Mousseau, 2003] Mousseau, V. (2003). Elicitation des préférences pour l'aide multicritère à la décision. http://www.lamsade.dauphine.fr/~mousseau/hdr-main.pdf. Mémoire présenté en vue de l'obtention de l'habiltation à diriger des recherches (HDR).
(cité aux pages 36, 58, 61, 75)

[Newell, 1980] Newell, A. (1980). The knowledge level. *AI Magazine*, 2(33) :1–20.
(cité aux pages 5, 22)

[Padgham and Lambrix, 2000] Padgham, L. and Lambrix, P. (2000). Agent capabilities : Extending BDI theory. In *proceedings of AAAI/IAAI*, pages 68–73, Menlo Park, CA, USA. AAAI Press.
(cité à la page 5)

[Pokahr et al., 2003] Pokahr, A., Braubach, L., and Lamersdorf, W. (2003). Jadex : Implementing a BDI-infrastructure for JADE agents. *exp : in search of innovation*, 3(3) :76–85.
(cité à la page 8)

[Pratt, 1976] Pratt, V. R. (1976). Semantical considerations on floyd-hoare logic. In *proceedings of the IEEE Symposium on Foundations of Computer Science*, pages 109–121.
(cité à la page 14)

[Rao and Georgeff, 1991a] Rao, A. S. and Georgeff, M. P. (1991a). Deliberation and intentions. In *proceedings of UAI'91 : the 7th Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA, USA.
(cité aux pages 6, 9)

[Rao and Georgeff, 1991b] Rao, A. S. and Georgeff, M. P. (1991b). Modeling rational agents within a BDI-architecture. In *proceedings of KR'91*, pages 473–484, San Mateo, CA, USA.
(cité aux pages ii, 3, 5, 9, 10, 11, 14, 20, 31, 157)

[Rao and Georgeff, 1995] Rao, A. S. and Georgeff, M. P. (1995). BDI-agents : from theory to practice. In *proceedings of the 1st Intl. Conference on Multiagent Systems*, San Francisco, Ca, USA.
(cité aux pages 6, 9, 11)

[Reiter, 1980] Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13 :81–132.
(cité à la page 26)

[Rossi et al., 2004] Rossi, F., Venable, K. B., and Walsh, T. (2004). mCP nets : representing and reasoning with preferences of multiple agents. In *proceedings of AAAI'04*, San Jose, CA, USA.
(cité aux pages 53, 58, 60, 61, 73, 124, 159)

[Roubens and Vincke, 1985] Roubens, M. and Vincke, P. (1985). *Preference Modeling*. Springer Verlag, Berlin. Lecture Notes in Economics and Mathematical Systems (LNEMS 250).
(cité aux pages 36, 88, 104)

[Russell and Norvig, 2003] Russell, S. and Norvig, P. (2003). *Artificial Intelligence : A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition edition.
(cité aux pages 4, 77)

[Sadek, 1991] Sadek, D. (1991). *Attitudes Mentales et Interaction Rationnelle : vers une théorie formelle de la Communication*. PhD thesis, Université de Rennes I, France.
(cité aux pages ii, 3, 4, 5, 11, 12, 13, 15, 16, 18, 20, 31, 66, 67, 143, 144, 157)

[Sadek, 1992] Sadek, D. (1992). A study in the logic of intention. In *proceedings of KR'92 : the 3rd conference on principles of Knowledge Representation and Reasoning*, pages 462–473, Cambridge, Massachusetts , USA. Kaufmann.
(cité aux pages 13, 16)

[Sadek, 1999] Sadek, D. (1999). Design considerations on dialogue systems : from theory to technology. the case of artimis. In *proceedings of IDS'99 (ETR Workshop on Interactive dialogue for multimedia systems)*, pages 173–187, Germany. Kloster Irsee.
(cité aux pages ii, 3, 12)

[Sadek et al., 1997] Sadek, D., Bretier, P., and Panaget, F. (1997). Artimis : Natural dialogue meets rational agency. In *proceedings of IJCAI'97*, volume 2, pages 1030–1035, NAGOYA, Aichi, Japan.
(cité aux pages ii, 3, 12)

[Sadek et al., 1996] Sadek, D., Ferrieux, A., Cozannet, A., Bretier, P., Panaget, F., and Simonin, J. (1996). Effective human-computer cooperative spoken dialogue : The ags demonstrator. In *proceedings of ICSLP'96 : the 4th Internationnal Conference on Spoken Language Processing*, pages 546–549, Philadelphia, PA, USA.
(cité aux pages 13, 16)

[Sadek and Mori, 1998] Sadek, D. and Mori, R. D. (1998). *Spoken dialogs with computers*, chapter Dialogue systems, pages 523–561. Academic Press, London, UK.
(cité aux pages ii, 3, 12)

[Searle, 1969] Searle, J. R. (1969). *Speech Acts : An essay in the philosophy of language.* Cambridge University Press.
(cité à la page 12)

[Shoham and Cousins, 1994] Shoham, Y. and Cousins, S. B. (1994). Logics of mental attitudes in ai : A very preliminary survey. In Lakemeyer, G. and Nebel, B., editors, *Foundations of Knowledge Representation and Reasoning*, pages 296–309. Springer, Berlin, Heidelberg.
(cité à la page 5)

[Simon, 1960] Simon, H. A. (1960). *The New Science of Management Decision.* Harper and Brothers, New York.
(cité aux pages 59, 152)

[Simon, 1982] Simon, H. A. (1982). *Models of bounded rationality.* The MIT Press.
(cité aux pages 61, 150)

[Singh, 1994] Singh, M. P. (1994). *Multiagent systems : a theoritical framework for intentions, knowhow, and communications.* Springer Verlag.
(cité à la page 25)

[Slantchev, 2005] Slantchev, B. L. (2005). Preferences and expected utility theory. http://www.polisci.ucsd.edu/~bslantch/courses/gt/02-preferences-expected-utility.pdf. Part of the Game Theory course of the Department of Political Science, University of California, San Diego, USA.
(cité à la page 32)

[Stratulat, 2002] Stratulat, T. (2002). *Systémes d'agens normatifs : concepts et outils logiques.* PhD thesis, Université de caen / Basse-Normandie, France.
(cité à la page 151)

[Thangarajah et al., 2002] Thangarajah, J., Padgham, L., and Harland, J. (2002). Representation and reasoning for goals in BDI agents. In Oudshoorn, M. J., editor, *proceedings of ACSC2002 : the 25th Australasian Computer Science Conference*, Melbourne, Australia. ACS.
(cité à la page 5)

[Thayse et al., 1989] Thayse, A., Gribomont, P., Hulin, G., and Pirotte, A. (1989). *Approche logique de l'intelligence artificielle. T.2.* Dunod Informatique. Dunod, Paris, France.
(cité aux pages 77, 144)

[Trapp, 1985] Trapp, R. (1985). Utility theory and preference logic. *Erkenntnis*, 22 :331–339.
(cité à la page 69)

[Tsoukias, 2002] Tsoukias, A. (2002). A first-order, four valued, weakly paraconsistent logic and its relation to rough sets semantics. *Foundations of Computing and Decision Sciences*, 27 :77 – 96.
(cité à la page 58)

[Tsoukias et al., 2002] Tsoukias, A., Perny, P., and Vincke, P. (2002). *Aiding Decisions with Multiple Criteria : Essays in Honor of Bernard Roy*, chapter From concordance/discordance to the modelling of positive and negative reasons in decision aiding., pages 147–174. Kluwer Academic, Dordrecht, Nederlands.
(cité à la page 58)

[Tsoukias and Vincke, 1995] Tsoukias, A. and Vincke, P. (1995). A new axiomatic foundation of partial comparability. *Theory and Decision*, 39 :79–114.
(cité à la page 58)

[van der Hoek et al., 1999] van der Hoek, W., van Linder, B., and Meyer, J.-J. (1999). *An integrated Modal Approach to Rational Agents*, chapter Foundations of Rational Agency, pages 37 – 75. Kluwer, Dordrecht, The Nederlands.
(cité aux pages 5, 25)

[van der Torre and Weydert, 1998] van der Torre, L. and Weydert, E. (1998). Goals, desires, utilities and preferences. In *proceedings of the ECAI'98 Workshop on Decision Theory meets Artificial Intelligence*.
(cité aux pages 5, 73)

[Vincke, 1989] Vincke, P. (1989). *L'aide multicritère à la décision,*. Editions de l'Université de Bruxelles - Editions Ellispses, Bruxelles - Belgique. English translation : Multicriteria decision-aid, Wiley, 1992.
(cité aux pages 58, 67, 75, 129, 161)

[von Neumann and Morgenstern, 1947] von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behaviour*. Princeton University Press, Princeton. 2nd edition.
(cité aux pages 36, 39, 42, 59)

[von Wright, 1963] von Wright, G. (1963). *The logic of Preference : an Essay*. Edinburg University Press.
(cité aux pages 42, 44, 46)

[von Wright, 1972] von Wright, G. (1972). The logic of preference reconsidered. *Theory and Decision*, 3 :140–169.
(cité aux pages iii, 42, 43, 44, 54, 69, 73, 74, 182)

[Wellman and Doyle, 1991] Wellman, M. and Doyle, J. (1991). Preferential semantics for goals. In *proceedings of AAAI'91 : the 9th National Conference on Artificial Intelligence*, volume 2, pages 698–703, Anaheim, California, USA. AAAI Press/MIT Press.
(cité aux pages 48, 50, 53, 54, 61, 73, 74)

[Wikipedia article, 2006] Wikipedia article (2006). Arrow's impossibility theorem. http://en.wikipedia.org/wiki/Arrow's_impossibility_theorem.
(cité à la page 201)

[Wooldridge, 1999] Wooldridge, M. (1999). *Multiagent*, chapter Intelligent Agents. The MIT Press.
(cité aux pages 4, 7)

[Öztürk et al., 2005] Öztürk, M., Tsoukiàs, A., and Vincke, P. (2005). *State of the Art in Multiple Criteria Decision Analysis*, chapter Preference Modelling, pages 27–72. Springer Verlag, Berlin, Germany.
(cité à la page 36)

Table des matières

Préambule i
Remerciements i
Introduction ii
Contexte de la thèse ii
Problématique ii
Notre proposition iii
Plan du document iii
Sommaire vi
Table des figures viii

I Problématique : agents rationnels et prise de décision 1

1 Contexte de la thèse : agents rationnels 3
1.1 Modèles cognitifs d'agents 4
1.1.1 La notion d'agent 4
1.1.2 Les modèles cognitifs d'agents 5
a) Fondements historiques et philosophiques 5
b) Le modèle *BDI* (aperçu) 6
c) La formalisation de Rao et Georgeff du modèle *BDI* 8
1.1.3 Le cas de la théorie de l'interaction rationnelle 11
a) Introduction 11
b) Attitudes mentales primitives : B et C 12
c) Modélisation de la dynamique de l'univers 14
d) Attitudes mentales composites et équilibre rationnel 15
e) L'agent et son environnement : 18
1.2 Une limitation : le choix de la réaction de l'agent 20
1.2.1 Situations adaptées aux formalismes classiques 20
1.2.2 Situations non adaptées aux formalismes classiques 21
1.2.3 Notre vision du cycle d'interaction d'un agent avec son environnement 23
1.2.4 Le formalisme de [Louis, 2002] 24
a) La notion de Stratégie 25
b) Axiomatisation 26
1.2.5 Conclusion 28

2 Prise de décision 31
2.1 La Théorie de la Décision 32
2.1.1 Le processus de décision 32
2.1.2 Qu'est-ce que la Théorie de la Décision ? 33
2.1.3 Représentations des préférences 34
a) Représentation des préférences par valeurs numériques (utilité) 34
b) Représentation des préférences par relations de comparaison 35
c) Rapport entre représentations logiques et numériques 36
2.2 Modélisations numériques des préférences 36
2.2.1 L'utilité espérée : la décision avec risque 37
a) Représentations classiques des problèmes de décision 37
b) Utilité espérée 38

2.2.2 L'utilité multi-critères 40
a) La notion d'indépendance 40
b) Utilisation de l'hypothèse MUI 40
2.2.3 Conclusion 41
2.3 Modélisations logiques des préférences 42
2.3.1 Hypothèses fréquentes 42
a) Expansion (ou Exclusion Mutuelle) : 43
b) Transitivité : 43
c) Dépendance au contexte : 43
d) Ceteris Paribus : 43
2.3.2 Formalisme proposé par von Wright dans [von Wright, 1972] 44
a) La vision des préférences de von Wright : 44
b) Formalisation : 45
c) Intérêts et critiques de la formalisation proposée par von Wright : 47
2.3.3 Formalisme proposé par Doyle et Wellman dans [Doyle and Wellman, 1994] 48
a) La vision des préférences de Doyle et Wellman 48
b) Formalisation 49
c) Travaux précédents 50
2.3.4 Formalisme proposé par Boutilier dans [Boutilier, 1994] 50
a) Logique proposée : 50
b) Le concept de préférence conditionnelle : 52
c) Le concept de normalité : 52
d) Le concept de but idéal 52
e) Rapports avec l'approche de Doyle et Wellman : 53
2.3.5 Les CP-nets : formalisme proposé dans [Boutilier et al., 1999] 53
a) Préliminaires 53
b) La notion d'indépendance 54
c) Formalisation 54
d) Schéma général d'élicitation 56
e) Raisonner avec les CP-nets 57
f) Critique 57
2.3.6 Autres travaux sur les préférences qualitatives 57
2.4 Conclusion 58
2.4.1 Limites de la théorie de la décision « classique » 58
2.4.2 Intérêts d'une représentation des préférences basée sur des comparaisons . . 60
2.4.3 Conclusion 61

II Fondements pour des préférences adaptées aux agents 63

3 Notre proposition dans les grandes lignes 65
3.1 Pourquoi proposer une nouvelle modélisation logique des préférences ? 65
3.2 Hypothèses de notre travail 67
3.2.1 Caractéristiques des informations initiales 67
a) Les alternatives 67
b) Informations sur la « désirabilité » des alternatives 67
3.2.2 Limite de l'hypothèses *Ceteris Paribus* 69
3.3 Le modèle proposé dans les grandes lignes 70

3.3.1 Idée directrice 70
3.3.2 Schéma général 71
3.4 Travaux connexes et intérêts de l'approche 73
3.4.1 Rapports avec les travaux existants 73
3.4.2 Intérêts 74

4 Langages, notations et exemple utilisés **77**
4.1 Le langage des prédicats du premier ordre $\mathscr{L}$ 77
4.1.1 Vocabulaire 77
4.1.2 Syntaxe 78
a) Terme, formule atomique et formule bien formée : 78
b) Abréviations syntaxiques usuelles : 78
c) Sous-formules et portée d'un quantificateur : 78
d) Variable libre/liée et formule ouverte/close : 79
e) Substitutions : 79
4.1.3 Sémantique 79
a) Modèle et assignation : 80
b) Satisfaction des formules : 80
c) Satisfiabilité d'une formule et formule vraie/fausse/valide : 80
d) Formules équivalentes et formule consistante avec une autre 81
e) Subsomption, théorie et conséquence logique 81
4.1.4 Axiomatique (à la Hilbert) 81
a) Système axiomatique 81
b) Notion de preuve et de théorème 82
c) Propriétés importantes de l'axiomatique et de la LPPO 82
4.2 La notation $\phi+\overline{\psi}$ 83
4.2.1 Définition 83
4.2.2 Propriétés remarquables 83
4.2.3 Notation $\phi + \overline{\psi}$ et opérateur de révision des croyances 86
4.3 Les concepts importants 87
4.3.1 Alternatives 87
4.3.2 Préférence 88
4.3.3 Indifférence 88
4.4 Exemple de scénario 89

III Formalisation des préférences en LPPO **93**

5 Préférences partielles primitives **95**
5.1 L'hypothèse *Ceteris Paribus* 96
5.1.1 Introduction : 96
5.1.2 Ceteris Paribus (Définition) : 97
5.1.3 Exemples 99
5.1.4 Propriétés remarquables : 99
a) Comportement avec l'équivalence logique 99
b) Comportement avec la propriété d'Expansion 100
c) Comportement avec la propriété de Spécialisation 100
d) Comportement avec les descriptions totales d'alternatives 100

e) Comportement avec les constantes $\top$ et $\perp$ . . . 100
f) Propriété *Ceteris Paribus* et égalité « Classique » . . . 101
5.2 La relation de *préférence partielle primitive* G_i . . . 101
5.3 Syntaxe de l'opérateur $\succeq_i$ . . . 101
5.4 Sémantique de l'opérateur $\succeq_i$ . . . 102
5.4.1 Les modèles du langage $\mathscr{L}^1$ . . . 102
5.4.2 Définition . . . 102
5.4.3 Corollaire 1 . . . 103
5.4.4 Corollaire 2 . . . 103
5.4.5 Lien entre l'opérateur $\succeq_i$ et la relation G_i . . . 103
5.5 Axiomatisation de l'opérateur $\succeq_i$ . . . 104
5.5.1 Défaut (D) . . . 104
5.5.2 Comportement de l'opérateur $\succeq_i$ avec l'équivalence logique . . . 104
5.5.3 Règle d'inférence imposant la réflexivité (R) . . . 105
5.5.4 Règle d'inférence imposant la transitivité (T) . . . 105
5.5.5 Règles d'expansion (E) . . . 105
5.5.6 Règle de spécialisation (S) . . . 106
5.5.7 Opérateur de préférence et constantes $\top$ et $\perp$ . . . 106
5.6 Propriétés remarquables . . . 107
5.6.1 Contraposition . . . 107
5.6.2 Représentation des préférences conditionnelles . . . 107
5.6.3 Addition des préférences . . . 108
5.6.4 Classes d'équivalences . . . 108
5.7 Exemple (suite 1) . . . 109

6 Préférences partielles étendues 111
6.1 L'hypothèse *Ceteris Imparibus* . . . 112
6.1.1 Introduction . . . 112
6.1.2 La relation de *préférence partielle étendue* G'_i . . . 112
6.1.3 Ceteris Imparibus (Définition) : . . . 113
6.1.4 Propriétés remarquables : . . . 113
6.2 Syntaxe de l'opérateur $\geq_i$ . . . 114
6.3 Sémantique de l'opérateur $\geq_i$ . . . 115
6.3.1 Les modèles du langage $\mathscr{L}^2$ . . . 115
6.3.2 Définition . . . 115
6.3.3 Corollaire 1 . . . 115
6.3.4 Corollaire 2 . . . 116
6.3.5 Lien entre l'opérateur $\geq_i$ et la relation G'_i . . . 116
6.4 Axiomatisation de l'opérateur $\geq_i$ . . . 116
6.4.1 Défaut (D) . . . 117
6.4.2 Règles de construction de l'opérateur $\geq_i$ . . . 117
6.4.3 Remarques . . . 117
6.4.4 Expansion (E) . . . 118
6.5 Propriétés remarquables . . . 118
6.5.1 Comportement de l'opérateur $\geq_i$ avec l'équivalence logique . . . 118
6.5.2 Réflexive (R) . . . 118
6.5.3 Spécialisation (S) . . . 119
6.6 Exemple (suite 2) . . . 119

7 Préférence globale et choix de la réaction **123**
7.1 La relation de *préférence globale* G . . . 124
7.2 Syntaxe de l'opérateur $\geq$ . . . 125
7.3 Sémantique de l'opérateur $\geq$ . . . 125
7.3.1 Les modèles du langage $\mathscr{L}^3$ . . . 125
7.3.2 Définition . . . 125
7.3.3 Corollaire 1 . . . 126
7.3.4 Corollaire 2 . . . 126
7.3.5 Lien entre l'opérateur $\geq$ et la relation G . . . 126
7.4 Politiques d'agrégation . . . 127
7.4.1 Objectif . . . 127
7.4.2 Méthode envisageable . . . 127
7.4.3 Méthode utilisé . . . 128
7.5 Axiomatisation des politiques . . . 130
7.5.1 Axiomatique de la politique *All* . . . 130
7.5.2 Axiomatique de la politique *Dictature j* . . . 131
7.5.3 Axiomatique de la politique *Lex* . . . 131
7.5.4 Axiomatique de la politique *Maj* . . . 132
7.6 Discussion sur les politiques agrégation . . . 133
7.6.1 Rappel des caractéristiques des situations à traiter . . . 133
7.6.2 Impossibilité d'exhiber une politique « optimale » . . . 133
7.6.3 Importance (relative) d'un point de vue . . . 134
7.6.4 Rapport entre les différentes politiques . . . 134
7.6.5 Politique et « personnalité » de l'agent . . . 135
7.7 Choix de la réaction . . . 135
7.8 Exemple (Fin) . . . 137

IV Extensions et conclusion générale **141**

8 Extensions possibles de notre formalisme **143**
8.1 Vers l'intégration de notre proposition à un formalisme d'agents . . . 143
8.1.1 Modification à apporter au langage $\mathscr{L}$ . . . 144
8.1.2 Modification à apporter au langage $\mathscr{L}^3$ . . . 145
8.1.3 Le langage $\mathscr{L}^4$ . . . 146
a) Syntaxe : . . . 146
b) Sémantique : . . . 147
c) Axiomatique . . . 148
d) Extension de la théorie de l'interaction . . . 148
8.1.4 Vers une autre extension possible . . . 149
8.2 Utilisation conjointe de différents types de désirabilités . . . 149
8.2.1 Origine du travail . . . 150
8.2.2 Proposition : l'utilisation séparée des données hétérogènes . . . 151
8.2.3 1^er^ problème : comment traiter chaque type d'information ? . . . 151
8.2.4 2^ème^ problème : comment agréger les différents résultats ? . . . 151
a) Chaque résultat est un sous-ensemble des alternatives *ALT* . . . 152
b) Chaque résultat est une relation binaire . . . 154
8.2.5 Conclusion . . . 155

9 Conclusion générale du document **157**
9.1 Contexte et problématique de la thèse 157
9.2 Notre proposition 158
9.3 Contributions 159
9.4 Perspectives 160

Références **163**

Table des matières **179**

Annexes **187**

A Preuves **189**
A.1 Preuves relatives à la notation $CP(a, b, \phi, \psi)$ 189
A.1.1 Si $\left[\vDash \phi_1 \Leftrightarrow \phi_2 \text{ et } \vDash \psi_1 \Leftrightarrow \psi_2\right]$ alors $\left[CP(a,b,\phi_1,\psi_1) \text{ ssi } CP(a,b,\phi_2,\psi_2)\right]$ 189
A.1.2 $CP(a, b, f_a, f_b)$ est toujours vérifiée 190
A.1.3 Les définitions (5.3) et (5.4) sont équivalentes 190
A.2 Preuves relatives à la préférence partielle primitive 191
A.2.1 Pourquoi la règle de spécialisation (5.23) a cette forme ? 191
A.2.2 La préférence partielle primitive est stable par équivalence logique 191
A.2.3 La réflexivité des opérateurs $\succeq_i$ implique celle des relations G_i 192
A.2.4 La transitivité des opérateurs $\succeq_i$ entraîne la transitivité des relations G_i . . 193
A.2.5 L'opérateur $\succeq_i$ vérifie la propriété d'addition 193
A.2.6 Classes d'équivalences 194
A.3 Preuves relatives à la préférence partielle étendue 194
A.3.1 Pourquoi faut-il bloquer la règle de construction (6.15) ? 194
A.3.2 Pourquoi l'opérateur de préférence partielle étendue ne doit pas être transitif ? 195
A.3.3 L'opérateur $\geq_i$ vérifie la propriétés de spécialisation S 195
A.3.4 L'opérateur $\geq_i$ vérifie partiellement la propriété d'expansion E 196

B Divers **197**
B.1 Quelques propriétés supplémentaires de $\phi{+}\overline{\psi}$ 197
B.1.1 La notation $\phi{+}\overline{\psi}$ et les constantes $\top$ et $\bot$ 197
B.1.2 La notation $\phi{+}\overline{\psi}$ lorsque ϕ et ψ ont une forme remarquable 197
B.1.3 La notation $\phi{+}\overline{\psi}$ lorsque ϕ et ψ entretiennent des rapports entre elles . . . 198
B.2 Rappels sur les relations 198
B.2.1 Ensemble 198
B.2.2 Relation n-aire 199
B.2.3 Relation binaire : 199
B.3 Le théorème d'impossibilité d'Arrow 201

Annexes

Annexe A

Preuves

A.1 Preuves relatives à la notation $CP(a, b, \phi, \psi)$

Dans cette annexe nous exposons quelques preuves relatives à la notation $CP(a, b, \phi, \psi)$. Plus précisément, nous montrons que si les formules ϕ_1 et ψ_1 sont respectivement équivalentes aux formules ϕ_2 et ψ_2 alors l'abréviation $CP(a,b,\phi_1,\psi_1)$ est équivalente à $CP(a,b,\phi_2,\psi_2)$. Nous montrons aussi que l'abréviation $CP(a, b, f_a, f_b)$ est toujours vérifiée et que les définitions (5.3) et (5.4) de cette abréviation sont équivalentes.

A.1.1 Si $\left[\models \phi_1 \Leftrightarrow \phi_2 \text{ et } \models \psi_1 \Leftrightarrow \psi_2\right]$ alors $\left[CP(a,b,\phi_1,\psi_1) \text{ ssi } CP(a,b,\phi_2,\psi_2)\right]$

Si $CP(a,b,\phi_1,\psi_1)$ est vérifiée alors, par définition (5.4) on a ,

$$\begin{bmatrix} \forall\alpha \in \mathscr{L}, & & \\ \text{soit} & \exists\alpha' \in \text{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi_1\psi_1}, & \models \gamma' \vee \neg\alpha') \\ \text{soit} & \exists\alpha' \in \text{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi_1\psi_1}, & \models \gamma' \vee \alpha') \\ \text{soit} & [\ a \models \alpha \quad \text{ssi} \quad b \models \alpha\ ] & \end{bmatrix}$$

De plus si $\models \phi_1 \Leftrightarrow \phi_2$ et $\models \psi_1 \Leftrightarrow \psi_2$, alors $\mathscr{L}^{\phi_1\psi_1}$ et $\mathscr{L}^{\phi_2\psi_2}$ désignent le même ensemble du fait des propriétés de la notation $\mathscr{L}^{\phi\psi}$. Par conséquent, l'équation précédente est équivalente à :

$$\begin{bmatrix} \forall\alpha \in \mathscr{L}, & & \\ \text{soit} & \exists\alpha' \in \text{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi_2\psi_2}, & \models \gamma' \vee \neg\alpha') \\ \text{soit} & \exists\alpha' \in \text{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi_2\psi_2}, & \models \gamma' \vee \alpha') \\ \text{soit} & [\ a \models \alpha \quad \text{ssi} \quad b \models \alpha\ ] & \end{bmatrix}$$

Il en résulte donc, au vue de la définition de (5.4) CP, que $CP(a,b,\phi_2,\psi_2)$ est aussi vérifiée.

A.1.2 $CP(a, b, f_a, f_b)$ est toujours vérifiée

En Remplaçant les formules ϕ et ψ respectivement par f_a et par f_b dans la définition (5.4) on obtient,

$$CP(a,b,f_a,f_b) \text{ ssi } \left[\begin{array}{lll} \forall\alpha \in \mathscr{L}, & & \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{f_a f_b}, \quad \vDash \gamma \vee \neg\alpha') & (i) \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{f_a f_b}, \quad \vDash \gamma \vee \alpha') & (ii) \\ \text{soit} & [\ a \vDash \alpha \ \text{ ssi } \ b \vDash \alpha\ ] & (iii) \end{array}\right]$$

Or, étant donné la définition de f_a, pour tout formule α, soit f_a implique α, soit f_a implique $\neg\alpha$. Par conséquent, pour tout α, soit le cas (i) est vrai, soit le cas (ii) est vrai. Pour s'en convaincre il suffit de prendre $\gamma = \neg f_a$ (même preuve en considérant $\gamma = \neg f_b$).

A.1.3 Les définitions (5.3) et (5.4) sont équivalentes

$$\left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{Si} & \left| \begin{array}{l} \forall\alpha' \in \mathrm{Subf}(\alpha), \forall\gamma \in \mathscr{L}^{\phi\psi}, \\ \exists M,\ M \vDash \gamma\wedge\alpha' \\ \exists M,\ M \vDash \gamma\wedge\neg\alpha' \end{array}\right. \\ \text{alors} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right] \quad \text{ssi} \quad \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{soit non} & \left(\begin{array}{l} \forall\alpha' \in \mathrm{Subf}(\alpha), \forall\gamma \in \mathscr{L}^{\phi\psi}, \\ \left| \begin{array}{l} \exists M,\ M \vDash \gamma\wedge\alpha' \\ \exists M,\ M \vDash \gamma\wedge\neg\alpha' \end{array}\right. \end{array}\right) \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right]$$

$$\text{ssi} \quad \left[\begin{array}{lll} \forall\alpha \in \mathscr{L}, & & \\ \text{soit} & \text{non} & [\forall\alpha' \in \mathrm{Subf}(\alpha), \forall\gamma \in \mathscr{L}^{\phi\psi}, \quad (\exists M,\ M \vDash \gamma\wedge\alpha')\ ] \\ \text{soit} & \text{non} & [\forall\alpha' \in \mathrm{Subf}(\alpha), \forall\gamma \in \mathscr{L}^{\phi\psi}, \quad (\exists M,\ M \vDash \gamma\wedge\neg\alpha')] \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] & \end{array}\right]$$

$$\text{ssi} \quad \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \forall M,\ \neg(M \vDash \gamma\wedge\alpha') \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \forall M,\ \neg(M \vDash \gamma\wedge\neg\alpha') \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right]$$

$$\text{ssi} \quad \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \forall M,\ M \vDash \neg(\gamma \wedge \alpha') \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \forall M,\ M \vDash \neg(\gamma \wedge \neg\alpha') \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right]$$

$$\text{ssi} \quad \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \vDash \neg\gamma \vee \neg\alpha') \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma \in \mathscr{L}^{\phi\psi}, \quad \vDash \neg\gamma \vee \alpha') \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right]$$

$$\text{ssi} \quad \left[\begin{array}{ll} \forall\alpha \in \mathscr{L}, & \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi\psi}, \quad \vDash \gamma' \vee \neg\alpha') \\ \text{soit} & \exists\alpha' \in \mathrm{Subf}(\alpha), \exists\gamma' \in \mathscr{L}^{\phi\psi}, \quad \vDash \gamma' \vee \alpha') \\ \text{soit} & [a \vDash \alpha \quad \text{ssi} \quad b \vDash \alpha] \end{array}\right]$$

A.2 Preuves relatives à la préférence partielle primitive

Dans cette annexe nous exposons quelques preuves relatives à la préférence partielle primitive. Plus précisément, nous montrons que la règle de spécialisation ne doit pas permettre de concaténer n'importe quelle formule aux deux parties d'une préférence. Nous montrons aussi que toute formule apparaissant dans une préférence peut être substituée par une formule sémantiquement équivalente, que la réflexivité et la transitivité des opérateurs de préférence partielle primitive implique la réflexivité et la transitivité des relations associées. Enfin, nous montrons que les opérateurs de préférence partielle primitive vérifient la propriété d'*addition des préférences* et qu'il est possibles de définir des classes d'équivalences vis-à-vis d'une préférence.

A.2.1 Pourquoi la règle de spécialisation (5.23) a cette forme ?

La règle de spécialisation (5.23) ne permet pas de concaténer n'importe quelle formule aux préférences déjà spécifiées. En particulier, une définition plus simple comme la suivante ne peut être utilisée : si $\vdash \phi \succeq_i \psi$ alors pour tout γ de $\mathscr{L}$, $\vdash (\phi\wedge\gamma) \succeq_i (\psi\wedge\gamma)$. En effet, une règle de spécialisation de cette forme implique entre autre, en identifiant γ avec $\neg\phi$, que si $\vdash \phi \succeq_i \psi$ alors $\vdash \phi\wedge\neg\phi \succeq_i \psi\wedge\neg\phi$ soit $\vdash \perp \succeq_i \psi\wedge\neg\phi$. Or cette dernière formule est fausse d'après la règle (5.25).

A.2.2 La préférence partielle primitive est stable par équivalence logique

La propriété (iii) de la notation $\phi{+}\overline{\psi}$ indique, lorsque la formule ϕ_1 est équivalente à ϕ_2 (i.e. $\vDash \phi_1 \Leftrightarrow \phi_2$) et que ψ_1 est équivalente à ψ_2 (i.e. $\vDash \psi_1 \Leftrightarrow \psi_2$), que la formule dénotée par la notation $\phi_1{+}\overline{\psi_1}$ est équivalente à la formule dénotée par la notation $\phi_2{+}\overline{\psi_2}$ (i.e. $\vDash \phi_1{+}\overline{\psi_1} \Leftrightarrow \phi_2{+}\overline{\psi_2}$) et que $\psi_1{+}\overline{\phi_1}$ est équivalente à $\psi_2{+}\overline{\phi_2}$ (i.e. $\vDash \psi_1{+}\overline{\phi_1} \Leftrightarrow \psi_2{+}\overline{\phi_2}$). Par conséquent, au vue de la propriété (5.6), Dans un telle situation l'abréviation $CP(a, b, \phi_2{+}\overline{\psi_2}, \psi_2{+}\overline{\phi_2})$ est vérifiée si et seulement si $CP(a, b, \phi_1{+}\overline{\psi_1}, \psi_1{+}\overline{\phi_1})$ l'est aussi.

La définition (5.11) indique de son côté, lorsque la formule ϕ_1 est préférée à ψ_1 (i.e. $(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} \phi_1 \succeq_i \psi_1$), que :

- $\nvDash \neg\phi_1 \quad \text{et} \quad \nvDash \neg\psi_1$
- $\left[\begin{array}{l} \forall(a,b)\in\mathrm{ALT}^2_{;} \quad \text{tel que} \left| \begin{array}{l} a \vDash \phi_1{+}\overline{\psi_1} \\ b \vDash \psi_1{+}\overline{\phi_1} \end{array}\right. \\ \text{si} \quad CP(a,b,\phi_1{+}\overline{\psi_1},\psi_1{+}\overline{\phi_1}) \quad \text{alors} \quad aG_ib \end{array}\right]$

Il résulte de tout ceci, lorsque la formule ϕ_1 est équivalente à ϕ_2, que ψ_1 est équivalente à ψ_2 et que ϕ_1 est préférée à ψ_1, que l'on a aussi que :

- $\not\models \neg\phi_2$ et $\not\models \neg\psi_2$
- $\left[\begin{array}{l} \forall(a,b)\in \mathrm{ALT}_i^2 \ \text{ tel que} \left| \begin{array}{l} a \models \phi_2 + \overline{\psi_2} \\ b \models \psi_2 + \overline{\phi_2} \end{array}\right. \\ \text{si} \quad CP(a,b,\phi_2+\overline{\psi_2},\psi_2+\overline{\phi_2}) \quad \text{alors} \quad aG_ib \end{array}\right]$

Par conséquent, étant donné la définition (5.11), on a $(\mathrm{ALT},\models,\mathscr{G}) \models_{\mathscr{L}1} \phi_2 \succeq_i \psi_2$: la formule ϕ_2 est préférée à ψ_2.

A.2.3 La réflexivité des opérateurs $\succeq_i$ implique celle des relations G_i

La définition (5.11) indique, une fois la formule ψ remplacée par la formule ϕ, que la formule ϕ est préférée à ϕ (i.e. $(\mathrm{ALT},\models,\mathscr{G}) \models_{\mathscr{L}1} \phi \succeq_i \phi$) si et seulement si :

- $\not\models \neg\phi$ et $\not\models \neg\phi$
- $\left[\begin{array}{l} \forall(a,b)\in \mathrm{ALT}_i^2 \ \text{ tel que} \left| \begin{array}{l} a \models \phi + \overline{\phi} \\ b \models \phi + \overline{\phi} \end{array}\right. \\ \text{si} \quad CP(a,b,\phi+\overline{\phi},\phi+\overline{\phi}) \quad \text{alors} \quad aG_ib \end{array}\right]$

Soit, en simplifiant puisque $\phi+\overline{\phi}$ est équivalent à ϕ, (propriété (i) de la notation $\phi+\overline{\psi}$), que :

- $\not\models \neg\phi$
- $\left[\begin{array}{l} \forall(a,b)\in \mathrm{ALT}_i^2 \ \text{ tel que} \left| \begin{array}{l} a \models \phi \\ b \models \phi \end{array}\right. \\ \text{si} \quad CP(a,b,\phi,\phi) \quad \text{alors} \quad aG_ib \end{array}\right]$

Cela est vrai en particulier lorsque a et b désignent la même alternative et que la formule ϕ est équivalente à f_a. (puisque f_a est consistante et que par définition l'alternative a est le seul modèle de la formule f_a). Par conséquent, la formule f_a est préférée à f_a si et seulement si l'alternative a est préférée à l'alternative a :

$$(\mathrm{ALT},\models,\mathscr{G}) \models_{\mathscr{L}1} f_a \succeq_i f_a \quad \text{ssi} \quad aG_ia$$

Il résulte de tout ceci que si l'opérateurs $\succeq_i$ est réflexif alors pour toute alternative a, l'alternative a est préférée à l'alternative a : la relation G_i est réflexive.

A.2.4 La transitivité des opérateurs $\succeq_i$ entraîne la transitivité des relations G_i

D'après la propriété numéro (5.16), si l'alternative a_1 est préférée à a_2 (i.e. $a_1 G_i a_2$) alors la propriété f_{a_1} est préférée à f_{a_2} :

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} f_{a_1} \succeq_i f_{a_2}$$

De même, d'après la propriété numéro (5.16), si l'alternative a_2 est préférée à a_3 (i.e. $a_1 G_i a_2$) alors la propriété f_{a_2} est préférée à f_{a_3} :

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} f_{a_2} \succeq_i f_{a_3}$$

Par conséquent, si l'opérateur $\succeq_i$ est transitif, que l'alternative a_1 est préférée à a_2 et que l'alternative a_2 est préférée à a_3 alors on a

$$(\mathrm{ALT}, \vDash, \mathscr{G}) \vDash_{\mathscr{L}^1} f_{a_1} \succeq_i f_{a_3}$$

Il en résulte que, dans un tel cas, on a aussi que l'alternative a_1 est préférée à a_3 (i.e. $a_1 G_i a_3$) d'après la propriété (5.16). La relation de préférence partielle G_i est alors transitive puisque les alternatives a_1, a_2, a_3 sont quelconques.

Remarquons que cette démonstration n'est pas valide dans le sens inverse (i.e. pour montrer que la transitivité de G_i implique la transitivité de $\succeq_i$) car les formules f_{a_1}, f_{a_2} et f_{a_3} ne sont pas quelconques.

A.2.5 L'opérateur $\succeq_i$ vérifie la propriété d'addition

Si la formule ϕ est préférée à ψ (i.e. $\vdash_{\mathscr{L}^1} \phi \succeq_i \psi$), que les formules $\phi{+}\overline{\psi}$ et γ sont consistantes (i.e. $\nvdash \neg[(\phi{+}\overline{\psi}) \wedge \gamma]$) et que les formules $\psi{+}\overline{\phi}$ et γ sont consistantes (i.e. $\nvdash \neg[(\psi{+}\overline{\phi}) \wedge \gamma]$), alors, du fait de la règle de spécialisation 5.5.6, la formule $\phi \wedge \gamma$ est préférée à $\psi \wedge \gamma$:

$$\vdash_{\mathscr{L}^1} \phi \wedge \gamma \succeq_i \psi \wedge \gamma$$

De même, si la formule γ est préférée à ω (i.e. $\vdash_{\mathscr{L}^1} \gamma \succeq_i \omega$), que les formules $\gamma{+}\overline{\omega}$ et ψ sont consistantes (i.e. $\nvdash \neg[(\gamma{+}\overline{\omega}) \wedge \psi]$) et que les formules $\omega{+}\overline{\gamma}$ et ψ sont consistantes (i.e. $\nvdash \neg[(\omega{+}\overline{\gamma}) \wedge \psi]$), alors, du fait de la règle de spécialisation 5.5.6, la formule $\gamma \wedge \psi$ est préférée à $\omega \wedge \psi$:

$$\vdash_{\mathscr{L}^1} (\gamma \wedge \psi) \succeq_i (\omega \wedge \psi)$$

Par conséquent, puisque l'opérateur de préférence partielle primitive $\succeq_i$ est transitif, dans de telles conditions on a aussi que la formule $\phi \wedge \gamma$ est préférée à $\omega \wedge \psi$:

$$\vdash_{\mathscr{L}^1} (\phi \wedge \gamma) \succeq_i (\psi \wedge \omega)$$

A.2.6 Classes d'équivalences

Si la formule ϕ est préférée à ψ (i.e. $\vdash_{\mathscr{L}^1} \phi \succeq_i \psi$), que les formules $\phi{+}\overline{\psi}$ et α sont consistantes (i.e. $\nvdash \neg[\phi{+}\overline{\psi}\wedge\alpha]$) et que les formules $\psi{+}\overline{\phi}$ et α sont consistantes (i.e. $\nvdash \neg[\psi{+}\overline{\phi}\wedge\alpha]$), alors, du fait de la règle de spécialisation 5.5.6, la formule $\phi\wedge\alpha$ est préférée à $\psi\wedge\alpha$ (i.e. $\vdash_{\mathscr{L}^1} (\phi\wedge\alpha)\succeq_i(\psi\wedge\alpha)$). Par conséquent, du fait de la règle de transitivité :

$$\text{si } \vdash_{\mathscr{L}^1} (\psi\wedge\alpha)\succeq_i\gamma \text{ alors } \vdash_{\mathscr{L}^1} (\phi\wedge\alpha)\succeq_i\gamma$$

De même, si la formule ψ est préférée à ϕ (i.e. $\vdash_{\mathscr{L}^1} \psi\succeq_i\phi$), que les formules $\psi{+}\overline{\phi}$ et α sont consistantes (i.e. $\nvdash \neg[\psi{+}\overline{\phi}\wedge\alpha]$) et que les formules $\phi{+}\overline{\psi}$ et α sont consistantes (i.e. $\nvdash \neg[\phi{+}\overline{\psi}\wedge\alpha]$), alors, du fait de la règle de spécialisation 5.5.6, la formule $\psi\wedge\alpha$ est préférée à $\phi\wedge\alpha$ (i.e. $\vdash_{\mathscr{L}^1} (\psi\wedge\alpha)\succeq_i(\phi\wedge\alpha)$). Par conséquent, du fait de la règle de transitivité :

$$\text{si } \vdash_{\mathscr{L}^1} \gamma\succeq_i(\phi\wedge\alpha) \text{ alors } \vdash_{\mathscr{L}^1} \gamma\succeq_i(\psi\wedge\alpha)$$

A.3 Preuves relatives à la préférence partielle étendue

Dans cette annexe nous exposons quelques preuves relatives à la préférence partielle étendue. Plus précisément, nous montrons pourquoi n'importe quelle préférence partielle primitive ne peut pas forcément être un argument et pourquoi l'opérateur de préférence partielle étendue ne doit pas être transitif. Nous montrons aussi que l'opérateur de préférence partielle étendue vérifie la propriété de spécialisation et la propriété d'expansion de façon partielle.

A.3.1 Pourquoi faut-il bloquer la règle de construction (6.15) ?

Supposons que lorsque l'on cherche un argument $\phi' \succeq_i \psi'$ pour que la propriété ϕ soit préférée à la propriété ψ, on ne teste pas s'il n'existe pas un contre argument évident (i.e. $\psi' \succeq_i \phi'$). Plus précisément, supposons que la définition de l'abréviation CI_i est la suivante :

$$\mathrm{CI}_i(\phi,\psi) \quad \text{ssi} \quad \left[\begin{array}{l|l} \multicolumn{2}{l}{\exists(\phi',\psi') \in \mathscr{L}\times\mathscr{L}} \\ \text{tel que} & \vdash_{\mathscr{L}^1}\phi'\succeq_i\psi' \\ & \vdash \quad \phi \Rightarrow \phi'{+}\overline{\psi'} \\ & \vdash \quad \psi \Rightarrow \psi'{+}\overline{\phi'} \end{array}\right]$$

Puisque l'opérateur de préférence partielle primitive est réflexif (propriété (5.19)), on a entre autre $\vdash_{\mathscr{L}^1} \top\succeq_i\top$. D'autre part, pour toutes formules ϕ et ψ on a $\vdash \phi \Rightarrow \top{+}\overline{\top}$ et $\vdash \psi \Rightarrow \top{+}\overline{\top}$.

Par conséquent, avec une telle définition, pour toutes formules ϕ et ψ, on aurait $\mathrm{CI}_i(\phi,\psi)$. Il en résulte, étant donné la règle de construction (6.15), que pour toutes formules ϕ et ψ, on aurait $\vdash_{\mathscr{L}^2} \phi\succeq_i\psi$.

A.3.2 Pourquoi l'opérateur de préférence partielle étendue ne doit pas être transitif ?

Supposons que la propriété ϕ est préférée strictement à la propriété α :

$$\vdash_{\mathscr{L}^1} \phi \succeq_i \alpha \quad \text{et} \quad \vdash_{\mathscr{L}^1} \neg(\alpha \succeq_i \phi)$$

Supposons aussi que la propriété β est préférée strictement à la propriété ψ :

$$\vdash_{\mathscr{L}^1} \beta \succeq_i \psi \quad \text{et} \quad \vdash_{\mathscr{L}^1} \neg(\psi \succeq_i \beta)$$

Dans de telles conditions, du fait de la règle de construction (6.15), on a donc, entre autres, à la fois que la propriété ϕ est préférée à la propriété $\alpha \wedge \beta$ et que la propriété $\alpha \wedge \beta$ est préférée à la propriété ψ :

$$\vdash_{\mathscr{L}^2} \phi \geq_i \alpha \wedge \beta \quad \text{et} \quad \vdash_{\mathscr{L}^2} \alpha \wedge \beta \geq_i \psi$$

Il en résulte, si l'opérateur de préférence partielle étendue est transitif, que la propriété ϕ est préférée strictement à la propriété ψ :

$$\vdash_{\mathscr{L}^2} \phi \geq_i \psi$$

Par conséquent, si la préférence partielle étendue était transitive alors il y aurait au plus deux nivaux dans la préférence.

A.3.3 L'opérateur $\geq_i$ vérifie la propriétés de spécialisation S

Du fait des règles de construction (6.14) et (6.15) de l'opérateur de préférence partielle étendue, si la formule ϕ est préférée à ψ via la préférence partielle étendue relative au point de vue i (i.e. $\vdash_{\mathscr{L}^2} \phi \geq_i \psi$) alors de deux choses l'une, (1) soit la formule ϕ est préférée à ψ via la préférence partielle primitive relative au point de vue i (i.e. $\vdash_{\mathscr{L}^1} \phi \succeq_i \psi$), (2) soit il existe un argument pour préférer ϕ à ψ (i.e. $CI_i(\phi, \psi)$).

Dans le premier cas, à la vue de la règle de spécialisation (5.23) de l'opérateur de préférence partielle primitive, cela implique que pour toute formule α consistante avec $\phi + \overline{\psi}$ et $\psi + \overline{\phi}$, on a $\vdash_{\mathscr{L}^1} (\phi \wedge \alpha) \succeq_i (\psi \wedge \alpha)$. Par conséquent, au vue de la règle de construction (6.14) de l'opérateur de préférence partielle étendue, cela implique aussi que la formule $\phi \wedge \alpha$ est préférée à $\psi \wedge \alpha$ via la préférence partielle étendue relative au point de vue i :

$$\vdash_{\mathscr{L}^2} (\phi \wedge \alpha) \geq_i (\psi \wedge \alpha)$$

Dans le second cas, du fait de la propriété (6.6) de la notation CI_i cela implique que l'on a aussi $CI_i(\phi \wedge \alpha, \psi \wedge \alpha)$ (si $\phi \wedge \alpha \not\equiv \perp$ et $\psi \wedge \alpha \not\equiv \perp$)[1]. Par conséquent, au vue de la règle de construction (6.15)l'opérateur de préférence partielle étendue, cela implique aussi que la formule $\phi \wedge \alpha$ est préférée à $\psi \wedge \alpha$ via la préférence partielle étendue relative au point de vue i :

$$\vdash_{\mathscr{L}^2} (\phi \wedge \alpha) \geq_i (\psi \wedge \alpha)$$

[1]($\phi \wedge \alpha \not\equiv \perp$ implique $\phi + \overline{\psi} \wedge \alpha \not\equiv \perp$ et $\psi \wedge \alpha \not\equiv \perp$ implique $\psi + \overline{\phi} \wedge \alpha \not\equiv \perp$).

A.3.4 L'opérateur $\geq_i$ vérifie partiellement la propriété d'expansion E

Du fait des règles de construction (6.14) et (6.15) de l'opérateur de préférence partielle étendue, si la formule ϕ est préférée à ψ via la préférence partielle étendue relative au point de vue i (i.e. $\vdash_{\mathscr{L}^2} \phi \geq_i \psi$) alors de deux choses l'une, (1) soit la formule ϕ est préférée à ψ via la préférence partielle primitive relative au point de vue i (i.e. $\vdash_{\mathscr{L}^1} \phi \succeq_i \psi$), (2) soit il existe un argument pour préférer ϕ à ψ (i.e. $CI_i(\phi, \psi)$).

Dans le premier cas, à la vue de la règle d'expansion (5.21), cela implique que la formule $\phi + \overline{\psi}$ est préférée à $\psi + \overline{\phi}$ via la préférence partielle primitive relative au point de vue i $\vdash_{\mathscr{L}^1} \phi + \overline{\psi} \succeq_i \psi + \overline{\phi}$. Par conséquent, au vue de la règle de construction (6.14), cela implique aussi dans que la formule $\phi + \overline{\psi}$ est préférée à $\psi + \overline{\phi}$ via la préférence partielle étendue relative au point de vue i

$$\vdash_{\mathscr{L}^2} \phi + \overline{\psi} \geq_i \psi + \overline{\phi}$$

Dans le second cas, du fait de la propriété (6.4) de la notation CI_i cela implique que l'abréviation $CI_i(\phi + \overline{\psi}, \psi + \overline{\phi})$ est aussi vérifiée. Par conséquent, au vue de la règle de construction (6.15), cela implique aussi que la formule $\phi + \overline{\psi}$ est préférée à $\psi + \overline{\phi}$ via la préférence partielle étendue relative au point de vue i

$$\vdash_{\mathscr{L}^2} (\phi + \overline{\psi}) \geq_i (\psi + \overline{\phi})$$

Par conséquent, l'opérateur $\geq_i$ est tel que si la formule $\phi \geq_i \psi$ est vérifiée alors la formule $\phi + \overline{\psi} \geq_i \psi + \overline{\phi}$ l'est aussi.

Annexe B

Divers

B.1 Quelques propriétés supplémentaires de $\phi + \overline{\psi}$

Dans cette annexe nous exposons quelques propriétés supplémentaires de la notation $\phi + \overline{\psi}$. Plus précisément, nous exposons quelques propriétés de cette notation relatives aux constantes $\top$ et $\bot$. Nous exposons aussi quelques propriétés de cette notation lorsque les formules ϕ et ψ ont une forme remarquable ou entretiennent des rapports entre elles.

B.1.1 La notation $\phi + \overline{\psi}$ et les constantes $\top$ et $\bot$

(xiii). $\top + \overline{\top} \equiv \top$

(xiv). $\top + \overline{\bot} \equiv \top$

(xv). $\bot + \overline{\bot} \equiv \bot$

(xvi). $\bot + \overline{\top} \equiv \bot$

(xvii). $\bot + \overline{\phi} \equiv \bot$

(xviii). $\top + \overline{\phi} \equiv \neg\phi$

(xix). $\phi + \overline{\top} \equiv \phi$

(xx). $\phi + \overline{\bot} \equiv \phi$

(xxi). $\phi + \overline{\psi} \equiv \bot$ ssi $\phi \equiv \bot$

B.1.2 La notation $\phi + \overline{\psi}$ lorsque ϕ et ψ ont une forme remarquable

(xxii). $(\phi + \overline{\psi}) + (\overline{\phi + \overline{\psi}}) \equiv \phi + \overline{\psi}$ Preuve évidente car de forme $\phi + \overline{\phi}$.

(xxiii). Si $\left[\ \phi \equiv C \wedge a_1 \wedge \ldots \wedge a_n \text{ et } \psi \equiv C \wedge \neg a_1 \wedge \ldots \wedge \neg a_n\ \right]$ alors $\phi + \overline{\psi} \equiv \phi$

Éléments de preuve :
Dans de telles conditions on a $\phi \wedge \neg\phi \equiv (C \wedge a_1 \wedge \ldots \wedge a_n) \wedge \neg(C \wedge \neg a_1 \wedge \ldots \wedge \neg a_n) \equiv (C \wedge a_1 \wedge \ldots \wedge a_n) \wedge (\neg C \vee a_1 \vee \ldots \vee a_n) \equiv C \wedge a_1 \wedge \ldots \wedge a_n \equiv \phi$.

(xxiv). Si $\left[\ \phi \equiv C \wedge a \text{ et } \psi \equiv C \wedge b\ \right]$ alors (si $\phi \Rightarrow \psi$ alors $\phi + \overline{\psi} \equiv \phi$, sinon $\phi + \overline{\psi} \equiv C \wedge a \wedge \neg b$).

Éléments de preuve :
Dans de telles conditions on a $\phi \wedge \neg\phi \equiv (C \wedge a) \wedge \neg(C \wedge b) \equiv (C \wedge a) \wedge (\neg C \vee \neg b) \equiv C \wedge a \wedge \neg b$. De plus, on a $(C \wedge a) \Rightarrow b$ si et seulement si $(C \wedge a) \Rightarrow (C \wedge b)$.

(xxv). $(\phi+\overline{\psi})+\overline{\psi} \equiv \phi+\overline{\psi}$.

Preuve similaire à la preuve de la propriété (ix).

(xxvi). $(\phi+\overline{\psi})+\overline{\phi} \equiv \phi+\overline{\psi}$.

Preuve similaire à la preuve de la propriété (ix).

(xxvii). $(\phi\wedge\gamma)+\overline{(\psi\wedge\gamma)} \equiv (\phi\wedge\gamma)+\overline{\psi}$

Élément de preuve :
Remarquer que l'on a $(\phi\wedge\gamma)\wedge\neg(\psi\wedge\gamma) \equiv (\phi\wedge\gamma)\wedge\neg(\psi)$.

B.1.3 La notation $\phi+\overline{\psi}$ lorsque ϕ et ψ entretiennent des rapports entre elles

(xxviii). Si $\vDash \phi\Rightarrow\psi$ alors $\phi+\overline{\psi} \equiv \phi$

Preuve :
Si $\vDash \phi\Rightarrow\psi$ alors $\vDash \phi\wedge\neg\psi \Rightarrow \psi\wedge\neg\psi \Rightarrow \bot$. Par conséquent, par définition, $\phi+\overline{\psi} \equiv \phi$.

(xxix). $\vDash \phi+\overline{\psi}\Rightarrow\psi$ si et seulement si $\vDash \phi\Rightarrow\psi$

Preuve en remarquant que de deux choses l'une :
- Soit $\phi+\overline{\psi} \equiv \phi$
  Dans ce cas, $\vDash \phi+\overline{\psi}\Rightarrow\psi$ si et seulement si $\vDash \phi\Rightarrow\psi$ (cas trivial).
- Soit $\phi+\overline{\psi} \equiv \phi\wedge\neg\psi$
  Dans ce cas $\vDash \phi+\overline{\psi} \Rightarrow \psi$ si et seulement si $\vDash \phi\wedge\neg\psi \Rightarrow \psi$, donc si et seulement si $\vDash \neg(\phi\wedge\neg\psi)\vee\psi$, donc si et seulement si $\vDash \neg\phi\vee\psi\vee\psi$, donc si et seulement si $\vDash \neg\phi\vee\psi$, donc si et seulement si $\vDash \phi\Rightarrow\psi$.

(xxx). Si $\vDash \phi+\overline{\psi}$ alors $(\neg\psi)+\overline{\neg\phi}$

Preuve en remarquant que de deux choses l'une :
- Soit $\phi\wedge\neg\psi \not\equiv \bot$
  Dans ce cas, $\phi+\overline{\psi} \equiv \phi\wedge\neg\psi$ et comme alors $\neg\psi\wedge\neg\neg\phi \not\equiv \bot$, on a $(\neg\psi+\overline{\neg\phi}) \equiv (\neg\psi\wedge\neg\neg\phi) \equiv \neg\psi\wedge\phi$. Par conséquent, $\phi+\overline{\psi} \Rightarrow \neg\psi+\overline{\neg\phi}$.
- Soit $\phi\wedge\neg\psi \equiv \bot$ (i.e. $\vDash \psi\Rightarrow\neg\phi$)
  Dans ce cas, $\phi+\overline{\psi} \equiv \phi$ et comme alors $\neg\psi\wedge\neg\neg\phi \equiv \bot$, on a $(\neg\psi+\overline{\neg\phi}) \equiv \neg\psi$. Par conséquent, là aussi $\phi+\overline{\psi}\Rightarrow\neg\psi+\overline{\neg\phi}$.

B.2 Rappels sur les relations

B.2.1 Ensemble

Définition : Un ensemble est une collection d'objets qui ne possède aucune structure particulière. En particulier, l'ordre de ces objets n'a aucune importance. On note $x \in A$ le fait que x est un élément de l'ensemble A. En général, un ensemble est désigné à l'aide d'accolades.

Un ensemble est défini *en extension* si la liste de ces éléments est donnée explicitement. Un ensemble est défini *en compréhension* (ou en intension) si la liste de ces éléments est donnée implicitement grâce à une liste de propriétés qui les caractérisent (et seulement ceux-ci).

Un ensemble S_1 est un sous-ensemble de S_2 (noté $S_1 \subset S_2$) si tous les éléments de S_1 sont aussi des éléments de S_2 : $\forall x$, si $x \in S_1$ alors $x \in S_2$. L'ensemble des *parties* d'un ensemble A, noté $P(A)$ est composé de tous les sous-ensembles de A. Si A contient k éléments, $P(A)$ en contiendra 2 puissance k.

Couples et n-tuples (séquences) : Une séquence est un ensemble dont les éléments sont numérotés : on peut parler du premier élément, du deuxième, du troisième, etc. Une séquence de deux éléments est appelée une paire ou un couple, une séquence de n éléments, un n-tuple ou simplement un tuple. En général, une séquence est désignée à l'aide de parenthèses. Par exemple, la notation (67, 1, 14) représente une séquence dont le premier élément est 67, le deuxième 1 et le troisième 14.

B.2.2 Relation n-aire

Une relation n-aire entre n ensembles $A_1, \ldots, A_n$ est un ensemble de n-tuples $(a_1, ..., a_n)$ où a_1 appartient à A_1, a_2 appartient à A_2, ... Si l'on considère l'ensemble de tous les n-tuples $(a_1, ..., a_n)$ possibles avec $a_1 \in A_1$, $a_2 \in A_2$, etc., on obtient le produit cartésien $A_1 \times A_2 \times \cdots \times A_n$ des ensembles n ensembles. Une relation n-aire sur $A_1, \ldots, A_n$ est donc un sous-ensemble du produit cartésien $A_1 \times A_2 \times \cdots \times A_n$.

B.2.3 Relation binaire :

Définition : Une relation (binaire) R entre deux ensembles A et B est un sous-ensemble de couples (a, b) où a appartient à A et b appartient à B (i.e. $R \subset A \times B$). On dira que a est en relation avec b si le couple (a, b) appartient à R. On le note fréquemment aRb. Souvent, A et B désignent le même ensemble ; on parle alors de la relation binaire R *sur* A.

Propriétés fréquentes des relations binaires : Une relation binaire est une structure mathématique utilisée fréquemment. Elle peut vérifier (ou ne pas vérifier) de nombreuses propriétés. Nous faisons ci-dessous la liste de propriétés fréquentes des relations binaires. Plus précisément, une relation (binaires) R sur A est dite :
- euclidienne si et seulement si $\forall (a, b, c) \in A^3, (aRb \wedge aRc) \Rightarrow bRc$
- réflexive si et seulement si $\forall a \in A, aRa$
- irréflexive si et seulement si $\forall a \in A, \neg(aRa)$
- sérielle si et seulement si $\forall a \in A, \exists b, aRb$
- transitive si et seulement si $\forall (a, b, c) \in A^3, (aRb \wedge bRc) \Rightarrow aRc$
- négativement transitive si et seulement si $\forall (a, b, c) \in A^3, (\neg(aRb) \wedge \neg(bRc)) \Rightarrow \neg(aRc)$
- semi-transitive si et seulement si $\forall (a, b, c, d) \in A^4, ((aRb) \wedge (bRc)) \Rightarrow (aRd \vee dRc)$

– de Ferrers si et seulement si $\forall(a,b,c,d) \in A^4, (aRb \wedge cRd) \Rightarrow (aRd \vee cRb)$
– symétrique si et seulement si $\forall(a,b) \in A^2, aRb \Rightarrow bRa$
– antisymétrique si et seulement si $\forall(a,b) \in A^2, (aRb \wedge bRa) \Rightarrow b = a$
– asymétrique si et seulement si $\forall(a,b) \in A^2, (aRb \Rightarrow \neg(bRa)$
– acyclique si et seulement si $\forall(a_1,\ldots,a_n) \in A^n, (aRb \Rightarrow \neg(bRa)$
– connexe si et seulement si $\forall(a,b) \in A^2$, si $a \neq b$ alors aRb ou bRa.
– totale si et seulement si deux éléments quelconques de A sont en relation : $\forall(a,b) \in A^2$ on a aRb ou aRb. (typiquement la relation $\geq$ dans $\mathbb{N}$). La relation R est aussi dite complète dans une telle situation.
– partielle si et seulement si il existe des couples d'éléments incomparables (typiquement la divisibilité dans $\mathbb{N}$) c'est-à-dire si elle n'est pas totale.

Ces propriétés ne sont pas indépendantes les unes des autres. Plus précisément, certaines sont des implications des autres ou de combinaisons des autres. En particulier, une relation est :

– asymétrique si et seulement si elle est irréflexive et antisymétrique
– totale si et seulement si elle est connexe et réflexive
– acyclique si et seulement si la clôture transitive est irréflexive

Pour plus de détails le lecteur pourra aller voir par exemple [Bouyssou and Vincke, 2003] ou le site `http://www.fact-index.com/b/bi/binary_relation.html`.

Représentation graphique d'une relation binaire : Une relation binaire R sur A peut être représentée par un graphe orienté $(A; R)$, comme celui de la figure B.1, où A est l'ensemble des sommets du graphe et R est l'ensemble des arcs du graphe. Ces derniers sont orientés : ce sont des couples de sommets. Les propriétés remarquables d'une relation

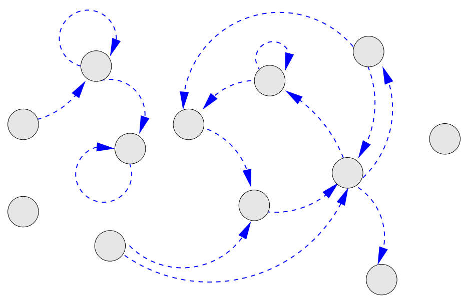


Fig. B.1 – Graphe représentant une relation binaire quelconque

binaire R sur A peuvent facilement s'exprimer avec le graphe $(A; R)$. La réflexivité de R se traduit par la présence d'une boucle en chaque sommet. La symétrie de R signifie que la présence d'un arc orientée d'un sommet a vers un sommet b implique l'existence d'un arc orienté de b vers a. La transitivité de R se traduit en termes de graphe par le fait que s'il existe un chemin de longueur 2 de a vers b, alors il existe un arc de a vers b. Prendre la relation inverse de R revient à inverser l'orientation de tous les arcs du graphe. Prendre la relation complémentaire consiste à ajouter tous les arcs manquants dans le graphe et à supprimer tous les arcs existants. Notons qu'une relation symétrique peut, plus commodément être représentée par un graphe non orienté dans lequel les couples (a, b) et (b, a) de la relation donnent lieu à une arête entre les sommets a et b.

Relation d'équivalence : Une relation binaire R sur un ensemble E est dite d'*équivalence* lorsqu'elle est à la fois réflexive, symétrique et transitive. Une *classe d'équivalence* pour une telle relation d'équivalence R est alors l'ensemble des éléments en relation avec un élément donné.

B.3 Le théorème d'impossibilité d'Arrow

En 1951, Kenneth Arrow à découvert un résultat qui peut sembler très négatif lorsqu'il tentait de trouver une moyen « équitable » d'agréger les préférences de différentes personnes. Plus précisément, il a prouvé dans [Arrow, 1951] que, étant donné n pré-ordres totaux sur l'ensemble d'alternatives ALT, il n'existe en général pas de méthode « équitable » (i.e. qui respecte les quatre conditions si dessous) pour obtenir un pré-ordre total sur ALT, si n est supérieur à un et ALT contient plus de deux alternatives. Pour Arrow, une méthode d'agrégation est « équitable » (dit « fair » en anglais) si elle respecte les quatre principes suivants :

- L'universalité : aucune restriction n'est faite sur les n pré-ordres totaux à agréger. En particulier, ceux ne sont pas des ordres et ils ne vérifient pas la propriété de cardinalité.
- L'indépendance face aux alternatives non pertinentes : le résultat de la comparaison entre deux alternatives (après agrégation) ne dépend que des comparaisons entre ces deux alternatives relatif aux n pré-ordres totaux à agréger.
- L'unanimité (pareto) : si les n pré-ordres à agréger indiquent unanimement qu'une alternative est meilleure qu'une autre, alors la comparaison entre ces deux alternatives (après agrégation) indique la même chose.
- la non dictature : il n'existe pas un des n pré-ordres à agréger qui est tel que pour tout couple d'alternatives, s'il indique qu'une alternative est meilleure que l'autre, alors la comparaison entre ces deux alternatives (après agrégation) indique la même chose.

Ce résultat est à la base de nombreux travaux. En particulier, Kenneth May a prouvé que lorsque ALT contient seulement deux alternatives, alors, quelque soit le nombre d'ordres à agréger, le système de simple majorité (et seulement lui) respecte les quatre conditions d'Arrow [May, 1952]. Par la suite, certains ont montré que l'on pouvait « sortir » de cette impossibilité en imposant plus de contraintes aux prés-ordres à agréger et en particulier s'il existe un pré-ordre objectif supportant les n pré-ordres à agréger [Black, 1948] ou si les pré-ordres à agréger vérifient la propriété de cardinalité [Harsanyi, 1955].

Pour une courte présentation un peu plus complète, le lecteur pourra se référer à [Albouy, 2004] et [Wikipedia article, 2006].