Résumé

Les limitations rencontrées par la théorie de l'utilité espérée pour rendre compte du comportement effectif des décideurs incitent à rechercher une modélisation de la décision qui prenne davantage en compte la façon dont les préférences sont prises en compte en fonction des raisonnements. L'idée est alors d'introduire le raisonnement par similarité comme mode de raisonnement dans un problème de décision : la solution à un problème courant est recherchée parmi des problèmes de décision mémorisés, qui lui sont similaires et qui sont maximaux au sens des préférences. La similarité de deux problèmes de décision est appréhendée, compte tenu de l'état en cours du raisonnement, par la recherche d'identités partielles entre les descriptions de problèmes, le reste des descriptions étant négligé. Cette identité partielle dépend de ce qui a été négligé dans le passé et est acceptée ou non en fonction des buts et de l'état du raisonnement. La relation de préférence est définie à partir des sextuplets permettant d'exprimer des préférences entre différentes identités partielles et différentes négligibilitées. Cette relation de préférence permettent alors de déterminer une solution maximale en fonction de ce qui est connu et de ce qui a été négligé. A partir de ces deux notions (négligeabilité et préférence), on introduit deux relations de déduction par similarité, dont on détermine une sémantique de point fixe permettant une implantation sur machine. La première de ces deux relations de déduction par similarité ou relation forte de satisfaction par similarité, permet une première formalisation du raisonnement par similarité, mais exige comme pré-requis que les préférences soient complètes, faute de quoi on peut ne pas savoir conclure. La deuxième de ces des deux relations de déduction par similarité, ou relation faible de satisfaction par similarité, permet de lever ce pré-requis. L'implantation dans le système Soar consiste à informellement rajouter un processus d'appariement partiel dans le processus d'inférence de Soar, et à rajouter la prise en compte des éléments négligés, les préférences proposées étant exprimables dans le formalisme de Soar. Cette implantation permet de prendre en compte la résolution de problème en incertitude, comme par exemple en photo-interprétation, en autorisant la prise en compte de l'incertitude par de la similarité.

Mots-clés: Décision, Logiques non standards, Similarité, Soar.

Abstract

The limitations encountered by expected utility theory in the modelisation of effective decision making led the path to a model of decision making which takes into account the way preferences are handled according to reasoning. The idea consists in introducing reasoning by similarity as the type of reasoning used in decision problems: The solution to a current problem is searched among maximal solutions (according to the preferences) of similar memorised decision problems. The similarity of two decision problems is defined, according to the current state of reasoning, by seeking partial identities between the descriptions of the two problems while neglecting descriptive features which distinguish the two problems. The partial identity depends upon previously used negligibilities, and is accepted or rejected with respect to the goals andstate of the reasoning. The preference relation is defined using tuples which allows the expression of preferences between different partial identities and different negligibilities. Using these two notions, negligibility and similarity, we introduce two relations of entailment by similarity, and two *fixed point semantics, the first relation being named strong relation of entailment allows a first formalisation of reasoning by similarity, but requires the completeness of the preference relation (or else the reasoning may not conclude). The second relation of entailment allows a second formalisation of reasoning by similarity without requiring the completeness of the preference relation. The implementation in the Soar system consists in transforming the inference mechanism from a pattern matching process to a partial one, and in taking into account negligible features, the preferences formally defined being expressible in the Soar system. This implementation can take into account problem solving under uncertainty, for example remote sensing image interpretation, by using similarity to handle uncertainty.

Keywords: Decision, Non monotonic logics, Similarity, Soar.

Remerciements

Je tiens en premier lieu à exprimer toute ma gratitude à *Jean-Charles Pomerol*, Professeur à l'Université Paris VI, et à *Régine Teulier*, Chargée de recherche au CNRS, qui m'ont encadré et conseillé judicieusement tout au long de cette thèse. Je leur suis particulièrement reconnaissant pour leurs encouragements, et pour le soutien et la confiance qu'ils m'ont accordés.

Je remercie par ailleurs tous les membres du jury pour l'intérêt dont ils ont fait preuve à l'égard de ce travail:

M Jean-Gabriel Ganascia, professeur à l'universitè Paris 6, qui m'a fait l'honneur de présider ce jury,

M Henri Prade, Directeur de recherche au CNRS, et *M Amedeo Napoli*, chargé de recherche au CNRS, tous deux rapporteurs de cette thèse, pour le temps qu'ils ont accepté de consacrer à la lecture de ce manuscript, et à le discuter avec moi,

M René Lalement, Maître de conférence, pour avoir accepté de regarder des versions antérieures de ce travail,

M Jean-Luc Dormoy, chercheur à EDF, pour l'intérêt qu'il a montré pour mon travail et son chaleureux support,

Tout particulièrement, je tiens à adresser ma plus profonde reconnaissance à *Jean-Philippe Lagrange*, pour les nombreuses discussions que nous avons pu avoir ensemble, pour son aide quotidienne, sa grande disponibilité malgré ses obligations. Le soutien scientifique et moral constant et irremplaçable qu'il m'a toujours témoigné ont rendu ce travail possible.

Je tiens naturellement à remercier toutes les personnes du laboratoire GIP et de la CEGN grâce auxquelles je garderais un bon souvenir de cette période. Je veux particulièrement remercier Éric pour avoir accepté de relire une partie de ce manuscript.

Je terminerais avec un remerciement à mes amis, Éric, Odile, Pascal, Philippe, Sabine, Slimane, ainsi qu'à toute ma famille pour m'avoir aidé moralement dans les moments difficiles; c'est grâce à eux que j'ai pu finir cette thèse dans de bonnes conditions.

"L'homme est plein de besoins. Il n'aime que ceux qui peuvent les remplir tous. 'C'est un bon mathématicien', dira-t-on. Mais je n'ai que faire de mathématiques, il me prendrait pour une proposition . . . "
Pascal, Pensées.

"L'esprit est le refus d'être quoi que ce soit."
Paul Valéry

Table des matières

Table des figures

Introduction

La décision est liée aux raisonnements, et l'une des définitions possibles de l'intelligence artificielle se réfère aux processus cognitifs et aux raisonnements. Il paraît donc naturel d'étudier les raisonnements dans la décision. On peut, tout comme [Pomerol 1997], admettre que toute décision a son origine dans une insatisfaction liée à une différence entre un état courant et un état futur plus désirable. Appelons *problème de décision* cette dissatisfaction. Notons qu'une dimension personnelle apparaît immédiatement, dans la mesure où ce qui est désirable pour un individu peut très bien ne pas être souhaité par un autre. Avant de décider, l'individu reconnaît l'état courant, connaissant une partie des événements advenus dans le passé et ayant sa propre perception de l'état courant: l'individu tente d'identifier l'état courant par rapport à son expérience. La première phase de la décision est donc de reconnaître un ou plusieurs états enregistrés proches de l'état courant. Mais la décision dépend aussi des états futurs (désirables) que l'on souhaite atteindre. Désignons par E une représentation des états futurs non contrôlés ni influencés par le décideur (les états de la nature en théorie de la décision) et soit A l'ensemble des actions possibles. On peut alors représenter l'ensemble des états accessibles lorsque l'état courant est S_i et l'action entreprise A_j par $\pi(S_i, A_j, E)$. Les préférences du décideur s'appliquent sur cet ensemble et déterminent par régression l'action préférée choisie, i.e. la meilleure décision au sens des préférences du décideur.

Ce type de modélisation est d'autre part rendu nécessaire par l'incapacité de la théorie de la décision classique de décrire les comportements des décideurs. En effet, en théorie de la décision, une deuxième notion se rajoute à celle de préférence, la notion de rationalité, qui permet de définir les contraintes, généralement de cohérence [1] que doit vérifier la relation de préférence pour que le qualificatif de rationnel puisse être appliqué à l'individu. Ces contraintes s'expriment généralement sous la forme d'une axiomatique ou de *contraintes nécessaires* que doit vérifier la relation de préférence. Cette idée de contrainte nécessaire a entraîné tout un courant de recherche dit normatif en théorie de la décision, que l'on peut faire dater du célèbre pari de Pascal, et qui va jusqu'aux axiomatiques de [Neumann and Morgenstern 1947] et [Savage 1954]. Ces axiomatiques montrent qu'un comportement idéal est compatible avec une hypothèse de classement des actions par une fonctionnelle d'utilité espérée.

Cette focalisation sur les contraintes **nécessaires** que doit satisfaire la relation de préférence du décideur a amené en retour une questionnement sur leur caractère **suffisant**: permettent elles de rendre compte des choix d'un décideur non plus idéal, mais concret ? L'ensemble des études menées depuis le célèbre Paradoxe de [Allais 1953], notamment par les psychologues [Kahneman and Tversky 1979][Tversky and Kahneman 1982], montre la faiblesse du support empirique de la théorie. Cette absence de validité empirique de la théorie a conduit en retour à diverses généralisations de [Neumann and Morgenstern 1947], tout en gardant le même cadre de formalisation, c'est à dire sans faire intervenir les raisonnements; en fait, seules les axiomes les plus controversés de la théorie de l'utilité espérée (axiome d'indépendance par exemple) sont affaiblis ou rejetés.

Ce n'est que très récemment que la théorie de la décision s'est intéressée aux raisonnements et donc à l'intelligence artificielle pour sortir de cette impasse, en commençant à introduire la mémoire des problèmes passés mémorisés similaires au problème courant. [Gilboa and Schmeidler 1995], par exemple, postulent que les individus choisissent des actions en fonction de leurs performances dans des problèmes passés mémorisés similaires au problème courant. Cette approche est intéressante dans la mesure où elle utilise un modèle reconnu en intelligence artificielle (le raisonnement à partir de cas) auquel les préférences sont ajoutées. En effet, le raisonnement à partir de cas est une activité essentielle dans la décision. En intelligence artificielle, le modèle de raisonnement à partir de cas [Schank 1982][Kolodner 1993] suppose que les individus raisonnent sur des cas complets, et non plus des faits isolés, comme dans les systèmes experts. Un cas est très similaire à un objet en programmation: il décrit de par ses attributs une partie

1. Requérir la transitivité de la relation de préférence est une exemple d'une telle contrainte.

de la connaissance. La première étape du raisonnement à partir de cas consiste alors à rechercher parmi les cas mémorisés pour voir s'il en existe un proche du cas courant. Si un cas est suffisamment proche du cas courant, alors les mêmes actions que celles entreprises dans le cas passés sont choisies. Si aucune similarité n'est trouvée, le cas est soumis soit à un nouveau raisonnement soit à une modification des cas passés par le biais d'un index par exemple. Cependant, ce modèle de raisonnement ne s'est jusqu'à maintenant pas intéressé aux préférences.

En dehors de cette forme particulière de similarité (le raisonnement à partir de cas s'intéresse surtout aux parties communes entre cas courant et cas passés), une autre forme de similarité a été implicitement formalisée en intelligence artificielle dans le cadre de certaines logiques non standards, comme par exemple les logiques des défauts: deux cas sont alors dits similaires si et seulement si les parties différençiant cas courant et cas passés sont **négligeables**: la notion de négligeable ayant dans certaines de ces logiques, comme par exemple la logique des défauts de [Reiter 1980], a le sens de *"un élément est négligeable si et seulement si sa négation ne peut être prouvée"*. Cette idée de cohérence inclut donc une notion très rudimentaire et opératoire de préférence. Cette notion de négligeable a également été indroduite dans d'autres approches, comme par exemple celle de [Lehmann 1996].

Objectif de la thèse

L'objectif principal de la thèse est de proposer un modèle formel pour l'introduction du raisonnement par similarité et des préférences dans un problème de décision: la solution à un problème courant est recherchée parmi des problèmes de décision mémorisés qui lui sont similaires, en choisissant l'action maximale au sens des préférences du décideur. Une relation de déduction par similarité est alors proposée dont les deux composantes principales sont:

- la relation de similarité. La similarité est définie, si on utilise comme langage l'ensemble des clauses de Horn, par l'ensemble des formules similaires à la fois à un problème courant (représenté par une conjonction de clauses unités B) et à un problème mémorisé (représenté par une clause définie C), c'est-à-dire l'ensemble des formules dont le corps subsume à la fois B et le corps de C, la tête de la formule similaire subsumant la tête de C. On appelle négligeable l'ensemble des parties non subsumées, qui représente donc les différences entre problème mémorisé et problème courant.

- la formalisation syntaxique des préférences. Les préférences sont introduites sous la forme d'un sextuplet:

(Pré-requis valide 1,	Pré-requis négligé 1, Fait 1)
	Préféré
(Pré-requis valide 2,	Pré-requis négligé 2, Fait 2)

  ce qui signifie intuitivement, "le fait 1 résultant de la connaissance de Pré-requis 1 et du fait de négliger Pré-requis négligé 1 est préféré au fait 2 résultant de la connaissance de Pré-requis 2 et du fait de négliger Pré-requis négligé 2".

La relation de déduction permet tout à la fois de raisonner par similarité et d'appliquer ces préférences qui permettent alors de déterminer une solution maximale en fonction de ce qui est connu et de ce qui a été négligé. Deux modes de raisonnement sont alors proposés et formalisés.

- La première de ces deux relations de déduction par similarité ou relation forte de déduction par similarité, permet une première formalisation du raisonnement par similarité, en supposant que tous les éléments de description des cas passés soient potentiellement négligeables. Ceci induit une relation de déduction qui peut s'avérer trop faible lorsque l'ensemble des préférences n'est pas complet vis à vis des éléments négligés.

- la deuxième de ces des deux relations de déduction par similarité, ou relation faible de satisfaction par similarité, permet de lever ce pré-requis, en permettant de spécifier des hypothèses de négligeabilités exclusives les unes des autres.

Ayant montré que les préférences introduites correspondent à des préférences d'un système d'intelligence artificielle existant, en l'occurrence le système Soar, on introduit alors la similarité dans ce système en lui permettant de faire des appariements partiels. Le but de notre modèle étant de prendre en compte le comportement réel des décideurs (il s'agit d'un modèle descriptif et non d'un modèle normatif), nous validant ensuite le raisonnement proposé sur un domaine particulier, en l'occurrence la photo interprétation. En effet, lorsqu'un photo-interprète travaille sur une image satellite, il n'a accès qu'à une description partielle des objets capturés dans la scène. S'appuyant alors sur son expérience (des cas similaires rencontrés dans le passé), il peut, en tenant compte de ses objectifs de photo-interprétation, établir une interprétation de la scène.

Le présent document s'articule en quatre grandes parties qui reflètent les objectifs majeurs de ce travail:

- Une première partie rappelle les formalisations normatives proposées en théorie de la décision, et propose une discussion autour des axiomes introduits et de leur validité empirique. Les généralisations du modèle de l'utilité espérée sont ensuite évoquées (une présentation complète en est donnée en annexe) et on montre que leur formalisation induit la nécessité de prendre en compte les raisonnements dans la décision. Dans le deuxième chapitre de cette première partie, on s'intéresse aux raisonnements dans la décision. Notre modèle de raisonnement est alors présenté, discuté autour d'exemples, et ses liens avec les formalismes existants en théorie de la décision sont discutés. La relation de déduction par similarité avec des préférences est informellement introduite.

- Une deuxième partie introduit et discute les propriétés essentielles que doit satisfaire une relation de déduction par similarité avec des préférences. Ces propriétés sont discutées et confrontées aux propriétés des différentes logiques non standards existantes. Le deuxième chapitre de cette partie effectue la comparaison de notre relation de déduction par similarité avec des préférences, ainsi formellement identifiée, avec les formalisations classiques du raisonnement à partir de cas.

- Une troisième partie formalise une relation de déduction par similarité avec des préférences qui satisfait les propriétés que la partie précédente a introduite. Le premier chapitre de cette partie donne une construction syntaxique et sémantique des préférences, le second chapitre formalise la relation de déduction forte par similarité avec des préférences, le troisième chapitre formalise la relation de déduction faible par similarité avec des préférences. Pour chacune de ces deux logiques, une sémantique de point fixe est proposée et les liens avec le système Soar sont présentés.

- Une quatrième partie s'intéresse au support empirique de la théorie ainsi définie. Le premier chapitre donne le modèle de l'implantation dans Soar de la logique ainsi définie, en proposant une relation de similarité correspondant intuitivement à de l'unification partielle. Nous présentons alors les modifications (pré-compilation de la similarité, ajout de la notion de négligeabilité) nessaires pour le raisonnement par similarité. Le deuxième chapitre de cette partie détaille une modélisation d'un raisonnement préférentiel réalisée dans le cadre de cette théorie, celle de la photo-interprétation.

Première partie

La théorie de la décision

L'analyse des choix dans l'incertain s'est concentrée sur la manière dont des actions doivent ou devraient être choisies par un décideur. Définir le sens de "choisir la meilleure action" dès que les conséquences des actions dépendent d'événements incertains [2], a constitué une partie de la théorie mathématique de la décision.

Cette analyse mathématique de la décision s'est concentrée autour de deux notions:

– L'idée de préférence, qui permet à un décideur de choisir entre plusieurs actions. C'est l'élément "directement observable" de la décision,

– L'idée de rationalité, qui consiste en un ensemble de postulats (mathématiquement d'axiomes) que doit satisfaire la relation de préférence du décideur.

Muni de ces deux éléments, il est possible de montrer, si l'on accepte les postulats les plus courants, que le classement des actions est compatible avec une hypothèse d'utilité espérée.

Cependant, il a pu être montré que cette théorie manquait de validité descriptive dans le sens où les décideurs ne satisfont pas toujours les axiomes ou postulats de rationalité. Bien que des affaiblissements de ces axiomes aient été proposés pour accroître leur validité empirique, ils ne permettent pas pour autant de rendre compte des comportements effectifs des décideurs.

L'hypothèse que nous avançons est que cette incapacité à rendre compte des comportements effectifs dans la décision provient de la focalisation de la théorie de la décision sur la structure des choix et non sur la manière dont les préférences sont élaborées, sur les raisonnements qui permettent à un individu de construire ses préférences. Cette constatation nous a amené à construire et proposer un modèle de décision qualitative, fondée sur une approche symbolique.

2. Nous ne nous intéresserons pas ici à la décision dans la certitude et renverrons le lecteur à [French 1988] ou [Fishburn 1970] pour le traitement de ce problème.

1

Théorie de la décision dans l'incertain

1.1 Introduction

La modélisation de la décision s'appuie sur trois éléments ou ensembles constitutifs de l'idée de décision:

- L'idée **d'action** représentant les différentes alternatives entre lesquelles un décideur doit trancher. L'ensemble des actions est noté A.
- L'idée que chaque action entraîne certaines **conséquences**. L'ensemble des conséquences est noté C.
- L'idée que les conséquences résultant du choix de chaque action sont déterminées non seulement par l'action elle-même, mais aussi par un certain nombre de facteurs extérieurs. Appelons **état de la nature** une description complète de chacun de ces facteurs extérieurs. L'ensemble des états de la nature est noté Ω. Chacun des éléments de Ω sont exclusifs les uns des autres, mais ils sont collectivement exhaustifs.

Pour chaque élément $E \in \Omega$ et $a \in A$, on note par $a(E)$ l'élément de C qui décrit les conséquences de l'acte a lorsque l'état de la nature est E. Comme indiqué dans la notation, chaque acte détermine une projection de E sur C.

Les choix sont effectués en fonction d'une relation de **préférence** sur l'ensemble des actions qui permet de déterminer la meilleure action. L'analyse de la décision s'est bâtie autour de deux interprétations de l'utilisation des préférences des individus comme fondement de la modélisation de leurs choix, une interprétation empirique ou descriptive et une interprétation normative [Shafer 1991]. Chacune de ces deux interprétations est incorporée dans un ensemble de postulats (plus formellement d'axiomes) que doit satisfaire la relation de préférence pour que le décideur soit **rationnel**. Chaque formalisation pose donc une définition de la rationalité.

Selon l'interprétation descriptive des préférences, celles-ci obéissent généralement aux postulats. Selon l'interprétation normative des préférences, de tels postulats sont un modèle de rationalité: ils décrivent les préférences d'une personne idéale, d'un individu capable de présenter un comportement constituant une norme; les préférences du décideur doivent obéir aux postulats. Remarquons que ces deux interprétations, étant donné qu'elles sont modèles du même système formel, sont indissociables l'une de l'autre.

Cette distinction entre interprétation descriptive et normative étant essentielle à la compréhension de la théorie de l'utilité espérée et de ses critiques, nous consacrerons la première partie de ce chapitre à en donner deux exemples, l'un sur l'interprétation normative concernant l'évaluation des éléments de l'ensemble des états de la nature par une mesure de probabilité, le second sur l'interprétation descriptive concernant l'évaluation des éléments de l'ensemble des conséquences par une fonction d'utilité. Ces deux notions étant alors bien posées, nous les utiliserons tant pour illustrer que pour critiquer les postulats de rationalité introduits, que ce soit dans le cas du risque ou de l'incertitude, postulats qui permettent de représenter la relation de préférence par une fonctionnelle d'utilité espérée. Ceci nous permettra de voir ensuite comment les critiques empiriques de ces théories ont donné lieu à d'autres postulats de rationalité, et donc à d'autres modèles. Ces derniers sont cependant, comme nous le verrons, incapables de répondre simultanément à toutes les critiques empiriques formulées à l'encontre de la théorie de l'utilité espérée.

Auparavant, effectuons deux remarques sur le type de rationalité que peut définir un ensemble de postulats en théorie de la décision:

- la rationalité définie est *substantive* et non *procédurale*: elle ne concerne que les résultats (comparaisons effectuées grâce à la relation de préférence) et non pas la procédure utilisée pour les produire,
- la rationalité modélise la structure des choix, et non pas leur contenu: c'est dans les prescriptions de choix cohérents entre eux, et non pas dans la prescription d'un choix particulier, qu'elle se situe.

Comme nous venons de le voir, approche descriptive et normative furent toutes les deux à la base de la théorie de l'utilité espérée. Pour donner un exemple de ces interprétations, nous verrons rapidement l'historique de la théorie de la décision, où tant l'évaluation de la vraisemblance des états de la nature (par une mesure de probabilité) que l'évaluation des conséquences (par une mesure d'utilité), résulte de constatations normatives pour la première, empiriques pour la seconde.

1.2 Lien entre interprétation normative et interprétation descriptive

La première modélisation de la décision fut celle de B. Pascal dans ce qui est connu comme étant le pari de Pascal. Dans ce tableau (tableau 1.1), les lignes représentent les décisions, les colonnes les événements et les conséquences les cellules du tableau.

Décisions \ Événements	Dieu existe (w_1)	Dieu n'existe pas (w_2)
Croire en Dieu	c_{11}	c_{12}
Ne pas croire en Dieu	c_{21}	c_{22}

TAB. 1.1: *Le Pari de Pascal*

Pascal, postulant sur les états de la nature l'existence de poids (w_i) représentant leur vraisemblance respective, pose comme critère de décision l'espérance mathématique des gains $\sum_j c_{ij} w_j$. Selon ce critère et l'argumentation de Pascal, on doit alors croire en Dieu, car si l'on juge son existence et inexistence équiprobable, on a cependant $c_{11} = +\infty$, c_{12} a une valeur finie positive, $c_{21} = -\infty$ et c_{22} a une valeur négative finie.

L'approche proposée par Pascal ouvre au moins trois domaines de recherche:

- L'étude des poids w_i sur l'ensemble des états de la nature (théorie des probabilités),
- L'étude des poids c_{ij} sur l'ensemble des conséquences (théorie de l'utilité),
- La combinaison de ces deux informations, i.e. l'hypothèse d'espérance de gains.

Ces trois domaines de recherche ont d'abord été abordés séparément, la recherche des poids w_i ou la recherche de l'évaluation des conséquences étant effectuées en supposant connus les deux autres points.

L'évaluation des états de la nature

L'évaluation des poids w_i, en fait par une mesure de probabilité, résulte d'une constatation normative de [De Finetti 1937], fondée sur l'idée qu'un individu rationnel n'accepterait pas un pari dans lequel il serait sûr de perdre, quoiqu'il arrive.

Considérons en effet le pari suivant: soit $(E_i)_{1\leq i\leq n}$ une partition de l'événement certain. Le décideur choisit des nombres $(p_i)_{1\leq i\leq n}$ et on choisit alors $(S_i)_{1\leq i\leq n}$. L'enjeu est $\sum_{1\leq i\leq n} p_i * S_i$ et le gain S_j si l'état de la nature est E_j. Le choix de $(p_i)_{1\leq i\leq n}$ est dit cohérent s'il n'existe pas de choix de $(S_i)_{1\leq i\leq n}$ tel que le décideur soit certain de perdre. Pour simplifier l'exposé, nous considérerons le cas $n = 3$. Les gains du décideur sont alors: $\begin{cases} G_1 &= S_1 - p_1S_1 - p_2S_2 - p_3S_3 \\ G_2 &= S_2 - p_1S_1 - p_2S_2 - p_3S_3 \\ G_3 &= S_3 - p_1S_1 - p_2S_2 - p_3S_3 \end{cases}$

Lemme 1. *[French 1988] Le choix de p_1, p_2, p_3 est dit cohérent si et seulement si $p_1 + p_2 + p_3 = 1$.*

Preuve. Écrivons les équations ci-dessus sous forme matricielle: $G = QS$ où:
$Q = \begin{pmatrix} 1-p_1 & -p_2 & -p_3 \\ -p_1 & 1-p_2 & p_3 \\ -p_1 & -p_2 & 1-p_3 \end{pmatrix}, G = \begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix}, S = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix}$.
Faire perdre le décideur consiste à choisir S tel que $S = Q^{-1}G$. Le choix de p_1, p_2, p_3 sera donc dit cohérent si et seulement si Q n'a pas d'inverse, soit $det\ Q = 1 - p_1 - p_2 - p_3 = 0$. ∎

Proposition 2. *[French 1988] p_1, p_2, p_3 satisfont les axiomes de Kolmogorov.*

Preuve. **i)** Considérons E_1, E_2, E_3 tels que $(E_1 \cup E_2) \cap E_3 = \emptyset$. Montrons que $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$, ce qui prouvera le troisième axiome de Kolmogorov.
On considère par q_{12}, q_3 les poids associés à $E_1 \cup E_2, E_3$, et p_1, p_2, p_3 les poids associés à respectivement E_1, E_2, E_3. Supposons les sommes S_1, S_2, S_3 fixées sur E_1, E_2, E_3 et T_{12}, T_3 les sommes fixées sur $E_1 \cup E_2, E_3$. Les gains du décideur sont alors:
$\begin{cases} G_1 &= S_1 + T_{12} - p_1S_1 - p_2S_2 - p_3S_3 - q_{12}T_{12} - q_3T_3 \\ G_2 &= S_2 + T_{12} - p_1S_1 - p_2S_2 - p_3S_3 - q_{12}T_{12} - q_3T_3 \\ G_3 &= S_3 + T_3 - p_1S_1 - p_2S_2 - p_3S_3 - q_{12}T_{12} - q_3T_3 \end{cases}$
Si le décideur choisit $q_{12} < p_1 + p_2$, soit $q_3 > p_3$, on peut alors choisir $S_1 = S_2 = T_3 = 1$ et $S_3 = T_{12} = -1$. Dans ce cas, $G_1 = G_2 = G_3 = (q_{12} - p_1 - p_2) + (p_3 - q_3) < 0$. Il en est de même avec le cas $q_{12} > p_1 + p_2$.
Reste donc le cas $q_{12} = p_1 + p_2$ et $q_3 = p_3$. Alors
$\begin{cases} G_1 &= (S_1 + T_{12}) - p_1(S_1 + T_{12}) - p_2(S_2 + T_{12}) - p_3(S_3 + T_3) \\ G_2 &= (S_2 + T_{12}) - p_1(S_1 + T_{12}) - p_2(S_2 + T_{12}) - p_3(S_3 + T_3) \\ G_3 &= (S_3 + T_3) - p_1(S_1 + T_{12}) - p_2(S_2 + T_{12}) - p_3(S_3 + T_3) \end{cases}$
D'après le lemme précédent, $p_1 + p_2 + p_3 = 1$ entraîne un choix cohérent. Ce qui montre donc le troisième axiome de Kolmogorov.

ii) $P(E) = 1$, où E est l'événement certain est alors immédiat, puisque $E = E_1 \cup E_2 \cup E_3$.

iii) Comme $\{E_1, E_2 \cup E_3, \emptyset\}$ forme une partition de l'évènement certain, il suffit de montrer que $p_1 \geq 0$ pour démontrer le troisième axiome de Kolmogorov.
Supposons donc $p_1 < 0$. Donc $p_2 + p_3 > 1$. En posant $S_1 = -1, S_2 = S_3 = 1$ alors $G_1 = G_2 = G_3 = p_1 - (p_2 + p_3) - 1 < 0$.

∎

Ainsi, pour que le décideur soit cohérent, dans le cadre de l'hypothèse de gain espéré, la vraisemblance qu'il attache aux états de la nature doit être compatible avec une mesure de probabilité.

L'évaluation des conséquences

L'évaluation des conséquences, par une mesure d'utilité, résulte d'une constatation empirique, à savoir que les individus associent des valeurs réelles à chacune des conséquences offrant une alternative probabiliste. Empiriquement, cette valeur de la conséquence correspond à la subjectivité propre de l'individu, à la manière dont il évalue les conséquences. Cette constatation repose sur le Paradoxe de St Petersbourg, suggéré à N. Bernouilli par R. De Montmort [Munier 1988]. Lançons une pièce et désignons par face le côté gagnant. Si face apparaît au $n^{\text{ième}}$ lancer de dé, le gain est de 2^n. Le gain espéré est donc de

$$\sum_{i=0}^{+\infty} \frac{1}{2^i} 2^i = \sum_{i=0}^{+\infty} 1$$

qui est infini alors que personne ne souhaite payer plus de 20 unités d'argent pour participer à ce jeu. La solution de ce paradoxe [Bernoulli 1738] consiste à considérer que l'évaluation d'une action ne consiste pas en une espérance de gain, mais en **l'espérance d'utilité** de ce même gain. Bernouilli définit cette utilité u comme proportionnelle aux richesses initiales X, soit $du = b.\frac{dX}{X}$, où b est une constante caractérisant l'individu, soit donc $u = b.log_2(\frac{\alpha+x}{\alpha})$, α caractérisant la richesse initiale de l'individu. Le paradoxe est alors résolu puisque, si v caractérise la valeur du gain espéré,

$$log_2(\frac{v+\alpha}{\alpha}) = \sum_{n=1}^{+\infty}(\frac{1}{2})^n * b * log_2(\alpha + \frac{2^n}{\alpha})$$

Soit

$$v = \prod_{n=1}^{+\infty}(\alpha + 2^n)^{\frac{1}{2^n}}$$

On vérifie alors que cette valeur est majorée par $(\alpha+2)^2 - \alpha$, ce dernier résultat reflétant l'attitude de la plupart des joueurs qui consiste à ne pas accepter de participer à ce jeu si la mise dépasse une certaine somme d'argent.

Ayant maintenant illustré les notions d'interprétations normatives et empiriques par un exemple, voyons maintenant comment ces notions ont permis tant de construire que de critiquer la théorie de l'utilité espérée.

1.3 Modélisation de la décision dans le risque et l'incertitude

Ces premières approches de la définition d'un critère de décision en ont entraîné d'autres, comme par exemple:

Le critère de Wald pour un optimiste, le critère est le maximax, l'acte préféré étant celui qui conduit au mieux à la conséquence la plus favorable; pour un pessimiste, c'est le maximin, l'acte préféré est celui qui conduit au pire à la conséquence la moins favorable.

Le critère d'Hurwicz $\alpha m + (1-\alpha)M$, uniquement fondé sur les mesures des deux conséquences la plus M ou la moins favorable m en fonction d'un coefficient de pondération $\alpha \in [0,1]$ traduisant le degré d'optimisme de l'individu,

Le critère de Laplace fondé sur la moyenne de toutes les conséquences possibles d'un acte.

Or cette multiplication des critères de décision entraîne la nécessité de les discriminer. Cette multiplication des critères a donc entraîné une première définition de la rationalité, [Milnor 1954] définissant un ensemble de contraintes que les préférences d'un décideur doivent satisfaire. Citons les plus importants de ces dix critères formulés sous forme de contraintes que doit satisfaire un critère de choix (voir aussi [Luce and Raiffa 1957]):

1. Le critère permet d'ordonner les actes et donc d'en trouver au moins un qui soit préféré.
2. L'ordre des actes qu'engendre le critère est indépendant de la manière dont on énumère les actes et les états.
3. Si deux actes sont tels que, pour chaque état, la conséquence du premier est supérieure ou égale à celle du second, le critère donne la préférence au premier acte. Cet axiome est mis en défaut par les critères de Wald et d'Hurwicz, puisque l'un et l'autre ne sont fondés que sur les conséquences extrêmes d'un acte.
4. Ajouter de nouveaux actes ne doit pas modifier l'ordre relatif des actes qui existaient déjà.
5. Si deux actes sont simultanément préférés vis-à-vis d'un critère, il doit en être de même pour toute combinaison linéaire de ces actes.

6. Ajouter ou retrancher un état identique à un autre état ne doit pas modifier la réponse du critère. Cet axiome est mis en défaut par le critère de Laplace dans la mesure où celui-ci suppose implicitement que tous les états de la nature sont équiprobables.

7. Le critère n'est pas modifié si on applique la même transformation linéaire à toutes les mesures des conséquences.

Ces contraintes, ayant un aspect normatif, ont permis de faire un tri entre les modèles de décision qui étaient alors proposés (critère de Walcz, critère de Hurwicz ...).

Les deux axiomatisations suivantes qui respectent ces critères ont alors été proposées, l'une dans le cas du risque, l'autre dans le cas de l'incertitude. Ces deux notions de risque et d'incertitude renvoient respectivement aux concepts de probabilités objectives et subjectives. Selon Maurice Allais, le premier est relatif à la perception de l'avenir à partir de la connaissance des distributions statistiques observées dans le passé: il s'agit d'une estimation subjective de la fréquence intrinsèque que les données de l'observation antérieure nous permettent d'inférer pour l'avenir. Le second concept concerne les prédictions qui portent sur des événements pour lesquels aucune analyse statistique antérieure n'est disponible: "*la probabilité subjective est alors le prix exprimé en francs par francs pour lequel l'individu serait prêt à échanger la possession d'une somme quelconque S, positive ou négative, subordonnée à l'arrivée d'un événement E, avec la possession de la somme pS*".

Chacune des deux axiomatisations, que ce soit dans le cas du risque ou de l'incertitude, propose alors de définir des axiomes de cohérence sur la relation de préférence (affirmer que cette relation est un pré-ordre est un exemple d'un tel axiome), et de fournir une méthode, par le biais d'autres axiomes, permettant de représenter cette relation qualitative par une fonction à valeur dans $\mathbb{R}$ ou dans un sous ensemble de $\mathbb{R}$.

La présentation de ces deux axiomatiques se fera en mettant en avant à la fois la formalisation mathématique et les constatations descriptives qui ont permis de définir ces axiomes. On s'efforcera, autant qu'il est possible, de présenter simultanément les expériences qui ont permis de mettre en doute, voire même d'invalider, ces axiomes.

1.3.1 La modélisation de la décision dans le risque

L'utilisation du concept de loterie (ensemble des lois de probabilité sur l'espace des résultats, muni d'une tribu contenant ces mêmes résultats) pour représenter les actions possibles auquel est confronté le décideur est particulièrement adaptée aux situations de risque. Supposons qu'il existe en effet une mesure de probabilité π sur Ω. Le passage de l'acte $f \in A$ à la loterie P_f qui lui correspond se fait par l'intermédiaire de la proposition suivante:

$$\forall c \in C, P_f(c) = \pi\{w \in \Omega, f(w) \in C\}$$

Intuitivement, la probabilité qu'associe la loterie P_f à un ensemble c de conséquences est donnée par la probabilité subjective de l'ensemble des états de la nature qui constituent un événement dont la réalisation entraîne par l'acte f, une conséquence appartenant à l'ensemble c.

Le décideur devant dans ce cas choisir parmi des décisions dont le résultat dépendra de la réalisation d'événements dont il connaît les probabilités, on associe à chaque décision une certaine loi de probabilité sur l'espace des résultats. Chaque acte correspond donc à une loterie sur l'ensemble des résultats.
Par exemple, soit le choix entre les décisions suivantes:

d_1: Miser un franc sur le jet d'une pièce de monnaie contre un gain de deux francs si l'issue est pile; la loi de probabilité correspondante sur l'espace des résultats est P_1 telle que $P_1(\{+1\}) = \frac{1}{2}$ et $P_1(\{-1\}) = \frac{1}{2}$;

d_2: Ne pas miser, d'où la certitude d'un gain net nul; la loi de probabilité correspondante sur l'espace des résultats est P_2 telle que $P_2(\{0\}) = 1$.

On prendra donc pour ensemble de choix l'ensemble $\mathcal{P}$ de toutes les lois de probabilité sur l'espace des résultats C, muni d'une tribu $\mathcal{G}$ contenant les singletons $\{c\}, c \in C$.

Définition 3. *(Combinaisons convexes de mesures)*
On appelle combinaison linéaire convexe ou mixage d'ordre n l'opération de composition externe dans $\mathcal{P}$

qui à tout n-uple $(P_1, P_2, \ldots, P_n) \in \mathcal{P}^n$ et tout n-uple $(\lambda_1, \ldots, \lambda_n) \in \mathcal{S}^n$, où $\mathcal{S}^n$ désigne le simplexe de $\mathbb{R}^n$, fait correspondre $Q \in \mathcal{P}$ tel que:

$$G \in \mathcal{G} \to Q(G) = \lambda_1 P_1(G) + \ldots + \lambda_n P_n(G)$$

On considérera donc à partir de maintenant que l'ensemble $\mathcal{P}$ est clos par combinaisons linéaires convexes.

On peut interpréter un mixage en considérant que, si $Q = \lambda_1 P_1 + \lambda_2 P_2 + \ldots + (1-\lambda_1-\lambda_2-\ldots-\lambda_{n-1})P_n$, le mixage modélise un tirage aléatoire selon la loi Q en deux temps:

- observer l'issue de la réalisation d'un tirage en n issues issues de probabilité, λ_1, λ_n, ..., $(1 - \lambda_1 - \lambda_2 - \ldots - \lambda_{n-1})$,
- si la j^{ime} issue est observée, tirer le résultat selon la loi P_j.

Cette définition va nous permettre de donner l'axiomatique de [Neumann and Morgenstern 1947], axiomatique que nous présenterons suivant les travaux de [Fishburn 1970] [Jaffray 1985]:

Axiomatique 4. *[Neumann and Morgenstern 1947]*

Axiome 1 *(Classement des loteries)*
L'ensemble de choix $\mathcal{P}$ est constitué de toutes les lois de probabilité sur $(C, \mathcal{G})$, où C est l'ensemble des conséquences et $\mathcal{G}$ une tribu contenant les singletons $\{c\}, c \in C$. La relation de préférence $\succeq$[3] est un pré-ordre large[4] dans $\mathcal{P}$.

Axiome 2 *(Continuité des préférences)*
Pour toutes les lois $P', P'', Q \in \mathcal{P}$, vérifiant $P' \succ Q \succ P''$, on peut trouver $\lambda, \mu \in]0,1[$ tel que $\lambda P' + (1-\lambda)P'' \succ Q$. et $Q \succ \mu P' + (1-\mu)P''$.

Axiome 3 *(Axiome d'indépendance)*
Pour toutes lois $P', P'', Q \in \mathcal{P}$ et tout $\lambda \in]0,1]$, $P' \succ P'' \Leftrightarrow \lambda P' + (1-\lambda)Q \succ \lambda P'' + (1-\lambda)Q$.

Avant de présenter le théorème de [Neumann and Morgenstern 1947], nous présenterons, pour chaque axiome, le support descriptif et normatif de celui-ci ainsi que les expériences qui ont permis d'invalider l'universalité de ce support descriptif et normatif.

Premier axiome

La justification normative du premier axiome est que le décideur n'attache pas d'importance à la manière dont un tirage selon une loi P est effectué, mais ne tient compte que de la loi elle-même, ce qui exclut tout plaisir à jouer.

Les contre-exemples de cet axiome sont au nombre de deux, ceux concernant l'invariance empirique de la relation de préférence, ceux concernant la transitivité de la relation de préférence et d'indifférence (partie symétrique de la relation de préférence).

Le premier axiome suppose que le décideur est toujours capable d'exprimer des préférences entre toutes les lois de probabilité. Empiriquement, cela suppose que les préférences du décideur sont invariantes en fonction du protocole expérimental permettant de les révéler. En effet, que signifie qu'un décideur possède des préférences, si ce n'est [Shafer 1991] que connaissant une ou des méthodes de mesure de ces préférences, il existe une invariance empirique dans les résultats d'application de ces méthodes? Dans le cas des préférences, la méthode de mesure consiste à demander au décideur ses préférences, qui ne satisfont pas en général l'invariance empirique requise comme le montre l'exemple suivant:

> On suppose que les États Unis se préparent au déclenchement d'une maladie asiatique, qui est supposée tuer 600 personnes en l'absence de tout programme préventif, et on demande de choisir entre deux programmes alternatifs.
> Dans le premier cas, les conséquences possibles des deux programmes sont construites comme suit:
> - Si le programme A est choisi, 200 personnes seront sauvées.

3. On notera respectivement $\succ$ et $\sim$ les parties asymétriques et symétriques de $\succeq$.
4. On rappelle qu'un pré-ordre large est une relation asymétrique et négativement transitive.

– Si le programme B est choisi, il y a une probabilité de 1/3 que 600 personnes seront sauvées, et de 2/3 que personne ne sera sauvé.

Dans l'autre cas, ils sont décrits de la manière suivante:

– Si le programme A est choisi, 400 personnes mourront.
– Si le programme B est choisi, il y a une probabilité de 1/3 que personne ne meure, et de 2/3 que 600 personnes meurent.

Les deux descriptions sont bien sûr équivalentes. Pourtant, dans le cas de la première description, une large majorité d'individus choisissent l'expérience A, alors que dans le deuxième cas, ils choisissent l'expérience B. Si l'on cherche à expliquer ce renversement de préférence, notons que la première description peut entraîner une argumentation en faveur du programme qui sauvera au moins quelques personnes, alors que la seconde description tend plutôt à entraîner une argumentation en faveur du programme qui n'entraîne aucune mort. L'incohérence s'atténue si l'on remarque que les décisions sont le fruit d'un raisonnement, et que la présentation offre différents arguments comme point de départ du raisonnement. (Exemple extrait de [Tversky and Kahneman 1981]).

Le premier axiome permet également de montrer que, à la fois la relation de préférence et la relation d'indifférence, doivent être transitives. Cette contrainte normative résulte du fait que, si l'on a $f \succ g \succ h \succ f$, il est alors possible, si le décideur possède f, de lui donner h moyennant une compensation monétaire (il le préfère à f), puis de lui échanger g et h moyennant la même compensation monétaire, et de même pour f et g. Le décideur se retrouve alors avec son choix de départ, mais appauvri: on dit alors qu'il fonctionne comme une pompe à monnaie. Cependant, il a pu être montré [Tversky 1969] que des préférences intransitives peuvent être exprimées par des individus:

Considérons une situation dans laquelle le décideur doit choisir entre des candidats pour un travail sur la base de leur QI et de leur expérience passée. Supposons qu'il préfère les candidats de plus grand QI, mais que le candidat le plus expérimenté sera choisi si la différence entre les QI est négligeable. Soit d la plus grande différence entre des QI que l'on considère négligeable, et soient A, B, C des candidats tels que:

	QI	Expérience
A	100	3
B	100+d	2
C	100+2d	1

Alors le décideur choisit $A \succ B, B \succ C$ et $C \succ A$. (Exemple extrait de [Tversky 1969]).

Ce premier axiome a également pour conséquence la transitivité de la relation d'indifférence. Si la question de la transitivité de la relation de préférence est normativement une condition raisonnable, celle de la transitivité de la relation d'indifférence est plus délicate. Commençons par définir formellement ce que nous entendons par indécision et indifférence [Shafer 1991], afin de clarifier le propos. Étant donné un individu muni d'une relation de préférence $\succ$, on dira;

– qu'il est indécis entre f et g si l'on n'a ni $f \succ g$ ni $g \succ f$,
– qu'il est indifférent entre f et g si et seulement si il est indécis et si il peut dans toute comparaison substituer f à g, i.e. $f \succ h$ ssi $g \succ h$ et $h \succ f$ ssi $h \succ g$.

Dans ces conditions, la transitivité de la relation d'indifférence signifie que le décideur est indifférent entre chaque paire d'actions entre lesquelles il est indécis. C'est ce point particulier qui n'est pas empiriquement toujours valide: un individu peut être indécis car il ne sait pas comparer les différentes alternatives, ne possédant pas assez de données pour choisir, ou pour toute autre raison.

Deuxième axiome

La justification empirique du deuxième axiome est la suivante [Jaffray 1985]: "il ne doit pas exister de résultat tellement préférable à tous les autres qu'il suffise qu'une loi le donne avec une probabilité si faible soit-elle, pour que cette loi soit préférée à toutes celles n'offrant pas la possibilité de ce résultat". Par exemple, les gens classent $P' \succ Q \succ P''$ où:

- $P'(\{c'\}) = 1$ où c'= vie de renoncement suivie du bonheur éternel,
- $P''(\{c''\}) = 1$ où c"= vie de renoncement suivie du néant,
- $Q(\{c\}) = 1$ où c = vie de plaisir suivie du néant,

et admettent cependant que, pour tout $\lambda \in]0,1[, \lambda P' + (1-\lambda)P'' > Q$.

Cet axiome de continuité est implicitement remis en cause car il suppose de la part du décideur des capacités cognitives infinies. Une conséquence immédiate en est en effet qu'il existe un unique $\lambda \in]0,1[$ tel que $P' \succ Q \succ P'' \Rightarrow \lambda P' + (1-\lambda)P'' \sim Q$. Si l'on prend par exemple:

- $P'(\{c'\}) = 1$ où c'= gain de 1000 Frs,
- $P''(\{c''\}) = 1$ où c"= gain de 0 Frs,
- $Q(\{c\}) = 1$ où c = gain de 400 Frs,

on peut très bien être indifférent pour $\lambda = 0.7$ entre Q et $\lambda P' + (1-\lambda)P''$, mais on le sera certainement aussi pour $\lambda = 0.69999$.

Troisième axiome

La justification empirique du troisième et dernier axiome est que, si le choix est indifférent en ce qui concerne la deuxième issue, il doit être déterminé par les préférences relatives à la première. Si l'on reprend l'interprétation que nous avons donnée du mixage d'ordre 2, cet axiome signifie que:

- si le tirage de probabilité $(1-\lambda)$ est observé, le choix préalable de P ou Q est sans effet sur la comparaison,
- si le tirage de probabilité λ est observé, selon les préférences préalables entre P et Q, l'un des deux sera observé.

Des contre-exemples systématiques de cet axiome peuvent être construits en considérant la manière différente dont un décideur traite certitude et incertitude: une somme certaine, acquise directement, est parfois plus attrayante que la perspective de gagner plus, tout en prenant plus de risques. L'exemple suivant illustre ce type de raisonnement:

Considérons les préférences entre les loteries suivantes:

<u>Préférence entre:</u>	$Q(500000)$	$=$	1	$P(2500000)$	$=$	0.10	
				$P(500000)$	$=$	0.89	
				$P(0)$	$=$	0.01	Il est
<u>Préférence entre:</u>	$R(500000)$	$=$	0.11	$S(2500000)$	$=$	0.10	
	$R(0)$	$=$	0.89	$S(0)$	$=$	0.90	

fréquent de trouver des décideurs tels que $P \prec Q$ et $R \prec S$.
Considérons maintenant les loteries suivantes, T et V:

$$\begin{array}{lll} T(2500000) & = & \frac{10}{11} \\ T(0) & = & \frac{1}{11} \end{array} \quad V(0) = 1$$

Alors

$$\begin{aligned} Q &= 0.11Q + 0.89Q \\ P &= 0.11T + 0.89Q \\ R &= 0.11Q + 0.89V \\ S &= 0.11T + 0.89V \end{aligned}$$

Or, d'après l'axiome d'indépendance, on a donc $P \prec Q \Rightarrow T \prec Q$ et $R \prec S \Rightarrow Q \prec T$. (Exemple extrait de [Allais 1953]).

Avant de donner le théorème de [Neumann and Morgenstern 1947], rappelons la définition suivante:

Définition 5. Une fonction numérique f, définie sur un ensemble mixable $\mathcal{P}$, est dite linéaire lorsque pour tous $P', P'' \in \mathcal{P}$ et $\lambda \in [0,1]$

$$f(\lambda P' + (1-\lambda)P'') = \lambda f(P') + (1-\lambda) f(P'')$$

Théorème de [Neumann and Morgenstern 1947]

Théoreme A. *[Neumann and Morgenstern 1947] Sous les axiomes 1, 2, 3, il existe sur $(\mathcal{P}, \succ)$ une fonction d'utilité linéaire, unique à une transformation affine strictement croissante près.*

Avant de donner la démonstration de ce théorème, remarquons qu'il repose sur une proposition intuitive de l'idée de mesure, à savoir que pour attribuer une représentation numérique à un ensemble d'objets, il suffit:

- de savoir comparer les éléments les uns par rapport aux autres, ce qui est une conséquence de l'axiome de continuité comme le montre les propositions iii et iv de la démonstration (voir annexe A.1).
- de se fixer un zéro et une unité de mesure ensuite, ce qui justifie intuitivement l'unicité à une transformation affine strictement croissante près de la fonction d'utilité.

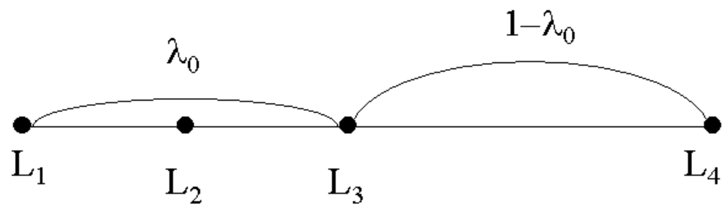


FIG. 1.1: *Comparaison des loteries*

Preuve. Voir annexe A.1. ■

La proposition suivante permet d'étendre le résultat précédent, avec donc la même interprétation intuitive que précédemment, à l'ensemble des conséquences C, en remarquant qu'une loterie "certaine" et une conséquence se correspondent.

Si l'on considère par $\mathcal{P}_0$ l'ensemble des lois de probabilité discrète à support fini, on peut exprimer facilement son utilité, $\mathcal{P}_0$ contenant en particulier toutes les lois donnant un résultat avec une probabilité 1. On note Φ_c la loi telle que $\Phi_c(\{c\}) = 1$.

Proposition 6. *Le pré-ordre total sur $\mathcal{P}$ induit sur l'ensemble des résultats C une relation de pré-ordre large $\succ_C$ définie par:*

$$c', c'' \in C;\ c' \succeq_C c'' \Leftrightarrow \Phi_{c'} \succeq \Phi_{c''}$$

A chaque fonction d'utilité linéaire U sur $\mathcal{P}$ pour $\succeq$ correspond une fonction d'utilité u sur C pour $\succeq_C$, dite <u>utilité de Von Neumann</u>, définie par $c \in C \to u(c) = U(\Phi_c)$

Proposition 7. *Étant donné un pré-ordre large total $\succeq$ sur $\mathcal{P}$, une fonction d'utilité linéaire U le représentant, u l'utilité de Von Neumann associée, l'utilité d'une loi $P \in \mathcal{P}_0$, prenant des résultats $(c_i)_{1 \leq i \leq n}$, avec des probabilités respectives $P(\{c_i\})$ de somme 1, peut s'exprimer par*

$$U(P) = P(\{c_1\})u(c_1) + \ldots + P(\{c_n\})u(c_n)$$

Preuve. Soit $G \in \mathcal{G}$.
Alors $P(G) = P(\{c_1, \ldots, c_n\} \cap G) = P(\{c_1\} \cap G) + \ldots + P(\{c_n\} \cap G)$ et:

- $c_i \in \mathcal{G} \Rightarrow P(\{c_i\} \cap G) = P(\{c_i\})$ et $\Phi_{c_i}(G) = 1$,
- $c_i \notin \mathcal{G} \Rightarrow P(\{c_i\} \cap G) = 0$ et $\Phi_{c_i}(G) = 0$

ce qui permet d'exprimer la loi P comme le mixage d'ordre n $P = P(\{c_1\}\Phi_{c_1} + \ldots + P(\{c_n\}\Phi_{c_n}$. D'où

$$\begin{aligned} U(P) &= P(\{c_1\}U(\Phi_{c_1}) + \ldots + P(\{c_n\}U(\Phi_{c_n}) \\ &= P(\{c_1\}u(c_1) + \ldots + P(\{c_n\}u(c_n) \end{aligned}$$

■

Effectuons une remarque sur cette fonction d'utilité de Von Neumann. Une propriété de la fonction d'utilité u, qui se conserve par transformations affines strictement croissantes, et qui est une propriété de $\succ_C$, concerne la concavité ou la convexité de u [Jaffray 1985]. En effet, u est concave lorsque, pour tous

$$c, c' \in C, \lambda \in [0,1], u(\lambda c + (1-\lambda)c') \succeq \lambda u(c) + (1-\lambda)u(c')$$

Il en résulte, si l'on considère les lois $\Phi_c, \Phi_{c'}$, et $\Phi_{\lambda c+(1-\lambda)c'}$ et le mixage $\lambda\Phi_c + (1-\lambda)\Phi_{c'}$, on a:

$$U(\Phi_{\lambda c+(1-\lambda)c'}) \succeq \lambda U(\Phi_c) + (1-\lambda)U(\Phi_{c'}) = U(\lambda\Phi_c + (1-\lambda)\Phi_{c'})$$

Le décideur préfère donc avoir avec une probabilité 1 le résultat $\lambda c + (1-\lambda)c'$ plutôt que d'avoir c ou c' avec les probabilités respectives λ ou $1-\lambda$. *A un résultat aléatoire d'espérance mathématique donnée, il préfère la certitude de cette espérance mathématique.* On dira donc:

Définition 8. Un décideur a de l'aversion pour le risque lorsque son utilité de Von Neumann est concave.

et réciproquement:

Définition 9. Un décideur a de l'inclinaison pour le risque lorsque son utilité de Von Neumann est convexe.

On retrouve donc ici un résultat bien connu, à savoir que des comportements satisfaisant les axiomes de la théorie de la décision sont compatibles avec une hypothèse de classement des actions par espérance d'utilité. Ce même résultat a été étendu, comme nous allons le voir maintenant, au cas de l'incertitude par [Savage 1954].

1.3.2 La modélisation de la décision en incertitude

La construction de Savage repose sur sept postulats à partir desquels il déduit l'existence d'une mesure de probabilité subjective sur l'ensemble des états de la nature, d'une mesure de préférence ou utilité sur l'ensemble des conséquences et d'une représentation numérique des préférences entre les actes, l'espérance mathématique d'utilité.

Avant d'énoncer les postulats de [Savage 1954], commençons par énoncer et rappeler quelques définitions. On note $\mathcal{E}$ une famille de parties de Ω (les événements). La théorie qui suit ayant besoin que l'ensemble de choix soit suffisamment riche, on pose donc les définitions préalables suivantes:

Définition 10. Un acte a est dit acte en escalier (resp. acte en escalier généralisé) lorsqu'il existe une partition finie (resp dénombrable) de Ω, $\{E_i, i \in I\}$ tel que pour tout $i, E_i \in \mathcal{E}$ et telle que $f(E_i) = \{c_i\}$, ensemble à un élément pour tout i.

On appelle greffe l'opération associant à deux actes a et b et à un événement E l'acte c défini par

$$c(w) = a(w) \text{ pour tout } w \in E;\ c(w) = b(w) \text{ pour tout } w \in E^C$$

On définit en outre les actes constants qui nous permettent de donner un sens aux préférences dans le certain du décideur. En effet, A contient comme actes en escalier particuliers, les actes constants et les actes à un seul résultat. On notera a_c l'acte constant c, tel que $a(\Omega) = c$. On définit dans C une relation de préférences $\succeq_C$ par $c' \succeq_C c'' \Leftrightarrow a_{c'} \Leftrightarrow a_{c''}$, qui s'interprète comme les préférences dans le certain du décideur.

Axiomatique 11. *[Savage 1954]*
Nous utiliserons la présentation de[Jaffray 1985]:

Axiome 1 *(Existence d'un ordre complet)*

- *$\mathcal{E}$ est une σ algèbre,*
- *A contient les actes en escalier, les actes en escalier généralisé et est fermé par greffe.*

– $\succeq$ *est un pré-ordre large total.*

Axiome 2 *(Axiome d'indépendance)*
Si des actes a, a', b, b' *et un événement* E *vérifient*

$$\begin{array}{ll} a_{|E} = a'_{|E} & a_{|E^C} = b_{|E^C} \\ b_{|E} = b'_{|E} & a'_{|E^C} = b'_{|E^C} \end{array}$$

Alors $a \succeq b \Leftrightarrow a' \succeq b'$. *Les préférences entre deux actes ne dépendent donc que des parties qui les distinguent. Définissons donc pour chaque événement* E *une relation de pré-ordre large* $\succeq_E$, *que l'on interprète comme les préférences du décideur sachant* E *réalisé, par:*

$$a \succeq_E b \Leftrightarrow [\forall a', b' \in A \ t.q, \quad \begin{array}{c} a_{|E} = a'_{|E} \\ b_{|E} = b'_{|E}, \\ a'_{|E^C} = b'_{|E^C} \end{array} \quad a' \succeq b']$$

Axiome 3 *(Ordonnancement des conséquences)*
Pour que cette interprétation de la relation de préférence sachant un événement E *réalisé soit compatible avec les préférences dans le certain du décideur, il faut que pour tous résultats* c', c'' *et tout événement* E *non nul,*

$$a_{c'} \succeq_E a_{c''} \Leftrightarrow c' \succeq_C c''$$

Axiome 4 *(Non trivialité de* $\succeq_C$*)*
Il existe $c, c' \in C$ *tels que* $c' \succ_C c$.

Axiome 5 *(Comparaison des événements)*
Pour tous résultats $c' \succ_C c''$ *et* $k' \succ_C k''$, *étant donné les actes* a_E *et* $a_{E'}$ *tels que:*

$$\begin{array}{ll} a_E(w) = c' \ pour \ w \in E, & a_E(w) = c'' \ pour \ w \in E^C \\ a_{E'}(w) = c' \ pour \ w \in E', & a_{E'}(w) = c'' \ pour \ w \in E'^C \end{array}$$

et les actes b_E *et* $b_{E'}$ *tels que:*

$$\begin{array}{ll} b_E(w) = k' \ pour \ w \in E, & b_E(w) = k'' \ pour \ w \in E^C \\ b_{E'}(w) = k' \ pour \ w \in E', & b_{E'}(w) = k'' \ pour \ w \in E'^C \end{array}$$

on a alors:

$$a_E \succeq a_{E'} \Leftrightarrow b_E \succeq b_{E'}$$

On interprète intuitivement cet axiome en introduisant la notion de prime. En effet, on dira qu'un acte a_E *offre une prime* (c'/c'') *lorsque* E *est réalisé si* a_E *est définit par:*

$$a_E(w) = c' \ pour \ w \in E, a_E(w) = c'' \ pour \ w \in E^C$$

L'axiome traduit l'intuition que le décideur qui préfère une prime (c'/c'') *lorsque l'évènement* E *est réalisé à la même prime si* E' *est réalisé est un décideur qui juge* E *plus probable que* E'. *Il faut pour cela que cette préférence ne dépende pas des montants de la prime, ce que traduit cet axiome. Cet axiome, combiné à la non-trivialité de* $\succeq$, *permet maintenant de définir une relation d'ordre* $\succeq_{\mathcal{E}}$ *dans* $\mathcal{E}$ *par:*

$$E \succeq_{\mathcal{E}} E' \Leftrightarrow \forall c, c', \ c \succeq_C c', a_E \succeq a_{E'}$$

avec a_E *(resp* $a_{E'}$*) acte offrant une prime* (c/c') *si* E *(resp si* B*).*

Axiome 6 *(Divisibilité et continuité)*
Pour tout couple d'actes a, b tels que $a \succ b$ et tout résultat c, on peut trouver une partition finie de l'événement certain $\{E_i \in \mathcal{E}, i \in I\}$ telle que pour tout $i \in I$:

1.
$$a_i \succ b, \qquad \text{où } a_i \text{ est l'acte défini par}$$
$$a_i(w) = a(w) \text{ pour } w \notin E_i, a_i(w) = c \text{ pour } w \in E_i$$

2.
$$a \succ b_i, \qquad \text{où } b_i \text{ est l'acte défini par}$$
$$b_i(w) = b(w) \text{ pour } w \notin E_i, b_i(w) = c \text{ pour } w \in E_i$$

Axiome 7 *(Axiome de dominance)*
Pour tout événement E.

1. *Si $a \succ_E a_c$ pour tout $c \in b(E)$, alors $a \succeq_E b$,*
2. *Si $a \prec_E a_c$ pour tout $c \in b(E)$, alors $a \preceq_E b$,*

Comme nous venons de le faire dans le cas de la théorie de [Neumann and Morgenstern 1947], nous présenterons tout d'abord le support descriptif de la théorie, ainsi que les contre-exemples qui permettent de mettre en doute leur validité empirique. Nous ne discuterons pas des axiomes 6, 7 qui sont des axiomes purement mathématiques, ni de l'axiome d'indépendance qui a déjà été abondamment discuté dans la section précédente.

Premier axiome

Le premier axiome suppose que le décideur est capable de classer toutes les actions. Cet axiome, auquel est associé la transitivité des relations de préférence et d'indifférence entre actions associées, renvoie à la discussion que nous avons déjà menée dans le cas du risque.

La contrainte supplémentaire que rajoute l'axiomatique de [Savage 1954] est la possibilité de comparer les actes constants, actes qui jouent un rôle important dans les axiomes 3, 4 et 7. Il est en effet possible d'observer que certains actes constants ne correspondent à aucune action concevable [Jaffray 1985]. Par exemple, si $E = \{Pluie,\ Beau\,Temps\}$, et que les actions sont "prendre ou non un parapluie", il est difficile d'imaginer une décision donnant s'il fait beau le résultat "être mouillé". Même s'il semble en effet possible que le décideur ait des préférences entre toutes les actions, il est peu probable qu'il ait des préférences sur des actions qui correspondent à des choix qu'il n'a jamais essayé. On peut toutefois résoudre ce problème en considérant que les éléments de C sont en fait les classes d'équivalence de l'ensemble des résultats vrais (ceux dont qui décrivent la réalité) pour la relation indifférent dans le certain du décideur [Jaffray 1985].

Troisième et quatrième axiome

Ces deux postulats (extraction de la relation d'ordre sur les conséquences et sur les états de la nature de la relation de préférence) présupposent tous les deux un même postulat empirique, à savoir la séparation entre vraisemblance et valeur. Ce postulat est valide dans le cas du joueur, où les deux sont initialement séparés et ne sont liés que par l'action de choisir un pari. La généralisation de ce postulat à tout problème de décision signifie que l'on doit construire l'ensemble des états de la nature et celui des conséquences de telle sorte que l'on puisse mettre des utilités sur l'ensemble conséquence sans savoir quel état de la nature est vrai, et des probabilités sur les états de la nature sans lien avec les conséquences.

Plus précisément, le troisième postulat affirme que si le décideur préfère c' à c'' en général, il le préfère dans toute situation E. De ce fait, la vérité d'un élément de l'ensemble des états de la nature est sans conséquence sur les préférences. [Savage 1954] cite lui-même un exemple dans lequel le postulat peut être violé:

> Before going on a picnic with friends, a person decides to buy a bathing suit or a tennis racket, not having at the moment enough money for both. If we call

> possession of the tennis racket and possession of the bathing suit consequences, then we must say that the consequences of his decision will be independant of where the picnic is actually held. If the person prefers the bathing suit, this decision would presumably be reversed, if he learned that the picnic were not going to be held near water.
> *(Extrait de [Savage 1954], p 25).*

Apparement, les préférences changent donc selon que l'on considère l'état de la nature comme le vrai état ou non (le décideur change de préférence s'il sait que le pique-nique ne se déroule pas près d'un plan d'eau). [Savage 1954] défend sa théorie face à ce contre-exemple en postulant une mauvaise analyse du problème de décision:

> under the interpretation of "act" and "consequence". I am trying to formulate, this is not the correct analysis of the situation. The possession of the tennis racket and the bathing suit are to be regarded as acts, not consequences. (...) The consequences relevant to the decision are such as these: a refreshing swim with friends, sitting on a shadeless beach twiddling a brand new tennis racket while other's friends swim, etc. It seems clear that, if the analysis is carried to its limit, the question at issue must be answered in the negative.
> *(Extrait de [Savage 1954], p 25).*

Il suffit donc en fait, pour résoudre le paradoxe, de supposer que l'on peut décrire les conséquences de manière plus fine, ce qui permet de rendre leur évaluation indépendante de l'état de la nature qui est vrai. Le problème de cette suggestion est que le raffinement de C peut entraîner un raffinement de S [Shafer 1991], chaque état de la nature devant être suffisamment détaillé pour savoir quel élément de C induit chaque acte. On doit donc satisfaire simultanément les deux propositions suivantes:

- chaque état de la nature détermine quelle conséquence de C va résulter du choix d'une action,
- la valeur de la conséquence ne doit pas dépendre de quel état de la nature est vrai

et que la première incite à limiter le détail dans C, et à accroître le détail dans S, alors que la seconde incite à accroître le détail dans C et à limiter celui de S. Il n'y a pas à-priori de raison de trouver un compromis entre les deux, et la seule réponse donnée par [Savage 1954] est que "l'analyse doit être conduite à ses limites".

Théorème de [Savage 1954]

La discussion des axiomes étant maintenant menée, nous présenterons maintenant le théorème de [Savage 1954], les points clefs de la démonstration du théorème de [Savage 1954] étant donnés en annexe. Avant de présenter ce théorème, nous avons cependant besoin d'admettre la proposition suivante:

Définition 12. Soit $\mathcal{E}$ une σ-algèbre de parties de Ω, et $\succeq_{\mathcal{E}}$ une relation binaire dans $\mathcal{E}$. On dit que $\succeq_{\mathcal{E}}$ est une probabilité qualitative sur $(\Omega, \mathcal{E})$ lorsqu'elle vérifie les conditions suivantes:

Q1 *(Classement total des événements)*

- $\succeq_{\mathcal{E}}$ est un pré-ordre large total,
- $\Omega \succeq_{\mathcal{E}} \emptyset$; $\forall A \in \mathcal{E}, \Omega \succeq_{\mathcal{E}} A \succeq_{\mathcal{E}} \emptyset$

Q2 *(Condition d'additivité)*
$\forall A_1, A_2, B_1, B_2 \in \mathcal{E}$,

- $[A_1 \cap A_2 = \emptyset, A_1 \succeq_{\mathcal{E}} B_1, A_2 \succeq_{\mathcal{E}} B_2] \Rightarrow A_1 \cup A_2 \succeq_{\mathcal{E}} B_1 \cup B_2$,
- $[A_1 \cap A_2 = \emptyset, A_1 \succ_{\mathcal{E}} B_1, A_2 \succeq_{\mathcal{E}} B_2] \Rightarrow A_1 \cup A_2 \succ_{\mathcal{E}} B_1 \cup B_2$,

Q3 *(Condition de continuité)*
$\forall A, B \in \mathcal{E}, \forall (A_n)_{n \in \mathbb{N}}, [A_n \uparrow A, \forall n \in \mathbb{N}, A_n \preceq_{\mathcal{E}} B] \Rightarrow A \preceq_{\mathcal{E}} B$

L'axiome $Q1$ exprime que les événements sont totalement classés, de façon non triviale, entre l'impossible et le certain. La seconde condition $Q2$ dit que le classement du tout résulte du classement des parties, et la dernière condition $Q3$ correspond à une propriété de σ-additivité.

Proposition 13. *On dit que $A \succ_{\mathcal{E}} \emptyset$ est un atome de $(\Omega, \mathcal{E}, \succeq_{\mathcal{E}})$ s'il n'existe aucun $B \subset A$ tel que $A \succ_{\mathcal{E}} B \succ_{\mathcal{E}} \emptyset$. $(\Omega, \mathcal{E}, \succeq_{\mathcal{E}})$ étant alors un espace de probabilité qualitative sans atome, il existe une loi de probabilité et une seule compatible avec $\succeq_{\mathcal{E}}$.*

Preuve. Le squelette de la démonstration est donné en annexe (Pour la la démonstration complète, voir [Savage 1954][Fishburn 1970][Jaffray 1985]). ■

Connaissant maintenant la définition d'un espace de probabilité qualitative sans atomes, on peut alors donner le théorème suivant:

Théoreme B. *[Savage 1954] Sous les axiomes 1 à 7, il existe une fonction d'utilité représentant $\succeq$, donnée par une expression de la forme*

$$a \in A \longmapsto U(a) = E_P U(a(.))$$

où E_P désigne l'opérateur mathématique d'espérance, P est une loi de probabilité sur $(E, \mathcal{E})$. $P(.)$ est alors nécessairement l'unique loi de probabilité compatible avec $\succeq_{\mathcal{E}}$ en se ramenant aux axiomes du théorème A. La fonction U est unique à une transformation affine strictement croissante près.

Preuve. On démontre en deux temps ce théorème de [Savage 1954], d'abord en montrant l'existence de la loi de probabilité compatible avec $\succeq_{\mathcal{E}}$, puis en démontrant que les actions peuvent être représentées par une hypothèse d'espérance d'utilité. Le schéma général de la démonstration est donné en A.2. ■

1.4 Décision humaine et modèles théoriques

L'impossibilité pour le modèle de l'utilité espérée de décrire la réalité des choix humains a amené un troisième concept en théorie de la décision, à savoir la nécessité de ne plus seulement distinguer entre deux interprétations d'un modèle de décision, descriptif et normatif, mais de distinguer une interprétation prescriptive. Là où le normatif consiste à examiner quelles axiomatiques sont liées à un comportement idéalement logique et rationnel, là où le descriptif consiste à examiner comment et pourquoi un individu agit comme il agit, le prescriptif consiste à rechercher, pour le compte d'un décideur et dans une situation donnée, une décision qui convienne le mieux possible à ce décideur, en choisissant une rationalité argumentée [Bell et al. 1991]. Cette notion fait en quelque sorte un lien entre descriptif et normatif: là où les modèles descriptifs sont évalués en fonction de qualités empiriques, où les modèles normatifs sont évalués en fonction de leur validité théorique (le degré avec lequel ils correspondent à des idéalisations de choix rationnels), les modèles prescriptifs sont évalués en fonction de leur valeur pragmatique, par leur capacité à aider les individus à faire de meilleurs choix en fonction de leurs comportements et à la lumière des idéalisations de ces choix.

A l'apparition de ce concept de modèle prescriptif, qui correspond à une réflexion pragmatique sur les manquements descriptifs de la théorie de l'utilité espérée, correspond également une réflexion théorique sous la forme de toute une série de nouveaux modèles de décision permettant de généraliser et de définir de nouvelles rationalités. Ces modèles ne répondent pas seulement aux expériences que nous avons présentées et qui contredisent les axiomes de l'utilité espérée, mais aussi à celles qui contredisent les conséquences des axiomes de cette même théorie:

Effet de proportionalité Il est issu de la comparaison entre des paires de loteries proportionnelles en probabilité. Par exemple, la comparaison entre les deux paires de loteries suivantes:

$$\begin{array}{lclllcl}
L_1(3000) & = & 1 & & L_2(4000) & = & 0.8 \\
 & & & & L_2(0) & = & 0.2 \\
L_3(3000) & = & 0.25 & & L_4(4000) & = & 0.2 \\
L_3(0) & = & 0.75 & & L_4(0) & = & 0.8
\end{array}$$

La théorie de l'utilité espérée impose $L_1 \succ L_2$ et $L_3 \succ L_4$, alors que la majorité des individus choisit $L_1 \succ L_2$ et $L_3 \prec L_4$ (voir [Kahneman and Tversky 1979]).

Effet de réflexion Il est issu de l'observation que les comportements face à un risque différent selon que les loteries engendrent des gains, des pertes ou les deux. Les individus ayant une aversion potentielle au risque, auront tendance, dans une certaine mesure, à prendre des risques devant des situations où ils pourraient perdre de l'argent. Par exemple, dans la comparaison entre les loteries suivantes:

$$\begin{array}{lclllcl} L_1(6000) & = & 0.25 & L_2(4000) & = & 0.25 \\ L_1(0) & = & 0.75 & L_2(2000) & = & 0.25 \\ L_3(-6000) & = & 0.25 & L_4(-4000) & = & 0.25 \\ L_3(0) & = & 0.75 & L_4(-2000) & = & 0.25 \end{array}$$

Les expériences montrent que la majorité des individus classe $L_2 \succ L_1$ et $L_3 \succ L_4$ (voir [Kahneman and Tversky 1979]). Ce qui a pour effet de faire de la fonction d'utilité une fonction concave pour les gains et convexes pour les pertes.

Surévaluation des petites probabilités Il est issu du fait que les individus ont tendance à surévaluer les loteries où les gains importants ont une probabilité très faible. Par exemple, dans la comparaison entre les loteries suivantes:

$$\begin{array}{lcllcl} L_1(5000) & = & 0.001 & L_2(5) & = & 1 \\ L_1(0) & = & 0.999 & & & \\ L_3(-5000) & = & 0.001 & L_4(-5) & = & 1 \\ L_3(0) & = & 0.999 & & & \end{array}$$

Les expériences montrent que la majorité des individus classe $L_1 \succ L_2$ et $L_4 \succ L_3$ (voir [Kahneman and Tversky 1979]), alors que ces couples de loteries ont les mêmes utilités espérées.

Effet d'isolement Il est issu de l'observation que les individus fondent leurs décisions entre des loteries sur les conséquences qui les distinguent. Ils isolent ainsi les conséquences communes aux loteries, afin de ne révéler des préférences que sur les conséquences restantes. Or des loteries formellement équivalentes peuvent de cette manière être décomposées de manière différente. Considérons par exemple le problème suivant:

> Considérons le jeu suivant. Dans un premier temps, il y a une probabilité 0.75 de terminer le jeu sans rien gagner, et une probabilité 0.25 d'accéder à la deuxième phase du jeu. Dans cette dernière, un choix est offert entre les loteries suivantes:
>
> $$\begin{array}{lcllcl} L_1(4000) & = & 0.80 & L_2(3000) & = & 1 \\ L_1(0) & = & 0.20 & & & \end{array}$$
>
> Le choix doit être fait avant que le résultat de la première phase de jeu soit connue.
>
> Considérons maintenant le choix entre les loteries suivantes:
>
> $$\begin{array}{lcllcl} L_3(4000) & = & 0.20 & L_4(3000) & = & 0.25 \\ L_3(0) & = & 0.80 & L_4(0) & = & 0.75 \end{array}$$
>
> (Exemple extrait de [Kahneman and Tversky 1979]).

On notera que les loteries L_1 et L_3 d'une part, L_2 et L_4 d'autre part sont identiques. Or les sujets choisissent $L_2 \succ L_1$ et $L_3 \succ L_4$.

Nous ne reviendrons pas dans cette partie sur l'ensemble des expériences qui démontrent que le raisonnement humain ne se fait pas sur la base du calcul mathématique des probabilités (le lecteur est pour ce faire renvoyé à [Tversky and Kahneman 1982][Hogarth 1988]). Ces expériences ont conduit à des généralisations du modèle de l'utilité espérée. Ces généralisations, et leurs comparaison, sont présentées en annexe B. Pour en donner un rapide aperçu, l'incapacité de la théorie de la décision à décrire effectivement la décision humaine a été étudiée principalement de deux manières:

L'intransitivité de la relation de préférence Elle est pour certains auteurs (voir section B.4.1) liée à une réaction psychologique devant le risque de prendre une mauvaise décision. Ces modèles modifient alors le paradigme de l'utilité espérée pour prendre en compte ce type de raisonnement,

L'affaiblissement ou le rejet de l'axiome d'indépendance Ces modèles rejettent la propriété liée à l'axiome d'indépendance, à savoir la linéarité en probabilités de la fonctionnelle d'indépendance. Chacun de ces modèles (voir section B.4.2 et section B.4.3) sont elles aussi fondées sur la modélisation d'un certain type de raisonnement dans la décision.

Chacune de ces généralisations étant fondée sur la modélisation d'un certain type de raisonnement qui guide le choix des axiomes, il ne peut modéliser des raisonnements qui ne sont pas directement conséquence de ces axiomes. C'est ce que démontre la comparaison expérimentale de ces généralisations (voir section B.5); en effet, des études cherchant à déterminer laquelle de ces généralisations est la plus proche de la décision humaine ont montré:

- qu'aucun de ces modèles n'est meilleur que celui de l'utilité espérée pour représenter l'ensemble des choix humains,
- que certains de ces modèles sont meilleurs que les autres pour représenter certains choix, choix qui sont fondés sur le type de raisonnement que la fonctionnelle d'utilité modélise.

L'idée qui ressort de ces expériences est que les individus utilisent une heuristique différente selon les problèmes de choix auxquels ils sont confrontés. Cette idée aborde au moins implicitement l'idée de représentation du résultat d'un *raisonnement* en changeant la fonctionnelle de préférence $f(x)$, où x représente les différents alternatives auxquelles le décideur est confronté, par une fonctionnelle $f[y](x)$, où y est un facteur lié au problème de décision déterminant quels sont les axiomes donnant la fonctionnelle $f[y]$ représentant les choix du décideur. Dit autrement, il est alors nécessaire, non plus d'avoir une unique axiomatique déterminant la représentation des résultats choisis, mais d'avoir une "mécanique" contextuelle permettant de manipuler les différentes axiomatiques représentant les différents raisonnements utilisés pour effectuer des choix. Il faut donc représenter les raisonnements dans la décision, ce qui nécessite une logique.

Cette idée de représenter les raisonnements dans la décision a déjà été proposée dans le cadre des approches multi-étapes de [Rubinstein 1988][Leland 1994]. Le but de ces auteurs est au moins d'expliquer le paradoxe de Allais, comme par exemple dans la loterie suivante, grâce à une heuristique fondée sur la similarité.

$$\begin{array}{lclllcl} L_1(3000) & = & 1 & L_2(4000) & = & 0.8 \\ & & & L_2(0) & = & 0.2 \\ L_3(3000) & = & 0.25 & L_4(4000) & = & 0.2 \\ L_3(0) & = & 0.75 & L_4(0) & = & 0.8 \end{array}$$

La théorie de l'utilité espérée impose $L_1 \succ L_2$ et $L_3 \succ L_4$, alors que la majorité des individus choisit $L_1 \succ L_2$ et $L_3 \prec L_4$. En effet, ce que [Rubinstein 1988] propose, pour la comparaison des loteries de la forme $(p, x;\ 1-p, 0)$[5], un mode de supposer que les décideurs raisonnent prioritairement en fonction de relations de similarités entre les probabilités et entre les conséquences. 0.25 et 0.20 étant similaires, le choix se porte naturellement sur la loterie de prix le plus élevé, expliquant ainsi la préférence $L_3 \prec L_4$. Les conséquences sont restreintes à $[0,1]$, et une fonction de similarité est alors définie sur les probabilités $\sim_p$ et une autre sur les conséquences $\sim_x$. L'heuristique de choix prend alors la forme suivante [Rubinstein 1988]:

1. Étape 1:
   - Si $x_1 > x_2$ et $p_1 > p_2$ alors $L_1 > L_2$.
   - Sinon, aller à l'étape 2.
2. Étape 2:
   - Si $x_1 \sim_x x_2$ et $\neg(p_1 \sim_p p_2)$ et $p_1 > p_2$ alors $L_1 > L_2$.
   - Si $p_1 \sim_p p_2$ et $\neg(x_1 \sim_x x_2)$ et $x_1 > x_2$ alors $L_1 > L_2$.
   - Sinon, aller à l'étape 3.

5. Lire probabilité de gagner x avec la probabilité p, probabilité de gagner 0 avec la probabilité $1-p$.

3. Étape 3: Non spécifiée.

[Rubinstein 1988] montre que cette procédure de décision, bien que non entièrement spécifiée, suffit à imposer une forme bien particulière pour que l'ensemble des relations de préférences soit compatible avec elle. La préférence produite peut alors être représentée par une fonction d'utilité de la forme $g(p)u(x)$, g étant une fonction de déformation des probabilités. Réciproquement, si la relation de préférence est représentée par $g(p)u(x)$, alors $\sim_p$ et $\sim_x$ sont des similarités de type ratio, i.e. telle que $a \sim b \Leftrightarrow \exists\lambda, \exists H, 1/\lambda < H(a)/H(b) < \lambda$.

Pour être cohérent avec la théorie de l'utilité espérée, la similarité sur les probabilités doit donc être de type ratio, ce que contredit à-priori le paradoxe de Allais puisque $0.8/1 = 0.2/0.25$ et que les probabilités 0.20 et 0.25 sont perçues comme étant similaires alors que les probabilités 0.8 et 1 sont perçues comme dissimilaires. Cette approche permet donc de justifier la non validité du paradoxe de Allais par le biais d'un raisonnement, mais sans cependant introduire la mémoire des cas passés dans la décision, alors qu'il a été montré par les psychologues [Klein 1993][Klein et al. 1993] que les cas passés jouaient un grand rôle dans la décision.

1.5 Conclusion

Les affaiblissements des axiomes de l'utilité espérée ne suffisent pas non plus à expliquer le comportement réel des décideurs. L'hypothèse que l'on pose, et que nous développerons dans les chapitres suivants, est que la modélisation d'un problème de décision par les trois ensembles (état de la nature, action, conséquence) proposée par la théorie de la décision est à la base de cette incapacité à décrire les comportements réels. Pour remédier à ce problème, nous proposons:

- de prendre en compte le contenu de la mémoire dans la formalisation; alors que dans la théorie de l'utilité espérée, l'ensemble des états de la nature doit résoudre toute l'incertitude, dans notre cas, seuls sont connus les problèmes de décision que le décideur a effectivement rencontrés. La mémoire sera structurée sous forme de cas, un cas représentant un problème de décision mémorisé décrit par un triplet (problème, action, résultat),

- de prendre en compte les interactions avec la mémoire dans la formalisation; confronté à un problème de décision, le décideur va rechercher dans sa mémoire des cas analogues au problème courant et entreprendre la meilleure action des problèmes similaires conformément à ses préférences. Ce type de modélisation a été avancée par [Klein et al. 1993] comme un modèle naturel de décision, et en particulier, la décision suscité par la reconnaissance d'un problème mémorisé est une caractéristique de la décision humaine, reconnue par les psychologues [Klein 1993][Klein et al. 1993][Lipshitz 1993] aussi bien que par les neurobiologistes [Montgomery 1983].

2

Approches symboliques pour la décision qualitative

2.1 Introduction

Nous avons vu dans le chapitre précédent les problèmes liés à la formalisation adoptée par la théorie de la décision et par ses généralisations. La décision humaine s'écarte sensiblement et sur plusieurs points de ce qu'attend la théorie. Ces écarts ne sont pas seulement dûs à de mauvaises évaluations de probabilités ou d'utilités, mais on doit prendre en compte l'influence des préférences sur le raisonnement. Ce chapitre va donc être consacré à définir le cadre dans lequel préférences et raisonnements vont être formalisés.

Pour ce faire, nous allons voir qu'il existe, outre les problèmes de formalisation en théorie de la décision que nous avons vus au chapitre précédent, des problèmes de modélisation liés à la manière dont un problème de décision est défini par le choix des ensembles états de la nature, actions, conséquences. Utilisant alors l'hypothèse de raisonnement à partir de cas que nous proposons, nous verrons la manière de définir les ensembles sur lesquels repose une formalisation descriptive d'un problème de décision. Ayant choisi une approche mathématico-symbolique, nous verrons ensuite la représentation proposée à la fois des connaissances des individus et des préférences.

2.2 La théorie de la décision: problèmes de formalisation

L'essentiel de la critique repose sur le fait que, dans un problème de décision dans l'incertain, le langage de la théorie de la décision est souvent mal approprié [Gilboa and Schmeidler 1995]. En effet, dans un problème de décision en milieu incertain, ni l'ensemble des états de la nature, ni celui des conséquences ne sont simples à énumérer ou générer.

Reprenons l'exemple de [Gilboa and Schmeidler 1995] pour nous en persuader. Le président Clinton doit décider d'une intervention militaire en Bosnie. L'ensemble des actions possibles est à peu près clair:

1. ne rien faire,

2. imposer des sanctions économiques,

3. utiliser une force militaire limitée (i.e. seulement des frappes aériennes),

4. opter pour une intervention massive.

Ni l'ensemble des états de la nature, ni celui des conséquences ne sont simples à imaginer. Par exemple, en ce qui concerne les conséquences, il est difficile d'estimer combien de blessés chaque option militaire peut occasionner et quelle serait la réaction de l'opinion publique. En ce qui concerne l'ensemble des états de la nature, il faut déterminer les capacités militaires de chaque faction militaire bosniaque, ou la réaction de la Russie face à chaque action militaire.

De ce fait, la théorie de l'utilité espérée n'est pas un bon modèle descriptif dans ce scénario. D'autant plus que la définition des trois ensembles n'est que le préalable à la formalisation: encore faut-il que le décideur détermine l'utilité associée à chaque résultat et forme une distribution de probabilité sur l'ensemble des états de la nature [Gilboa and Schmeidler 1995]. Par exemple, quelles sont les chances qu'une intervention militaire dégénère en véritable conflit armé, ce qui ne serait peut être pas le cas avec quelques frappes aériennes?

Notons également deux autres problèmes de modélisation liés à la théorie de la décision et à ses généralisations:

- le premier est lié au fait que l'ensemble des états de la nature doit résoudre toute l'incertitude. Considérons en effet la manière dont l'information évolue au cours du temps. Dans la théorie de l'utilité espérée, une nouvelle information est modélisée comme un événement (i.e. un ensemble d'états de la nature) et les probabilités sont mises à jour grâce au théorème de Bayes. Ainsi, implicitement, la théorie de l'utilité espérée suppose que les individus naissent avec une connaissance de chacun des événements et de leur probabilité, et que l'apprentissage consiste à éliminer (via le théorème de Bayes) les événements qui ne sont plus valides,

- le deuxième est lié au fait que le théorème de [Savage 1954] est intéressant dans la mesure où, si l'on fait des hypothèses de choix rationnel, on peut alors en déduire qu'il existe des probabilités sur l'ensembles des états de la nature, et que les actes sont classés selon une hypothèse d'utilité espérée. Le problème est que la réponse à la question empirique: *"est-ce que les préférences des individus sont suffisamment détaillées et définies, et s'accordent suffisamment bien avec les postulats de [Savage 1954] pour que probabilités et utilités les résument?"* reçoit, d'après le contenu expérimental que nous avons donné (voir chapitre précédent), une réponse négative.

La prise en compte de ces remarques passe par une théorie plus adaptée à la manière dont les individus pensent. Cette approche a déjà été avancée par [Gilboa and Schmeidler 1995]. En effet, à la question "comment les individus pensent-ils dans un problème de décision?", [Gilboa and Schmeidler 1995] propose d'en revenir à la proposition de Hume: *"De causes qui apparaissent similaires, on attend des effets similaires. Telle est la somme de toutes nos conclusions expérimentales"*. La technique de raisonnement proposée est donc le raisonnement par analogie. Reprenons par exemple le cas de l'intervention militaire en Bosnie, où le raisonnement par rapport à des cas passés a déjà été utilisé par des experts[6]:

- ceux favorables à une intervention militaire ont cité la guerre du Golfe comme un cas de succès: ils se fondaient sur la similarité entre ces deux conflits locaux et en post guerre froide,

- ceux opposés à une intervention militaire citaient eux la guerre du Vietman où l'intervention militaire est généralement considérée comme une erreur: ils se fondaient eux aussi sur la similarité entre les deux cas, i.e. la possibilité de guérilla à cause d'un terrain difficile.

En termes triviaux, la question que nous voudrions étudier dans ce travail est la possibilité de définir un cadre formalisé dans lequel on puisse dire si la situation bosniaque est plus proche de la situation vietnamienne ou de la guerre du Golfe. Comme l'intervention militaire en Bosnie et celle au Vietnam ou dans le Golfe ont des éléments permettant de les différencier, la question qui se pose est: ces différences sont elles ou non négligeables en regard de la décision à prendre? Par exemple, un décideur qui estime que les liens entre la Russie et le pays où se déroule l'intervention influent sur la similarité, va choisir de ne pas les négliger dans le cas de la Bosnie si son modèle est le Vietnam et ses préférences vont vers la non intervention. Les préférences vont alors avoir pour conséquence qu'au moins cette différence entre une intervention dans le Golfe et en Bosnie ne peut être négligée, ce qui conduit à ne pas effectuer cette similarité.

Pour nous résumer, deux points sont donc à noter dans la démarche que nous proposons:

- la similarité est définie par les caractéristiques que partagent cas courant et cas en mémoire, mais aussi par la *négligeabilité* des parties qui les distinguent,

- la similarité dépend des préférences individuelles, et donc transitivement, du contexte défini par le raisonnement en cours.

6. Les critères des experts étaient bien sûr plus affinés, mais l'existence des conclusions divergentes demeure.

2.3 La modélisation proposée

Nous avons illustré dans le chapitre précédent comment la prise en compte d'un raisonnement dans la décision pouvait expliquer certaines violations expérimentales des axiomes de l'utilité espérée. Pour proposer un modèle descriptif, commençons par poser la définition d'un raisonnement dans la décision, et reconstruisons les ensembles définissant un problème de décision.

Si on reprend donc la formulation d'un problème de décision en s'appuyant sur [Pomerol 1997], on notera qu'un problème de décision a sa source dans une différence entre un état courant et un état futur plus désirable, le but de la décision étant de réduire cette différence. Considérons qu'il s'agit là de notre premier ensemble et appelons P l'ensemble des problèmes de décision. Un élément p de P est alors défini par un but et un état courant.

De ce fait, le décideur commence par reconnaître l'état courant. Rétablissons donc la distinction entre états courants et états futurs, distinction abolie par la théorie de la décision où chaque état de la nature comprend indistinctement présent et futur. Si on admet que l'état courant contient des informations sur le passé et sur le présent, sur des situations antérieures déjà rencontrées, la première phase de la décision est alors de reconnaître un ou plusieurs états proches de l'état courant perçu: on retrouve ici la partie de l'ensemble des états de la nature concernant le présent. On appellera cette phase de reconnaissance de l'état courant appariement (*pattern-matching*).

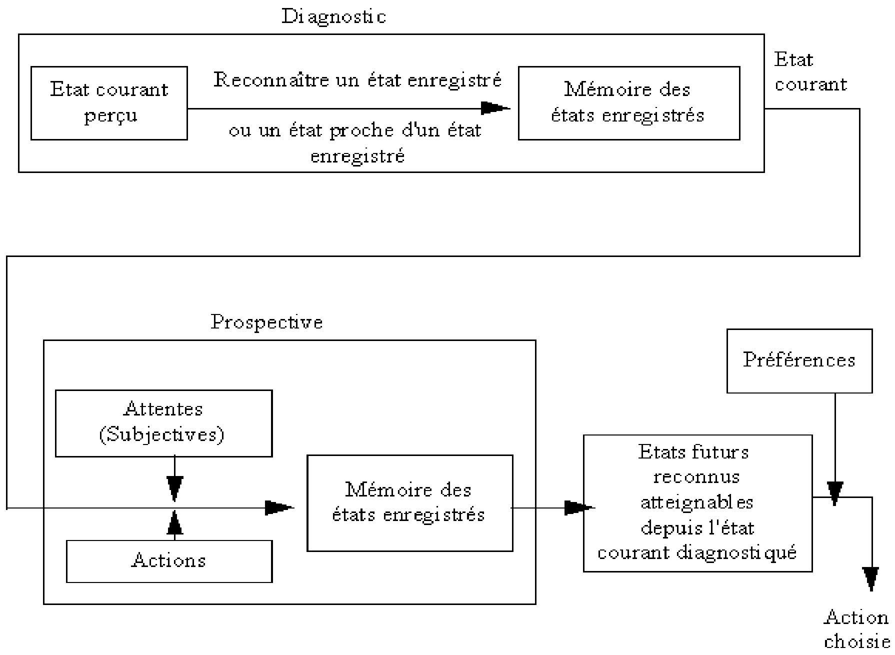


FIG. 2.1: *Une modélisation descriptive de la décision*

On peut ensuite supposer, suivant toujours en cela la démarche de [Pomerol 1997], que la phase suivante, ou phase de prospective consiste à explorer, à partir de cet état courant, les états futurs accessibles. Ces états futurs, correspondant à l'ensemble des conséquences, sont le fruit d'un mélange entre les représentations des événements futurs non contrôlés et non influencés par le sujet (la deuxième partie de l'ensemble des états de la nature en théorie de la décision, partie concernant le futur) et entre les actions possibles, états futurs, actions et représentations étant extraits de la mémoire de l'individu. La décision se modélise alors selon le schéma suivant:

On retrouve donc nos trois ensembles E, C et A et un quatrième ensemble des problèmes de décision (représentés implicitement dans la figure 2.1) par les rectangles entourant les phases de diagnostic et de prospective. L'accès aux ensembles E, C et A se fait au travers de cet ensemble qui les structure et grâce à une certaine relation de similarité entre problèmes.

Dans ce schéma, la notion de similarité est présente, à la fois dans la reconnaissance de l'état courant, et dans la recherche d'actions entreprises dans des problèmes passés. Notre hypothèse est que ces similarités vont dépendre des caractéristiques que partagent problèmes courant et problèmes mémorisés, mais aussi des parties qui les distinguent ou *négligeables*. Ces négligeables, associés aux conséquences possibles de l'action, déterminent l'action choisie par application des préférences. Deux fonctionnelles permettent alors de formaliser un tel schéma:

- *une fonctionnelle de déduction par similarité* permettant, en recherchant dans les cas passés, de déduire les actions à entreprendre dans le cas courant, puis de déduire les conséquences de telles actions,
- *une fonctionnelle de préférence* permettant de choisir une action en fonction des similarités effectuées et des préférences sur les résultats.

Ayant choisi une approche mathématico-logique, nous allons maintenant voir comment formaliser similarité et préférence. Pour ce faire, nous commençerons par étudier une adaptation [Langlotz and Shortliffe 1989] d'un exemple classique, celui de $Tweety$ [Langlotz and Shortliffe 1989]. Tout comme ces auteurs, nous nous intéressons en effet au fait que ce sont des préférences sur les conséquences du raisonnement qui permettent de raisonner par défaut en considérant qu'il n'existe pas de connaissances contradictoires avec le fait que $Tweety$ vole. En effet, il a été proposé dans la littérature [Shoham 1988][Langlotz and Shortliffe 1989][Doyle 1985] que la notion d'inférences non monotones était liée à des notions de préférences, et non plus seulement aux notions de vraisemblance et de probabilité. [Langlotz and Shortliffe 1989] propose que le fait d'agir comme si une assertion était vraie (ce qui est implicite dans un raisonnement non monotone) ne dépend pas seulement de la vraisemblance de cette proposition, mais dépend aussi des conséquences qu'induit le fait d'agir comme si une assertion était vraie. Ainsi, par exemple, si l'on se demande [Langlotz and Shortliffe 1989], dans l'exemple classique

$$\frac{\begin{array}{l} Bird(Tweety) \\ Bird(X): \quad M[Fly(X)] \end{array}}{Fly(X)}$$

Ceci exprime la connaissance "Si X est un oiseau, et en l'absence d'informations contradictoires, alors X peut voler".

de $Tweety$, quelles sont les conséquences du raisonnement si l'on suppose que $Tweety$ vole alors qu'il ne vole pas. S'il s'agit d'acheter une cage à $Tweety$, il est aisé de raisonner comme s'il volait. Si, par contre, $Tweety$ est un oiseau apprécié, serions nous aussi également prêt à raisonner comme s'il volait si on doit le lâcher du haut d'une falaise? Plus les conséquences sont importantes, plus il est difficile de raisonner comme si $Tweety$ volait, et ce, bien que la vraisemblance que $Tweety$ vole reste parfaitement constante.

2.3.1 Tweety: un exemple désormais classique

Considérons la décision de relâcher $Tweety$ ou de le laisser en cage sans cependant savoir s'il vole. On considérera de plus que le décideur a les deux cas suivants en mémoire (voir aussi figures 2(a) et 2(b)):

X est un oiseau. S'il ne vole pas et qu'il reste en cage, il risque de dépérir.	X est un oiseau. S'il vole et qu'on le relâche, il apprécie la liberté.

ce qui se traduit en logique d'ordre un par les formules suivantes [7]:

Cas numéro 1		**Cas numéro 2**	
$Oiseau(X) \wedge \neg Vole(X)$	$\Rightarrow MettreEnCage(X)$	$Oiseau(X) \wedge Vole(X)$	$\Rightarrow Relache(X)$
$Oiseau(X)$	$\wedge MettreEnCage(X)$	$Oiseau(X)$	$\wedge Relache(X)$
	$\Rightarrow Deperir(X)$		$\Rightarrow ApprecieLiberte(X)$

7. On notera que ces cas sont des raisonnements, et non seulement des "frames" ou des règles uniques isolées

(a) Mettre Tweety en cage

(b) Relâcher Tweety sachant qu'il apprécie la liberté

FIG. 2.2: *Les aventures de Tweety*

Bien qu'ayant des connaissances incomplètes, le décideur ne connaît que le fait: "Tweety est un oiseau", i.e. son cas courant est $Oiseau(Tweety)$, il souhaite quand même raisonner et peut alors le faire de deux manières:

- ou bien il raisonne comme si $Tweety$ volait, ce qui traduit le fait que la différence entre le cas courant et le cas mémorisé numéro 2 est négligeable. Le décideur en déduit que $Tweety$ appréciera la liberté,
- ou bien il raisonne comme si que $Tweety$ ne volait pas, ce qui traduit le fait que la différence entre le cas courant et le cas mémorisé numéro 2 est négligeable. Il raisonne alors comme si $\neg Vole(Tweety)$ était vrai, et en déduit qu'en étant en cage, $Tweety$ va dépérir.

Notons que les éléments qui sont négligés ont une valeur de vérité **incertaine**: *les éléments connus (donc vrais) ne peuvent être négligés.*

Les préférences du décideur vont s'exprimer par le fait que l'intérêt que l'on porte à $Tweety$ et la préférence que l'on possède sur "$Tweety$ vivant" par rapport à "$Tweety$ en train de dépérir" vont faire que l'on va négliger ou non le fait que $Tweety$ pourrait bien ne pas voler. **Les similarités étant calculées, ce sont les préférences qui vont déterminer si l'une ou l'autre des similarités va être choisie.** On suppose que le décideur a une préférence entre les résultats $Deperir(Tweety)$ et $Mort(Tweety)$, i.e. $AppecieLiberte(Tweety) \succ Deperir(Tweety)$ où $\succ$ désigne une relation binaire partielle, asymétrique[8] signifiant *"est plus préféré que"*. $\succ$ représente une relation de préférence entre des descriptions partielles du monde. Notre hypothèse est que ces préférences peuvent se mettre sous la forme d'un sextuplet:

(Pré-requis valide 1, Pré-requis négligé 1, Fait 1)
Préféré
(Pré-requis valide 2, Pré-requis négligé 2, Fait 2)

sextuplet qui signifie intuitivement, "le fait 1 résultant de la connaissance de Pré-requis 1 et du fait de négliger Pré-requis négligé 1 est préféré au fait 2 résultant de la connaissance de Pré-requis 2 et du fait de négliger Pré-requis négligé 2".Pour toujours rester informel dans cette partie de présentation intuitive, notons par $\perp$ un symbole signifiant faux (i.e. traduisant l'action ne rien faire). La préférence que nous avons mentionnée prend alors la forme syntaxique suivante:

$$\begin{gathered}(Oiseau(X), Vole(X), AppecieLiberte(X)) \\ \succ \\ (Oiseau(X), \neg Vole(X), Deperir(X))\end{gathered}$$

8. Nous avons vu dans les chapitres précédents que la validité descriptive d'un modèle exige que sa relation de préférence ne soit nécessairement ni complète, ni transitive.

Dans notre exemple, cette préférence permet de négliger $Vole(Tweety)$ plutôt que $\neg Vole(Tweety)$, et alors l'appariement est effectué comme si $Tweety$ volait, et on le relâche.

Ces six éléments (Pré-requis valide 1, Pré-requis négligé 1, Pré-requis valide 1, Pré-requis négligé 2, Fait 1, Fait 2) permettent de construire un raisonnement par similarité bâti sur les préférences. On peut cependant noter que l'ensemble des éléments négligés pris en compte dans les préférences peut très bien être vide (on préfère alors toujours le fait 1 au fait 2 dès que les pré-requis validés 1 et 2 sont établis). Les cas sont incomplètement décrits, mais complétés par des raisonnements non montones guidés par les préférences.

D'un point de vue représentation des connaissances et des préférences pour un problème de décision, les connaissances seront représentées par des séquents, qui sont des listes de symboles de la forme:

$$\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m$$

exprimant que de $\varphi_1, \ldots, \varphi_n$, on peut déduire ψ_1 ou ψ_2 ou ... ou ψ_m. Dans le cas d'une décision, ces formules correspondent respectivement à l'une des deux représentations suivantes:

- A partir de la description de problème $\varphi_1, \ldots, \varphi_n$, on peut entreprendre les actions ψ_1 ou ψ_2 ou ... ou ψ_m; dans la figure 2.1, cette formule correspond à la reconnaissance de l'état courant et au lien entre cet état courant et l'action,

- A partir de l'action ψ_1, et de la description de problème $\varphi_2, \ldots, \varphi_n$, on obtient les conséquences ψ_1 ou ψ_2 ou ... ou ψ_m; dans la figure 2.1, cette formule correspond à la relation de déduction par similarité entre d'une part, les actions et les attentes, et d'autre part, l'ensemble des états atteignables à partir de cette action.

Un sextuplet $(p_1 \wedge a_1, \varphi_1, \Phi_1) \succ (p_2 \wedge a_2, \varphi_2, \Phi_2)$ représente une préférence de la manière suivante:

- Lorsque l'action a_1 peut être entreprise en sachant au moins la description de problème p_1, résultant en au moins la conséquence Φ_1, et en négligeant φ_1,

- et si l'action a_2 peut être entreprise en sachant au moins la description de problème p_2, résultant en au moins la conséquence Φ_2, et en négligeant φ_2,

- alors la conséquence Φ_1 est préféré à la conséquence Φ_2.

La relation de similarité va s'effectuer sur les séquents, l'objet syntaxique préférence permettant lui de modéliser les préférences du décideur. Voyons quelques autre exemples d'application de l'approche que nous proposons, l'un dans le cadre de la photo-interprétation, le second dans le cadre de problème de forages pétroliers.

2.3.2 L'exemple de la photo-interprétation

L'interprétation d'images est un cas courant où le raisonnement s'appuie sur des données incomplètes, les caractéristiques permettant de reconnaître un objet étant variables d'une image à une autre en fonction du capteur, de l'angle de prise de vue, de la résolution .... On peut penser modéliser les connaissances et les préférences du décideur par les formules suivantes:

- Cas mémorisés:
  Si à un endroit donné de l'image (endroit ici représenté par le symbole l_1) se trouvent à la fois une piste d'atterrissage, un taxiway, une tour de contrôle, alors un aéroport se trouve en l_1.

  $$Piste(l_1) \wedge Taxyway(l_1) \wedge TourControle(l_1) \Rightarrow Aeroport(l_1)$$

  Si à un endroit donné de l'image se trouvent un aéroport et un aérogare, alors c'est un aéroport civil.

  $$Aeroport(l_1) \wedge Aerogare(l_1) \Rightarrow AeroportCivil(l_1)$$

Si à un endroit donné de l'image se trouvent un aéroport et des aires de stationnement d'avions non alignés, alors c'est un aéroport militaire.

$$Aeroport(l_1) \wedge StationnementNonAligne(l_1) \Rightarrow AeroportMilitaire(l_1)$$

Si à un endroit donné de l'image se trouvent des formes circulaires, d'éloignements relatifs faibles, clôturées et contenant un bâtiment de gardiennage, alors c'est un site à munitions.

$$FormesCirculaires(l_1) \wedge EloignementRel(V < 50, l_1) \wedge Cloture(l_1)$$
$$\wedge Gardiennage(l_1) \Rightarrow SiteMunition(l_1)$$

Si à un endroit donné de l'image se trouvent des formes circulaires, d'éloignements relatifs importants, clôturées, alors c'est un site à carburant.

$$FormesCirculaires(l_1) \wedge EloignementRel(V > 50, l_1) \wedge Cloture(l_1)$$
$$\Rightarrow SiteCarburant(l_1)$$

– Préférences:
  On doit voir la ou les pistes pour s'autoriser à déduire qu'il s'agit d'un aéroport: la préférence exprime donc que l'on doit rejeter tout aéroport civil ou militaire déduit en négligeant la piste.

$$(\bot, \bot, \bot) \succ (\top, Piste(l_1), AeroportCivil(l_1))$$
$$(\bot, \bot, \bot) \succ (\top, Piste(l_1), AeroportMilitaire(l_1))$$

Ainsi, les préférences permettent de prendre en compte l'incomplétude des données. Ce ne sont pas les seuls types de préférences disponibles. Par exemple, un photo-interprète militaire peut avoir comme but de détecter tous les aéroports militaires dans une image. Il peut alors avoir une préférence du type:

$$(Piste(l_1)\ \bot, AeroportMilitaire(l_1))$$
$$\succ$$
$$(Piste(l_1), \bot, AeroportCivil(l_1))$$

qui correspond à: "plutôt reconnaître un aéroport militaire en lieu et place d'un aéroport civil que d'oublier un aéroport militaire".

2.3.3 Un exemple de décision dans le cas d'un forage pétrolier

Nous reprendrons ici l'exemple de [Charreton and Bourdaire 1985] p93. Une entreprise pétrolière, en explorant une région, a mis en évidence une accumulation d'hydrocarbures liquides. Elle hésite pour le développement et la mise en production de ce champ, entre deux stratégies, qui sont les stratégies usuelles dans ce cadre (cas en mémoires):

– mettre en place immédiatement des structures légères et peu onéreuses permettant la mise en production rapidement; en contrepartie, le taux de soutirage sera limité à 200 000 tonnes par an,

– poursuivre l'appréciation du jugement par un puits supplémentaire ce qui permettra alors, soit de prendre la solution de la structure légère et peu onéreurse, soit de choisir une architecture lourde adaptée à un gisement plus important, permettant un taux de soutirage de 500 000 tonnes par an.

Les coûts du projet sont les suivants:

– Coûts réserves faibles et architecture lègère: 100,

– Coûts réserves faibles, forage du second puits et architecture lourde: 10,

- Coûts réserves importantes et architecture légère: 200,
- Coûts réserves importantes, forage du second puits et architecture lourde: 310,

Ce projet se formalise alors de la manière suivante:

$$\begin{aligned}
&UnPuits(L) \wedge Hydrocarbures(L) \wedge Quantite(L, Faible)\\
&\qquad\qquad \Rightarrow MiseEnProduction(L, StructureLegere)\\
&MiseEnProduction(L, StructureLegere) \Rightarrow TauxSoutirage(L, StructureLegere, 200000)\\
&UnPuits(L) \wedge MiseEnProduction(L, StructureLegere) \wedge Quantite(L, Faible)\\
&\qquad\qquad \Rightarrow Cout(L, 100)\\
&UnPuits(L) \wedge MiseEnProduction(L, StructureLegere) \wedge Quantite(L, Important)\\
&\qquad\qquad \Rightarrow Cout(L, 200)\\
&PremierPuits(L) \wedge Hydrocarbures(L) \wedge Quantite(L, Important) \Rightarrow DeuxPuits(L)\\
&DeuxPuits(L) \wedge Hydrocarbures(L) \wedge Quantite(L, Importante)\\
&\qquad\qquad \Rightarrow MiseEnProduction(L, StructureLourde)\\
&MiseEnProduction(L, StructureLourde) \Rightarrow TauxSoutirage(L, StructureLourde, 500000)\\
&MiseEnProduction(L, StructureLourde) \wedge Quantite(L, Important) \Rightarrow Cout(L, 310)\\
&SecondPuits(L) \wedge Hydrocarbures(L) \wedge Quantite(L, Faible)\\
&\qquad\qquad \Rightarrow MiseEnProduction(L, StructureLegere)\\
&SecondPuits(L) \wedge MiseEnProduction(L, StructureLegere) \wedge Quantite(L, Faible)\\
&\qquad\qquad \Rightarrow Cout(L, 10)
\end{aligned}$$

Nous supposerons que les préférences du décideur sont les suivantes:

- Le décideur préfère ne faire qu'un seul forage pour un gisement peu important:

$$\begin{gathered}
(PremierPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000))\\
\succ\\
(PremierPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000))
\end{gathered}$$

- Le décideur préfère un gisement à moindre coût en ayant supposé un jugement de grande quantité à un gisement ayant coûté plus cher

$$\begin{gathered}
(PremierPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Faible), Cout(L, 100))\\
\succ\\
(SecondPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Importante), Cout(L, 200))
\end{gathered}$$

Le cas courant est bien sûr $UnPuits(l_1), Hydrocarbures(l_1)$.

Ayant maintenant défini intuitivement à l'aide d'exemples l'approche que nous proposons, voyons maintenant sa comparaison avec les approches existantes pour le raisonnement par similarité dans la décision.

2.4 Comparaison avec les approches décisionnelles existantes

Il existe deux types d'approches dans lesquelles les notions de similarité ou de négligeabilité jouent un rôle en théorie de la décision:

les approches qualitatives: notamment celle de [Lehmann 1996] introduisant la notion de négligeabilité entre événements dans le cadre d'un affaiblissement des axiomes de [Savage 1954]; celle de [Dubois et al. 1998a] qui introduisent une notion implicite de négligeabilité entre états dans une version possibiliste de la théorie de l'utilité espérée, celle de [Boutilier 1994], qui développe un calcul logique qui utilise les éléments de base de la théorie classique de la décision, à savoir les probabilités, les utilités et les actions, mais qui exploitent ces éléments de manière qualitative et celle de [Pearl 1993] qui formalise un modèle de décision qualitative pour modéliser certaines expressions conditionnelles,

les approches quantitatives: dont celle de [Gilboa and Schmeidler 1995] postule que les décideurs choisissent des actions en fonction de leur performance dans des problèmes passés similaires; on peut également formaliser la notion de similarité en termes d'inférences probabilistes, et exploiter alors la sémantique numérique de [Dubois and Prade 1989][Hadj-Ali et al. 1998],

2.4.1 Les approches qualitatives

L'approche de [Lehmann 1996] définit elle aussi la notion de négligeable. Deux affaiblissements de l'axiomatique de [Savage 1954] y sont considérés:

- d'une part, le remplacement de l'axiome qui fait de la relation de préférence un ordre complet par un axiome qui en fait un ordre partiel,
- d'autre part, la considération d'événements négligeables par rapport à d'autres événements.

Nous présenterons rapidement l'axiomatique de [Lehmann 1996] et discuterons ensuite le lien entre la notion de négligeabilité que l'auteur introduit et la nôtre. Pour ce faire, on conserve les notations du premier chapitre, où Ω désigne un ensemble d'états, dont les sous ensembles notés $A, B, C \ldots$ sont des événements; C est un ensemble de conséquences notées $c, d, \ldots$, et les actions sont des projections de Ω sur C notées $f, g, \ldots$. D'autre part, on désigne par $\preceq_A$ une relation de préférence ou indifférence sachant l'événement A.

Avant d'introduire l'axiomatique de [Lehmann 1996], posons les quelques définitions suivantes:

Définition 14. Deux actes f et g sont dits équivalents en l'événement A si et seulement si, pour tout $s \in A, f(s) = g(s)$, ce qui est noté $f =_A g$

Définition 15. Soient A, B deux événements tels que $A \cap B = \emptyset$. On dit que A est négligeable comparé à B, ce que l'on note $A \mathcal{N} B$ si, $\forall h, h', h \preceq_{A \cup B} h' \Leftrightarrow h \preceq_A h'$.

Définition 16. Un acte f est dit constant si et seulement si $\forall s, t \in S, f(s) = f(t)$.

Ces définitions nous permettent alors de présenter l'axiomatique suivante:

Axiomatique 17. *(Axiomatique de [Lehmann 1996])*

Q 0 $f \preceq_A f$,

La relation de préférence sachant un événement est réflexive.

Q 1 $f \preceq_A g, g \succeq_A h \Rightarrow f \preceq_A h$

La relation de préférence sachant un événement est transitive. Combinée à la proposition précédente, cet axiome fait de $\succeq_A$ un pré-ordre partiel.

Q 2 $f =_A g \Rightarrow f \sim_A g$

Deux actes qui ont les mêmes résultats en l'événement A sont indifférents entre eux.

Q 3 $A \cap B = \emptyset, f \preceq_A g, f \sim_B g \Rightarrow f \preceq_{A \cup B} g$

Connaissant les préférences sachant A, on peut déduire celles sachant $A \cup B$ en imposant que f et g soient indifférents sachant B.

Q 4 $A \cap B = \emptyset, f \preceq_{A \cup B} g, f \sim_B g \Rightarrow f \preceq_A g$ ou $A \mathcal{N} B$

Lorsque $A \cap B = \emptyset, f \preceq_{A \cup B} g, f \sim_B g$, alors ou bien $f \preceq_A$, et par Q 3, cela explique $f \preceq_{A \cup B} g$, ou bien $A \mathcal{N} B$.

Q 5 *si c et d sont constants, si A est un événement non nul, et $c \preceq_A d$, alors pour tout événement B, $c \preceq_B d$. On note alors $\preceq$ la relation de préférence entre les actes constants.*

Les préférences sur les actes constants sont indépendants des événements considérés.

Q 6 *Il existe des actes constants c, d tels que que $c \prec d$.*

C'est une condition de non trivialité.

Ces axiomes permettent alors de définir une notion de probabilité qualitative sur les événements en posant:

$w_A^{c,d}(s) = \begin{cases} \bullet\ c \text{ si } s \in A \\ \bullet\ d \text{ sinon} \end{cases}$ (c, d) étant des actes constants (conséquences) tels que $c \succ d$.

Définition 18. On dit que A est tout autant plausible que B, ce que l'on note $A \preceq B$ si et seulement si, pour tous actes constants c, d tels que $c \prec d$, on a $w_A^{c,d} \preceq_{A\cup B} w_B^{c,d}$. La relation $\preceq$ sur les événements est alors définie à partir de la notion de préférence.

On peut alors dire que deux événements $A \cup B$ et A sont similaires si et seulement si A est négligeable devant B. Cependant, la définition posée par [Lehmann 1996] n'a pas la vertu descriptive que nous souhaitons d'une théorie de la décision, sa principale faiblesse résidant dans la non prise en compte des raisonnements qui peut induire des violations de la définition 15, i.e. un événement A peut être négligeable devant un événement B sans que la propriété que pose cette définition sur les préférences soit vérifiée. Considérons l'exemple de la photo-interprétation, et désignons par A l'événement "présence d'une aérogare", par B l'événement "présence d'une piste, d'un taxiway et d'une tour de contrôle", et désignons par:

- h l'action reconnaître une aéroport militaire,
- h' l'action reconnaître un aéroport civil.

Alors $h \succeq_{A\cup B} h'$ (car la description de l'aéroport civil est complète) et $h' \succeq_B h$ (on préfère reconnaître un aéroport civil à un aéroport militaire, i.e. en cas d'incomplétude dans les deux descriptions, on préfère bombarder un aéroport militaire plutôt que bombarder un aéroport civil). Le postulat de [Lehmann 1996] impose une sorte de monotonie de choix de l'action par rapport au fait de négliger ou non, ce que l'exemple précédent dément.

La seconde approche [Dubois et al. 1998a] vise à établir un critère de décision dans le cas où l'information sur les événements et sur les conséquences n'est que qualitative, et que l'on peut représenter:

- les connaissances incomplètes sur les états de la nature par une distribution de possibilité $\pi : \Omega \mapsto L$, L étant une échelle qualitative[9],
- les préférences sur les conséquences par une autre distribution de possibilité $\mu : C \mapsto T$, T étant une échelle qualitative de préférence,

L'utilité de la décision f dont la conséquence dans l'état s est $c = f(s) \in C$ pour tout état s, peut être évaluée en combinant les plausibilités $\pi(s)$ et les utilités $\mu(c)$. L'idée est que si, si l'incertitude sur l'état est représentéepar π, chaque décision induit sur l'ensemble des conséquences une distribution de possibilité telle que $\pi_f(c) = \pi(f^{-1}(c))$, et ordonner les décisions en fonction de préférences revient alors à ordonner des distributions de possibilités sur Ω. Si l'on suppose que le décideur fournit un ordre sur les distributions de possibilités sur Ω, la question est alors de savoir les axiomes que l'on doit poser pour que l'ordre des distributions de possibilités soit représentable par un critère de choix. [Dubois et al. 1998a] propose un tel ensemble d'axiomes, ayant défini par $(\lambda/\pi, \mu/\pi')$ une "loterie qualitative" donnant π avec la plausibilité λ et π' avec la plausibilité μ:

Axiomatique 19. *(Axiomatique de [Dubois et al. 1998a])*

Ordre *L'ensemble des distributions de possibilités est une structure de pré-ordre complet, et $\succeq$ désigne la relation de préférence, $\sim$ la relation d'indifférence,*

9. On conserve les mêmes notations que précédemment, et donc Ω désigne un ensemble d'états de la nature, C un ensemble de conséquences, et A un ensemble d'actions.

Indépendance $\pi_1 \sim \pi_2 \Rightarrow (\lambda/\pi_1, \mu/\pi) \sim (\lambda/\pi_2, \mu/\pi)$

Continuité *Pour toute distribution* π*, il existe* $\lambda \in T$*, tel que* $\pi \sim (1/\underline{c}, \lambda/\overline{c})$*, où* $\underline{c}$ *et* $\overline{c}$ *sont les éléments minimaux et maximaux de* C*.*

Aversion au risque $\pi \geq \pi' \Rightarrow \pi' \succeq \pi$.

Cet axiome stipule que plus le degré d'information modélisé par π *est faible, plus la situation est risquée: l'état le plus mauvais est celui de l'ignorance totale.*

[Dubois et al. 1998a] montrent alors que sous les axiomes précédents, la relation de préférence peut être représentée par le critère suivant:

Théorème 20. *Sous les axiomes précédents, une décision d est évaluée par le critère suivant:*

$$v(d) = inf_{s \in \Omega} max(n(\pi(s)), \mu(d(s))$$

où n *est une fonction de projection de* L *dans* T *renversant l'ordre.*

Il s'agit d'un modèle que ces auteurs ont qualifiés de pessimiste[10], puisque les décisions sont évalués sur la base de leur moins bonne conséquence (on peut la voir comme une généralisation du critère de Wald en l'absence de connaissances probabilistes).

Les auteurs [Dubois et al. 1997][Dubois et al. 1998b] ont également proposé une syntaxe pour les critères de décision. Pour ce faire, ils considèrent deux bases de connaissance stratifiées pour modéliser des degrés de connaissance et de préférence. La stratification des bases de connaissance est effectuée en fonction d'une distribution de possibilité. Dans une logique propositionnelle, [Dubois et al. 1997][Dubois et al. 1998b] utilisent une échelle ordonnée linéaire identique pour les connaissances et les préférences. Deux bases de connaissances sont utilisées :

- une base de connaissance $K = \{(\phi_i, \alpha_i)\}$ où $\alpha_i \in S$ désigne un degré de certitude et ϕ_i est une formule du langage L qui représente une connaissance sur l'état du monde,
- une base de connaissance $P = \{(\psi_i, \beta_i)\}$, où β_i désigne un niveau de priorité et ψ_i est une formule de L représentant un but (les buts sont liés aux états préférés du monde).

On désigne par K_α (resp. P_β) l'ensemble des formules de degré de certitude au moins égal à α (resp. les formules de priorité au moins égale à β). Une bonne décision, d'un point de vue pessimiste, est alors une conjonction $d^\wedge$ de littéraux qui permettent de dériver (associés à la partie la plus certaine de $K^\wedge$) la satisfaction des buts, même ceux de priorité faible:

$$K_\alpha\,\hat{}\wedge d\,\hat{}\vdash P_\beta\,\hat{}$$

avec donc α grand et β petit. On peut alors montrer [Dubois et al. 1997][Dubois et al. 1998b] que cette recherche peut être ramenée à celle de α, β étant donné par $n(\alpha)$. A cette recherche correspond une maximisation d'utilité syntaxique donnée par [Dubois et al. 1997][Dubois et al. 1998b]:

$$U_\star(d) = max_{\alpha, K_\alpha \wedge d \vdash P_{n(\alpha)}} \alpha$$

D'un point de vue plus optimiste, on peut juste rechercher des décisions d qui sont cohérentes avec la base de connaissance et les buts:

$$K_\alpha\,\hat{}\wedge d\,\hat{}\wedge P_\beta\,\hat{}\ \neq \perp$$

avec α aussi grand que possible et β aussi petit que possible: les états préférés sont parmi les plus plausibles et sont également cohérents avec la décision. De la même manière, cette recherche correspond à une maximisation d'utilité syntaxique donnée par [Dubois et al. 1997][Dubois et al. 1998b]:

$$U_\star(d) = max_{\alpha, K_\alpha \wedge d \vdash P_{n^{-1}(\alpha)}} \alpha$$

10. Il existe également une version optimiste de leur critère de choix donnée par $v(d) = sup_{s \in \Omega} min(m(\pi(s)), \mu(d(s))$, où m est une fonction de projection de T dans L préservant l'ordre.

[Dubois et al. 1997][Dubois et al. 1998b] montrent alors que, si l'on interprète les α_i comme des degrés de nécessité des formules dans les niveaux de $K \cup d$, on peut alors calculer une distribution de possibilité Π_{K_d} sur Ω (ensemble de toutes les interprétations) exprimant la sémantique de $K \cup d$ par, $\forall \omega \in \Omega$:

$$\begin{aligned}
\Pi_{K_d}(\omega) = &\quad min_{(\phi_i,\alpha_i)\in K, w\models\neg\phi_i} n(\alpha_i) \text{ si } w \models d^\wedge \\
\Pi_{K_d}(\omega) = &\quad \bot \text{ si } \{\phi_i / w \models \neg\phi_i\} = \emptyset \text{ et } w \models d^\wedge \\
\Pi_{K_d}(\omega) = &\quad\quad 0 \text{ si } w \not\models d^\wedge
\end{aligned}$$

En ce qui concerne la fonction d'utilité, on peut construire une fonction d'utilité μ sur Ω de manière similaire par:

$$\begin{aligned}
\mu(\omega) = &\quad min_{(\psi_i,\beta_i)\in P, w\models\neg\psi_i} n(\beta_i) \\
\mu(\omega) = &\quad \bot \text{ si } \{\psi_i / w \models \neg\psi_i\} = \emptyset
\end{aligned}$$

Les expressions qualitatives de l'utilité d'une décision données par le théorème 20 avec ces distributions permettent de donner une sémantique à la maximisation d'une utilité syntaxique déja présentée. Cette approche [Dubois et al. 1998b] a été implémentée sous la forme d'un ATMS modifié.

Dans cette sémantique, les états les moins plausibles sont négligés; leur négligeabilité est le fait d'un seuil déterminé par la comparaison des distributions de possibilités par la projection n. Cependant, par rapport à l'approche que nous proposons, cette négligeabilité n'est pas déterminée par un raisonnement symbolique (et donc aisément manipulable par l'expert), et les axiomes utilisés (notamment la complétude du pré-ordre) nous font envisager l'existence de paradoxes tels que ceux que nous avons mentionnés dans le premier chapitre.

La troisième approche de [Boutilier 1994] utilise une représentation qualitative de l'information, incluant des préférences qualitatives sur les mondes, et des probabilités qualitatives décrivant les vraisemblances des différents mondes, et ajoutant la notion de capacité de contrôle ou d'influence sur les mondes. [Boutilier 1994] propose une logique dotée d'une sémantique des mondes possibles pour représenter et raisonner avec ces quantités. En ce qui concerne les préférences qualitatives sur les résultats, elles sont représentées par la notion de préférence conditionnelle. On écrit $I(B \mid A)$, lu "idéalement B connaissant A" pour indiquer que si on sait A, alors on préfère que A soit vrai. Pratiquement, $I(B \mid A)$ signifie que si l'individu connait A, et si la valeur de vérité de A est fixé (sous son contrôle), alors l'individu doit rechercher B. Une logique bimodale CO est proposée pour exprimer ces préférences conditionnelles.

Étant donné un langage propositionnel L, deux opérateurs modaux $\Box, \vec{\Box}$, on définit une sémantique des mondes possibles de la forme $(W, \preceq, \varphi)$, où W est un ensemble de mondes possibles, φ est une fonction d'évaluation et $\preceq$ est une relation de pré-ordre sur W. La relation $\preceq$ représente un ordonnancement lié aux préférences: $v \preceq w$ signifie que v est au moins autant préféré que w. Les symboles modaux sont interprétés par:

- $M \models_w \Box\alpha$ ssi $\forall v \preceq w, M \models_v \alpha$, signifiant que $\Box\alpha$ est vrai dans le monde w si et seulement si il est vrai dans tous les mondes qui sont au moins autant préféré que w,
- $M \models_w \vec{\Box}\alpha$ ssi $\forall v \prec w, M \models_v \alpha$, signifiant que $\Box\alpha$ est vrai dans le monde w si et seulement si il est vrai dans tous les mondes qui sont au moins strictement préféré à w,

Ceci permet à l'auteur de définir un opérateur conditionnel $I(- \mid -)$ pour exprimer les préférences conditionnelles. Intuitivement, $I(B \mid A)$ est vrai si et seulement si B est vrai dans tous les mondes préférés de A, ce qui se définit par:

$$I(B \mid A) \equiv \vec{\Box}\neg A \vee \vec{\Diamond} A \wedge (\Box(A \Rightarrow B))$$

En dehors de cette préférence sur les résultats certains, [Boutilier 1994] ajoute une relation de vraisemblance en définissant une sémantique des mondes possibles de la forme $(W, \preceq_P, \preceq_N, \varphi)$, où $\preceq_P$ est la relation de préférence qualtitative qui vient juste d'être définie, et $\preceq_N$ est la relation de vraisemblance. On interprète la relation $v \preceq_N w$ par le fait que w est au moins aussi normal que v. Une relation de vraisemblance conditionnelle est définie de la même manière que la relation de préférence conditionnelle qui vient d'être définie, et deux opérateurs modaux $\Box_N, \vec{\Box}_N$ sont définis. Un opérateur $\Rrightarrow$ est ajouté signifiant, si $A \Rrightarrow B$, que normalement tous les A sont B et est défini par:

$$A \Rrightarrow B \equiv \vec{\Box}_N \neg A \vee \vec{\Diamond}_N \wedge (\Box_N(A \Rightarrow B))$$

Étant donné une base de connaissance KB, on définit sa fermeture déductive par $Cl(KB) = \{\alpha \in L, KB \Rrightarrow \alpha\}$, c'est à dire l'ensemble des propositions qui sont normalement vraies connaissant les connaissances de l'individu. C'est sur cet ensemble que les préférences vont être appliquées en définissant les buts idéaux de l'agent par, étant donné un modèle M, l'ensemble des α tel que $M \models I(\alpha \mid cl(KB))$. Cette approche offre deux avantages sur celle que nous proposons: d'une part, la prise en compte de la vraisemblance et la prise en compte de la capacité à agir (influencer ou contrôler) la valeur de vérité de certains faits.

Enfin, la quatrième et dernière approche de [Pearl 1993] permet une formalisation d'un modèle de décision qualitative pour modéliser des expressions conditionnelles de la forme *"si l'on connait, croit ou observe C, alors l'utilité espérée résultant de l'action A est supérieure à l'utilité espérée de ne pas faire A"*. Pour mesurer les utilités, [Pearl 1993] utilise une valuation des conséquences sur une échelle entière d'utilités. Cette évaluation des utilités, combinée avec la notion développée par l'auteur de probabilité infinitésimale, permet de mesurer les actions par l'utilité espérée de leur conséquence. L'auteur ajoute à ce schéma la prise en compte des relations causales entre actions et conséquences, par le biais de réseaux causaux [Goldszmidt and Pearl 1992]. Cette approche, si elle offre l'avantage de prendre en compte les relations causales entre actions et conséquences, ne permet pas de prendre en compte des préférences autres que celles entre faire A et ne pas faire A.

2.4.2 Les approches quantitatives

L'idée de prendre en compte les préférences dans le raisonnement à partir de cas a été mise en oeuvre par [Gilboa and Schmeidler 1995]. Les deux auteurs postulent que les décideurs choisissent des actions en fonction de leurs performances dans des problèmes passés mémorisés similaires au problème courant. Pour être plus précis, ces auteurs supposent qu'il existe un ensemble de problèmes mémorisés par le décideur sur lequel il existe une mesure de similarité. Confronté à un nouveau problème, la similarité des problèmes induit la remémoration de problèmes anciens similaires, ainsi que les actions qui y ont été choisies et le résultat qui s'en est suivi.

Plus précisément, un cas est un triplet (p, a, r) où p est un problème, a est une action et r un résultat. Soit M la mémoire des cas passés. Un décideur est caractérisé par deux fonctions:

- une fonction d'utilité u, qui affecte une valeur numérique aux résultats,
- une fonction de similarité qui affecte une valeur positive ou nulle à des paires de problèmes.

Confronté à un nouveau problème p, le décideur choisit un acte qui maximise alors $\sum_{(p,a,r)\in M} s(p,q)u(r)$.

2.4.3 La modélisation

Soit P un ensemble de problèmes, A un ensemble d'actions, R un ensemble de résultats. L'ensemble des cas est défini par $C = P \times A \times R$. Pour simplifier les notations, on suppose que tous les actes sont disponibles pour tous les problèmes $p \in P$, et qu'il existe un résultat $r_0 \in R$ signifiant "cet acte n'a pas été choisi".

Soit $M \subseteq C$, et H sa projection sur P, H ensemble des problèmes de M:

$$H = H(M) = \{q \in P, \exists a \in A, \exists r \in R, (q, a, r) \in M\}$$

[Gilboa and Schmeidler 1995] impose alors à M les deux conditions suivantes pour que l'on puisse dire qu'il s'agit d'une mémoire:

1. $\forall q \in H(M), \forall a \in A, \exists! r \in R, (q, a, r) \in M$

2. $\forall q \in H(M), \exists! a \in A, \exists r \in R, (q,a,r) \in M$ et $r \neq r_0$

conditions qui permettent de spécifier que pour toute mémoire, et pour tout problème $q \in H(M)$ appartenant à cette mémoire, il existe un et seul acte choisit en q qui a conduit à un seul et unique résultat.

Pour simplifier la théorie, on suppose que l'ensemble des résultats est l'ensemble des réels (i.e. qu'il est déjà mesuré en utilités) et on suppose que $r_0 = 0$. Une mémoire est alors une matrice où les colonnes représentent les problèmes, les lignes les actions et les résultats sont à l'intersection d'une ligne et d'une colonne. Par la contrainte 2, chaque colonne contient au plus une entrée non nulle. On suppose de plus, qu'ayant la mémoire M, ayant le problème P, le décideur a une relation de préférence $\succeq_{p,H(M)} \subseteq A \times A$, et on suppose que l'évaluation d'un acte n'est fonction que de ses conséquences. Sur la matrice représentant la mémoire, cette relation s'exerce donc comme une relation d'ordre sur les lignes.

D'autre part, on demande que les préférences entre deux vecteurs ne dépendent pas de la mémoire qui les contient. Formellement, soit M, M' deux mémoires telles que $H(M) = H(M')$, et soit x, y deux actions. Alors $x \succeq_{p,H(M)} y \Leftrightarrow x \succeq_{p,H(M')} y$.

Remarquons de plus que $\succeq_{p,H(M)}$ n'est pas un ordre complet sur R^H. Si l'on considère en effet deux actions $x, y \in R^H$, ayant pour résultat $x(q), y(q)$ différents de r_0 pour $q \in H$. Étant donné la contrainte 2, x et y ne peuvent apparaître dans la même mémoire M. On pose donc une restriction sur l'ordre $\succeq_{p,H(M)}$ en lui imposant de ne pouvoir comparer des paires qui ne peuvent apparaître que dans la même mémoire M, paires que l'on qualifiera dorénavant de *compatibles*. Formellement, si $x, y \in A$ et si $x * y$ définit le produit coordonnées à coordonnées des lignes de la matrice M. alors x, y sont compatibles si et seulement si $x * y = 0$ ou $x = y$.

2.4.4 Axiomatique

Ces définitions étant données, on peut maintenant définir l'axiomatique de [Gilboa and Schmeidler 1995]:

Axiomatique 21. *Axiomatique de [Gilboa and Schmeidler 1995]. (Case Based Decision Theory)*

A1: Comparabilités des profils compatibles $\forall p \in P, \forall M \in C, \forall (x,y) \in R^{H(M)}$, *x et y sont compatibles si et seulement si* $x \succeq_{p,H(M)} y$ *ou* $y \succeq_{p,H(M)} x$.

A2: Monotonie $\forall p \in P, \forall M \in C, x \succeq y$ *et* $x * y = 0$ *implique* $x \succeq_{p,H(M)} y$.

A3: Continuité $\forall p \in P, \forall M \in C, \forall x \in R^{H(M)}$, *les ensembles* $\{y \in R^{H(M)}, y \succeq_{p,H(M)} x\}$ *et* $\{y \in R^{H(M)}, x \succeq_{p,H(M)} y\}$ *sont des fermés dans la topologie standard de* $R^{H(M)}$.

A4: Séparabilité $\forall p \in P, \forall M \in C, \forall (x,y,z,w) \in R^{H(M)}, (x+z) * (y+w) = 0$
$x \succeq_{p,H(M)} y$ *et* $z \succeq_{p,H(M)} w \Leftrightarrow (x+z) \succeq_{p,H(M)} (y+w)$

[Gilboa and Schmeidler 1995] montrent alors que la fonction de préférence prend la forme suivante:

Théoreme C. *Les deux affirmations suivantes sont équivalentes:*

- *Les axiomes A1, A2, A3, A4 sont valides,*
- *Pour tout* $p \in P, \forall M \in C$, *il existe une fonction* $s_{p,H(M)} : H(M) \longmapsto \mathbb{R}$ *telle que:*

$$x \succeq_{p,H(M)} y \Leftrightarrow \sum_{q \in H(M)} s_{p,H(M)}(q)x(q) \succ \sum_{q \in H(M)} s_{p,H(M)}(q)y(q)$$

La fonction de similarité est alors unique à la multiplication par un scalaire positif près.

Preuve. Voir [Gilboa and Schmeidler 1995]. ∎

Le problème est que cette fonction de similarité s peut dépendre de la mémoire M. Pour relâcher cette contrainte, [Gilboa and Schmeidler 1995] ajoutent un axiome permettant d'assurer que la fonction de similarité est indépendante de M, en imposant que l'importance relative de deux problèmes q_1, q_2 dans les mémoires M_1, M_2 soit proportionnelle:

Axiomatique 22. *L'importance relative de deux problèmes q_1, q_2 dans les mémoires M_1, M_2 est proportionnelle ssi:*

Q5: Proportionnalité *$\forall p \in P, \forall M_1, M_2 \in C, \forall q_1 \in H(M_1), \forall q_2 \in H(M_2), p \notin H(M_1), p \notin H(M_2)$ et soit v_j^i le vecteur unité de $R^{H(M_i)}, i = 1, 2$ correspondant à $q_j, j = 1, 2$, i.e. v_j^i est un vecteur dont la $j^{ième}$ composante est 1 et dont les autres composantes sont 0. Alors, si $x, y \in R^{H(M_1)}, z, w \in^{H(M_2)}$, (x, y) et (z, w) compatibles et telles que $x \sim_{p,H(M_1)} y$ et $z \sim_{p,H(M_2)} w$. Alors*

$$\begin{array}{ccc} x + \alpha v_1^1 & \sim_{p,H(M_1)} & y + \beta v_2^1 \\ & \textit{entraîne} & \\ z + \alpha v_1^2 & \sim_{p,H(M_2)} & w + \beta v_2^2 \end{array}$$

On peut alors définir une fonction de similarité indépendante de la mémoire M:

Théoreme D. *Les deux affirmations suivantes sont logiquement équivalentes:*

- *Les axiomes $A1, A2, A3, A4, A5$ sont valides,*
- *Pour tout $p \in P, \forall M \in C, p \notin H(M)$, il existe une fonction $s : P \times P \longmapsto \mathbb{R}$ telle que:*

$$x \succeq_{p,H(M)} y \Leftrightarrow \sum_{q \in H(M)} s(p, q)x(q) \succ \sum_{q \in H(M)} s(p, q)y(q)$$

La fonction de similarité est alors unique à la multiplication par un scalaire positif près.

Notons qu'en dehors de l'aspect mathématique très ressemblant des formules de la CBDT [11] et de l'EUT [12], les deux théories ont peu de choses en commun. Dans l'EUT, la sommation se fait sur l'ensemble des états de la nature, qui permet de résoudre toute l'incertitude, alors que dans la CBDT, l'ensemble des problèmes ne concerne que ceux que le décideur a effectivement rencontrés. D'autre part, dans la théorie de l'utilité espérée, l'unité d'analyse est typiquement un seul problème de décision, alors que dans le cas de la CBDT, le problème de décision est analysé par rapport à d'autres problèmes: la relation de préférence est dépendante de la mémoire et du contexte. Enfin, les deux théories modélisent le *raisonnement* différemment: dans le cas de l'EUT, le décideur raisonne de manière purement hypothétique par des propositions de la forme *"si l'état de la nature est w et si je choisis a, alors j'aurais le résultat r"*. Dans la CBDT, au contraire, le décideur ne raisonne que par rapport à ses propres actions et non plus par rapport à celles de la nature: *"si je choisis a, alors j'ai le résultat r"*.

Cependant, tout comme dans la théorie de l'utilité espérée, il s'agit, sous contrainte de rationalité, de déterminer les conditions **nécessaires** à l'existence de fonctions de représentation des résultats (le fait que la relation de préférence soit un pré-ordre constitue certainement expérimentalement un problème). Or l'ensemble des expériences que nous avons présenté dans le chapitre précédent montre que le mode de construction d'une théorie alternative doit plutôt être, sous contraintes empiriques donc sous conditions expérimentalement **suffisantes**, de construire une fonction de représentation des raisonnements.

2.4.5 Règles à exceptions en raisonnement plausible

Une autre approche possible est l'utilisation d'inférences probabilistes liées à une interprétation probabiliste du connecteur d'implication $A \rightarrow B$, et signifiant que la probabilité conditionnelle $p(B/A)$ est

11. Case-Based Decision Theory.
12. Expected Utility Theory.

infiniment proche de 1. Pour représenter la notion de négligeabilité, elle peut se décrire par la décomposition d'une formule $a, d \Rightarrow_S c$ en une règle d'inférence probabiliste $a \rightarrow d$ et une règle classique $a \wedge d \Rightarrow c$. [Adams 1975] propose un système d'inférence probabiliste fondé sur les trois axiomes suivants:

Triangularité $A \rightarrow B, B \rightarrow C \Rightarrow (A \cap B) \rightarrow C$
Règle de Bayes $A \rightarrow B, (A \cap B) \rightarrow C \Rightarrow A \rightarrow C$
Disjonction $A \rightarrow C, B \rightarrow C \Rightarrow (A \cup B) \rightarrow C$

Ces règles d'inférence sont complètes vis à vis de l'interprétation donnée des probabilités conditionnelles. [Hadj-Ali et al. 1998] propose une contrepartie de ces règles avec une sémantique non infinitésimale pour prendre en compte une notion de gradualité. Il s'agit de comprendre $A \rightarrow B$ comme "presque tous les A sont des B", en d'autre termes $P(B/A)$ est proche de 1 au sens d'une relation de voisinage Vo. Une sémantique fondée sur les voisinages Vo est utilisée [Dubois and Prade 1989][Hadj-Ali et al. 1998]. La relation Vo est définie par la relation floue entre les deux nombres réels positifs a et b et exprimée par [Dubois and Prade 1989]:

$$\mu_{Vo}(x, y) = \mu_M(x/y)$$

où la fonction d'appartenance μ_M est celle d'un nombre flou proche de 1, tel que $\mu_M(1) = 1$. La notation $Vo(M)$ signifie alors que Vo est définie à partir de la fonction d'appartenance précédente. Une relation de négligeabilité Ne est alors définie par:

$$(a, b) \in Ne \Leftrightarrow (a + b, b) \in Vo$$

Revenant alors à la formalisation de l'implication $A \Rightarrow B$, on aurait alors, utilisant les règles données par [Hadj-Ali et al. 1998]:

$$(a, b) \in Vo(M) \Leftrightarrow (max(a, b) - min(a, b), min(a, b)) \in Ne(M)$$

En terme de négligeabilité, l'inférence $a \rightarrow d$ correspond donc à $(1 - P(D/A), P(D/A)) \in Ne(M)$, ce qui signifie bien que l'on néglige les cas où les d ne sont pas des a. On retrouve donc en numérique une sémantique proche de celle que nous proposons. Cette approche n'intégre néanmoins pas les préférences.

2.5 Conclusion

A la lumière de la discussion menée dans ce chapitre, il apparaît donc que la relation de déduction par similarité va syntaxiquement s'exprimer sur des séquents. Un séquent est une paire $\Gamma : \Delta$ de listes de formules $\Gamma =< \varphi_1, \ldots, \varphi_n >$ et $\Delta =< \psi_1, \ldots, \psi_m >$. Intuitivement, il signifie que si l'on connaît le problème désigné par la conjonction des $\varphi_1, \ldots, \varphi_n$, alors on peut effectuer ou bien l'action ψ_1, ou ..., ou bien l'action ψ_m. Les préférences sont définies par un sextuplet $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ signifiant que *"ϕ_1 résultant de la connaissance de φ_1 et de la négligeabilité de φ_2 est préféré à ϕ_2 résultant de la connaissance de ψ_1 et de la négligeabilité de ψ_2"*. Le raisonnement par similarité en fonction de préférences consiste alors, sachant un problème courant désigné par un ensemble de séquents $(: \varphi'_1), \ldots, (: \varphi'_p)$, pour chaque séquent représentant un problème mémorisé à:

- calculer les parties communes entre cette conjonction de séquents et la partie gauche du séquent mémorisé, et à déduire de la partie droite du séquent mémorisé l'action à effectuer, en l'adaptant au besoin,

- à calculer les parties différenciant cette conjonction de séquents et la partie gauche du séquent mémorisé, parties que l'on qualifiera désormais de négligeables,

- à appliquer les préférences pour déterminer la meilleure action à entreprendre.

Ces négligeables correspondent à des éléments de la description de problème mémorisé qui sont incertains dans le problème courant: les éléments vrais ne peuvent être négligés, seuls les éléments incertains le peuvent. Ayant ainsi défini formellement le raisonnement par similarité en fonction de préférences, nous allons maintenant étudier quelles sont les propriétés formelles que la relation de déduction par similarité va devoir satisfaire, et nous allons les comparer tant avec les propriétés formelles des logiques non monotones existantes qu'avec celles des logiques de descriptions existantes. Nous verrons ensuite comment formaliser intégralement et définir une sémantique point fixe pour les logiques que nous avons défini informellement dans ce chapitre.

Deuxième partie

Raisonnement à partir de cas et décision

L'analyse d'un raisonnement par similarité en fonction de préférences peut se faire de deux manières différentes, connaissant un ensemble P de problèmes mémorisés et sachant que q est le problème courant:

- par une représentation plutôt orientée **données**: à chaque ensemble de problèmes mémorisés correspond via une fonction f un ensemble d'actions à entreprendre, i.e. $f(q)$ désigne l'action à entreprendre pour le problème courant q. Dans cette approche, dans le cas où q n'appartient pas à l'ensemble des problèmes mémorisés, la similarité est appréhendée par le biais d'une fonction ω, et les actions à entreprendre dans le problème q sont alors $\omega'(f(\omega(q)))$, où ω' est une fonction de modification des actions correspondant à ω sur l'ensemble des actions. C'est cette approche qui est utilisée dans les formalisations du raisonnement à partir de cas utilisant les logiques de descriptions. Dans cette approche, la similarité est rejetée "en dehors" de la manipulation des connaissances "*description du problème - résultat*".
- par une représentation plutôt orientée **fonction**: les problèmes mémorisés sont des descriptions de problèmes, auxquels une fonction d'ensemble f (lue relation de déduction par similarité) anti-monotone [13] associe un ensemble d'actions à entreprendre. Ici, la similarité est directement intégrée dans la manipulation des connaissances, et c'est un raisonnement similaire que l'on applique, alors que l'approche précédente recherchait un problème similaire.

Nous étudierons dans cette partie ces deux formalismes, et justifierons le choix du deuxième en recherchant quelles bonnes propriétés une relation de satisfaction par similarité doit posséder. Nous nous servirons alors de ces propriétés pour étudier l'approche orientée données.

13. L'anti-monotonie de f dérive du fait que plus le problème courant est décrit avec précision, plus le nombre de similarités sera faible, et donc moins il y aura d'actions déclenchables.

3

Propriétés de la relation de déduction par similarité avec des préférences

3.1 Introduction

Comme nous l'avons dit dans le chapitre précédent, le raisonnement décisionnel dans un contexte d'incomplétude conduit à prendre en compte des logiques non standards. Reste maintenant à déterminer quelle forme cette logique non standard doit prendre, le fait qu'elle ne satisfasse pas une propriété de monotonie ne suffisant en rien à la spécifier. Nous allons donc maintenant étudier cette relation de satisfaction par similarité connaissant des préférences en étudiant l'ensemble des propriétés que doit satisfaire l'opérateur d'inférence par similarité, i.e. en déterminant les méta-axiomes (système syntaxique) qu'il doit satisfaire. Soit donc $\vdash_S$ l'opérateur d'inférence par similarité, et soit Γ un ensemble de séquents, Π un ensemble de préférences; Γ, Π, lu Γ et Π décrit donc un ensemble de problèmes de décision mémorisés, et Π un ensemble de préférences mémorisées. Les propriétés de $\vdash_S$ seront étudiées de deux manières:

- d'une part, vis à vis de la similarité, i.e. si Γ désigne un ensemble de séquents représentant l'ensemble des problèmes de décision mémorisés, et Π un ensemble de préférences mémorisées, alors on étudiera les propriétés de la relation de déduction par similarité $\vdash_S$ à "Π *constant*",
- d'une part, vis à vis des préférences, i.e. si Γ désigne un ensemble de séquents représentant l'ensemble des problèmes de décision mémorisés, et Π un ensemble de préférences, alors on étudiera les propriétés de la relation de déduction par similarité $\vdash_S$ à "Γ *constant*",

Cette approche ayant permis de fédérer et d'expliciter les propriétés des logiques non monotones connues, nous verrons également en quoi notre approche s'en distingue.

3.2 Les méta axiomes liés à la notion de similarité

Désignons par $\vdash$ la relation de déduction de la logique du premier ordre. Cette relation satisfait plusieurs propriétés, dont on distingue les trois suivantes:

Inclusion $\quad X, a \vdash a$

Monotonie $\quad \dfrac{X \vdash a}{X, y \vdash a}$

Coupure $\quad \dfrac{X \vdash y \quad X, y \vdash a}{X \vdash a}$

Désignons alors par $\models_S$ notre relation de déduction par similarité connaissant des préférences. Elle ne satisfait alors pas la propriété de monotonie. Considérons en effet l'exemple de la photo-interprétation, et l'exemple suivant: un aéroport civil est défini par la présence à un endroit donné de l'image d'une tour de contrôle, d'une piste et d'un taxiway; un aéroport militaire est défini par la présence à un endroit donné d'une image d'une tour de contrôle, d'une piste et d'un ensemble de stationnements non alignés; sur une image, on ne voit qu'une tour de contrôle et une piste. On suppose de plus que le décideur a une préférence lui disant qu'il préfère conclure à un aéroport civil plutôt qu'à un aéroport militaire. Raisonnant par similarité, le décideur va alors déduire qu'il s'agit d'un aéroport civil; par ailleurs, si on ajoute la connaissance qu'il existe sur l'image des stationnements non alignés, le décideur déduit alors qu'il s'agit d'un aéroport militaire. L'opérateur d'inférence par similarité ne satisfait donc pas la propriété de monotonie.

La propriété de coupure semble une propriété raisonnable d'une relation de déduction par similarité sachant des préférences. Elle permet une sorte de transitivité de la relation de similarité. En effet, supposons que la présence d'un site à munitions induise par similarité celle d'un site stratégique qu'il alimente, et que la présence d'un site stratégique induise par similarité celle d'un site missile qui le protège. Ainsi il semble logique de conclure que la présence d'un site à munitions induise par similarité celle d'un site missile. Cette propriété peut cependant s'avérer indésirable dans la mesure où la transitivité des similarités implique un affaiblissement à chaque pas dans une chaîne de similarité.

La propriété d'inclusion est généralement requise des différents formalismes non monotones proposés dans la littérature (voir par exemple [Krauss et al. 1990][Brewka et al. 1997]). Une logique non monotone doit en effet inclure la logique classique, et l'identité semble bien un cas raisonnable de similarité.

Les logiques non monotones, si elles ont abandonné la propriété de monotonie, ne le font pas cependant totalement [Makinson 1989]. Elle est en effet remplacée par une propriété plus faible, dite de monotonie prudente "(*Cautious Monotony*)". Dans la règle de monotonie, la formule y est en effet arbitraire et peut induire des informations contredisant a. Le principe de monotonie prudente évite cet écueil en imposant que $X \models_S y$.

Une relation de déduction non monotone satisfaisant ces méta axiomes a été qualifiée de cumulative dans la littérature [Krauss et al. 1990][Makinson 1989]. Ayant adapté ces axiomes au raisonnement par similarité en fonction de préférences (i.e. nous avons introduit une représentation syntaxique pour les préférences), nous obtenons donc:

Définition 23. *(Relation d'inférence cumulative)*
Une relation d'inférence par similarité $\vdash_S$ est dite cumulative si et seulement si elle satisfait les axiomes suivants:

Supra Classique $$\frac{\Gamma \vdash a}{(\Gamma, \emptyset) \vdash_S a}$$

Inclusion $$(\Gamma \cup a, \emptyset) \vdash_S a$$

Monotonie prudente $$\frac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma, \Pi) \vdash_S a}{(\Gamma \cup y, \Pi) \vdash_S a}$$

Coupure $$\frac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma \cup y, \Pi) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$$

Ces règles d'inférence décrivant la dérivation de formules sans prendre en compte leur forme syntaxique, il reste encore à préciser les règles permettant de définir les propriétés de dérivation pour les symboles logiques usuels. Les règles régissant les systèmes non monotones existant sont les règles suivantes, dont il nous faut étudier l'adéquation à la modélisation d'un raisonnement par similarité sachant des préférences:

$$\wedge \quad \frac{(\Gamma,\Pi) \vdash_S a \quad (\Gamma,\Pi) \vdash_S b}{(\Gamma,\Pi) \vdash_S a \wedge b}$$

$$\vee \quad \frac{(\Gamma \cup y,\Pi) \vdash_S a \quad (\Gamma \cup z,\Pi) \vdash_S a}{(\Gamma \cup (y \vee z),\Pi) \vdash_S a}$$

$$\neg \quad \frac{(\Gamma,\Pi) \vdash_S a \quad (\Gamma,\Pi) \nvdash_S \neg y}{(\Gamma \cup y,\Pi) \vdash_S a}$$

- La première règle ($\wedge$) est une règle dérivée de la règle de coupure et de la règle supra classique.
- ($\vee$) est une sorte de règle d'apprentissage en raisonnement à partir de cas. Si un aéroport civil se déduit par similarité d'une piste, d'une tour de contrôle, et se déduit d'une piste et d'une aérogare, alors il se déduit par similarité d'une piste et d'une tour de contrôle ou d'une aérogare.
- La troisième règle ($\neg$) de rationalité, affirme que toute proposition qui n'est pas explicitement niée par l'ensemble X peut lui être ajoutée sans changer ses conclusions. C'est une propriété qui n'est pas toujours valide pour la modélisation d'un raisonnement par similarité sachant des préférences: elle suppose en effet implicitement que la préférence consiste dans la consistance du raisonnement. Lorsqu'on raisonne sachant des préférences quelconques, il faut donc modifier cette règle en une règle de rationalité pour exprimer que toute formule qui n'induit pas de contradiction avec les préférences peut être ajoutée à la théorie. Nous verrons dans la section suivante les propriétés que doit satisfaire la relation de déduction par similarité en ce qui concerne les préférences.

Avant de nous intéresser aux autres propriétés de la relation d'inférence préférentielle par similarité, considérons le lien qu'ont les propriétés que nous venons d'annoncer avec les logiques non monotones existantes. Ces logiques n'utilisent pas en général une telle formulation syntaxique des préférences[14], il faut entendre la relation de déduction comme liant des ensembles de formules du premier ordre. Nous verrons dans les sections suivantes en quoi ces logiques, bien que certaines satisfont les bonnes propriétés que nous avons mentionnées pour définir un raisonnement par similarité, n'ont pu être utilisées.

3.2.1 Les relations d'inférence cumulatives

Nous avons vu que les logiques permettant de définir un raisonnement devaient être cumulatives. Parmi l'ensemble des relations d'inférence cumulatives, certaines d'entre elles occupent une place particulière parmi les théories cumulatives, à savoir les logiques définies sémantiquement à l'aide d'une relation de préférence sur les modèles. Leur première définition sous cette forme date de [Shoham 1988]. Soit donc L un langage, et $\prec$ un ordre strict bien fondé sur ses interprétations. Alors:

14. Sans une formulation syntaxique telle que nous l'avons proposé des préférences, il faut alors entendre de la manière suivante les propriétés de la relation d'inférence non monotone que l'on notera $\Vdash$ pour ne pas la confondre avec la relation d'inférence $\models_S$:

Définition 24. *(Relation d'inférence cumulative)*
Une relation d'inférence non monotone $\Vdash$ est dite cumulative si et seulement si elle satisfait les axiomes suivants:

Supra Classique $\quad \dfrac{\Gamma \Vdash a}{\Gamma \Vdash a}$

Inclusion $\quad \Gamma, a \Vdash a$

Monotonie prudente $\quad \dfrac{\Gamma \Vdash y \quad \Gamma \Vdash a}{\Gamma, y \Vdash a}$

Coupure $\quad \dfrac{\Gamma \Vdash y \quad \Gamma, y \Vdash a}{\Gamma \Vdash a}$

Définition 25. Une interprétation $\mathcal{I}$ satisfait préférentiellement une formule A (noté $\mathcal{I} \models_{\prec} A$) si et seulement si $\mathcal{I}$ satisfait A et il n'existe pas d'interprétation $\mathcal{I}'$ telle que $\mathcal{I}' \prec \mathcal{I}$. Dans ce cas, $\mathcal{I}$ est un modèle préféré de A. A satisfait préférentiellement B (noté $A \models_{\prec} B$) ssi tous les modèles préférés de A satisfont B.

Dans ces logiques, un axiome d'ordre supérieur ou une écriture syntaxique particulière de certaines règles, permettent de prendre en compte une notion de préférence équivalente à la notion de préférence sur les modèles, comme nous allons le voir.

L'hypothèse du monde clos

La première logique préférentielle est liée à la notion de monde clos en programmation logique. La similarité de deux descriptions est appréhendée en ajoutant aux descriptions incomplètes d'un problème la négation de tous les éléments de description non explicitement référencés. En termes syntaxiques, l'hypothèse du monde clos vise à prendre en compte la négation en programmation logique: en effet, l'approche de [Kowalski and van Emden 1974] donne un plus petit modèle de Herbrand à un programme logique dans lequel tous les atomes sont vrais, excluant ainsi la possibilité de la négation. La prise en compte de la négation s'est tout d'abord effectuée par omission, en posant comme hypothèse que tout fait non explicitement annoncé est faux. Cette hypothèse peut s'exprimer par une augmentation d'un programme logique en un programme $\hat{P} = P \cup \{\neg A, A \in At_\Sigma \text{ et } P \not\models A\}$ et revient [Lalement 1990] à étendre la relation d'inférence classique en une règle d'inférence non locale permettant de déduire $\neg\varphi$ si φ n'est pas déductible des données:

$$\text{MC: } \frac{\Gamma \not\models \varphi}{\Gamma \models \neg\varphi}$$

[Clark 1978] a donné une notion de complétion permettant cette extension d'un programme logique. Pour ce faire, rappelons quelques définitions: un programme défini est un ensemble de clauses de Horn et un programme est dit défini pour le prédicat P si et seulement si chaque clause contenant P contient au plus une occurrence positive de P. [Clark 1978] donne la notion de complétion en fermant la théorie par rapport au prédicat P: chaque clause définie pour le prédicat P est réécrite en $\forall X A(X) \Rightarrow P(X)$, puis la théorie est complétée avec l'axiome $\forall X P(X) \Rightarrow A(X)$.

D'un point de vue sémantique, cette complétion peut s'interpréter en terme de modèles préférentiels. Soit en effet T un programme défini dans le prédicat P, et soit Z l'ensemble des autres prédicats de T. La complétion de T vise alors à restreindre les modèles de la théorie à ceux qui minimisent les extensions du prédicat P, l'interprétation de Z pouvant elle varier parmi les modèles comparés. Soient donc $\mathcal{M}_1, \mathcal{M}_2$[15] deux modèles de T, et K^M l'interprétation du symbole K dans $\mathcal{M}$. Soit alors la relation d'ordre suivante sur les modèles de T:

Définition 26. On définit la relation suivante entre les modèles de la théorie T:

$$\mathcal{M}_1 \preceq^{P;\,Z} \mathcal{M}_2 \Leftrightarrow \begin{cases} \bullet\ M_1 \subseteq M_2 \\ \bullet\ K^{\mathcal{M}_1} = K^{\mathcal{M}_2} \text{ pour tout symbole } K \text{ qui n'est pas dans } P \text{ ou } Z \\ \bullet\ P^{\mathcal{M}_1} \subseteq P^{\mathcal{M}_2} \end{cases}$$

Alors

Théorème 27. *[Reiter 1982]*
Soit T une théorie définie pour le prédicat P. La complétion de P est vraie dans tous les modèles minimaux pour $\preceq^{P;\,Z}$ de T, où Z désigne l'ensemble des autres prédicats de T.

Il s'ensuit donc que la complétion de [Clark 1978] correspond à une logique à modèles préférentiels pour la relation d'ordre $\prec$ définie par:

$$\mathcal{M}_1 \prec \mathcal{M}_2 \Leftrightarrow \mathcal{M}_1 \preceq^{P;\,Z} \mathcal{M}_2 \text{ et } \neg(\mathcal{M}_2 \preceq^{P;\,Z} \mathcal{M}_1)$$

D'un point de vue raisonnement par similarité, l'approche est extrêmement pauvre puisque les éléments de description manquants sont inférés comme faux. C'est la seule similarité possible. Il n'y a pas de notion de préférence.

15. Si $\mathcal{M}$ est une algèbre, M l'ensemble de définition de l'algèbre $\mathcal{M}$.

La circonscription

Tout comme l'hypothèse du monde clos, la circonscription est une méthode pour calculer une fermeture déductive d'une théorie en restreignant ses modèles à ceux qui sont des extensions minimales de certains prédicats ou fonctions. Ici, considérant deux descriptions, il s'agit de minimiser les différences entre les deux descriptions, ce qui dans cette logique prendra le concept d'anormalités. Depuis ses premières formulations [McCarthy 1980], la théorie a pris plusieurs formes diverses, dont la plus utilisée "*Parallel predicate circumscription*" que nous présenterons ici [McCarthy 1980][Lifschitz 1985] [Besnard 1989][Brewka et al. 1997]. La dénomination particulière de ce type de circonscription tient au fait qu'un ensemble de prédicats est minimisé en parallèle. Pour ce faire, définissons:

$$U \preceq V \text{ par } \forall x\ U(X) \Rightarrow V(X)$$
$$U \prec V \text{ par } U \preceq V \text{ et } V \not\preceq U$$
$$U \approx V \text{ par } U \preceq V \text{ et } V \preceq U$$

Définition 28. [Lifschitz 1985] *(Parallel Predicate Circumscription)*
Soit L un langage, P un ensemble de symboles de prédicats que l'on souhaite minimiser et soit Z un autre ensemble de symboles de prédicats. Soit $A(P, Z)$ une formule contenant les symboles P et Z. Alors:

$$Circum(A(P,Z);\ P;\ Z) = A(P,Z) \wedge \neg\exists pz(A(p,z) \wedge p \prec P)$$

Bien que la circonscription soit donc un schéma pour ajouter de nouvelles formules à une théorie, elle peut aussi être décrite par une restriction de l'ensemble des modèles d'une théorie. La circonscription préfère en effet les modèles de $A(P, Z)$ qui sont minimaux pour le prédicat P, en supposant que tous ces modèles ont les mêmes interprétations pour tous les symboles qui ne sont pas dans $A(P, Z)$.

Définition 29. Soit $\mathcal{M}_1, \mathcal{M}_2$ deux modèles de $A(P, Z)$. Alors

$$\mathcal{M}_1 \preceq^{P;\,Z} \mathcal{M}_2 \Leftrightarrow \begin{cases} \bullet\ M_1 \subseteq M_2 \\ \bullet\ K^{\mathcal{M}_1} = K^{\mathcal{M}_2} \text{ pour tout symbole } K \text{ qui n'est pas dans } P \text{ ou } Z \\ \bullet\ \forall P_i \in P, P_i^{\mathcal{M}_1} \subseteq P_i^{\mathcal{M}_2} \end{cases}$$

$\preceq^{P;\,Z}$ est une relation réflexive et transitive. Les modèles préférés sont les modèles maximaux pour cet ordre.

[Lifschitz 1985] a montré que cette définition était équivalente à la définition de la "*Parallel predicate circumscription*". De même, cette logique correspond à une logique à modèles préférentiels pour la relation d'ordre $\prec$ définie par:

$$\mathcal{M}_1 \prec \mathcal{M}_2 \Leftrightarrow \mathcal{M}_1 \preceq^{P;\,Z} \mathcal{M}_2 \text{ et } \neg(\mathcal{M}_2 \preceq^{P;\,Z} \mathcal{M}_1)$$

La circonscription peut être utilisée pour représenter un raisonnement dans un milieu incomplet par le biais d'une théorie anormale [McCarthy 1986]. Ainsi, par exemple, pour représenter le fait qu'un dépôt à carburant (site à carburant) situé à l'endroit l est en général éloigné d'une ville située en l_1:

$$\forall x, y, l\ SiteCarburant(x,l) \wedge Ville(y,l_1) \wedge \neg ab_1(x,y) \Rightarrow Eloigne(x,y)$$

où la signification de $ab_1(x, y)$ est "*x est anormal par rapport aux villes éloignées d'un site à carburant*". Notons qu'il peut y avoir plusieurs anormalités, en autorisant les variations des autres prédicats. Ainsi, par exemple si A représente la conjonction de la formule précédente et de $SiteCarburant(a, l_1), Ville(b, l_2)$, alors on peut déduire:

$$Circum(A, ab_1, (SiteCarburant, Ville), Eloigne) \models Eloigne(a,b)$$

Si par contre on connaît $\neg Eloigne(a, b)$, alors la conclusion $ab_1(a, b)$ s'ensuit dans tous les modèles. La prise en compte des préférences dans cette logique peut se faire par la notion de priorité [Lifschitz 1985].

Considérons en effet la théorie suivante:

$$\forall x,y,l,l_1\ SiteCarburant(x,l) \wedge Ville(y,l_1) \wedge \neg ab_1(x,y) \Rightarrow Eloigne(x,y)$$
$$\forall x,y,l,l_1\ Raffinerie(x,l) \wedge Port(y,l_1) \wedge \neg ab_2(x,y) \Rightarrow \neg Eloigne(x,y)$$
$$\forall x,y,l\ Raffinerie(x,l) \Rightarrow SiteCarburant(x,l)$$
$$\forall x,y,l\ Port(x,l) \Rightarrow Ville(y,l)$$
$$Port(b,l_1), Raffinerie(c,l_1)$$

Il y a alors deux classes de modèles minimaux pour la circonscription, l'un dans lequel $ab_1(a,b)$ est vrai et dans lequel le site à carburant est éloigné de la ville, l'autre dans lequel $ab_2(a,b)$ est vrai et le site à carburant est proche de la ville. Le premier type de modèles est à rejeter puisque la connaissance qu'il s'agit d'une ville portuaire où se trouve une raffinerie est plus spécifique que la connaissance générale qui affirme que les sites à carburant sont éloignés des villes. Une manière générale de résoudre ce type de problème est d'ajouter une priorité aux conclusions de la logique, en définissant un ordre sur les prédicats à minimiser [McCarthy 1986][Lifschitz 1985]. Le prédicat P est alors divisé en ensembles exhaustifs $P^1, \ldots, P^n$, avec le prédicat P^1 ayant la plus haute priorité pour la minimisation, puis en cas d'égalité, la minimisation est conduite par rapport au prédicat P^2 ...

Définition 30.

$$Circum(A(P,Z);\ P^1 \succ \ldots \succ P^n;\ Z) = \bigwedge_{i=1}^{n} Circum(A(P,Z), P^i, P^{i+1}, P^n;\ Z)$$

Le problème pour représenter un raisonnement par similarité avec cette approche est qu'elle nécessite de connaître explicitement toutes les anormalités possibles et que celles-ci ne peuvent être inferées.

Outre ces logiques que nous venons de présenter, [Krauss et al. 1990][Brewka et al. 1997] introduisent la règle $\vee$ que nous rappelons ici:

$$\vee\colon \frac{\Gamma, y \Vdash a \quad \Gamma, z \Vdash a}{\Gamma, y \vee z \Vdash a}$$

Considérant alors des logiques **sans représentation logique des préférences**, avec seulement un opérateur d'inférence cumulatif et satisfaisant la règle ($\vee$), [Krauss et al. 1990] donne un théorème de validité et de complétude pour cette relation d'inférence associée à une légère modification des logiques à modèles préférentiels[16] pour la partie sémantique. L'ajout que nous effectuons d'une syntaxe pour les préférences ne permet plus de conserver cette équivalence, et donc les théorèmes de validité et de complétude associés.

16. Pour être plus précis, c'est la notion de modèle qui est modifiée en considérant qu'intuitivement, un modèle préférentiel représente un ensemble d'états et une relation d'ordre sur ces états. Supposons de plus que chaque état est, au sein d'un modèle, étiqueté par un ensemble de mondes, i.e. l'ensemble des mondes que le décideur considère comme possibles dans cet état. Alors considérer une relation binaire sur ces états étiquetés induit un degré de liberté supplémentaire pour construire ces modèles: un même ensemble de mondes peut apparaître à plusieurs états. Plus formellement,

Définition 31. [Krauss et al. 1990]
Un modèle préférentiel est un triplet $(W, v, \prec)$ où:

1. W est un ensemble de mondes possibles,
2. $v : W \mapsto 2^{\mathcal{U}}$ qui marque chaque état avec un ensemble non vides de mondes,
3. $\prec$ est un ordre partiel strict lisse sur W.

Pour toute formule $A \in L$, si $\hat{A}$ est l'ensemble de tous les mondes $w \in W$ tels que $v(w) \models A$; B est alors une conséquence préférentielle de A si tous les éléments maximaux de $\hat{A}$ satisfont B.

Ne nous reste dès lors plus qu'à définir le concept de relation lisse:

Définition 32. La relation $\prec$ est dite lisse si et seulement si pour tout $w \in \hat{A}$, ou bien w est minimal en $\hat{A}$, ou bien il existe un autre w' minimal en A avec $w' \prec w$.

Les relations de déduction possibiliste

Avant de voir les relations de déduction possibilites, commençons par rappeler les fondements de la théorie des possibilités. La théorie des possibilités (voir [Dubois and Prade 1998]) est fondée sur une notion de distribution de possibilité π qui est un classement complet des interprétations d'un langage donné fini propositionnel par une projection sur l'intervalle $[0, 1]$, où les interprétations les plus plausibles ont la valeur 1. $\pi(\omega) > 0$ signifie que ω est plausible, alors que $\pi(\omega) = 0$ signifie que ω est impossible. $\pi(\omega) > \pi(\omega')$ signifie que ω est plus plausible que ω'. Deux ensembles peuvent être associés à π:

- le degré de possibilité $\Pi(\phi) = sup\{\pi(w)/w \models \phi\}$, qui évalue le degré avec lequel ϕ est consistent avec les connaissances exprimée par π. Π satisfait les propriétés suivantes:
  - $\forall\phi, \forall\psi, \Pi(\phi \vee \psi) = max(\Pi(\phi), \Pi(\psi))$
  - $\Pi(\perp) = 0$
- La notion de degré de nécessité satisfait les propriétés duales:

$$N(\phi) = 1 - \Pi(\neg\phi) = inf\{1 - \pi(w)/w \models \neg\phi\}$$

La théorie des possibilités est guidée par un principe de spécificité minimale: une distribution de possibilité est dite au plus aussi spécifique qu'une distribution π' si et seulement si, pour toute interprétation ω, $\pi(\omega) \leq \pi'(\omega)$. Étant donné un ensemble de contraintes restreignant un ensemble de distributions de possibilités, le meilleur représentant est celui qui est le moins spécifique, qui assigne le plus haut degré de possibilité à chaque algèbre.

L'intervalle unité représentant une échelle ordinale, on peut associer à chaque distribution de possibilité la relation comparative $>_\pi$, définie par $\omega >_\pi \omega'$ ssi $\pi(\omega) > \pi(\omega')$, qui induit une partition "bien ordonnée" $(E_i)_{1 \leq i \leq n+1}$ de l'ensemble des interprétations [Benferhat et al. 1997b], i.e. telle que:

$$\forall\omega \in E_i, \forall w' \in E_j, \pi(\omega) > \pi(\omega') \text{ ssi } i < j \ (i \leq n+1, j \geq 1)$$

E_{n+1} est alors un sous ensemble d'interprétations impossibles telles que $\pi(\omega) = 0$. Par convention, E_1 représente l'état de fait le plus normal possible. Une distribution de possibilité partitionne donc l'ensemble des interprétations en classes d'interprétations également possibles. Pour toute distribution de possibilité comparative $>_\pi$, une relation de possibilité comparative peut aussi être définie pour toute paire de formules ϕ et ψ par:

$$\phi \geq_\Pi \psi \Leftrightarrow \exists w \models \phi, \forall w' \models \psi, w >_\pi w' \Leftrightarrow \Pi(\phi) \geq \Pi(\psi)$$

Ces relations peuvent également être définies de manière équivalente par les propriétés suivantes [Benferhat et al. 1997b] $\forall\alpha, \forall\beta, \forall\gamma$:

- complétude $\alpha \geq_\Pi \beta$ ou $\beta \geq_\Pi \alpha$,
- transitivité
- $\top >_\Pi \perp$,
- $\top \geq_\Pi \alpha \geq_\Pi \perp$,
- $\alpha \geq_\Pi \beta \Rightarrow \alpha \vee \gamma \geq_\Pi \beta \vee \gamma$,
- si $\alpha \models \beta$, alors $\beta \geq_\Pi \alpha$.

La notion de déduction possibiliste peut alors être introduite, relation proche de la notion de modèle préféré que nous avons introduite précédemment:

Définition 33. *(Extrait de [Benferhat et al. 1997b])*
Soit $>_\pi$ une distribution de possibilité comparative et $\{E_1, \ldots, E_n, E_{n+1}\}$ une partition bien ordonnée induite par $>_\pi$. Une interprétation w est un modèle π préféré d'une formule ϕ par rapport à $>_\pi$ si et seulement si:

1. $w \models \phi$,

2. $w \notin E_{n+1}$,

3. $\neg\exists w', w' \models \phi$ et $w' >_\pi w$

La relation de déduction possibiliste $\models_\pi$ peut alors définie:

Définition 34. *(Extrait de [Benferhat et al. 1997b])*
Une formule ψ est une conséquence possibiliste de ϕ pour la relation $>_\pi$, ce qui est noté $\phi \models_\pi \psi$ ssi $\phi \wedge \psi >_\Pi \phi \neg\psi$.

Cette relation de déduction satisfait les bonnes propriétés que nous avons mentionnées pour la similarité. [Benferhat et al. 1992][Benferhat et al. 1997b] donnent un ensemble de méta propriétés que satisfait cette relation de déduction, ce qui leur permet de fournir un théorème de validité et de complétude pour cette logique, en montrant d'une part, que $\alpha >_\pi \beta \Leftrightarrow \alpha \vee \beta \models_\pi \neg beta$ et qu'une relation satisfaisant un système syntaxique comprenant les règles que nous avons cités et celles proposées par [Benferhat et al. 1997b], est telle que la relation $>$ définie par $\alpha > \beta$ ssi $\alpha \vee \beta \Vdash \neg\beta$ est une relation de possibilité comparative. Cependant, elle n'autorise pas une formulation déclarative des préférences.

3.2.2 Les relations d'inférence non cumulatives

L'approche que nous avons présentée des relations d'inférence non monotone n'est pas celle qui a été initialement utilisée pour leur spécification. Initialement, elles ont plutôt été définies par une propriété de consistance, et ce, de trois manières au moins:

- dans le premier cas, des règles d'inférence avec une syntaxe particulière (les défauts) permettent de définir des symboles méta-logiques représentant les croyances d'un individu (les extensions): c'est le cas des logiques des défauts,

- dans le deuxième cas, les extensions restent à l'intérieur de logique, et c'est la signature de celle-ci qui est augmentée de symboles permettant de prendre en compte les croyances: c'est le cas des logiques modales non monotones,

- dans le troisième cas, on reste dans le cadre de la logique classique et c'est un ensemble de théories du premier ordre sur lequel une relation d'ordre est définie qui permet ainsi de rendre compte de la notion de croyances.

Ces trois approches ne vérifient pas, dans leur cadre le plus général, la notion de cumulativité. Pour que ce soit le cas, il faut alors réduire fortement la syntaxe considérée.

Les logiques de type point fixe

L'une des premières, et certainement l'une des plus utilisées de ces logiques, est la logique des défauts de [Reiter 1980]. Cette logique pose comme hypothèse la représentation des connaissances sous la forme d'une théorie des défauts, c'est à dire d'une paire (D, W), où W est un ensemble de formules du premier ordre représentant les faits connus avec certitude, et D un ensemble de défauts représentant des faits incertains. Un ensemble de défauts est un ensemble de formules de la forme $\alpha : \beta_1, \ldots, \beta_n/\omega$, où α (le pré-requis), β_i (les justifications) et ω (la conclusion) sont des formules du premier ordre. Le défaut a la signification intuitive que si α est connu, alors ω l'est sachant que l'ensemble des β_i doit être consistant avec ce qui est connu.

Ces défauts permettent de générer un ensemble d'ensembles de formules, chaque ensemble représentant les croyances que peut posséder un individu raisonnant sachant W et D. D'un point de vue plus formel:

Définition 35. Soit (D, W) une théorie des défauts, et soit S un ensemble de formules. Soit $\Gamma(S)$ le plus petit ensemble tel que:

- $W \subseteq \Gamma(S)$,

- $Th(\Gamma(S)) = \Gamma(S)$, où Th désigne la fermeture déductive d'une théorie,

– Si $\alpha : \beta_1, \ldots, \beta_n/\omega \in D, \alpha \in \Gamma(S), \forall 1 \leq i \leq n, \neg\beta_i \notin \Gamma(S)$, alors $\omega \in \Gamma(S)$.

E est une extension de (D, W) si et seulement si $\Gamma(E) = E$.

[Reiter 1980] donne aussi une définition semi-inductive des extensions, de la manière suivante:

Proposition 36. *Soit (D, W) une théorie des défauts. Soit E un ensemble de formules, et soit $(E_i)_{i\in\mathbb{N}}$ la suite suivante:*

– $E_0 = W$,

– $E_{i+1} = E_i \cup \{\omega, \alpha : \beta_1, \ldots, \beta_n/\omega \in D, \alpha \in E_i \text{ et } \neg\beta_i \in E\}$

Alors E est une extension du couple (D, W) si et seulement si $E = \bigcup_{i=0}^{\infty} E_i$.

Deux types de relations de déduction peuvent alors être proposés à partir de ces définitions: ou bien les formules démontrables à partir d'une théorie des défauts sont les formules appartenant à une extension de la théorie, ou bien les formules démontrables appartiennent à l'intersection des extensions de cette théorie.

D'un point de vue sémantique, cette définition correspond aussi à une relation de préférence seulement **asymétrique** (elle n'est plus transitive) entre des mondes, qui sont des classes d'algèbres validant un même ensemble de formules:

Définition 37. Soit $\mathcal{M}$ un ensemble d'algèbres du langage construit sur la signature Σ. On définit par $//\mathcal{M}//_L = \{\varphi \in L_\Sigma / \forall M \in \mathcal{M}, M \models \varphi\}$, l'ensemble des formules qui sont validées par la classe d'algèbres $\mathcal{M}$. Un monde est alors défini par l'ensemble des algèbres qui valident les mêmes formules:

$$\mathcal{M}^+ = \{M \in \mathbf{Mod}_L / \forall \varphi \in //\mathcal{M}//_L, M \models_L \varphi\}.$$

Commençons par spécifier la notion d'ordre sur les mondes induite par une règle des défauts:

Définition 38. *Adapté de [Etherington 1989]*
Soit $\delta \equiv \alpha : \beta_1, \ldots, \beta_n$ un défaut de D. On définit par $\sqsubset_\delta$ la relation de pré-ordre partiel strict définie sur l'ensemble des mondes de la logique par, pour tout couple de mondes $\mathcal{M}, \mathcal{N}$, par $\mathcal{M} \sqsubset_\delta \mathcal{N}$ si et seulement si:

1. $\mathcal{N} \models \alpha$,
2. $\neg(\exists i, \exists \beta_i \in, \mathcal{N} \models \neg\beta_i)$
3. $\mathcal{M} = \mathcal{N} \setminus \{M / M \models \omega\}$

Intuitivement, $\mathcal{M} \sqsubset_0 \mathcal{N}$ signifie que la description de l'environnement $\mathcal{M}$ est prioritaire face à la description $\mathcal{N}$. La définition se généralise à une théorie des défauts de la manière suivante:

Définition 39. [Etherington 1989][Voorbraak 1993]
Dans le cas d'une théorie des défauts (W, D), on note $\prec_D$ la fermeture transitive de l'union des relations $\sqsubset_D$ restreinte aux classes de modèles dans lesquelles W est valide. La classe d'algèbres $\mathcal{M}$ est dite D-stable ssi:

$$\exists D' \subseteq D, \mathcal{M} \prec_{D'} \backslash\backslash W \backslash\backslash \text{ et}$$

$$\text{chaque justification d'un défaut de } D' \text{ est vraie dans un } M \in \mathcal{M}$$

Intuitivement, un ensemble de modèles est stable si c'est une spécialisation maximale de l'ensemble des modèles de W, et si celle-ci ne réfute aucun des défauts (éléments négligés) utilisés dans la spécialisation. La définition de la stabilité peut alors être prise en compte pour donner une sémantique en terme de mondes préférés à la logique des défauts de [Reiter 1980]:

Définition 40. Adapté de [Voorbraak 1993]
La relation asymétrique sur les mondes telle que tout élément minimal de cette relation est modèle d'une extension de la théorie des défauts (D, W) est définie par l'agrégation lexicographique[17] de:

- la restriction de la fermeture transitive de $\prec_\delta$ aux classes de modèles dans lesquelles W est valide,
- et de la relation $\succ_0$ définie comme suit:

$$\mathcal{M} \prec_0 \mathcal{N} \text{ si et seulement si } \mathcal{M} = \emptyset \neq \mathcal{N}$$
$$\text{et il n'existe aucun } \prec_D \text{ minimum } \mathcal{N}' \text{ stable tel que } \mathcal{N}' \prec_D \mathcal{N},$$

Ayant maintenant les définitions des logiques des défauts, voyons les propriétés formelles de la relation de déduction qu'elles engendrent:

Définition 42. Soit D un ensemble de défauts, W un ensemble de formules du premier ordre. On définit la relation $\Vdash_D$ la relation telle que $W \Vdash_D p$ si et seulement si p est contenue dans toutes les extensions de (D, W).

Existence La première propriété attendue des extensions est leur existence. Elle n'est cependant pas toujours garantie. Considérons en effet la théorie suivante:

$$(\{: \neg A/A\}, \emptyset)$$

Cette théorie n'a aucune extension. Une forme syntaxique particulière de théorie des défauts admet toujours des extensions, à savoir les théories des défauts normales. Dans ces théories, tous les défauts sont de la forme $\alpha : \beta/\beta$. Ces théories ont de plus la propriété d'être semi-monotones, c'est à dire que si E est une extension de (D, W), alors toute théorie des défauts de la forme $(D \cup D', W)$ a une extension E' telle que $E \subseteq E'$.

Cumulativité La deuxième propriété attendue de cette relation de déduction par défaut est sa cumulativité. Elle n'est cependant pas toujours vérifiée. Considérons l'exemple suivant [Brewka et al. 1997]:

$$: B/B$$
$$A \wedge B : \neg B/\neg B$$

Il est aisé de se convaincre que cette théorie n'a qu'une extension contenant $A \wedge B$. Si cependant, règle de coupure, $A \wedge B$ est ajouté à cette théorie des défauts, une seconde extension contenant $\neg B$ est générée et A n'est alors plus déductible pour la relation de déduction par défaut. [Dix 1992] a cependant montré qu'en restreignant encore la forme syntaxique des défauts de telle sorte qu'il existe une formule B telle que tout défaut soit de la forme : $B \wedge C/C$ la logique des défauts était cumulative.

Règle $\vee$ La troisième propriété attendue de cette logique est la satisfaction de la règle $\vee$. Considérons maintenant la théorie des défauts suivante:

$$A : B/B$$
$$C : B/B$$
$$A \vee C$$

Il est aisé de voir que la seule extension de cette théorie est $Th(A \vee C)$. Ramené à la relation de déduction, cet exemple montre que cette relation de conséquence ne satisfait pas la règle $\vee$.

17. Rappelons la définition d'une agrégation lexicographique:

Définition 41. Soit une famille de n relations asymétriques $(<_i)_{1\leq i\leq n}$, et soit $(\sim_i)_{1\leq i\leq n}$ les relations d'équivalence associées à ces n relations binaires. On définit par $\lambda(<_1, \ldots, <_n)$ la fonction d'agrégation lexicographique des $(<_i)_{1\leq i\leq n}$ par récurrence de la manière suivante:

- $\lambda() = \emptyset$
- $\lambda(<_1, \ldots, <_{k+1}) = \lambda(<_1, \ldots, <_k) \cup (\sim_1 \cap \ldots \cap \sim_k \cap <_{k+1})$

Comme on peut aisément l'imaginer, les logiques des défauts sont utilisables pour représenter un raisonnement par similarité. Cependant, les propriétés formelles que nous venons d'étudier montrent que cette logique n'est pas bien adaptée aux raisonnement par similarité en fonction de préférence, puisqu'elle n'est pas cumulative et ne satisfait pas non plus la règle $\vee$.

Les logiques de type logique modale non monotone

Face à cette approche, où les conséquences sont définies par un symbole méta logique (les extensions), les logiques modales non monotones [McDermott and Doyle 1980][Moore 1985][Marek 1991] utilisent un opérateur modal pour exprimer directement dans le langage le fait qu'une formule soit crue ou non. On se restreindra ici dans la présentation de la logique épistémique de [Moore 1985], puisque c'est la logique la plus utilisée et que ses liens avec la logique des défauts de [Reiter 1980] sont bien connus.

[Moore 1985] ajoute un opérateur modal $\mathcal{L}$ dans le langage logique; sa signification est que si p est une proposition, alors $\mathcal{L}p$ signifie que *"p est cru"*. L'idée, utilisant cette construction syntaxique, est de modéliser un agent idéal ayant des capacités complètes d'introspection. Cela signifie que:

- si p appartient aux croyances de l'agent, alors $\mathcal{L}p$ doit aussi appartenir à ces croyances,
- si p n'appartient pas aux croyances de l'agent, alors $\neg\mathcal{L}p$ doit aussi appartenir aux croyances de cet agent.

Les ensembles de formules qui satisfont ces deux ensembles de propriétés sont appelés des ensembles stables. Pour donner un exemple de non monotonie dans cette logique, si l'on considère la formule $Piste(X,l_1) \wedge Taxiway(Y,l_1) \wedge TourControle(Z,l_1) \neg\mathcal{L}\neg Aerogare(W,l_1)/AeroportMilitaire(T,l_1)$ et que l'on suppose que l'individu connaisse seulement $Piste(a,l), Taxiway(b,l), TourControle(c,l)$, alors comme $\neg Aerogare(w,l)$ n'est pas connu, il s'ensuit que $\neg\mathcal{L}\neg Aerogare(w,l)$ est dans l'ensemble des croyances, et que l'on peut donc inférer $AeroportMilitaire(t,l)$. Si on obtient une information sur le fait qu'un aérogare est dans l'aéroport, alors $\neg\mathcal{L}\neg Aerogare(w,l)$ sera retiré de l'ensemble des croyances de l'individu, ainsi que la conclusion précédente. La définition de ces ensembles de croyances est fondée sur une équation de point fixe qui, si l'on reprend la définition de [Konolige 1988] est:

Définition 43. Soit A un ensemble de formules. T est une extension de A si et seulement si

$$T = \{p/A \cup Bel(T) \cup Disbel(T) \models p\}$$

où $Bel(T) = \{\mathcal{L}\ q/q \in T\}$ et $Disbel(T) = \{\neg\mathcal{L}\ q/q \notin T\}$.

On peut s'interroger sur la relation de déduction qu'induit cette logique. Il est assez facile d'étudier ces propriétés puisqu'il existe une traduction directe d'une logique des défauts dans une telle logique épistémique modale [Konolige 1988]:

$$A : B_1, \ldots, B_n/C \Rightarrow \mathcal{L}A \wedge \neg\mathcal{L}\neg B_1 \wedge \ldots \wedge \neg\mathcal{L}\neg B_1 \Rightarrow C$$

Comme d'autre part, toute formule modale peut, par le même moyen, être transformée en une règle des défauts, la traduction fonctionne aussi dans le sens inverse. [Konolige 1988] montre que les extensions d'une théorie des défauts (D,W) correspondent exactement à la partie objective (c'est à dire celles ne contenant pas l'opérateur $\mathcal{L}$) de certaines des extensions de la traduction en logique modale. Les extensions qui ne correspondent pas à celles de la logique des défauts associée sont des extensions contenant des dépendances entre les croyances, comme par exemple les extensions $\{\mathcal{L}p, p\}$ et $\{\neg\mathcal{L}p\}$ de la théorie $\{\mathcal{L}p \Rightarrow p\}$. D'autres systèmes modaux non monotones ont été proposés, comme par exemple [Lin and Shoham 1992] qui, utilisant deux opérateurs modaux indépendants, ont pu montrer que leur logique subsumait la logique précédente et la logique des défauts de [Reiter 1980]. Étant donné les équivalences entre les formalismes, les remarques données sur l'application des logiques des défauts au raisonnement par similarité aux théories des défauts s'appliquent ici.

Les logiques de type consistance maximale

Dans les logiques des défauts, les défauts sont représentés par des règles d'inférence non standard, et dans les logiques modales par des formules modales spéciales. Ce type de logique vise à modéliser les défauts comme des formules logiques classiques qui sont traitées de manière spéciale. Si l'on représente les défauts comme des règles classiques, alors ces règles sont en général inconsistantes. Une manière de résoudre ce problème est de considérer non plus la théorie entière, mais certains de ses sous ensembles de formules maximaux consistants. Il est alors possible de définir une relation de préférence entre ces ensembles de théories et d'utiliser les préférences pour définir une notion de prouvabilité. Par exemple, si l'on a:

1. $Piste(X, l_1) \wedge Taxyway(Y, l_1) \wedge TourControle(Z, l_1) \wedge AireStationnementNonAlignees(R, l_1)$ $\Rightarrow AeroportMilitaire(T, l_1)$
2. $Piste(X, l_1) \wedge Taxyway(Y, l_1) \wedge TourControle(Z, l_1) \wedge Aerogare(R, l_1) \Rightarrow AeroportCivil(T, l_1)$
3. $AeroportCivil(T, l_1) \wedge AeroportMilitaire(W, l_1) \Rightarrow \perp$,
4. $Piste(a, l), Taxiway(b, l), TourControle(c, l)$

on obtient des sous théories contenant deux de ces formules. L'idée en est que 1. et 2. contiennent des informations par défaut qui sont prises en compte en préférant les sous théories qui ne contiennent pas 1. ou 2.

Définition 44. [Brewka et al. 1997]
Soit T un ensemble de formules (pouvant être inconsistent), $Sub(T)$ l'ensemble des sous ensembles consistants de T, et $\prec$ un ordre partiel strict sur $Sub(T)$. Une formule p est alors fortement (resp. faiblement) démontrable de $(T, \prec)$ si et seulement si p est prouvable de tous les (resp. de certains des) éléments minimaux de $Sub(T)$ pur $\prec$. On appelle extensions de T les ensembles E tels que $E = Th(S)$, où S est un ensemble minimal pour $\prec$ de $Sub(T)$.

Une façon simple de définir un ordre $\prec$ sur les sous théories est alors d'introduire une structure sur T et d'induire la relation d'ordre de cette structure. Par exemple, [Poole 1988] divise T en:

- un ensemble consistant F de formules représentant les faits connus,
- un ensemble Δ de formules représentant des informations hypothétiques ou par défaut.

Une sous théorie est préférée si et seulement si elle contient F. [Poole 1988] a alors pu montrer que cette approche simple était équivalente à un cas spécial des théories des défauts de [Reiter 1980] dans lesquels tous les défauts sont de la forme $(: p/p)$. Les propriétés formelles de cette relation de déduction se déduisent alors de cette équivalence. C'est la raison pour laquelle cette logique ne saurait convenir à une représentation de la similarité.

3.2.3 Les logiques conditionnelles

Une autre approche utilisable pour représenter un raisonnement par similarité aurait pu être d'utiliser les logiques conditionnelles et de traiter l'opérateur d'inférence par similarité comme un connecteur logique $\Rrightarrow$ dans une certaine logique. Un cas aurait alors été représenté par une formule $A \Rrightarrow B$, la formule A représentant un problème, la formule B une action. Ceci nous amène à considérer les logiques conditionnelles qui utilisent justement cet opérateur.

Nous ne nous intéresserons ici qu'à la plus représentative de ces logiques conditionnelles (le lecteur est renvoyé à [Nute 1984] pour une présentation complète de ce domaine) et nous ne détaillerons ici que la logique C_4 [Boutilier 1990][Lamarre 1991][Brewka et al. 1997] ici:

Définition 45. La logique conditionnelle C_4 a les axiomes suivants:

(PC) Toutes les tautologies,

(ID) $A \Rrightarrow A$,

(CS) $A \wedge B \Rightarrow (A \Rrightarrow B)$

(NEC) $(\neg A \Rrightarrow A) \Rightarrow A$

(MP) $(A \Rrightarrow B) \Rightarrow (A \Rightarrow B)$

(CC) $(A \Rrightarrow B) \wedge (A \Rrightarrow C) \Rightarrow (A \Rrightarrow B \wedge C)$

(AD) $(A \Rrightarrow C) \wedge (B \Rrightarrow C) \Rightarrow (A \vee B \Rrightarrow C)$

(ASC) $(A \Rrightarrow B) \wedge (A \Rrightarrow C) \Rightarrow (A \wedge B \Rrightarrow C)$

et les règles d'inférence suivantes:

(Mp) De A et de $A \Rightarrow B$ inférer B,

(RCM) De $B \Rightarrow C$ inférer $(A \Rrightarrow B) \Rightarrow (A \Rrightarrow C)$

On retrouve alors les propriétés que nous avons requises pour notre opérateur d'inférence par similarité pour lequel les préférences ne sont pas exprimées dans la logique:

- ID est le principe d'inclusion,
- ASC est la règle de monotonie prudente,
- AD est la règle $(\vee)$,
- CC est la règle $(\wedge)$,
- $((A \wedge B) \Rrightarrow C) \wedge (A \Rrightarrow B) \Rightarrow (A \Rrightarrow C))$ est la règle de coupure, qui est donc un théorème de C_4,
- De $A \Rightarrow B)$, inférer $A \Rrightarrow B$, on déduit que la règle de coupure est un théorème de C_4.

On retrouve donc là une logique qui satisfait les bonnes propriétés que nous avons définies pour un raisonnement par similarité sachant des préférences. De plus, cette logique a un lien qui n'est pas sans intérêt avec les logiques à modèles préférentiels:

Théorème 46. *[Boutilier 1990][Lamarre 1991]*
Soient A, B deux propositions qui ne contiennent pas d'opérateurs conditionnels $(\Rrightarrow)$. Alors, si $\vdash_{C_4} A \Rrightarrow B$, pour tout ordre strict partiel bien fondé $\prec$, $A \vdash_{\prec} B$.

De ce fait, les théorèmes de C_4 de la forme $A \Rrightarrow B$ sont appliquables à toute logique préférentielle, incluant la circonscription. L'utilisation de cette logique pour faire des raisonnements par similarité en fonction de préférences pose deux problèmes:

- d'une part, la non existence d'axiomes se référant directement aux préférences,
- d'autre part, la non possibilité d'écrire que A entraîne B de manière conditionnelle si C, ce qui revient à dire que A et C sont similaires.

Ces logiques ne satisfont donc pas les pré-requis dont nous avons besoin.

3.3 Les méta-axiomes liés à la notion de préférence

Ayant les méta-axiomes désirables pour la notion de similarité, nous allons maintenant analyser les méta-axiomes désirables pour la notion de préférence. Rappelons que nous avons défini une préférence comme étant un objet syntaxique de la forme $(\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2)$ où $\varphi_1, \varphi_2, \phi_1, \psi_1, \psi_2, \phi_2$ sont des formules du premier ordre. Une préférence a la signification intuitive "*le fait ϕ_1 résultant de la connaissance de φ_1 et de la négligeabilité de φ_2 est préféré au fait ϕ_2 résultant de la connaissance de ψ_1 et de la négligeabilité de ψ_2*".

Commençons par caractériser la relation de déduction par similarité sachant des préférences, indépendamment de la forme syntaxique des formules de la préférence:

Anti-Monotonie $$\frac{(\Gamma, (\Pi \cup p)) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$$

Monotonie conditionnelle aux préférences $$\frac{(\Gamma, \Pi) \vdash_S a \quad \Gamma \nvdash \varphi_1, \phi_1}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

La première de ces deux règles (l'anti-monotonie) vise à formaliser le fait que si l'on a moins de préférence, on peut déduire au moins autant. La deuxième règle (la monotonie conditionnelle aux préférences) est le pendant de la règle sur la négation que nous avons rejetée dans la section précédente: elle spécifie en effet que toute préférence qui n'est pas contradictoire avec la connaissance de la théorie ne change pas les conclusions.

Les règles correspondant aux propriétés de la relation de dérivation pour les symboles usuels sont donnés en annexe D. Dans ces règles, le lien entre préférence et similarité n'apparaît pas non plus directement dans les règles. Ceci est dû au fait que ce lien dépend uniquement de la fonction de similarité choisie, et donc de sa syntaxe. Si l'on veut dans un premier temps décrire les propriétés liées au raisonnement par similarité en fonction de préférences sans imposer une relation de similarité particulière, il faut s'en tenir là et renoncer au théorème de complétude: on ne donne qu'une série de *bonnes propriétés* que doit satisfaire la relation de déduction.

On notera également que les règles que nous avons définies pour la similarité induisent certainement une relation de préférence entre les mondes de la logique considérée, comme nous l'avons mentionné dans la définition 31. Tel quel, rien ne garantit donc que ces préférences ne soient pas contradictoires avec celles exprimées syntaxiquement par les sextuplets.

Bien qu'à notre connaissance, aucun méta-axiome n'aît jamais été posé pour la notion de préférence, cela ne signifie pas pour autant que l'importance de l'utilisation de la notion de préférence, ou tout au moins de la notion de priorité, n'ait pas été évoquée dans les logiques que nous avons mentionnées. De fait, au moins trois types d'approches ont été utilisées pour prendre en compte une telle notion:

- Certaines approches utilisant des notions de préférences qui doivent être spécifiées par l'utilisateur, mais qui ne font pas partie du langage logique; c'est le cas par exemple de la circonscription avec priorités que nous avons mentionnée de [Lifschitz 1985], de [Konolige 1988] pour les théories épistémiques ou de [Brewka 1989] pour les logiques des défauts,

- Certaines approches utilisant *implicitement* des informations de priorité liées à la notion de spécificité, comme par exemple [Geffner and Pearl 1992],

- Des approches utilisant *explicitement* des informations de préférences, permettant ainsi au système de raisonner sur ces priorités: c'est le cas des approches de [Brewka 1994][Brewka 1996].

Ayant déjà abondamment discuté des deux premières approches dans la section précédente, nous ne nous intéresserons ici qu'au cas où l'on peut raisonner sur les préférences. L'approche de [Brewka 1994] est une modification de la logique des défauts de [Reiter 1980] pour prendre en compte les préférences, mais seulement avec des défauts normaux, donc trop limitée du point de vue de la puissance d'expression du langage. Nous ne nous intéresserons ici qu'à l'approche de [Brewka 1994] qui est une modification de la logique des défauts de [Reiter 1980] avec l'introduction d'une relation d'ordre $\prec$ entre des défauts normaux. Elle a été généralisée par [Brewka 1996], qui utilise des programmes généraux avec à la fois négation et négation par échec, cette dernière forme de négation correspondant à des défauts non normaux.

Commençons par définir une logique des défauts avec priorités avant de définir un raisonnement sur ces priorités. Reprenant les notations de [Brewka 1989], un défaut $A : B/B$ est noté $A \to B$, et l'on défini la notion d'activité d'un défaut:

Définition 47. Soit E un ensemble de formules, $\delta \equiv A \to B$ un défaut. On dit que δ est actif dans E ssi 1) $A \in E$, 2) $B \notin E$, 3) $\neg B \notin E$.

On peut alors définir les extensions d'une logique des défauts avec priorités (PDL) de la manière suivante:

Définition 48. *(Extrait de [Brewka 1994])*
Soit $(D, W, \prec)$ une théorie des défauts prioritaire (D ensemble de défauts, W ensembles de formules fermées du premier ordre), et $\lll$ un ordre strict contenant $\prec$. On dit que E est une extension prioritaire de Δ générée par $\lll$ si et seulement si $E = \bigcup_i E_i$ avec:

- $E_0 = Th(W)$, où Th désigne la fermeture déductive d'une théorie,
- $E_{i+1} = \begin{cases} \bullet\ E_i \text{ si aucun défaut n'est actif dans } E_i \\ \bullet\ Th(E_i \cup B) \text{ sinon, où B est la conséquence d'un défaut minimal pour } \lll \text{ actif dans } E_i \end{cases}$

Il s'ensuit donc

Définition 49. *(Extrait de [Brewka 1989])*
Soit $(D, W, \prec)$ une théorie des défauts prioritaire. E est une extension prioritaire de Δ si et seulement si il existe un ordre strict total contenant $\prec$ qui génère E au sens de la définition précédente.

Ainsi, par exemple, si l'on considère la théorie suivante,

1. $Bird \to Fly$	3. $Penguin \to \neg Fly$
2. $Penguin \to Bird$	4. $Penguin$

Dans la logique des défauts de [Reiter 1980], cette théorie a deux extensions, en l'occurrence:

$$E = Th(\{Penguin, Bird, Fly\})$$
$$E' = Th(\{Penguin, Bird, \neg Fly\})$$

où Th désigne toujours la fermeture déductive d'une théorie. En définissant l'ordre $3 < 1$, trois ordres totaux respectent $<$ qui sont:

$$2 < 3 < 1$$
$$3 < 2 < 1$$
$$3 < 1 < 2$$

Il est aisé de vérifier que seule E' est compatible avec ces ordres.

Cette approche permet de poser la définition d'un raisonnement sur les priorités ainsi définies. Pour ce faire, on doit pouvoir se référer directement à un défaut en le nommant. Le langage logique est donc étendu de deux manières:

- [Brewka 1994] utilise des défauts nommés dans le sens de [Poole 1988] de la forme $d_i : A \to b$. On suppose que des défauts différents ont des noms différents. Logiquement parlant, les noms des défauts sont traités comme des constantes,
- [Brewka 1994] utilise un prédicat binaire spécial $\prec$ pour représenter les priorités des défauts. Par exemple, si d_1 et d_2 sont des noms de défauts, la formule $d_2 \prec d_1$ signifie que d_2 a priorité sur d_1.

Une théorie des défauts est alors une paire de défauts nommés. On suppose de plus que W contient les axiomes qui assurent que $\prec$ est un ordre strict partiel (i.e. irréflexivité et transitivité). Étant donné un ensemble de défauts D, on note D^0 l'ensemble de défauts correspondant sans noms. L'idée est de considérer les extensions de cette théorie des défauts avec priorités sans prendre en compte dans un premier temps la relation d'ordre entre les défauts; on ajoute ensuite à chaque extension l'information sur l'ordre des défauts et toutes les extensions qui sont incompatibles avec cet ordre (elle utilise des défauts moins préférés que d'autre alors que les deux étaient actifs) sont supprimées. Plus formellement,

Définition 50. Soit $\Delta = (D, W)$ une théorie des défauts nommée, E une extension de la théorie (D^0, W) et soit $\lll$ un ordre strict total sur D^0. On dit que $\lll$ est compatible avec E si et seulement si:

$$E \cup \{d_i \prec d_k / d_i : r_i \in D, d_k, r_k \in D, r_i \lll r_k\}$$

est consistent.

Définition 51. Soit $\Delta = (D, W)$ une théorie des défauts nommée, E une extension de la théorie (D^0, W). On dit que E est une extension avec priorités de Δ si et seulement si c'est une extension prioritaire de Δ générée par un ordre total $\lll$ compatible avec E.

Pour reprendre l'exemple précédent, considérons la théorie suivante:

$$\begin{gathered} d_1 : Bird \rightarrow Fly \\ d_2 : Penguin \rightarrow \neg Fly \\ Penguin \\ Bird \\ d_2 \prec d_1 \end{gathered}$$

Négligeant l'information de priorité, cette théorie a deux extensions: Dans la logique des défauts de [Reiter 1980], cette théorie a deux extensions, en l'occurrrence:

$$\begin{gathered} E = Th(\{Penguin, Bird, Fly\}) \\ E' = Th(\{Penguin, Bird, \neg Fly\}) \end{gathered}$$

mais E' est telle que l'ordre total la générant $(Bird \rightarrow Fly) \lll (Penguin \rightarrow \neg Fly)$ n'est pas compatible avec l'information de priorité $d_2 \prec d_1$.

Si cette approche se révèle intéressante puisqu'elle introduit une notion de préférence dans la logique et permet de raisonner sur celle-ci, elle a au moins un inconvénient majeur qui consiste a reposer sur la logique des défauts de [Reiter 1980], dont nous avons vu dans les sections précédentes qu'elle ne satisfait pas les méta axiomes que nous avons requis pour une relation d'inférence par similarité

Ces deux problèmes nous conduisent à choisir une autre logique pour la représentation des raisonnements par similarité en fonction de préférences. Nous rappelons ci-après les propriétés que doit satisfaire cette logique.

3.4 Conclusion

Nous avons vu dans ce chapitre les propriétés que devait satisfaire une relation de déduction par similarité avec des préférences. Nous avons également vu que les logiques non standards usuelles ne suffisaient pas à représenter un tel raisonnement, soit parce qu'elle ne satisfont pas ces propriétés, soit parce qu'il leur manque une représentation syntaxique formelle des préférences.

Pour nous résumer, deux catégories de "*bonnes*" propriétés sont attendues d'une relation de déduction par similarité:

• Celles concernant la similarité proprement dite:

Supra Classique	$\dfrac{\Gamma \vdash a}{(\Gamma, \emptyset) \vdash_S a}$
Inclusion	$(\Gamma \cup a, \emptyset) \vdash_S a$
Monotonie prudente	$\dfrac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma, \Pi) \vdash_S a}{(\Gamma \cup y, \Pi) \vdash_S a}$
Coupure	$\dfrac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma \cup y, \Pi) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$
$\wedge$	$\dfrac{(\Gamma, \Pi) \vdash_S a \quad (\Gamma, \Pi) \vdash_S b}{(\Gamma, \Pi) \vdash_S a \wedge b}$
$\vee$	$\dfrac{(\Gamma \cup y, \Pi) \vdash_S a \quad (\Gamma \cup z, \Pi) \vdash_S a}{(\Gamma \cup (y \vee z), \Pi) \vdash_S a}$

- Celles concernant les préférences:

Anti-Monotonie	$\dfrac{(\Gamma, (\Pi \cup p)) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$
Monotonie conditionnelle aux préférences	$\dfrac{(\Gamma, \Pi) \vdash_S a \quad \Gamma \nvdash \varphi_1, \phi_1}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$

Avant de voir une telle bonne relation de déduction par similarité, nous verrons dans la chapitre suivant les modélisations usuellement proposées pour le raisonnement à partir de cas. Nous verrons alors comment notre approche s'en distingue, et nous verrons dans la chapitre suivant une telle "*bonne*" relation de déduction par similarité.

4

Raisonnement par similarité avec des préférences et raisonnement à partir de cas

4.1 Introduction

Nous présentons dans ce chapitre les formalisations usuelles du raisonnement à partir de cas, principalement à partir des logiques de descriptions. Après une brève présentation des principes de raisonnement à partir de cas, nous verrons alors les propriétés formelles des logiques de descriptions utilisées pour effectuer des jugements de similarité, et leur comparaison avec l'approche que nous proposons. La similarité n'ayant au départ été définie que par l'ensemble des éléments communs à un cas courant et à un cas mémorisé, elle a ensuite été étendue pour prendre en compte les éléments différençiant cas courant et cas mémorisé, grâce principalement à des adaptations des logiques des défauts. Ces dernières approches, si elles permettent comme la nôtre de faire dépendre les appariements des préférences, proposent une interprétation trop restrictive de celles-ci, qui ne satisfont pas les propriétés que nous avons défini au chapitre précédent.

4.2 Les systèmes de raisonnement à partir de cas

Les systèmes de raisonnement à partir de cas ont pour objectif de résoudre un cas (appelé aussi cas cible) en recherchant dans la base des problèmes similaires (cas sources) qui ont déjà été traités et en utilisant (modulo bien sûr d'éventuelles adaptations) les solutions de ces anciens problèmes. Généralement, le processus de fonctionnement d'un système de raisonnement à partir de cas (RàPC) est découpé en une suite d'étapes:

- Recherche des cas similaires mémorisés,
- Réutilisation d'une ou plusieurs solutions mémorisées,
- Évaluation des solutions adoptées,
- Apprentissage de ces nouveaux cas,

Seules les trois premières étapes nous intéressent par rapport à l'approche que nous proposons. La recherche des anciens cas similaires consiste à effectuer des similarités entre le cas courant et les cas sources en mémoire. Si le nouveau cas coïncide parfaitement avec un cas déjà enregistré, on parle de *filtrage complet*, sinon on parle de *filtrage partiel*. Le degré de similarité peut varier en fonction de préférences: il y a donc des cas plus similaires que d'autres. Plusieurs approches ont été proposées en RàPC pour la recherche des anciens cas et pour les critères de préférences [Schank 1982] [Kolodner 1993]. Ces approches

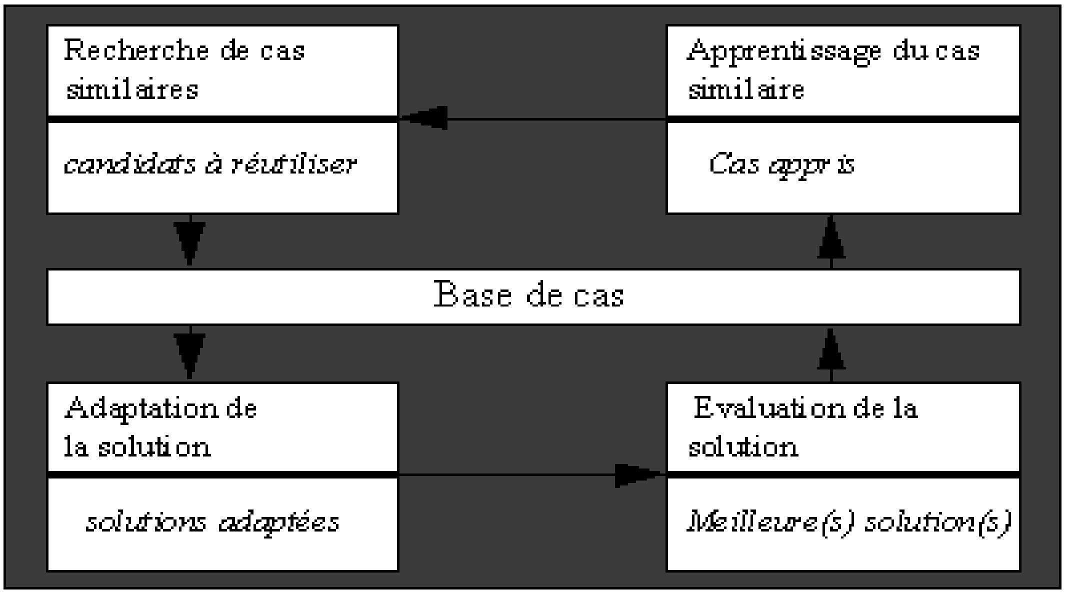


FIG. 4.1: *Le principe de fonctionnement d'un CBR*

reposent principalement sur les théories psychologiques de la cognition et ont amené un grand nombre de propositions sur la représentation, l'indexation et l'organisation des cas mémorisés. Ces systèmes de raisonnement à partir de cas raisonnent par approximation et similarité. Pour résoudre un nouveau cas en utilisant des cas existants en mémoire, plusieurs modes de raisonnement sont utilisés: tout d'abord, un index est construit à partir du nouveau cas en extrayant les caractéristiques qui sont *suffisamment abstraites* pour rendre ce cas utile dans une grande variété de situations aussi bien que *suffisamment concrètes* pour être facilement reconnaissable dans des situations futures. Dans la plupart des approches, le vocabulaire de l'index est un sous ensemble du vocabulaire utilisé pour la représentation symbolique du cas [Kolodner 1993]. Ensuite, l'index de ce nouveau cas est utilisé comme une clé de recherche sur laquelle la récupération des anciens cas est fondée. Le but de la récupération est de déterminer les *"bons cas"* dans la base de cas, i.e. ceux qui permettent de bien résoudre le cas courant. Ce problème de récupération implique la résolution d'un problème d'appariement [Slade 1989] [Kolodner 1993]. Comme on ne peut toujours supposer que les caractéristiques des différents cas coïncident complètement, des appariements partiels doivent être calculés et les cas avec la meilleure qualité d'appariement sont récupérés.

Les recherches en raisonnement à partir de cas proposent des solutions pour ces problèmes de récupération, d'indexation et d'appariement. Cependant, ces approches sont pour la plupart informelles et hétérogènes, et très liées aux applications décrites, ce qui a pour effet:

- de rendre difficile d'évaluer les propriétés formelles de ces problèmes, et des solutions apportées, comme le bien fondé des résultats obtenus ou leur complétude,

- d'utiliser ces solutions dans d'autres applications,

- de comparer ces solutions d'un système à l'autre,

- de calculer la complexité des algorithmes utilisés

Nous présenterons donc dans un premier temps les formalisations proposées du raisonnement à partir de cas, nous comparerons chacune d'entre elles avec l'approche que nous proposons en déterminant les propriétés formelles que satisfait la relation de similarité dans les logiques de descriptions.

4.3 La formalisation du raisonnement à partir de cas

Cette formalisation résulte de la proposition de certains chercheurs [Koehler 1994][Koehler 1996] [Lieber and Napoli 1997] d'utiliser les logiques de description pour formaliser la tâche de recherche des anciens cas. En effet, les logiques de descriptions ont une représentation structurée des connaissances abstraites. Par contraste avec les formalismes précédemment évoqués, les logiques de descriptions possèdent une sémantique formelle. La sémantique de Tarsky, qui leur a été donnée par [Brachman and Levesque 1984], permet de les considérer comme des parties du langage des prédicats du premier ordre. De plus, les logiques de descriptions fournissent des algorithmes d'inférence de subsomption et de classification, suggérant ainsi leur utilisation dans le raisonnement à partir de cas.

Un cas, représenté dans une base de cas, est constitué de trois parties principales [Kolodner 1993]:

une situation initiale (pre) une description (ensemble de propriétés) spécifie les pré-conditions sur lesquelles la solution représentée dans le cas repose,

une situation finale (but) une description (ensemble de propriétés) spécifie les états finaux à atteindre,

une solution (S) la solution permet d'atteindre l'état final à partir de l'état initial, et de résoudre la spécification du problème $C \equiv < pre, but >$,

Le nouveau cas est donné sous la forme d'une spécification de problème $C_{new} \equiv < pre_{new}, but_{new} >$, pour laquelle une solution doit être trouvée dans la base de cas. Pour déterminer cette solution, une clé de recherche est dérivée de la spécification de problème, qui doit refléter les propriétés principales de celui-ci. Usuellement, cette recherche est effectuée dans l'ensemble des états mémorisés au lieu de l'être dans l'ensemble des solutions [Koehler 1994]. L'implication logique $S_{old} \Rightarrow S_{new}$ peut être généralisée à $C_{old} \models C_{new}$ [Koehler 1994] [Napoli et al. 1996], avec la signification intuitive: *chaque solution $S(C_{new})$ de C_{new} est une solution de C_{old}*; dans ce contexte, une solution peut alors être considérée comme un modèle. Cette relation est donc telle que C_{old} est suffisant pour dériver C_{new}.

Cependant, il apparaît que la recherche dans la base de cas à partir de la relation $\models$ est trop restrictive, puisqu'on ne fait alors que récupérer des solutions dans la base de cas. L'idée est donc d'introduire un *index* dérivé de C_{new} qui sera directement utilisé dans la recherche. Un index est donc calculé à partir du cas $C_{new} \equiv < pre_{new}, but_{new} >$, index qui est $w(C_{new}) \equiv < w(pre_{new}), w(but_{new}) >$, w étant le schéma d'encodage. Ce dernier réalise une opération d'abstraction: à partir d'une spécification d'un cas particulier, un index donne ses propriétés essentielles. Ce schéma d'encodage vérifie une propriété de monotonie [Koehler 1994][Koehler 1996]:

$$\text{Si } C_{old} \Rightarrow C_{new} \text{ alors } w(C_{old}) \Rightarrow w(C_{new})$$

Cette propriété assure que le schéma d'encodage soit telle qu'une solution entre cas reste une solution lorsqu'on passe aux indices. Bien entendu, la propriété inverse n'est en général pas vraie. Un cas récupéré de la base de cas grâce à l'index ne fournit pas alors *nécessairement* une solution du nouveau cas: ceci est dû au fait que le raisonnement à partir de cas est un raisonnement par approximation. La définition d'un schéma d'encodage particulier ne dépend plus alors que de trois facteurs [Koehler 1994]:

- le formalisme de représentation pour les cas,
- le formalisme de représentation pour les indices,
- le domaine d'application.

Pour le second facteur, il est fréquent que les indices soient représentés par un concept dans une logique de descriptions [Koehler 1994][Koehler 1996][Coupey and Salotti 1997][Napoli 1992]. Les concepts sont ordonnés par une relation de subsomption [18]. Des concepts primitifs étant données, les concepts définis

18. Informellement parlant, un concept A subsume un concept B si et seulement si A est plus général que B. Formellement parlant, A subsume B si et seulement si $B \models A$, i.e. l'extension de A est incluse dans celle de B.

sont formés à partir des concepts primitifs. Un classifieur calcule alors les subsomptions entre un nouveau concept défini et les concepts existants afin de l'insérer à la meilleure place possible dans la hiérarchie déterminée par la relation de subsomption: un problème est classifié dans la hiérarchie en déterminant les concepts les plus spécifiques dont il est instance, et les concepts les plus généraux qui en sont une instance.

Le second aspect du raisonnement à partir de cas, qui est celui qui nous intéresse ici, est le raisonnement par similarité. Les systèmes de raisonnement à partir de cas calculent la similarité des cas en comparant leur placement dans la hiérarchie induite par la relation d'abstraction ou en calculant leur distance sur une échelle qualitative ou quantitative [Kolodner 1993]. La fonction d'encodage ne permet pas une formalisation complète de la similarité, mais permet de déterminer quand un cas est plus spécifique qu'un autre:

Définition 52. Un cas C_1 est défini comme étant plus spécifique qu'un cas C_2 ssi $w(C_1) \models w(C_2)$.

Comme un cas comprend trois parties (situation initiale, situation finale et solution), la relation de déduction entre les cas peut être réduite à une relation de déduction entre états initiaux et finaux. Un cas est une instance d'un cas mémorisé ssi:

- Le nouvel état permet de déduire l'ancien état $pre_{new} \models pre_{old}$ (la solution S_{old} est applicable à la nouvelle situation),
- L'ancien but permet de déduire le nouveau but $but_{old} \models but_{new}$ (la solution résout le nouveau problème).

De ce fait, le test entre $w(C_{old}) \models w(C_{new})$ peut être réduit à un test sur les spécifications de problème, et deux types de recherche sont alors possibles, *forte* et *faible*:

$$\begin{array}{rll} \textbf{Forte}: & w(Pre_{new}) \models_L w(Pre_{old}) & \text{et } w(Goal_{old}) \models_L w(Goal_{new}) \\ \textbf{Faible}: & w(Pre_{new}) \models_L w(Pre_{old}) & \text{ou } w(Goal_{old}) \models_L w(Goal_{new}) \end{array}$$

Ayant ainsi défini la notion de similarité, il nous faut maintenant comparer ses propriétés formelles avec celles que nous avons établies pour le raisonnement par similarité au chapitre précédent. On désigne par $\models_S$ la relation de similarité qui, d'après la définition 52 que nous avons donnée est telle qu'un cas C_1 est similaire à (plus spécifique qu'un) cas C_2 ssi $w(C_1) \models w(C_2)$, ce que l'on note $C_1 \models_S C_2$. Cette relation est par définition de la monotonie du schéma d'encodage supra-classique (puisque si $C_1 \models C_2$, alors $w(C_1) \models w(C_2)$). Les autres propriétés formelles qui permettent d'avoir la cumulativité de la relation de déduction par similarité dépendent elles des propriétés formelles du schéma d'encodage. Il est cependant un cas où cette relation de similarité est cumulative, c'est le cas où les indices sont représentés par un indice dans une logique de descriptions. L'indexation d'un concept représentant le concept le plus abstrait auquel le cas courant est similaire, il nous importe donc d'étudier plus particulièrement ces logiques de descriptions et leurs propriétés formelles pour la similarité. Avant d'effectuer cette étude, mentionnons une autre formalisation du raisonnement par cas proposée par [Lieber and Napoli 1997][Lieber and Napoli 1998] qui définissent un cas par un couple $(pb, Sol(pb)$, où pb est un problème et $Sol(pb)$ sa solution. [Lieber and Napoli 1998] propose une approche permettant de déterminer un chemin de similarité entre un cas source et un cas cible, noté $sim(src, cbl)$, en supposant que l'ensemble des problèmes est structuré par un ensemble de relations de similarité entre problème R. Appelant espace de problèmes l'ensemble des problèmes muni des relations R, un chemin de similarité $sim(src, cbl)$ est un chemin dans l'espace de problème:

$$sim(src, tbl) = (pb_0 r_1 pb_1 \dots r_k pb_q)$$

avec $pb_0 = src$, $pb_q = cbl$, et l'ensemble des relations $(r_i)_{1 \leq i \leq q} \in R$. Ceci permet de déterminer un chemin entre les problèmes grâce aux relations de similarité. Il ne reste dès lors plus qu'à adapter les solutions des problèmes ainsi récupérés pour donner l'ensembles des opérations à appliquer pour résoudre le problème cible, ce qui est fait par le biais d'une fonction d'adaptation spécifique à chaque relation r_i. Ces relations de similarité peuvent être des relations de généralisations, comme par exemple la subsomption. L'approche de ces auteurs [Lieber and Napoli 1997][Lieber and Napoli 1998] ne satisfait cependant pas les bonnes propriétés que nous avons proposées pour la relation de similarité: seule la propriété supra-classique est satisfaite, par une propriété de la relation de généralisation similaire à la propriété de monotonie d'un schéma d'encodage.

4.4 Les logiques de description

Les logiques de descriptions sont utilisées pour représenter des hiérarchies de concepts ou terminologies. Elles ont pour origine le langage KL-ONE [Brachman and Schmolze 1985] et formalisent l'idée de définitions de concepts et de raisonnement sur ces concepts. Ce langage a été proposé par [Brachman and Schmolze 1985] dont les travaux s'inspiraient des réseaux sémantiques et des systèmes de frames. Les premiers réseaux sémantiques étaient des graphes orientés dont les noeuds représentaient des concepts, et les liens des relations entre ces concepts. Les concepts sont organisés en une hiérarchie selon des liens "est-un", et le principal mécanisme de raisonnement est un héritage utilisant cette hiérarchie.

[Minksy 1975] propose de regrouper toutes les informations relatives à un concept sous un frame, mais sans pour autant ajouter sémantique formelle ou propriétés de consistance. C'est ce que critique [Woods 1975], arguant également que les liens étaient indifféremment utilisés pour exprimer des définitions, des assertions et des propriétés nécessaires ou suffisantes, typiques ou incidentes, sans distinction claire entre individu ou classe. L'idée de [Brachman and Schmolze 1985] a alors été de concevoir un système de représentation des connaissances remédiant à ces défauts. Pour ce faire ont été introduits les notions de concepts et de rôle:

Les concepts Alors que dans les réseaux sémantiques, un concept était essentiellement vu comme une structure de données, dans KL-ONE, un effort a été fait pour voir dans un concept une structure abstraite, indépendante d'une structure de données particulière. Un concept décrit des conditions nécessaires et suffisantes pour l'appartenance d'un individu à l'extension de ce concept: tout individu qui satisfait la définition d'un concept C appartient à la classe dénotée par C, et réciproquement,

Les rôles un rôle décrit des liens de propriétés afférentes à un concept en décrivant les liens entre les concepts. Il correspond à la notion de liens entre des noeuds dans des réseaux sémantiques.

Les deux mécanismes de raisonnement fondamentaux dans KL-ONE sont la détermination de la relation de subsomption entre les concepts et la classification. Un concept C subsume un concept D si et seulement si tous les individus de la classe dénotée par D sont des individus de la classe dénotée par C. Le classifieur calcule la relation de subsomption entre un nouveau concept et les concepts existants, afin de l'intégrer à la meilleure place dans la hiérarchie. A partir de ce premier travail ont été développés un certain nombre de systèmes qui gardèrent les idées de KL-ONE, tout en cherchant à y intégrer une sémantique formelle. Citons KL-ONE [Brachman and Schmolze 1985], KANDOR [Patel-Schneider 1984], KRYPTON[Brachman et al. 1985], LOOM [MacGregor and Bates 1987], BACK [Nebel 1990], KRIS [Baader and Hollunder 1991] parmi les plus connus de ces systèmes.

4.4.1 Un exemple de langage terminologique

Afin de pouvoir étudier les propriétés de la relation de déduction par similarité proposée dans ce cadre, commençons par introduire un langage terminologique standard, à savoir $\mathcal{ALN}$ qui est un langage décrit dans [Donimi et al. 1991] dont le calcul de subsomption reste cependant polynômial. Ce langage est défini de manière inductive à partir d'un ensemble de rôles primitifs, un ensemble de concepts primitifs et du concept $\top$ (qui décrit l'ensemble de tous les individus) avec les règles de syntaxe suivantes:

$C,$	$D \rightarrow T$	le concept le plus général
	$\mid P$	concept primitif
	$\mid C \sqcap D$	conjonction de deux concepts
	$\mid \forall R : C$	restriction de valeur
	$\mid \exists_{\geq n} R$	restriction de cardinalité sur R (il existe au plus n valeurs pour le rôle R)
	$\mid \exists_{\leq n} nR$	restriction de cardinalité sur R (il existe au moins n valeurs pour le rôle R)
	$\mid \neg P$	négation d'un concept primitif

La sémantique formelle des concepts construits selon les règles ci-dessus évoquées est donnée de manière classique par une sémantique extensionnelle par l'interprétation $I = (\mathbf{D}, \|\ .\ \|)$. D (le domaine) désigne un ensemble non vide de concepts, et $\|\ .\ \|$ est une fonction d'interprétation telle que tout concept est projeté sur un sous ensemble de $\mathbf{D}$ et tout rôle sur un sous ensemble de $\mathbf{D} \times \mathbf{D}$. La dénotation d'un concept est donnée de la manière suivante:

$$\|\top\|^I = \mathbf{D}$$
$$\|C \sqcap D\|^I = \|C\|^I \cap \|D\|^I$$
$$\|\neg A\|^I = \mathbf{D} \setminus \|A\|^I$$
$$\|\forall R : C\|^I = \{a \in \mathbf{D} / \forall b,\ \text{si } (a,b) \in \|R\|^I,\ \text{alors } b \in \|C\|^I\}$$
$$\|\exists_{\geq n} nR\|^I = \{a \in \mathbf{D} / card(b/(a,b) \in \|R\|^I) \geq n\} \text{ où } card(.) \text{ est la fonction cardinal,}$$
$$\|\exists_{\leq n} nR\|^I = \{a \in \mathbf{D} / card(b/(a,b) \in \|R\|^I) \leq n\}$$

Une interprétation I est un modèle d'un concept C si $\|C\|^I$ est non vide. Se fondant sur cette sémantique, C est subsumé par D, noté $C \sqsubset D$ si et seulement si $\|C\|^I \subseteq \|D\|^I$ pour toute interprétation I. C est équivalent à D si et seulement si $C \sqsubset D$ et $D \sqsubset C$.

4.4.2 Les propriétés de la relation de similarité

L'idée de la relation de similarité est qu'il doit exister deux notions pour la définir: une relation d'ordre correspondant à la similarité (la subsomption), et une relation de congruence qui partitionnent les termes en ensembles équivalents pour la relation de similarité (l'égalité). Ces deux relations doivent être reliées par la relation suivante: si $\sqsubset$ désigne la relation de subsomption et $=$ la relation de congruence, si A et B sont deux termes, $A \sqsubset B \Leftrightarrow A \sqcap B = A$. Généralement, on présente les propriétés de la relation de similarité par celles de $=$, qui exprime des relations que la similarité doit satisfaire.

On considère l'ensemble des équations Eq suivantes, où A, B, C désignent des concepts, et R est un rôle:

- $\sqcap$

Associativité	$(A \sqcap B) \sqcap C$	$= A \sqcap (B \sqcap C)$
Commutativité	$A \sqcap B$	$= B \sqcap A$
Idempotence	$A \sqcap A$	$= A$
Élément Neutre	$\top \sqcap A$	$= A$

- $\forall$

Distributivité	$\forall R : (A \sqcap B)$	$= (\forall R : A) \sqcap (\forall R : B)$
	$\forall R : \top$	$= \top$

Les deux premières règles indiquent que la similarité ne dépend pas de la manière d'écrire les règles, l'idempotence qu'un concept est similaire à lui-même, l'élément neutre que vrai est similaire à tout et la distributivité que les concepts liés à une conjonction de concepts sont la conjonction des concepts qui leurs sont liés.

Les algèbres qui satisfont ces équations sont appelées algèbres de concepts, que l'on note C-algèbres. Étant donné un ensemble de variables $\{x_1, \ldots, x_n\}$, on note $\mathcal{AL}[X]$ l'ensemble des termes sur X et sur la signature du langage. Le quotient $\mathcal{AL}[X]_{/Eq}$ est une algèbre libre dans laquelle les seules égalités valides sont celles qui le sont dans les algèbres C. $\mathcal{AL}[\emptyset]$ est l'algèbre initiale, et pour toute algèbre $\mathcal{A}$, il existe un unique homomorphisme de $\mathcal{AL}[\emptyset]$ dans $\mathcal{A}$ [Dionne et al. 1993b][Dionne et al. 1993a].

Revenant à la base de connaissance $\Delta = \{x_1 \equiv C_1, \ldots, x_n \equiv C_n\}$, où $C_i \in \mathcal{AL}[X]$, on considère la plus petite congruence $\equiv_\Delta$ sur $\mathcal{AL}[X]_{/Eq}$ satisfaisant Δ. Cette algèbre quotient, notée $D_\Delta[X]$ est appelée une algèbre descriptive, puisqu'elle fournit la description des classes de termes "respectant" la base de connaissances.

Dans l'implémentation des systèmes terminologiques, c'est la subsomption structurale qui est testée, équivalente à la subsomption descriptive [Dionne et al. 1993b]. Un concept en subsume un autre si et seulement si:

1. tous les composants primitifs le définissant sont inclus dans ceux de l'autre,
2. ses rôles constituent un sous ensemble des rôles de l'autre,
3. les restrictions de valeur sur ses rôles subsume les restrictions de valeur sur les rôles de l'autre concept.

Alors la relation de similarité est cumulative lorsque les index sont représentés dans une logique de description: nous avons vu qu'elle est supra-classique quel que soit le schéma d'encodage choisit, et les autres propriétés se dérivent du fait que la subsomption est dérivée de la relation classique de déduction (inclusion entre des classes d'algèbres).

Ainsi, lorsque les index sont représentés par des concepts dans une logique de description, la relation de similarité satisfait les bonnes propriétés que nous avons requis au chapitre précédent pour la similarité. Cependant, deux points sont à souligner par rapport à l'approche que nous proposons:

- La similarité est définie par l'ensemble des parties communes à la description d'un cas courant et d'un cas mémorisé. Aucun mécanisme ne permet de prendre en compte les parties distinguant cas courant et cas mémorisé.
- Les préférences ne sont pas introduites et ne permettent donc pas de contrôler la similarité.

Considérons en effet l'exemple de la photo-interprétation que nous avons introduit au chapitre 2:

$$\begin{aligned}
Aeroport(T, l_1) \equiv \quad & Piste(X, l_1) \sqcap Taxiway(Y, l_1) \sqcap TourControle(Z, l_1) \\
AeroportCivil(X, l_1) \equiv \quad & Aeroport(T, l_1) \sqcap Aerogare(W, l_1) \\
AeroportMilitaire(X, l_1) \equiv \quad & Aeroport(T, l_1) \sqcap StationnementNonAligne(W, l_1) \\
AeroportMilitaireChasse(Z, l_1) \equiv \quad & AeroportMilitaire(X, l_1) \sqcap AireAlerte(Y, l_1) \\
AeroportMilitaireBombardiers(T, l_1) \equiv \quad & AeroportMilitaire(X, l_1) \sqcap Piste(Y, l_1) \sqcap \\
& LongueurPiste(Y, L > 1000) \sqcap Hangars(Z, l_1)
\end{aligned}$$

Le cas courant étant donné par la dscription : $: Piste(A, l_1), : : Taxiway(B, l_1)$. Il est donc possible de déduire que l'objet qui se trouve en l_1 est à la fois un aéroport civil, un aéroport militaire, un aéroport militaire pour avions de chasse et un aéroport militaire pour bombardiers, ce qui est impossible, un aéroport ne pouvant être les quatre à la fois. L'absence de préférences ne permet pas de choisir entre ces ensembles de conclusions. Dans notre cas, l'ajout de la préférence

$$\begin{gathered}
P_1 \equiv (Piste(W, l_1), \perp, AeroportCivil(X, l_1)) \\
\succ \\
(Piste(Y, l_1), \perp, AeroportMilitaire(Z, l_1))
\end{gathered}$$

permet de décider que l'aéroport est civil.

Ces deux éléments sont partiellement introduits par l'ajout de la possibilité de raisonner par défaut dans les logiques terminologiques. Les défauts permettent en effet de contrôler les similarités effectuées par une relation de préférence purement opérationnelle, la consistance syntaxique. Cette dernière permet d'exprimer une préférence entre:

- un raisonnement dans lequel des faits ont été négligés et tels que leur négation peut simultanément être prouvée (violation de la contrainte induite par le défaut),

- et un raisonnement dans lequel des faits ont été négligés sans que cependant leur négation puisse simultanément être prouvée (pas de violation de la contrainte induite par le défaut).

Du point de vue des logiques terminologiques, cela exige d'envisager le cas où la définition partielle d'un concept permet de reconnaître ses instances. Dès lors, deux types d'approches ont été proposées:

- la première de ces deux approches consiste à prendre en compte les défauts en dehors de la syntaxe de la logique, i.e. en dehors de la définition des concepts: ce sont les approches dites contingentes.
- la deuxième de ces deux approches intègre l'incomplétude dans la définition des concepts: ce sont les approches dites définitionnelles.

Approches contingentes

Plusieurs propositions ont été faites pour étendre les systèmes terminologiques avec une forme limitée de raisonnement par défaut [Heinsohn 1991][Quantz and Royer 1992][Padgham and Zhang 1993]. [Heinsohn 1991] propose une approche probabiliste pour prendre en compte l'incertitude en introduisant une implication probabiliste séparée du langage terminologique. [Quantz and Royer 1992] et [Padgham and Zhang 1993] introduisent les défauts comme des règles contingentes. [Quantz and Royer 1992] fonde sa sémantique sur la théorie des mondes préférentiels. La préférence est définie de trois manières distinctes:

- le premier critère (*exception preference*) minimise les exceptions (les modèles préférés sont ceux qui ont le moins d'exceptions),
- le second critère (*fulfilment preference*) minimise les reconnaissances (les modèles préférés sont ceux où le plus de défauts se sont déclenchés. On retrouve là l'idée de la définition semi-inductive des extensions dans les logiques des défauts [Reiter 1980],
- le critère (*anteriority preference*) consiste à appliquer un ordre entre les défauts qui dépend de la hiérarchie de subsomption induite par les connaissances strictes (les modèles préférés sont ceux qui contiennent le plus de connaissances spécifiques)

Cette approche n'est pas comparable avec celle que nous proposons dans la mesure où les préférences (il faudrait ici parler plus rigoureusement de priorités) sont définies complètement en dehors de la logique, tout comme dans la logique des défauts de [Reiter 1980]. La relation de déduction par similarité sachant des préférences souffre alors des mêmes problèmes que nous avons signalés lors de la présentation de la logique des défauts de [Reiter 1980].

[Padgham and Zhang 1993] proposent eux une approche unifiant l'héritage non monotone dans les réseaux sémantiques et dans les logiques de description avec la logique des défauts de [Reiter 1980]. Pour ce faire, ils étendent le langage des logiques de description afin d'attacher des règles par défaut à un concept. Chaque concept est alors séparé en deux parties:

- un noyau représentant un ensemble de propriétés strictes nécessaires et suffisantes pour l'appartenance d'une instance à un concept,
- un défaut représentant l'ensemble de toutes les propriétés strictes et par défaut.

Par exemple, adaptant l'approche qu'ils proposent au cas de la photo-interprétation,

$$AeroportCivil \equiv Piste \sqcap TourControle/Aerogare$$

décrit l'ensemble des instances qui sont exactement constituées d'une piste et d'une tour de contrôle, et qui ont comme propriété contingente d'être aussi constituée d'une aérogare. A chaque concept C correspond alors deux noeuds dans la taxinomie, le noeud C_{noyau} et le noeud C_{defaut}. Ainsi:

$$AeroportCivil_{noyau} \equiv Piste \sqcap TourControle$$
$$AeroportCivil_{defaut} \equiv Piste \sqcap TourControle \sqcap Aerogare$$

Les noeuds noyau et défaut sont classés dans la taxinomie à l'aide de l'algorithme de classification standard. [Padgham and Zhang 1993] mettent principalement en avant l'intérêt des défauts pour la classification d'instances. Le processus de fonctionnement d'un tel système est le suivant:

1. A partir de la description B d'un individu A, on détermine un premier ensemble de concepts dont A est reconnu instance
2. On infère pour A les propriétés par défaut attachées à ces concepts, ce qui donne une description plus précise de A,
3. On recommence l'opération. Le cycle se termine lorsqu'un point fixe est atteint.

Pour le raisonnement par similarité en fonction de préférences, cette approche souffre des mêmes inconvénients que celle de [Reiter 1980] et de [Quantz and Royer 1992]. Les mêmes propriétés formelles de la relation de déduction n'y sont pas satisfaites.

Approches définitionnelles

La différence est que dans cette approche, les défauts font partie de la définition des concepts, ce qui permet de donner une définition complète (i.e. nécessaire et suffisante) à des concepts qui sinon ne seraient que partiellement définis. La conséquence en est qu'avec une définition complète, le classfieur peut reconnaître automatiquement les instances d'un concept.

Pour prendre en compte cette forme définitionnelle, il faut pouvoir intégrer le point de vue classique du raisonnement par défaut que "A est par défaut B" signifie que *en général* les A sont des B, mais qu'il peut exister des A qui sont exceptionnels et qui ne sont pas des B. Le problème avec le classifieur est qu'il ne peut prendre en compte de telles inférences probables pour classer un concept dans la terminologie. De ce fait, les auteurs [Coupey and Salotti 1997][Coupey et al. 1998] séparent ce qui est vrai dans toutes les interprétations sur les connaissances par défaut et les exceptions et leur assigne deux niveaux: la subsomption standard pour le premier, l'héritage pour le second. Pour ce faire, les auteurs rajoutent deux connecteurs dans la logique pour représenter les défauts et les exceptions:

$C,$	$D \rightarrow T$	le concept le plus général
	$\mid P$	concept primitif
	$\mid C \sqcap D$	conjonction de deux concepts
	$\mid \forall R : C$	restriction de valeur
	$\mid \exists_{\geq n} R$	restriction de cardinalité sur R (il existe au plus n valeurs pour le rôle R)
	$\mid \exists_{\leq n} nR$	restriction de cardinalité sur R (il existe au moins n valeurs pour le rôle R)
	$\mid \neg P$	négation d'un concept
	$\mid \delta C$	concept par défaut
	$\mid C^{\epsilon}$	exception au concept C

Ces nouveaux connecteurs sont appréhendés par les équations suivantes donnant leurs propriétés par rapport à la similarité:

- $\epsilon : (\delta A)^{\epsilon} = A^{\epsilon}$
- δ:

$$\begin{array}{lll} \delta 1 & \delta(A \sqcap B) & = (\delta A) \sqcap (\delta B) \\ \delta 2 & A \sqcap \delta A & = A \\ \delta 3 & A^{\epsilon} \sqcap \delta A & = A^{\epsilon} \\ \delta 4 & \delta\delta A & = A \end{array}$$

Les équations ($\delta 2$) et ($\delta 3$) sont les deux plus importantes. Elles expriment une relation de subsomption entre A et δA (A est subsumé par δA) et entre A^{ϵ} et δA (A^{ϵ} est subsumé par δA). L'équation ϵ présuppose qu'une exception n'a de sens que si c'est l'exception d'un concept par défaut. L'équation ($\delta 4$) permet aux chaînes redondantes d'être effacées, et l'équation ($\delta 1$) est une propriété de distributivité classique. [Coupey and Fouqueré 1997] étendent alors la définition de la subsomption descriptive pour prendre en compte ces nouveaux connecteurs, puis déterminent la sémantique structurale du langage en incluant ces deux connecteurs, et prouvent son équivalence avec la subsomption descriptive du langage ainsi étendu.

Avec cette extension, une définition de concept C (incluant des connaissances par défaut) représente toujours une condition nécessaire et suffisante pour reconnaître une instance: un objet est une instance de C si et seulement si il satisfait la définition conceptuelle stricte de C et est au moins exceptionnel vis à vis des concepts par défaut de C. Comme nous l'avons dit, les auteurs [Coupey and Fouqueré 1997] séparent alors ce qui est vrai dans toutes les interprétations sur les défauts et les exceptions et ce qui est seulement probable (i.e. pas vrai dans toutes les interprétations) en assignant au premier la subsomption et la classification et l'héritage au second. Le point de vue définitionnel des défauts induit des subsomptions sûres: un concept par défaut δC est un ensemble d'instances qui inclue les instances de C et les instances exceptionnelles par rapport à C $(C^{\epsilon}, C^{\epsilon^{\epsilon}}, \delta(C^{\epsilon}), \delta(C^{\epsilon^{\epsilon}}))$ ... Se fondant alors sur ces subsomptions sûres, le processus de classification usuel permet alors d'insérer ces concepts dans la terminologie. Le niveau héritage correspond lui à complèter la subsomption en interprétant les défauts de manière classique, i.e. une exception inhibe un défaut. Par exemple, si la subsomption a permis de déduire que A est subsumé par δC, mais que A n'est pas subsumé par C^{ϵ}, alors l'héritage infère que A est en héritage *probable* avec C. Si un concept B est subsumé par δC ou δC^{ϵ} ou $C^{\epsilon^{\epsilon}}$, alors l'héritage infère que C est en contre-exception avec B, donc B est en héritage *"probable"* avec C. Plus formellement, les auteurs [Coupey et al. 1998] déterminent différents types d'héritage, qui permettent de distinguer différents types de proximité pour les cas:

héritage sûr un concept C est en relation d'héritage sûr avec un concept D si et seulement si C est subsumé par D,

héritage probable un concept C est en relation d'héritage probable avec un concept D si et seulement si C n'est pas subsumé, ni par D, ni par $\neg D$, et C est subsumé par δD et ni D, ni aucun concept en héritage avec D n'est excepté par C,

héritage typique un concept C est en relation d'héritage typique avec D quand les concepts hérités par D sont aussi hérités par C,

héritage exceptionnel un concept C est en relation d'héritage exceptionnel avec D quand il existe au moins un concept herité par D qui n'est pas hérité par C.

Les auteurs [Coupey et al. 1998] utilisent alors ces différents formes d'héritage pour donner une relation de proximité plus ou moins grande (d'un point de vue symbolique) entre le cas courant et les cas mémorisés. Par rapport à l'approche que nous proposons, cette logique a l'inconvénient de ne pas prendre en compte la notion de préférence autrement que par des liens d'héritage.

4.5 Conclusion

Les approches que nous avons présenté sur le raisonnement à partir de cas se divisent en deux parties:

- celles qui ne prennent en compte dans la similarité **que les parties communes** des descriptions respectives du problème courant et des problèmes passés. C'est le cas des formalisations du raisonnement à partir de cas utilisant des logiques de description dans le cas où les concepts sont définis par des conditions nécessaires et suffisantes,
- celles qui prennent en compte, outre l'identité entre les parties communes que partagent cas courant et cas passés, une forme limitée de dissimilarité:
  - que ce soit en admettant des formes contingentes de dissimilarité, liées aux défauts, mais qui souffrent alors des lacunes en terme de propriétés formelles que nous avons signalés lorsque nous avons étudié cette théorie,
  - que ce soit en admettant des formes définitionnelles de dissimilarité, mais par rapport à l'approche que nous proposons, elles souffrent du fait que la dissimilarité, contrairement à la similarité, n'est pas inférée, mais explicitement écrite dans la syntaxe de la logique et elle n'admet pas non plus une spécification déclarative simple des préférences.

De plus, du point de vue des préférences, aucune de ces logiques n'offre un cadre formel suffisamment souple pour, par exemple, appréhender des concepts tels que l'aversion ou l'inclinaison au risque. En effet, ces logiques s'appuient essentiellement sur une formalisme des défauts. Dans un tel cadre conceptuel et syntaxique, la prise en compte des préférences dépend de façon masquée des défauts syntaxiquement fixés. A l'inverse, l'expression de la connaissance que nous souhaitons obtenir doit permettre de formuler explicitement des préférences, indépendamment des connaissances purement déductives qui, elles, ont vocation à rester stable. Les éléments négligés (les défauts) dépendent alors du raisonnement, c'est à dire à la fois du cas concret examiné, des connaissances mémorisées et des préférences exprimées. L'aversion ou l'inclinaison au risque sont ainsi formulées de façon déclarative, dans les préférences, et indépendamment des clauses déductives (connaissances mémorisées).

Troisième partie

Inférences par similarité

Nous proposons dans ce document deux modélisations du raisonnement par similarité en fonction de préférences. Dans chacune de ces deux modélisations, la similarité de deux cas s'appréhende par le partage de caractéristiques communes et par le fait de considérer comme **négligeables** les parties qui les différencient. Les caractéristiques communes sont le fait, dans chacun des deux cas, d'une subsomption entre description du cas courant à résoudre et une partie de la description des cas mémorisés.

Les préférences sont prises en compte par un objet syntaxique appelé préférence s'écrivant sous la forme d'un sextuplet:

(Pré-requis valide 1, Pré-requis négligé 1, Fait 1)
Préféré
(Pré-requis valide 2, Pré-requis négligé 2, Fait 2)

sextuplet qui signifie intuitivement, "le fait 1 résultant de la connaissance de Pré-requis 1 et du fait de négliger Pré-requis négligé 1 est préféré au fait 2 résultant de la connaissance de Pré-requis 2 et du fait de négliger Pré-requis négligé 2". Ces deux éléments sont ensuite pris en compte par une sémantique formelle permettant alors de construire un raisonnement par similarité en fonction de ces préférences. Cette relation se décline en deux versions:

satisfiabilité forte par similarité : le décideur ne connaît alors aucun négligeable potentiel, et raisonne par similarité en fonction du cas courant connus avec certitude,

satisfiabilité faible par similarité : le décideur fait des hypothèses de négligeabilité qui sont exclusives les une des autres. Il raisonne alors par similarité en fonction de ces négligeables et des préférences.

Pour chacun de ces deux modes de raisonnement, une implantation sous forme d'une sémantique de point fixe est proposée.

5

Syntaxe et sémantique des préférences

5.1 Définitions - Notations

5.1.1 Syntaxe

Les définitions rappelées sont extraites de [Lalement 1990] et les notations de l'auteur sont conservées. Soit Σ une signature finie, $\Sigma = \Sigma_f \cup \Sigma_R$, Σ_f étant la signature fonctionnelle, Σ_R la signature relationnelle. Soit X un ensemble dénombrable de variables, on forme l'ensemble $T_{\Sigma_f}[X]$ des termes finis sur la signature Σ_f et l'ensemble $At_\Sigma[X]$ des formules atomiques par les $Rt_1 \dots t_n$ avec $R \in \Sigma_R$ d'arité n.

Rappelons la définition de l'ensemble des formules des prédicats du premier ordre [Lalement 1990]:

Définition 53. On définit par $L_\Sigma[X]$ l'ensemble des formules des prédicats du premier ordre sur la signature Σ comme étant constitué par la plus petite partie E du monoïde libre $(\Sigma \cup X \cup \{\vee, \wedge, \Rightarrow, \bot, \forall\})$ [19] telle que:

i) $At_\Sigma[X] \in E$

ii) $\bot \in E$, et si $\varphi, \psi \in E$, alors $\varphi \vee \psi \in E$, $\varphi \wedge \psi \in E$ et $\varphi \Rightarrow \psi \in E$

iii) Si $x \in X$, $\varphi \in E$, alors $\forall x \varphi \in E$

Désignons par $var(\varphi)$ l'ensemble des variables libres de la formule φ. On note $L[x_1, \dots, x_n]$ l'ensemble des formules $\varphi \in L_\Sigma[X]$ telle que $var(\varphi) \subseteq x_1, \dots, x_n$ avec $(x_1, \dots, x_n) \in X$. On note L_Σ l'ensemble des formules fermées, i.e. l'ensemble des formules φ telles que $var(\varphi) = \emptyset$. L'égalité syntaxique de deux termes sera noté $\equiv$.
Si φ est une formule quelconque de $L_\Sigma[x_1, \dots, x_n]$, on note $\forall \vec{x} \varphi$ la formule $\forall x_1, \dots, x_n \varphi \in L_\Sigma[X]$

Définition 54. Un séquent est une paire $\Gamma : \Delta$ de listes de formules $\Gamma = < \varphi_1, \dots, \varphi_n >$ et $\Delta =< \psi_1, \dots, \psi_m >$ [Gallier 1986]. On appellera Γ la prémisse du séquent et Δ sa conclusion. On notera désormais $S_\Sigma[X]$ l'ensemble des séquents, et si $S \equiv \Gamma : \Delta$ est un séquent, S^- désigne la conjonction des éléments de Γ et S^+ désigne la disjonction des éléments de Δ.

5.1.2 Sémantique

On rappelle la définition suivante ainsi que les notations utilisées:

Définition 55. Une Σ-algèbre $\mathcal{A}$ est la donnée d'un ensemble A non vide et pour chaque symbole de

- constante $c \in \Sigma$, d'un élément $c^{\mathcal{A}}$ de A.
- fonction n-aire $f \in \Sigma_f$, d'une application $f^{\mathcal{A}} : A^n \to A$,

19. Les connecteurs gardent bien entendu leur sens usuel.

– relation n-aire $R \in \Sigma_R$, d'une relation $R^{\mathcal{A}} : A^n \to \{0,1\}$.

On note A^X l'ensemble des applications de X dans A, et ξ un élément de A^X. On note par $\hat{\xi}$ l'unique Σ-morphisme prolongeant ξ de X à $T_\Sigma[X]$. Un terme $t \in T_\Sigma[X]$ est interprété par une application $t^{\mathcal{A}} : A^X \mapsto A$, définie par induction sur les termes:

1. $x^{\mathcal{A}}(\xi) = \xi(x)$ si $x \in X$,

2. $c^{\mathcal{A}}(\xi) = c^{\mathcal{A}}$ si c est un symbole de Σ d'arité 0,

3. $(ft_1 \ldots t_n)^{\mathcal{A}}(\xi) = f^{\mathcal{A}}(t_1^{\mathcal{A}}(\xi), \ldots, t_n^{\mathcal{A}}(\xi))$

Une formule est interprétée par une application $\varphi^{\mathcal{A}} : A^X \mapsto \{0,1\}$ définie par les règles d'induction usuelle sur les formules. Et pour $\varphi \in L_\Sigma$, on note $\mathbf{Mod}\{\varphi\}$ la classe des algèbres qui sont modèles de φ.

Enfin, il nous reste à interpréter les séquents. Intuitivement, une algèbre valide un séquent si et seulement si elle valide la formule [Gallier 1986]:

$$(\varphi_1 \wedge \ldots \wedge \varphi_n) \Rightarrow (\psi_1 \vee \ldots \vee \psi_m)$$

Définition 56. Une Σ-algèbre valide le séquent $\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m$, où $(\varphi_1, \ldots, \varphi_n, \psi_1, \ldots, \psi_m)$ sont des formules de $L_\Sigma[X]$ si et seulement si:

$$\mathcal{A} \models_L (\varphi_1 \wedge \ldots \wedge \varphi_n) \Rightarrow (\psi_1 \vee \ldots \vee \psi_m)$$

5.2 La notion de monde épistémique

Intuitivement, elle correspond à la notion que les connaissances sont constituées de l'ensemble des formules que l'individu connaît. Sémantiquement, elles se modélisent donc par l'ensemble des algèbres qui valident ces formules. Cependant, ces deux notions, syntaxiques et sémantiques, ne se correspondent pas exactement: il y a plus de modèles que de formules. Si l'individu connaît n formules, il y a au maximum 2^n ensemble de formules et donc 2^{2^n} algèbres. De ce fait, un même ensemble d'algèbres représente les mêmes connaissances [Levesque 1990]: la sémantique ne doit donc pas distinguer entre des algèbres qui valident les mêmes formules. On appellera désormais *monde épistémique* un tel ensemble d'algèbres.

Intuitivement, un individu n'a qu'une connaissance approximative du monde qui l'entoure. Les mondes épistémiques représentent une description partielle du monde vu par l'individu. Chaque monde modèle d'une théorie (= les connaissances de l'individu) représente une part de l'incertitude de celui-ci sur le monde qui l'entoure. Chaque ensemble de formules validées par une algèbre du monde épistémique modèle de la théorie qui ne l'est pas par une autre représente des connaissances non détenues par l'individu: il est donc incertain sur leur validité. L'ensemble des mondes modèles d'une théorie est donc une représentation à la fois de ce que l'individu connaît, et de ce sur quoi il n'a pas d'informations.

La définition d'un monde épistémique nécessite de définir la satisfaction d'une formule par une classe d'algèbre. Reprenant en cela les définitions de [Voorbraak 1993], on note $\models_{h,L} \subseteq \mathcal{P}(\mathbf{Mod}_L) \times L_\Sigma$ la relation de satisfaction définie par:

$$\mathcal{M} \models_{h,L} \varphi \Leftrightarrow \forall M \in \mathcal{M}, M \models_L \varphi$$

où $\models_L$ est la relation de satisfaction classique.

D'autre part, pour chaque ensemble $\mathcal{M}$ de $\mathbf{Mod}_L$, et pour chaque théorie Γ de L_Σ, on définit [Voorbraak 1993]:

– $//\mathcal{M}//_L = \{\varphi \in L_\Sigma / \mathcal{M} \models_{h,L} \varphi\}$, l'ensemble des formules qui sont validées par la classe d'algèbres $\mathcal{M}$,

– $\backslash\backslash\Gamma\backslash\backslash_L = \{M \in \mathbf{Mod}_L / M \models_L \Gamma\}$, l'ensemble des algèbres qui sont modèles de Γ.

On est dès lors amené à ne s'intéresser qu'à des classes d'algèbres particulières, à savoir celles qui valident les mêmes formules. L'adéquation entre ensembles de formules et modèles se fait donc en définissant un quotient de l'ensemble des classes d'algèbres par une relation d'équivalence entre des ensembles d'algèbres qui valident les mêmes formules. Pour ce faire, on définit à l'instar de [Levesque 1990] l'ensemble:

$$\mathcal{M}^+ = \{M \in \mathbf{Mod}_L / \forall \varphi \in //\mathcal{M}//_L, M \models_L \varphi\}.$$

[Levesque 1990] montre que $\mathcal{M}^+$ est la borne supérieure au sens de l'inclusion de l'ensemble des classes de modèles équivalents à $\mathcal{M}$. Intuitivement, $\mathcal{M}^+$ a la même propriété exigée par [Levesque 1990], à savoir celle de ne connaître que certaines connaissances. On ne peut alors discriminer parmi les modèles qui valident les mêmes ensembles de formules.

D'autre part, on définit l'ensemble des mondes possibles de L (comme dans les logiques épistémiques) par $WD_L = \{\backslash\backslash\Gamma\backslash\backslash_L / \Gamma \subseteq L_\Sigma\}$.

Proposition 57. *On peut montrer [Voorbraak 1993] que les notions introduites satisfont les propriétés suivantes:*

- $\backslash\backslash(//\mathcal{M}//)\backslash\backslash = \mathcal{M}^+$
- $\forall \mathcal{M} \in WD_L, \mathcal{M} = \mathcal{M}^+$

5.3 Définition de la syntaxe et de la sémantique des préférences

Nous définissons formellement dans cette section les éléments de syntaxe permettant d'exprimer des préférences, puis nous donnerons leur sémantique dans la section suivante.

5.3.1 Syntaxe des préférences

Avant de donner la syntaxe des préférences, commençons par introduire la notion de relation binaire asymétrique:

Définition 58. On introduit un symbole $\succ, \succ \subseteq \mathcal{P}(\mathcal{P}(L_\Sigma[X]) \times \mathcal{P}(L_\Sigma[X]))$ représentant une relation asymétrique, i.e. telle que si ϕ_1, ϕ_2, sont des séquences de formules de L, telles que $(\phi_1 \neq \phi_2)$ aux propriétés algébriques des connecteurs près, alors:

$$(\phi_1, \phi_2) \in \succ \Rightarrow (\phi_2, \phi_1) \notin \succ$$

Définition 59. On définit par $P_\Sigma[X]$ l'ensemble des formules purement préférentielles du premier ordre sur la signature Σ comme étant constitué par la plus petite partie P du monoïde libre $(\Sigma \cup X \cup \{\vee, \wedge, \Rightarrow, \perp, \forall, \succ\})$ telle que:

-
$$\begin{gathered}\forall \varphi_1 \in L_\Sigma[X], \varphi_1 \neq \perp, \forall \varphi_2 \in L_\Sigma[X], \forall \psi_1 \in L_\Sigma[X],\\ \forall \psi_2 \in L_\Sigma[X], \forall \phi_1 \in L_\Sigma[X], \phi_1 \neq \perp, \forall \phi_2 \in L_\Sigma[X],\\ (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2) \in P,\end{gathered}$$

-
$$\begin{gathered}\forall \varphi_2 \in L_\Sigma[X], \forall \psi_1 \in L_\Sigma[X], \psi_1 \neq \perp, \forall \psi_2 \in L_\Sigma[X], \forall \phi_2 \in L_\Sigma[X], \phi_2 \neq \perp\\ (\perp, \varphi_2, \perp) \succ (\psi_1, \psi_2, \phi_2) \in P\end{gathered}$$

Cette définition que l'on peut réécrire comme:

(Pré-requis valide 1, Pré-requis négligé 1, Fait 1)
préféré
(Pré-requis valide 2, Pré-requis négligé 2, Fait 2)

s'interprète intuitivement comme: "le fait 1 résultant de la connaissance de Pré-requis 1 et du fait de négliger Pré-requis négligé 1 est préféré au fait 2 résultant de la connaissance de Pré-requis 2 et du fait de négliger Pré-requis négligé 2". Ainsi, par exemple:

$$(Piste(l_1), Aerogare(l_1), AeroportCivil(l_1)$$
$$\succ$$
$$(Piste(l_1), StationnementNonAligne(l_1), AeroportMilitire(l_1))$$

correspond à une préférence disant que, dans un problème de décision, si à partir d'un cas mémorisé, on peut déduire un aéroport civil à partir d'au moins une piste et en négligeant une aérogare, et si à partir d'un autre cas mémorisé, on peut déduire *au même endroit*, un aéroport militaire à partir de la reconnaissance d'au moins une piste et en négligeant des stationnements non alignés, alors on préfère le premier cas au second, i.e. on préfère croire qu'il s'agit d'un aéroport civil plutôt que d'un aéroport militaire.

On peut donc définir:

Définition 60. On définit par $LP_\Sigma[X]$ l'ensemble des formules préférentielles du premier ordre par $L_\Sigma[X] \cup P_\Sigma[X]$. On définit par $SP_\Sigma[X]$ l'ensemble des séquents préférentiels du premier ordre par $S_\Sigma[X] \cup P_\Sigma[X]$.

5.3.2 Sémantique des préférences

Reste maintenant à déterminer la sémantique des objets syntaxiques que nous venons de construire. La première hypothèse est d'utiliser, pour représenter les préférences que nous venons d'introduire, la notion épistémique de préférence entre des mondes introduite par [Voorbraak 1993] pour représenter des connaissances par défaut.

Exemple 61. (Extrait de [Voorbraak 1993])
Considérons un individu avec les croyances suivantes:

1. Typiquement, il ne pleut pas en Californie
2. Il pleut maintenant en Californie.

La première croyance l'incite à considérer comme préféré tous les mondes épistémiques dans lesquels il ne pleut pas en Californie, induisant une notion de préférence entre les mondes. Cette préférence est supplantée par celle induite par la deuxième croyance qui affirme bien qu'il pleut maintenant en Californie. Par ailleurs, comme l'individu n'a aucune croyance sur les conditions atmosphériques au Kansas, il considère comme préféré les mondes épistémiques qui ne se réfèrent pas à la question de pluie au Kansas.

◀

Cette notion épistémique est une généralisation du formalisme proposé par [Shoham 1988] qui a permis d'unifier les différentes logiques non monotones jusqu'alors proposées en définissant une logique non monotone à partir d'une logique classique (i.e. monotone) et d'une relation de pré-ordre partiel strict sur l'ensemble des modèles de cette logique.

Considérant cette approche comme trop restrictive, puisqu'un individu ne possède qu'une connaissance imparfaite du monde qui l'entoure, [Voorbraak 1993] propose une généralisation de cette approche en considérant une relation seulement asymétrique sur des ensembles de modèles d'une logique standard. Cette approche lui permet alors d'exprimer dans ce formalisme la logique des défauts de [Reiter 1980], alors que [Shoham 1988] n'avait pu exprimer que des variantes de la définition originale de Reiter dans son formalisme.

Malheureusement, l'idée n'est pas ici directement utilisable puisque nous devons comparer des mondes pouvant résulter d'hypothèses de négligeabilités différentes. Nous devons donc établir une relation binaire entre des couples (mondes, négligeabilités) pour représenter les préférences. Avant de donner ces relations binaires, nous devons d'abord établir la définition de la relation de satisfaction d'une préférence par un couple de monde:

Définition 62. *(Validation partielle d'une préférence)*
Soit $\mathcal{A}, \mathcal{B}$ [20] un couple d'algèbres, $\sigma : X \mapsto T_\Sigma$ une substitution. Une préférence $P \in P_\Sigma[X]$ est interprétée partiellement par une application

$$p^{\mathcal{A},\mathcal{B},\sigma} : A^X \times B^X \times (X \mapsto T_\Sigma) \mapsto \{0,1\}$$

dont la définition par induction est donnée par le tableau suivant:

- $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)^{\mathcal{A},\mathcal{B},\sigma}(\xi, \xi', \sigma) = inf((\sigma\varphi_1)^{\mathcal{A}}(\xi), (\sigma\psi_1)^{\mathcal{B}}(\xi'), (\sigma\phi_1)^{\mathcal{A}}(\xi), (\sigma\phi_2)^{\mathcal{B}}(\xi'))$

Comme dans le cas du premier ordre, on note, si $\mathcal{A}, \mathcal{B}$ sont deux algèbres, σ une substitution $\mathcal{A}, \mathcal{B}, \sigma \models_{Pref} (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ lorsque pour tout couple de valuation $\xi \in A^X, \xi' \in B^X, (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)^{\mathcal{A},\mathcal{B},\sigma}(\xi, \xi', \sigma) = 1$.

Cette définition va nous permettre d'interpréter une préférence par un ensemble de relations binaires asymétriques sur des ensembles de couples (mondes, négligeables) de la manière suivante:

Définition 63. *(Interprétation d'une préférence)*
Une préférence $P \equiv (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ de $P_\Sigma[X]$ s'interprète par un ensemble de relations binaires asymétriques sur des couples (mondes, négligeables), relations binaires notées $Pref(P) \in \mathcal{P}(WD_L \times L_\Sigma[X])$, chacune d'entre elle étant définie pour $\sigma : X \mapsto T_\Sigma$ et telle que:

- $((\mathcal{M}, D), (\mathcal{N}, D')) \in Pref(\sigma P) \Leftrightarrow \begin{cases} \bullet\ (\mathcal{M}, \mathcal{N}, \sigma) \models_{Pref} (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2) \\ \bullet\ D \models_L \sigma\varphi_2 \text{ et } D' \models_L \sigma\psi_2 \end{cases}$
- $((\mathcal{M}, D), (\mathcal{N}, D')) \in Pref(\sigma P) \Rightarrow ((\mathcal{N}, D'), (\mathcal{M}, D)) \notin Pref(\sigma P)$

Dorénavant, on notera $Pref(\sigma P)$ la préférence obtenue en remplaçant les variables universellement quantifiées par la substitution σ.

On définit également la 1 (respectivement 2) applicabilité d'une préférence:

Définition 64. *(1-applicabilité d'une préférence)*
Soit $\mathcal{A}$ une d'algèbre. Une préférence $P \in P_\Sigma[X]$ est dite 1-applicable si et seulement si il existe une substitution $\sigma : X \mapsto T_\Sigma$ telle que l'application $p_1^{\mathcal{A},\sigma} \mapsto \{0,1\}$ vaut 1, application définie par induction par:

- $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)_1^{\mathcal{A},\sigma}(\xi, \sigma) = inf((\sigma\varphi_1)^{\mathcal{A}}(\xi), (\sigma\phi_1)^{\mathcal{A}}(\xi))$ et $D \models \sigma\varphi_2$

Respectivement, il est facile de définir la 2-applicabilité d'une préférence.

Comme on peut le constater, ces définitions font la part belle à une utilisation différente des variables dans l'interprétation des formules du cas classique. La raison en est la volonté de prendre en compte une notion contextuelle dans son interprétation en supposant que toutes les préférences sont définies dans un contexte particulier. Le rajout des variables dans la définition des préférences doit donc être considéré

20. On rappelle que par convention de notation, une algèbre est notée $\mathcal{A}$ et qu'alors A désigne son ensemble de définition.

comme étant du *"sucre syntaxique"*, permettant de ne pas écrire cette préférence pour tous les termes du langage.

Dorénavant, et dans **toute** la suite de ce document, nous ne considérerons plus que des préférences instantiées, et ce afin de simplifier les notations et définitions. D'autre part, nous considérerons également dans toute la suite de ce document des ensembles **finis** de formules préférentielles.

On remarquera que les préférences que nous introduisons ici existent déjà dans un système d'intelligence artificielle, le système Soar [Laird et al. 1986][Newell 1990], sans sémantique formelle cependant. En effet,

- une préférence du type $(\bot, \bot, \bot) \succ (\varphi_1, \varphi_2, \phi_2)$, signifie intuitivement: si φ_1 et ϕ_2 sont vraies, si l'on a la possibilité de négliger φ_2, alors préférer ne rien faire plutôt que de faire ϕ_2, ou dit autrement rejeter l'action ϕ_2: c'est la préférence *reject* de Soar,
- les préférences du type $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ correspondent quant à elles intuitivement à un jugement du type: si ϕ_1 résulte de la connaissance de φ_1 et de la négligeabilité de φ_2, et si ϕ_2 résulte de la connaissance de ψ_1 et de la négligeabilité de ψ_2, alors préférer ϕ_1 à ϕ_2: c'est la préférence *better* de Soar,
- Le système Soar définit aussi la préférence *worse* qui est la préférence duale de la précédente, ainsi que les préférences *best* et *worst*, qui correspondent à une quantification sur ϕ_1 dans le premier cas, sur ϕ_2 dans le deuxième cas, donc sont des préférences d'ordre 2.

5.4 Discussion sur les capacités de représentation des préférences

Ces préférences peuvent être utilisées pour la représentation de plusieurs types de jugements que peut effectuer un décideur dans le monde réel, à savoir des jugements de classification et des jugements de décision, jugements que nous allons maintenant détailler:

Les jugements de classification Soit $(S_i)_{1\leq i\leq n}$ un ensemble fini d'états, et $(A_j)_{1\leq j\neq n}$ un ensemble d'actions. Si l'on suppose que l'ensemble des actions que peut accomplir le décideur est fini, tout se passe alors comme si il existait une fonction f telle que $f^{-1}(A_j)$ pour $j = 1, \ldots, n$ réalise une partition de $(S_i)_{1\leq i\leq n}$ et le décideur doit choisir une action telle que $f(S_i) = A_j$, ce qui revient à dire [Pomerol 1997] qu'il doit déterminer si $S_i \in f^{-1}(A_j)$. Un système tel que le nôtre propose alors une telle fonction f au décideur, en donnant le séquent représentant la décision en fonction des états. Il permet de plus d'augmenter à l'aide de préférences la classification obtenue. C'est le cas du problème de photo-interprétation que nous avons mentionné, si l'on considère par exemple la préférence suivante:

$$(\bot, \bot, \bot) \succ (\top, Piste(l_1), AeroportCivil(l_1))$$

Les jugements de décision Les problèmes de décision correspondent eux aussi à une incertitude sur l'état réel de la nature, incertitude ici traduite par le biais de la négligeabilité d'un élément de description. A cette négligeabilité est associée une action, qui conduit à un résultat. La relation de préférence peut alors représenter des préférences sur des actions, comme dans l'exemple suivant où elle permet de représenter des préférences sur actions sachant des résultats et des négligeabilités:

$$\begin{gathered}(PremierPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Faible), Cout(L, 100)) \\ \succ \\ (SecondPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Importante), Cout(L, 200))\end{gathered}$$

6

Similarité - Satisfaction préférentielle forte

Ce chapitre est consacré à poser la définition de la relation de déduction par similarité en prenant en compte les préférences. Dans ce chapitre, nous supposerons que le décideur peut négliger toute différence potentielle entre cas courant et cas passé, sans à priori d'incompatibilités. Souhaitant poser la définition d'un raisonnement par similarité avec des préférences, nous commençerons par poser la définition de la fonction de similarité. Puis, utilisant la définition des préférences données au chapitre précédent, nous définissons la relation de déduction sémantique par similarité prenant en compte les préférences. Cette définition étant posée et discutée, nous fournissons ensuite une sémantique de point fixe permettant une implantation dans un système d'intelligence artificielle. Les liens entre la fonctionnelle de point fixe et le système Soar sont discutés.

6.1 Définition de la notion de similarité

Nous allons raisonner à partir de séquents, qui sont des listes de symboles de la forme:

$$\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m$$

exprimant que de $\varphi_1, \ldots, \varphi_n$, on peut déduire ψ_1 ou ψ_2 ou ... ou ψ_m. Nous allons définir une relation de similarité en posant que si l'on connaît une généralisation des prémisses $\varphi_1, \ldots, \varphi_n$, on peut déduire des conclusions généralisées de $\psi_1, \ldots, \psi_m$. Cette généralisation sera prise en compte par la fonction de similarité que nous présenterons.

Nous allons maintenant définir sémantiquement un raisonnement par similarité. Nous allons l'appréhender comme étant une subsomption entre ce que le décideur connaît du problème à résoudre et ses connaissances représentées par des séquents.

6.1.1 Définition de la fonction de similarité

Intuitivement, la notion de similarité s'appréhende alors de la manière suivante:

> Soit $(: \psi'_1), \ldots, (: \psi'_k)$ un ensemble de séquents désignant le problème courant à résoudre, et soit $(\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m)$ un séquent désignant les solutions pouvant être entreprises sachant le problème désigné par $\varphi_1, \ldots, \varphi_n$. On peut adapter le séquent $(\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m)$ pour résoudre le problème $(: \psi'_1), \ldots, (: \psi'_k)$, adaptation qui prend la forme du séquent $(\varphi''_1, \ldots, \varphi''_l : \psi''_1, \ldots, \psi''_m)$ si et seulement si:
>
> - **Généralisation des prémisses**
>   - $\varphi''_1 \wedge \ldots \wedge \varphi''_l$ est plus général que $\psi'_1 \wedge \ldots \wedge \psi'_k$ et que $\varphi_1 \wedge \ldots \wedge \varphi_n$,
> - **Généralisation des conclusions**
>   - $\psi''_1 \vee \ldots \vee \psi''_m$ est plus général que $\psi_1 \vee \ldots \vee \psi_m$,

Cette notion de généralisation sera pour nous appréhendée par une proposition de subsomption. Dans la mesure où nous utilisons une généralisation, des éléments trop particuliers n'apparaissent plus dans les formules généralisées. On qualifiera dorénavant ces éléments non pris en compte par la relation de généralisation de **négligeables**. La théorie que nous proposons ici suppose que le décideur ne connaît pas du tout les éléments pouvant être négligés pendant le raisonnement: tous les éléments des formules sont donc potentiellement négligeables.

Nous terminerons cette section d'introduction au raisonnement par similarité par un exemple concret de similarité, en utilisant des clauses de Horn (séquents intuitionnistes sur des formules atomiques):

Exemple 65. Considérons le problème de photo-interprétation suivant donné par l'image et les cas en mémoire suivants:

- Cas mémorisé:

$$\begin{aligned}
C_1 &\equiv \quad Piste(l_1), Taxiway(l_1), TourControle(l_1) : Aeroport(l_1) \\
C_2 &\equiv \quad Aeroport(l_1), Aerogare(l_1) : AeroportCivil(l_1) \\
C_3 &\equiv \quad Aeroport(l_1), StationnementNonAligne(l_1) : AeroportMilitaire(l_1) \\
C_4 &\equiv \quad AeroportMilitaire(l_1), AireAlerte(l_1) : AeroportMilitaireChasse(l_1) \\
C_5 &\equiv \quad AeroportMilitaire(l_1), Piste(l_1), LongueurPiste(l_1, L > 1000), \\
&\qquad Hangars(l_1) : AeroportMilitaireBombardiers(l_1)
\end{aligned}$$

- Préférences mémorisées:

$$\begin{gathered}
P_1 \equiv (Piste(l_1), \perp, AeroportMilitaire(l_1)) \\
\succ \\
(Piste(l_1), \perp, AeroportCivil(l_1))
\end{gathered}$$

$$\begin{gathered}
P_2 \equiv (AeroportMilitaire(l_1), AireAlerte(l_1), AeroportMilitaireChasse(l_1)) \\
\succ \\
(AeroportMilitaire(l_1), Hangars(l_1), AeroportMilitaireBombardiers(l_1))
\end{gathered}$$

- Cas courant

$$Piste(l), Taxiway(l), LongueurPiste(l, 1500)$$

Dans cet exemple, il nous semble souhaitable de proposer comme formule similaire à la fois aux cas passés et au cas courant la formule

$$C_1' \equiv \; Piste(l) \wedge Taxiway(l) \Rightarrow Aeroport(l)$$

sachant que $TourControle(l)$ est négligé. D'autres similarités sont bien sûr possibles avec les autres éléments de la base de faits.

◀

6.1.1.1 Fonction de similarité

Comme nous l'avons discuté dans l'introduction de ce chapitre, la relation de généralisation que nous utiliserons est la subsomption, dont nous rappelons brièvement la définition:

Définition 66. Soient $\varphi, \psi \in L_\Sigma$. On dit que ψ "subsume" φ si et seulement si $\varphi \models_L \psi$, c'est à dire si et seulement si $\mathbf{Mod}\{\varphi\} \subseteq \mathbf{Mod}\{\psi\}$. On note ψ "subsume" φ par $\psi\mathbf{S}\varphi$.

La subsomption va nous permettre de calculer les parties communes de deux formules. En ce qui concerne les parties négligeables, nous allons tirer profit d'une caractéristique des mondes qui justifie ici leur utilisation, à savoir leur propriété essentielle de tri-valuation. Nous n'avons en effet utilisé jusqu'à présent que la relation $\models_{h,L}$, mais on peut définir la valuation $V_i : L_\Sigma \mapsto \{0,1,U\}$, définie par, si $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$, $\mathbf{M}$ une algèbre de $\mathcal{M}$, M son ensemble de définition, et M^X l'ensemble des applications de X dans M:

$$\begin{array}{ll} V_i^{\mathcal{M}}(\varphi) = 1 & \text{si } \forall \mathbf{M} \in \mathcal{M}, \forall \xi \in M^X, \xi^M(\varphi) = 1 \\ V_i^{\mathcal{M}}(\varphi) = 0 & \text{si } \forall \mathbf{M} \in \mathcal{M}, \forall \xi \in M^X, \xi^M(\varphi) = 0 \\ V_i^{\mathcal{M}}(\varphi) = U & \text{sinon} \end{array}$$

La première relation est bien sûr la relation $\models_{h,L}$. Un monde a la propriété que les formules qui sont valides dans certaines algèbres du monde, mais invalides dans d'autres sont incertaines. Cette incertitude va nous permettre de poser la fonction de similarité (les éléments négligeables seront incertains).

La définition suivante va nous permettre de prendre en compte l'ensemble des formules similaires à la fois à un séquent donné et à un monde:

Définition 67. (*Fonction de similarité*)
On définit une fonction

$$\begin{array}{llll} Sim: & S_\Sigma \times WD_L & \longmapsto & \mathcal{P}(S_\Sigma \times L_\Sigma) \\ & (S, \mathcal{M}) & \longmapsto & \{(S'_k, \Delta_k)_{k \in K}\} \end{array}$$

définissant l'ensemble des séquents S'_k similaires à un séquent donné S dans le contexte déterminé par le monde $\mathcal{M}$ sachant les négligeables déterminés par Δ_k, fonction qui satisfait les propriétés suivantes :

1. **Calcul des éléments similaires**
   Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

   $$\forall (S', \Delta') \in Sim(S, \mathcal{M}), \left\{ \begin{array}{l} \bullet\ S'^- \text{ subsume } B \text{ et } S^-, \\ \bullet\ S'^+ \text{ subsume } S^+ \end{array} \right.$$

   où B est l'ensemble des séquents de la forme suivante $(: \varphi)$ validés par $\mathcal{M}$.

   Intuitivement, cette définition signifie que les prémisses du séquent S' sont plus générales que ce qui est connu à la fois dans le monde et dans le séquent S, et que ses conclusions sont également plus générales que celles de S.

2. **Calcul des éléments dissimilaires**
   Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

   $$\forall (S', \Delta') \in Sim(S, \mathcal{M}), \left\{ \begin{array}{l} \bullet\ S^- \models_L \Delta' \text{ et } S'^- \not\models_L \Delta' \\ \bullet\ \forall \varphi \in S^-, S'^- \not\models_L \varphi \Rightarrow \exists \phi \in \Delta', \varphi \models_L \phi \\ \bullet\ \forall \varphi' \in \Delta', \exists \varphi \in S^-, S'^- \not\models_L \varphi \text{ et } \varphi \models_L \varphi' \end{array} \right.$$

   Intuitivement, cette condition signifie que ce qui a été négligé ne se déduit pas du nouveau séquent (première condition), que ce qui a été négligé se déduit de l'ancien séquent (seconde condition), et que ce sont les seuls éléments qui sont négligés (troisième condition).

3. **Condition aux limites**
   Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

   $$\text{Si } \mathcal{M} \models_{h,L} S, \text{alors } Sim(S, \mathcal{M}) = \{(S, \emptyset)\}$$

   Lorsque le monde satisfait le séquent sans similarité, alors sa seule formule similaire est lui-même.

4. **Incertitude des éléments négligés**
   Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

   $$\forall (S', \Delta') \in Sim(S, \mathcal{M}), \left\{ \begin{array}{l} \bullet\ \forall \varphi \in \Delta', V_i^{\mathcal{M}}(\varphi) = U \\ \bullet\ \forall \varphi \in S^-, S'^- \not\models_L \varphi \Rightarrow V_i^{\mathcal{M}}(\varphi) = U \end{array} \right.$$

6.1.2 Exemple de fonction de similarité

On peut considérer comme fonction de similarité ce que l'on peut intuitivement considérer comme une unification partielle, et qui correspond à une notion de similarité maximale en terme syntaxique. Plus formellement parlant, il s'agit de montrer que la fonction de similarité qui sur les prémisses d'un séquent correspond à la borne supérieure au sens du pré-ordre de subsomption sur l'ensemble des formules fermées du premier ordre correspond bien en terme syntaxique à une unification partielle. La tête du séquent similaire est alors la tête du séquent initial auquel on a appliqué la substitution que la subsomption a définie. La démonstration complète étant donnée en annexe C, nous n'en esquisserons ici que l'ossature.

La première partie de la démonstration consiste à déterminer l'existence de cette borne supérieure, donc à montrer que l'ensemble des formules du premier ordre muni du pré-ordre de subsomption est un treillis (voir lemme 105). Cette structure de treillis est en quelque sorte un héritage de la structure de treillis des classes d'algèbres finiment axiomatisable pour l'inclusion.

Ensuite, partant d'une famille de formules universelles du premier ordre, il s'agit de déterminer la forme syntaxique de leur borne supérieure au sens du pré-ordre de subsomption. Pour ce faire, on travaille sur la forme syntaxique de cette borne supérieure, en réduisant sa classe parmi les formules du premier ordre, c'est à dire en montrant que c'est une formule universelle (voir annexe 109), et en montrant que l'on peut se ramener au calcul d'une borne supérieure de formules ne contenant pas le connecteur de disjonction (voir proposition 107). Cette démonstration ayant permis de ramener à une classe suffisamment restreinte de formules, on peut alors aborder une démonstration constructive, qui prouve que "l'unification partielle" des formules considérées est la borne supérieure au sens du pré-ordre de subsomption de ces formules (voir proposition 112). Ceci nous conduit alors à un calcul syntaxique effectif de la solution (voir théorème 115) pour le cas des formules ne contenant pas la disjonction comme connecteur binaire, et voir théorème 116 pour l'extension à l'ensemble des formules universelles du premier ordre.

Ainsi, en supposant que les séquents considérés sont constitués de formules universelles, la fonction de similarité réalise un "appariement partiel" entre les séquents de type $(:\psi_1), \ldots, (:\psi_n)$ validés par le monde $\mathcal{M}$ et le séquent S dont on recherche une similarité. Lorsqu'on se restreint aux clauses de Horn, la fonction de similarité prend la forme d'une unification partielle entre une clause définie et une conjonction de clauses unités.

6.1.3 Discussion sur la notion de similarité

Deux points peuvent être évoqués concernant cette fonction de similarité. Le premier point que nous souhaitons aborder est le lien avec les logiques de descriptions et leur utilisation pour le raisonnement à partir de cas. En effet, comme nous l'avons précisé dans le chapitre précédent, la modification des anciens cas se fait par l'intermédiaire d'un *schéma d'encodage* [Kolodner 1993][Koehler 1994] [Koehler 1996][Lieber and Napoli 1997]. Informellement parlant, le résultat de l'application d'un schéma d'encodage à un cas, appelé *indice*, doit représenter "*les caractéristiques du cas qui le rendent aussi bien suffisamment abstrait pour être utile dans une grande variété de situations que concret pour être facilement reconnaissable*" [Koehler 1994]. Notre fonction de similarité correspond à ce schéma d'encodage, à ceci près qu'elle permet de déterminer les négligeables associés à la généralisation du séquent.

Enfin, par rapport aux logiques des défauts, cette fonction de similarité permet, situation par situation (i.e. monde par monde) de déterminer le lien entre les négligeabilités et ce qui est connu, et donc d'inférer en fonction du monde les éléments négligeables.

6.2 Relation de satisfaction par similarité

Dans cette section, nous commencerons par proposer la définition de la relation de conséquence sémantique par similarité.

6.2.1 La relation de conséquence sémantique

Avant de nous intéresser à la prise en compte des préférences, commençons par nous intéresser à la prise en compte des négligeables sans préférences, et commençons par introduire une relation de satisfaction sachant des éléments négligeables:

Définition 68. *(Satisfaction sachant des négligeables)*
Soit $\mathcal{M} \in WD_L, \varphi \in L_\Sigma$, et soit Γ un ensemble de séquents de $S_\Sigma[X]$. On étend la valuation V_i à la valuation V_S, lue valuation par similarité sachant Γ de la manière suivante:

$$
\begin{array}{lll}
V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(\varphi) & = 1 \Leftrightarrow & \left\{ \begin{array}{l} \bullet\ \forall \mathbf{M} \in \mathcal{M}, V_i^{\mathcal{M}}(\varphi) = 1 \\ \bullet\ \text{ou } V_i^{\mathcal{M}}(\varphi) = U \text{ et } \exists \Delta', (S', \Delta') \in Sim(\Gamma, \mathcal{M}) \\ \quad \text{et } \varphi \text{ subsume } \Delta' \end{array} \right. \\
& = 0 \Leftrightarrow & V_i^{\mathcal{M}}(\varphi) = 0 \\
& = U & \text{sinon}
\end{array}
$$

On étend la définition à un ensemble de formules de manière usuelle.

On définit une relation de satisfaction, dite satisfaction sachant des négligeables déterminés par $\Gamma \subseteq S_\Sigma[X]$, $\models_{neg} \subseteq (\mathcal{P}(WD_L) \times \mathcal{P}(S_\Sigma[X])) \times S_\Sigma[X]$ [21] la relation de satisfaction sémantique définie par, si $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$:

$$
\mathcal{M} \models_{neg/\Gamma} S \Leftrightarrow \left\{ \begin{array}{l} \bullet \text{ si } S^- = \emptyset \text{ ou } S^+ = \emptyset, V_i^{\mathcal{M}}(S) = 1 \\ \bullet \text{ si } S^- \neq \emptyset \text{ et } S^+ \neq \emptyset, V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(S) = 1 \end{array} \right.
$$

On a alors bien sûr:

Remarque 69. Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X], \Gamma \subseteq S_\Sigma[X]$.

$$
\mathcal{M} \models_{neg/\Gamma \cup S} S \Leftrightarrow \forall (S', \Delta') \in Sim(S, \mathcal{M}), \mathcal{M} \models_L S'
$$

La relation de conséquence sémantique par similarité a pour but d'exprimer qu'en cas d'incomplétude d'une théorie constituée par un ensemble de séquents, la similarité permet "d'élargir" l'ensemble des conséquences de cette théorie, à condition qu'il n'y ait pas de préférences contradictoires. Dit autrement, on cherche donc l'ensemble des conséquences maximales au sens des préférences. Avant de poser la définition de la relation de conséquence sémantique, rappelons la définition suivante:

Définition 70. *(Rappel: notion d'admissibilité)*
Soit Y un ensemble, et $\succ$ un ordre large sur Y. On appelle ensemble des admissibles le sous ensemble A de Y défini par:

$$
a \in A \Leftrightarrow \text{il n'existe aucun } y \in Y \text{ vérifiant } y \succ a.
$$

Bien que nous n'ayons qu'une simple relation binaire asymétrique en lieu et place d'un pré-ordre, convenons cependant que l'ensemble des éléments admissibles pour une relation binaire R est l'ensemble des éléments a tels qu'il n'existe pas de b tels que $(a, b) \in R$: dit autrement, R ne définit pas de "meilleur" élément que a. Ré-exprimons alors la notion de conséquences d'une théorie incomplète: *"les conséquences d'un programme sont les conséquences admissibles au sens des préférences du décideur"*:

Définition 71. Soit WD_L l'ensemble des mondes sur la logique L, soit S un séquent de $S_\Sigma[X]$, soit Γ une théorie de $S_\Sigma[X]$ et soit $P = (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ une préférence de $P_\Sigma[X]$. On appelle ensemble des admissibles pour S sachant les connaissances déterminées par Γ selon les préférences définies par P l'ensemble suivant, noté $Admiss_{/\Gamma}(P, S) \subseteq WD_L$:

$$
\begin{array}{ll}
\mathcal{M} \in Admiss_{/\Gamma}(P, S) & \Leftrightarrow \mathcal{M} \models_{\subset, neg/\Gamma \cup S} S \text{ et} \\
& \neg(\exists \gamma \in \Gamma, \mathcal{M} \models_{\subset, neg/\Gamma \cup S} \Gamma \text{ et} \\
& \exists (S', \Delta') \in Sim(S, \mathcal{M}), \exists (\gamma'', \Delta'') \in Sim(\gamma, \mathcal{M}), \\
& \Delta' \neq \emptyset, \Delta'' \neq \emptyset, ((\mathcal{M}, \Delta''), (\mathcal{M}, \Delta')) \in Pref(P))
\end{array}
$$

où $\models_{\subset, neg} \subseteq \mathcal{P}(WD_L) \times \mathcal{P}(S_\Sigma[X])) \times S_\Sigma[X]$ est la relation de déduction définie par:

$$
\mathcal{M} \models_{\subset, neg/\Gamma} S \Leftrightarrow \mathcal{M} \models_{neg/\Gamma} S \text{ et } \neg(\exists \mathcal{M}', \mathcal{M}' \models_{neg/\Gamma} S \text{ et } \mathcal{M}' \subset \mathcal{M})
$$

L'ensemble des admissibles est défini par les mondes qui satisfont la formule sachant des négligeabilités et tels qu'il n'existe pas de "meilleure formule" des connaissances du décideur pour ces préférences.

21. La notation $\mathcal{M} \models_{neg/\Gamma} \varphi$ prise en lieu et place d'une notation déduction sachant les négligeables déterminés par Γ.

Considérons à l'appui de cette définition l'exemple suivant:

$$\begin{array}{ll} C_1 \equiv & Piste(l_1), Taxiway(l_1), TourControle(l_1) : Aeroport(l_1) \\ C_2 \equiv & Aeroport(l_1), Aerogare(l_1) : AeroportCivil(l_1) \\ C_3 \equiv & Aeroport(l_1), StationnementNonAligne(l_1) : AeroportMilitaire(l_1) \\ C_4 \equiv & AeroportMilitaire(l_1), AireAlerte(l_1) : AeroportMilitaireChasse(l_1) \\ C_5 \equiv & AeroportMilitaire(l_1), Piste(l_1), LongueurPiste(l_1, L > 1000), \\ & Hangars(l_1) : AeroportMilitaireBombardiers(l_1) \end{array}$$

$$P_1 \equiv (Piste(l_1), \bot, AeroportMilitaire(l_1)) \succ (Piste(l_1), \bot, AeroportCivil(l_1))$$

$$P_2 \equiv (Piste(l_1) \wedge LongueurPiste(l_1, L), AireAlerte(l_1), AeroportMilitaireChasse(l_1)) \succ (Piste(l_1), Hangars(l_1), AeroportMilitaireBombardiers(l_1))$$

$$B \equiv Piste(l), Taxiway(l), LongueurPiste(l, 1500)$$

On considère dans cet exemple que la fonction de similarité est définie par une unification partielle. Si l'on considère les mondes admissibles de la clause C_2 sachant les connaissances déterminées par cette théorie, tous les mondes validant $AeroportMilitaire(l_1)$ en négligeant ou non $StationnementNonAligne(l_1)$ sont des mondes admissibles de cette théorie. Par contre, si l'on regarde maintenant les mondes admissibles de la clause C_3 sachant cette même théorie, tous les mondes validant $AeroportCivil(l_1)$ en négligeant $Aerogare(l_1)$ ne sont pas admissibles pour la préférence P_1.

Pour justifier l'introduction de la relation $\models_{\mathsf{C},neg}$, considérons ce même programme amputé du fait $LongueurPiste(l, 1500)$, et retirons la condition de minimalité sur $\mathcal{M}$. On s'aperçoit alors qu'on ne peut déduire $AeroportMilitaireBombardiers(l)$, car la préférence P_2 reste appliquable dans les mondes qui satisfont le programme, ainsi que $AeroportMilitare(l), Piste(l), Taxiway(l), LongueurPiste(l, 1500)$. La notion de minimalité nous permet d'encoder dans le formalisme que **tout ce que le décideur connaît est dans la théorie** Γ.

Remarquons bien cependant que la définition de l'admissibilité n'est pas suffisante pour donner la relation de déduction par similarité. Un élément manque en effet pour que la notion de similarité ait bien la signification intuitive qu'on lui prête, et que la relation de déduction par similarité puisse permettre de déduire plus qu'en ordre un classique. Parmi les mondes admissibles de C_2 sachant les connaissances déterminées par cette théorie, et sachant la préférence P_1, considérons les mondes validant $\neg StationnementNonAligne(l)$ et ceux validant $B, AeroportMilitaire(l)$. S'ils sont bien tous les deux admissibles, on préfère, sachant les préférences et les mondes admissibles, le second au premier. Ici donc, on préfère valider la tête du séquent plutôt que la négation de l'une de ses prémisses.

Définition 72. *(Préférence pour la négligeabilité)*
Soit WD_L l'ensemble des mondes sur la logique L, soit S un séquent de $S_\Sigma[X]$, soit Γ une théorie de $S_\Sigma[X]$ et soit $P = (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ une préférence de $P_\Sigma[X]$. On introduit une relation binaire asymétrique notée $\succ_0$ telle que, si $\mathcal{M}, \mathcal{N}$ sont deux mondes de $Admiss_{/\Gamma}(P, S)$:

$$\mathcal{M} \succ_0 \mathcal{N} \Leftrightarrow \left\{ \begin{array}{l} \bullet\ \exists S' \in S_\Sigma[X], \exists \Delta' \in L_\Sigma[X], (S', \Delta') \in Sim(S, \mathcal{M}), \\ \bullet\ \exists S'' \in S_\Sigma[X], (S'', \emptyset) \in Sim(S, \mathcal{N}) \end{array} \right. \quad \exists \varphi \in \Delta', \mathcal{N} \models \neg\varphi$$

et on note $AdmissNeg_{/\Gamma}(P, S)$ l'ensemble des mondes de $Admiss_{/\Gamma}(P, S)$ admissibles pour la relation $\succ_0$

Intuitivement , le monde $\mathcal{M}$ est préféré au monde $\mathcal{N}$ car le second valide la négation d'une prémisse alors que le premier considére cette prémisse comme négligeable. Il ne nous reste plus maintenant qu'à intégrer cette condition à la définition du raisonnement par similarité:

Définition 73. *(Satisfaction par similarité)*
Soit WD_L l'ensemble des mondes sur la logique L, soit S un séquent de $S_\Sigma[X]$, soit Γ une théorie de $S_\Sigma[X]$ et soit $P = (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ une préférence de $P_\Sigma[X]$. On introduit une relation $\models_S \subseteq (\mathcal{P}(WD_L) \times \mathcal{P}(S_\Sigma[X])) \times SP_\Sigma[X]$, lue relation de déduction par similarité sachant des connaissances (ici déterminées par Γ), la relation définie par:

$$\mathcal{M} \models_{\mathcal{S}/\Gamma} (P, S) \Leftrightarrow \mathcal{M} \in AdmissNeg_{/\Gamma}(P, S)$$

On élargit à une théorie de la manière suivante:

Définition 74. Soit WD_L l'ensemble des mondes sur la logique L, soit (Γ, Π) une théorie de $SP_\Sigma[X], \Gamma \subseteq S_\Sigma[X], \Pi \subseteq P_\Sigma[X]$. On élargit la définition de la relation de déduction sémantique par similarité de la manière suivante:

$$\mathcal{M} \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Leftrightarrow \forall S \in \Gamma, \forall P \in \Pi, \mathcal{M} \models_{\mathcal{S}/\Gamma} (S, P)$$

Cette définition va nous permettre de poser la définition d'un modèle épistémique similaire:

Définition 75. (*Modèle épistémique similaire*)
Un modèle *épistémique similaire* est formé par un triplet $(WD_L, Sim, \models_S)$ où $\models_S \subseteq (\mathcal{P}(WD_L) \times \mathcal{P}(S_\Sigma[X])) \times SP_\Sigma[X]$ est définie par:

1. Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X], P \in P_\Sigma[X], \Gamma \in \mathcal{P}(S_\Sigma[X])$:

$$\mathcal{M} \models_{\mathcal{S}/\Gamma} (P, S) \Leftrightarrow \mathcal{M} \in AdmissNeg_{/\Gamma}(P, S)$$

2. Soit $\mathcal{M} \in WD_L, \Gamma \in \mathcal{P}(S_\Sigma[X]), \Pi \in \mathcal{P}(P_\Sigma[X])$:

$$\mathcal{M} \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Leftrightarrow \forall S \in \Gamma, \forall P \in \Pi, \mathcal{M} \models_{\mathcal{S}/\Gamma} (S, P)$$

3. Soit $\Gamma \in \mathcal{P}(S_\Sigma[X]), \Pi \in \mathcal{P}(P_\Sigma[X]), S \in S_\Sigma[X]$,

$$(\Gamma, \Pi) \models_{\mathcal{S}/\Gamma} S \Leftrightarrow \forall \mathcal{M} \in WD_L, \mathcal{M} \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Rightarrow \mathcal{M} \models_L S$$

Les conséquences d'une théorie sont donc les conséquences valides dans tous les mondes admissibles des théories qui lui sont similaires.

On notera alors $\mathcal{M} \models_{\mathcal{S}} (\Gamma, \Pi)$ en lieu et place de la notation $\mathcal{M} \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi)$, Γ ensemble de séquents, Π ensemble de préférences, une telle relation pour éviter les surcharges de notations.

6.2.2 Propriétés de la relation de déduction par similarité

Nous avons dans le chapitre II donné l'ensemble des propriétés que doit satisfaire une "bonne" relation de déduction par similarité avec des préférences. Il nous faut maintenant vérifier que la relation de similarité $\models_{\mathcal{S}}$ que nous avons défini satisfait ces bonnes propriétés.

Définition 76. La relation d'inférence par similarité $\models_{\mathcal{S}}$ satisfait les propriétés suivantes:

Supra Classique $\dfrac{\Gamma \vdash a}{(\Gamma, \emptyset) \vdash_{\mathcal{S}} a}$

Inclusion $(\Gamma \cup a, \emptyset) \vdash_{\mathcal{S}} a$

$\wedge$ $\dfrac{(\Gamma, \Pi) \vdash_{\mathcal{S}} a \quad (\Gamma, \Pi) \vdash_{\mathcal{S}} b}{(\Gamma, \Pi) \vdash_{\mathcal{S}} a \wedge b}$

Monotonie prudente $\dfrac{(\Gamma, \Pi) \vdash_{\mathcal{S}} y \quad (\Gamma, \Pi) \vdash_{\mathcal{S}} a}{(\Gamma \cup y, \Pi) \vdash_{\mathcal{S}} a}$

$\vee$ $\dfrac{(\Gamma \cup y, \Pi) \vdash_{\mathcal{S}} a \quad (\Gamma \cup z, \Pi) \vdash_{\mathcal{S}} a}{(\Gamma \cup (y \vee z), \Pi) \vdash_{\mathcal{S}} a}$

Coupure $\dfrac{(\Gamma, \Pi) \vdash_{\mathcal{S}} y \quad (\Gamma \cup y, \Pi) \vdash_{\mathcal{S}} a}{(\Gamma, \Pi) \vdash_{\mathcal{S}} a}$

Commençons par effectuer la remarque suivante:

Remarque 77. Soit $\Delta, \Pi \subseteq SP_\Sigma[X], y \in S_\Sigma[X]$ tels que $(\Delta, \Pi) \models_S y$. Alors

$$\forall \mathcal{M} \in WD_L, \mathcal{M} \models_S (\Delta, \Pi), \forall a \in S_\Sigma[X], V_{S/\Gamma}^{\mathcal{M}}(a) = 1 \Leftrightarrow V_{S/\Gamma \cup y}^{\mathcal{M}}(a) = 1$$

Preuve. Considérons la théorie $\Gamma \cup y$. Alors $Sim(\Gamma \cup y, \mathcal{M}) = Sim(\Gamma, \mathcal{M}) \cup (y, \emptyset)$ puisque par définition $\mathcal{M} \models_{h,L} y$. Dans ces conditions, les éléments négligés dans le monde $\mathcal{M}$ sont les mêmes pour les théories $\Gamma \cup y$ et Γ. L'équivalence s'ensuit immédiatement. ■

Nous pouvons maintenant prouver les propriétés de la relation $\models_{\mathcal{S}}$:

Preuve. Les propriétés supra classique et d'inclusion reposent sur la remarque suivante:

- la propriété ne fait pas intervenir les préférences,
- la valuation par similarité $V_{S/\Gamma}$ ne fait qu'étendre une valuation V_i (elle ne change pas la valuation des formules φ telles que $V_i^{\mathcal{M}}(\varphi) = 1$).

Les propriétés de monotonie prudente et de coupure repose trivialement sur la remarque 77. En ce qui concerne la démonstration de la règle $\vee$, considérons un monde $\mathcal{M}$ tel que $\mathcal{M} \models_{\mathcal{S}} (\Gamma \cup (y \vee z), \Pi)$. Comme $y \vee z$ subsume y et z, par exemple $\mathcal{M} \models_{\mathcal{S}} (\Gamma \cup y, \Pi)$, et donc $\mathcal{M} \models_{\mathcal{S}} a$. ■

6.2.3 Exemples de déduction par similarité

Nous considérerons dans cette section deux exemples de raisonnement par similarité en fonction de préférences, un exemple dans un problème de photo-interprétation et un exemple dans un problème de décision dans le cas d'un forage pétrolier.

6.2.3.1 La photo-interprétation

Considérons la théorie préférentielle suivante:

$$\begin{aligned}
C_1 &\equiv Piste(l_1), Taxiway(l_1), TourControle(l_1) : Aeroport(l_1)\\
C_2 &\equiv Aeroport(l_1), Aerogare(l_1) : AeroportCivil(l_1)\\
C_3 &\equiv Aeroport(l_1), StationnementNonAligne(l_1) : AeroportMilitaire(l_1)\\
C_4 &\equiv AeroportMilitaire(l_1), AireAlerte(l_1) : AeroportMilitaireChasse(l_1)\\
C_5 &\equiv AeroportMilitaire(l_1), Piste(l_1), LongueurPiste(l_1, L > 1000),\\
&\quad Hangars(l_1) : AeroportMilitaireBombardiers(l_1)
\end{aligned}$$

$$P_1 \equiv (Piste(l_1), \perp, AeroportMilitaire(l_1)) \succ (Piste(l_1), \perp, AeroportCivil(l_1))$$

$$P_2 \equiv (Piste(l_1) \wedge LongueurPiste(l_1, L), AireAlerte(l_1), AeroportMilitaireBombardiers(l_1)) \succ (Piste(l_1), Hangars(l_1), AeroportMilitaireChasse(l_1))$$

$$B \equiv Piste(l), Taxiway(l), LongueurPiste(l, 1500)$$

On suppose que la fonction de similarité est donnée par une unification partielle, et que $Aeroport, AeroportMilitaire$ ne sont jamais négligeables dans les séquents où ils apparaissent. On considère bien entendu les conséquences préférentielles de cette théorie sachant les connaissances déterminées par cette même théorie. L'ensemble des mondes admissibles pour la clause C_2 se résume à ceux qui valident $\neg Aerogare(l)$ où à ceux qui valident $Aerogare(l)$ et $AeroportCivil(l)$. L'ensemble des mondes admissibles pour la première clause (C_1) sont les mondes qui valident $Aeroport(l)$ ou $\neg TourControle(l)$, et par similarité, les mondes qui en sont modèles sont tous ceux qui permettent de déduire à un aéroport en l. L'ensemble des mondes modèles par similarité de la clause C_3 sont tous ceux qui valident $AeroportMilitaire(l)$, alors que les mondes admissibles, et donc modèles par similarité de la clause C_2 sont tous ceux qui valident $\neg Aerogare(l)$. Par application de la première préférence, on en déduit donc qu'il y a un aéroport militaire en l. La seconde préférence, associée aux cas en mémoire, permet de déduire qu'il y a un aéroport militaire pour bombardiers en l.

6.2.3.2 Les décisions de forage pétrolier

Considérons la théorie préférentielle suivante:

$$PremierPuits(L), Hydrocarbures(L), Quantite(L, Faible) : MiseEnProduction(L, StructureLegere)$$
$$MiseEnProduction(L, StructureLegere) : TauxSoutirage(L, StructureLegere, 200000)$$
$$PremierPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Faible) : Cout(L, 100)$$
$$PremierPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Important) : Cout(L, 200)$$
$$PremierPuits(L), Hydrocarbures(L), Quantite(L, Important) : DeuxPuits(L)$$
$$SecondPuits(L), Hydrocarbures(L), Quantite(L, Importante) : MiseEnProduction(L, StructureLourde)$$
$$MiseEnProduction(L, StructureLourde) : TauxSoutirage(L, StructureLourde, 500000)$$
$$MiseEnProduction(L, StructureLourde), Quantite(L, Important) : Cout(L, 310)$$
$$SecondPuits(L), Hydrocarbures(L), Quantite(L, Faible) : MiseEnProduction(L, StructureLegere)$$
$$SecondPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Faible) : Cout(L, 10)$$
$$PremierPuits(l_1), Hydrocarbures(l_1)$$

$$(PremierPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000)) \succ (SecondPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000))$$

$$(PremierPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Faible), Cout(L, 100)) \succ (SecondPuits(L) \wedge TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Importante), Cout(L, 200))$$

On suppose toujours que la fonction de similarité est donnée par une unification partielle, qui de plus ne permet pas de négliger les prédicats $MiseEnProduction, SecondPuits$. On en déduit alors que l'on peut mettre le puits en production avec une structure légère, directement sans second forage, ou le mettre en production avec une structure lourde après un second forage.

Si on rajoute de plus la préférence suivante:

$$(SecondPuits(L), Quantite(L, Importante), TauxSoutirage(L, StructureLegere, 500000)) \succ (PremierPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000))$$

Alors on choisit de mettre le puits en production avec une structure lourde après un second forage.

6.3 La sémantique de point fixe

Le but de cette section est de définir une fonction implantable sur machine permettant de "calculer" l'ensemble des conséquences d'une théorie préférentielle particulière, dans le sens où l'on aurait restreint l'ensemble des formules permettant de former ladite théorie. Pour ce faire, on établit un certain nombre de propriétés formelles qui font l'objet de la section suivante.

6.3.1 Clauses de Horn, préférences de Horn

Définition 78. *(Rappel).* Une formule est dite clausale si elle est de la forme $\forall\vec{x}\varphi_0$, où φ_0 est une disjonction de littéraux [22]. La fermeture universelle d'une disjonction de littéraux, dont au plus un est un atome, est appelée une clause de Horn.

Désormais, afin d'alléger les notations, on désignera par L_Σ l'ensemble des clauses de Horn formées sur la signature Σ. On distingue usuellement trois formes de clauses de Horn:

$$\forall\vec{x}A \text{ (clause unité ou fait)}$$
$$\forall\vec{x}(A \vee \neg A_1 \vee \ldots \vee \neg A_p) \text{ (clause définie, ou règle)}$$
$$\forall\vec{x}(\neg L_1 \vee \ldots \vee \neg L_q) \text{ (clause négative ou but)}$$

Bien entendu, on considère également chaque clause comme un séquent $A_1, \ldots, A_p : A$ qui a la même signification déclarative. Dorénavant, on notera, si $\mathcal{M} \in WD_L$, $B(\mathcal{M})$ l'ensemble des clauses unités validées par $\mathcal{M}$. Enfin, si C est la clause définie $A : - A_1 \ldots A_p$, on note C^+ l'atome A (la tête de C) et C^- la suite de littéraux $(A_1, \ldots, A_p)$ (le corps de C).

Dans le cadre restreint des clauses de Horn, l'idée de raisonnement que nous avons proposé s'appréhende en supposant que le corps de chaque clause représente des caractéristiques d'un problème mémorisé. Une conjonction de faits décrit des caractéristiques du problème courant à résoudre et le raisonnement par similarité consiste à rechercher des subsomptions entre le corps d'une clause définie et cette conjonction de faits. Ces subsomptions fournissent alors le corps d'une règle dont la tête est constituée de la tête de la clause définie, que l'on adapte au besoin, pour prendre en compte la subsomption effectuée sur le corps de la clause.

On restreint la classe des formules utilisées pour définir une relation de préférence de la manière suivante:

Définition 79. *(Préférences de Horn)*
Une préférence $P \in P_\Sigma[X]$ est dite de Horn si et seulement si P est une préférence universellement quantifiée, et si $P = (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$, les formules $(\varphi_1, \varphi_2, \psi_1, \psi_2, \phi_1, \phi_2)$ étant des conjonctions de clauses unités.

Dorénavant, on utilisera dans les définitions les notations $L_\Sigma[X]$ pour désigner l'ensemble des clauses de Horn sur la signature Σ, $P_\Sigma[X]$ pour désigner l'ensemble des formules préférentielles de Horn sur la signature Σ.

6.3.2 Algèbres de Herbrand étendues

La valuation par similarité d'une formule par une algèbre de Herbrand sachant des négligeabilités est une adaptation directe de la définition donnée dans le cas des mondes (voir définition 68). Une algèbre de Herbrand étendue va donc être déterminée par un ensemble d'atomes vrais et un ensemble d'atomes incertains. Plus précisément, on associe à tout monde $\mathcal{M}$ une algèbre de Herbrand similaire déterminée par l'ensemble des atomes fermés $\mathcal{H}_1, \mathcal{H}_2$ tels que $\forall\alpha$, $\begin{array}{ll} \alpha \in \mathcal{H}_1, & V_i^{\mathcal{M}}(\alpha) = 1 \\ \alpha \in \mathcal{H}_2, & V_i^{\mathcal{M}}(\alpha) = U \end{array}$
En effet, l'inclusion entre mondes signifie que, ou bien l'un des deux mondes valide moins que l'autre, ou bien que moins de formules sont incertaines dans l'un que dans l'autre. L'équivalence entre satisfiabilité par un monde et par une algèbre de Herbrand étendue permet alors de conclure.

Alors, par définition de V_s (voir définition 68), une formule C est vraie sachant les négligeables Δ dans un monde $\mathcal{M}$ si et seulement si elle est vraie dans l'algèbre de Herbrand étendue sachant les mêmes négligeables Δ (ceci résulte du théorème de Herbrand et du fait que si $S(\mathcal{H}_2)$ désigne l'ensemble des éléments incertains vrais par similarité, alors $\mathcal{H}_1 \cup S(\mathcal{H}_2)$ satisfait de manière classique C).

Pour définir une sémantique de point fixe, nous allons avoir besoin de donner une équivalence entre satisfaction par similarité avec des préférences dans un monde et satisfaction par similarité avec des préférences dans une algèbre de Herbrand étendue. Avant de donner cette équivalence, la sémantique de

22. On rappelle qu'un littéral est un atome (littéral positif) ou la négation d'un atome (littéral négatif)

point fixe exige d'imposer une propriété supplémentaire de monotonie à cette fonction, condition que l'on présente ici sur des algèbres de Herbrand étendues, mais qui peut tout aussi bien se poser sur les mondes:

Définition 80. (*Fonction de similarité*)

On définit une fonction $\begin{array}{llll} Sim: & L_\Sigma \times (\mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma)) & \longmapsto & \mathcal{P}(L_\Sigma \times At_\Sigma) \\ & (C, (\mathcal{H}_1, \mathcal{H}_2)) & \longmapsto & \{(C'_k, \Delta_k)_{k\in K}\} \end{array}$

définissant l'ensemble des clauses C'_k similaires à une clause donnée C dans le contexte déterminé par l'algèbre de Herbrand étendue $(\mathcal{H}_1, \mathcal{H}_2)$, fonction qui satisfait les propriétés suivantes :

1. **Calcul des éléments similaires**
   Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X]$
   $$\forall (C', \Delta') \in Sim(C, (\mathcal{H}_1, \mathcal{H}_2)), \left\{ \begin{array}{l} \bullet\ C'^- \text{ subsume } B \text{ et } C^-, \\ \bullet\ C'^+ \text{ subsume } C^+ \end{array} \right.$$
   où B est la conjonction des clauses unités validées par $\mathcal{H}_1$.

   Intuitivement, cette définition signifie que les prémisses de la clause C' sont plus générales à la fois que ce qui est connu dans l'algèbre de Herbrand étendue et dans la clause C, et que ses conclusions sont également plus générales que celles de C.

2. **Calcul des éléments dissimilaires**
   Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X]$
   $$\forall (C', \Delta') \in Sim(C, (\mathcal{H}_1, \mathcal{H}_2)), \left\{ \begin{array}{l} \bullet\ C^- \models_L \Delta' \text{ et } C'^- \not\models_L \Delta' \\ \bullet\ \forall A \in C^-, C'^- \not\models_L A \Rightarrow \exists A' \in \Delta' A \models_L A' \\ \bullet\ \forall A' \in \Delta', \exists A \in C^-, C'^- \not\models_L A \text{ et } A \models_L A' \end{array} \right.$$

   Intuitivement, cette condition signifie que ce qui a été négligé ne se déduit pas du nouveau séquent (première condition), que ce qui a été négligé se déduit de l'ancien séquent (seconde condition), et que ce sont les <u>seuls</u> éléments qui sont négligés (troisième condition).

3. **Condition aux limites**
   Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X]$
   $$\text{Si } \mathcal{H}_1 \models_{h,L} C, \text{alors } Sim(C, (\mathcal{H}_1, \mathcal{H}_2)) = \{(C, \emptyset)\}$$

   Lorsque la clause est satisfaite par l'algèbre de Herbrand étendue, sa seule formule similaire est elle-même.

4. **Condition de monotonie**
   Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), (\mathcal{H}'_1, \mathcal{H}'_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma$.
   Si $\mathcal{H}'_1 \subseteq \mathcal{H}_1$ alors $(C'', \Delta'') \in Sim(C, (\mathcal{H}'_1, \mathcal{H}'_2))$
   $$\Rightarrow \exists (C', \Delta') \in Sim(C, (\mathcal{H}_1, \mathcal{H}_2)), \left\{ \begin{array}{l} \bullet\ C''^+ \text{ subsume } C'^+ \\ \bullet\ C''^- \text{ subsume } C'^- \\ \bullet\ \Delta' \text{ subsume } \Delta'' \end{array} \right.$$

   Intuitivement, cette définition signifie que moins l'on connaît, plus l'on doit négliger et plus la clause similaire est "générale".

5. **Incertitude des éléments négligés**
   Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X]$
   $$\forall (C', \Delta') \in Sim(C, (\mathcal{H}_1, \mathcal{H}_2)), \left\{ \begin{array}{l} \bullet\ \forall \varphi \in \Delta', V_i^{(\mathcal{H}_1, \mathcal{H}_2)}(\varphi) = U \\ \bullet\ \forall \varphi \in C^-, C'^- \not\models_L \varphi \Rightarrow V_i^{(\mathcal{H}_1, \mathcal{H}_2)}(\varphi) = U \end{array} \right.$$

La seule difficulté pour reposer la définition 66 et la définition 67 (validation d'une préférence) en utilisant des algèbres de Herbrand étendues en lieu et place des mondes consiste à donner l'équivalent sur les algèbres de Herbrand de la relation $\models_{\subset, neg/\Gamma}$:

$$\mathcal{M} \models_{\subset, neg/\Gamma} S \Leftrightarrow \mathcal{M} \models_{neg/\Gamma} S \text{ et } \neg(\exists \mathcal{M}', \mathcal{M}' \models_{neg/\Gamma} S \text{ et } \mathcal{M}' \subset \mathcal{M})$$

Une algèbre de Herbrand étendue satisfaisant la proposition équivalente sera alors dit modèle minimal sachant des négligeabilités d'un programme P, ce qui est le cas ssi:

$$\begin{array}{ll} (\mathcal{H}_1, \mathcal{H}_2) \models_{\subset, neg/\Gamma} P & (\mathcal{H}_1, \mathcal{H}_2) \models_{neg/\Gamma} P \text{ et} \\ & \neg(\exists (\mathcal{H}'_1, \mathcal{H}'_2), (\mathcal{H}'_1, \mathcal{H}'_2) \models_{neg/\Gamma} P \text{ et } \mathcal{H}'_1 \subseteq \mathcal{H}_1) \\ & \text{et } \neg(\exists (\mathcal{H}'_1, \mathcal{H}'_2), (\mathcal{H}'_1, \mathcal{H}'_2) \models_{neg/\Gamma} P \text{ et } \mathcal{H}'_1 = \mathcal{H}_1 \text{ et } \mathcal{H}'_2 \subseteq \mathcal{H}_2) \end{array}$$

On peut alors définir:

Définition 81. (*Modèle épistémique similaire de Herbrand*)
Un modèle *épistémique similaire de Herbrand* est formé par un triplet $(At_\Sigma \times At_\Sigma, Sim, \models_S)$ où $\models_S \subseteq (\mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma)) \times \mathcal{P}(L_\Sigma[X]) \times LP_\Sigma[X]$ est définie par:

1. Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X], \Gamma \subseteq S_\Sigma[X], P \in P_\Sigma[X]$:

$$(\mathcal{H}_1, \mathcal{H}_2) \models_{\mathcal{S}/\Gamma} (P, C) \Leftrightarrow (\mathcal{H}_1, \mathcal{H}_2) \in AdmissNeg_{/\Gamma}(P, C)$$

2. Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), \Gamma \subseteq L_\Sigma[X], \Pi \subseteq P_\Sigma[X]$:

$$(\mathcal{H}_1, \mathcal{H}_2) \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Leftrightarrow \forall C \in \Gamma, \forall P \in \Pi, (\mathcal{H}_1, \mathcal{H}_2) \models_{\mathcal{S}/\Gamma} (C, P)$$

3. Soit $\Gamma \subseteq L_\Sigma[X], \Pi \subseteq P_\Sigma[X], C \in L_\Sigma[X]$:

$$(\Gamma, \Pi) \models_{\mathcal{S}/\Gamma} C \Leftrightarrow \forall (\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), (\mathcal{H}_1, \mathcal{H}_2) \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Rightarrow \mathcal{H}_1 \models_L C$$

Théorème 82. *Soit* $(WD_L, Sim, \models_S)$ *un modèle épistémique similaire et* $(At_\Sigma \times At_\Sigma, Sim, \models_S)$ *un modèle épistémique similaire de Herbrand. Alors les formules universelles préférentielles valides dans le modèle épistémique similaire sont valides dans le modèle épistémique similaire de Herbrand associé (et réciproquement). Les deux modèles seront alors dits associés.*

Preuve. Immédiat puisque, une formule universelle admet un modèle sachant des négligeabilités si et seulement si elle admet un modèle sachant des négligeabilités de Herbrand, et que les clauses de Horn sont des formules universelles et les formules préférentielles formées à partir de clauses unités. ∎

6.3.3 Plus petit modèle de Herbrand similaire

On se concentrera donc d'abord sur le cas du (des) plus petit(s) modèle (s) de Herbrand étendu d'un programme sans préférences.

6.3.3.1 Interactions entre négligeabilités et vérité

Ce qui pose problème dans la définition et le calcul d'un plus petit modèle de Herbrand étendu similaire d'un programme, ce sont les interactions entre négligeabilité et vérité. Prenons le programme (instantié) suivant pour illustrer notre propos:

$$\begin{aligned} C_1 &\equiv A(c) \\ C_2 &\equiv A(c) \wedge B(c) \Rightarrow C(c) \\ C_3 &\equiv C(c) \Rightarrow B(c) \end{aligned}$$

Pour illustrer notre propos, on utilisera comme fonction de similarité ce qui se réduit informellement à de l'unification partielle. *Dans notre cas, il faut pouvoir prendre en compte que $B(c)$ est négligeable dans la deuxième règle pour une première instantiation, et le reste pour cette instantiation, même après application de la troisième règle qui fournit une autre instantiation complète de la deuxième*, le plus petit modèle de Herbrand étendu modèle par similarité de ce programme étant $(\{A(c), C(c)\}, \{B(c)\})$.

Avant de voir comment prendre formellement ce problème, donnons un second exemple moins "linéaire":

$C_1 \equiv \quad A(c)$
$C_2 \equiv \quad A(c) \wedge B(c) \Rightarrow C(c) \quad C_4 \quad \equiv A(c) \wedge D(c) \Rightarrow E(c)$
$C_3 \equiv \quad C(c) \Rightarrow T(c) \quad C_5 \quad \equiv E(c) \Rightarrow F(c)$
$C_6 \equiv \quad T(c) \wedge F(c) \Rightarrow B(c)$

Ce programme a un plus petit modèle de Herbrand étendu qui est $(\{A(c), C(c), E(c), F(c), T(c)\}, \{B(c), D(c)\})$. Si, par contre, on substitue à la règle C_6 la règle $C_7 \equiv F(c) \Rightarrow B(c)$, on obtient alors comme plus petit modèle de Herbrand étendu $(\{A(c), C(c), E(c), F(c), T(c), B(c)\}, \{D(c)\})$. Il nous font donc pouvoir prendre en compte les dépendances entre atomes vrais et négligés, pour pouvoir appréhender le cas où un atome a été déduit de sa propre négligeabilité.

6.3.3.2 Modèles similaires stables

On considère donc dorénavant un programme défini P, et on s'intéresse aux conséquences de ce programme par similarité. On suppose de plus que chaque règle de P a été remplacée par toutes ses instantiations, de telle sorte que tous les atomes de P sont liés.

Les modèles similaires stables que nous présentons sont des adaptations des modèles stables de [Gelfond and Lifschitz 1988]. Pour chaque ensemble d'atomes (M_1, M_2) de P, on définit par P_{M_1} le programme obtenu à partir d'une fonction $Sim(P, (M_1, \complement M_1))$, où $\complement M_1$ désigne l'ensemble des atomes de P non contenus dans M_1 [23], en effaçant tous les couples $(C', \Delta') \in Sim(P, (M_1, \complement M_1))$ tels qu'il existe une suite de couples $(C_i, \Delta_i)_{1 \leq i \leq n}$ de $Sim(P, (M_1, \complement M_1))$ avec $(C_n, \Delta_n) \equiv (C', \Delta')$ telle que:

(1) $\forall 1 \leq i \leq n, \forall A \in C'^{-}_i, A \in M_1$ et $C'^{+}_i \in M_1, C'^{+}_i \in C'^{-}_{i+1}$,

(2) $C^{+}_n \in \Delta_1$

On considère alors le programme constitué par l'ensemble des règles C' (telles que $(C', \Delta') \in Sim(P, (M_1, \complement M_1))$) restantes. C'est alors un programme sans similarité. Il a donc un plus petit modèle de Herbrand. Si ce modèle $\mathcal{H}_1$ est tel que $(\mathcal{H}_1, \mathcal{H}_2)$ (où $\mathcal{H}_2$ est constitué de l'ensemble des atomes appartenant aux Δ') coïncide avec (M_1, M_2), on dit que (M_1, M_2) est un modèle similaire stable de P. De tels ensembles peuvent également être décrits comme étant les points fixes de l'opérateur T_P défini par la condition suivante:

Définition 83. Soit M_1, M_2 un ensemble d'atomes de P, et soit $P_{(M_1, M_2)}$ le programme à partir d'une fonction $Sim(P, \complement M_1))$ et de Π par les conditions (1) et (2). On définit alors l'opérateur:

$$\begin{array}{llll} T_{(P,\Pi)}: & \mathcal{P}(At_\Sigma[X]) \times \mathcal{P}(At_\Sigma[X]) & \longmapsto & \mathcal{P}(At_\Sigma[X]) \times \mathcal{P}(At_\Sigma[X]) \\ & (M_1, M_2) & \longmapsto & (M'_1, M'_2) \end{array}$$

tel que M'_1 est le plus petit modèle de Herbrand du programme obtenu à partir de P en appliquant les conditions (1) et (2), M_2 représentant les éléments négligés pour obtenir cette similarité, c'est à dire que M_2 est constitué de l'ensemble $\{\Delta' / (C', \Delta') \in Sim(C, (\mathcal{H}_1, \complement \mathcal{H}_1))\}$.

Ces ensembles similaires stables ont de plus la propriété suivante:

Théorème 84. *Tout ensemble similaire stable de P est modèle par similarité minimal de Herbrand étendu de P.*

Avant de donner la preuve de ce théorème, effectuons la remarque suivante:

Remarque 85. Soit $\mathcal{M} \in WD_L, C \equiv A :- A_1, \ldots, A_n$ une clause définie, soit Γ un ensemble de clauses définies et soit $C' \equiv A' :- A'_1, \ldots, A'_m$ une clause définie similaire à C pour le monde $\mathcal{M}$ et non syntaxiquement équivalente à C, i.e. $\exists \Delta' \neq \emptyset, (C', \Delta') \in Sim(C, \mathcal{M})$. Alors:

$$\mathcal{M} \models_{\mathcal{S}/\Gamma} C \Rightarrow V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(A') = 1.$$

23. Cette considération est expliquée intuitivement par le fait que dans cette logique, on ne connaît initialement pas les éléments négligeables, comme ce serait le cas dans une logique des défauts par exemple. Ici, au contraire, tout fait non connu est potentiellement négligeable.

Preuve. Par définition, $\mathcal{M} \models_{\mathcal{S}/\Gamma} C \Rightarrow V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(C') = 1$, Or C'^{-} subsume à la fois C^{-} et l'ensemble des clauses unités validées par chaque algèbre de $\mathcal{M}$. Donc $\forall i, 1 \leq i \leq m, V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(A'_i) = 1$. Comme $V^{\mathcal{M}}_{\mathcal{S}/\Gamma}(C') = 1$, $V_{\mathcal{S}/\Gamma}(A')^{\mathcal{M}} = 1$. Ce qui achève la démonstration. ■

Montrons maintenant le théorème 84:

Preuve. Considérons un ensemble similaire stable (M_1, M_2). Montrons tout d'abord qu'il est modèle similaire de P.

Soit $R \in P$. Si une règle R a été effacée par la condition du théorème, c'est que cette règle est vraie puisqu'un élément négligé valide sa tête. En ce qui concerne les autres règles non effacées, M_1 est le plus petit modèle de Herbrand des règles qui leur sont similaires, et M_2 représente les négligeabilités nécessaires pour obtenir cette similarité. (M_1, M_2) est donc bien modèle par similarité de P.

Pour maintenant montrer qu'il est minimal, nous allons raisonner par l'absurde et considérer qu'il existe un modèle plus petit que (M_1, M_2). Soit (M'_1, M'_2) le modèle plus petit que (M_1, M_2). Alors nécessairement $M'_1 \subseteq M_1$ et $M'_2 \subseteq M_2$. Commençons d'abord par montrer que ces deux modèles valident par similarité les mêmes atomes, grâce à la remarque 85 précédente. En effet,

- les règles similaires non syntaxiquement équivalente à leur règle originelle dans l'un et l'autre de ces modèles ont la même tête de clause (on a instantié toutes les règles et la fonction de similarité est monotone) valide,

- si l'on suppose qu'une règle C du programme qui admet une règle similaire (C', Δ') dans M_1 n'en admet pas dans M'_1, il s'ensuit alors que M'_1 valide la négation de l'une des prémisses de C. Or, par la définition de la relation de satisfaction par similarité (voir définition 72), (M'_1, M'_2) n'est alors pas modèle par similarité de P, ce qui est absurde.

En ce qui concerne les règles similaires, ils valident donc les mêmes atomes. En ce qui concerne les autres règles, ils sont tous les deux modèles de Herbrand minimaux en ordre 0 de ces règles (sachant qu'ils valident en plus les éléments similaires): ils valident donc là aussi par similarité les mêmes atomes.

Montrons maintenant que ces deux modèles sont forcément identiques. Soit donc $A \in M_1 \setminus M'_1$, $A \in M'_2 \setminus M_2$. Comme A est vrai parce qu'incertain dans M'_2, il existe donc une clause de P telle que $(C', \Delta') \in Sim(C, (M'_1, M'_2))$ et $A \in \Delta'$. Si cet élément n'appartient à aucune tête de clause de C, alors il est absurde que M_1 soit modèle minimal du programme similaire à P puisque $A \in M_1$. Cet élément A appartient donc à une tête d'une clause de P:

- Si cette clause est une clause unité, alors A doit être classiquement vrai par définition de la relation de validation par similarité, donc A doit appartenir à M'_1, ce qui est absurde.

- Si cette clause est une clause définie:

  - et qu'elle ne satisfait pas la propriété du théorème, alors le plus petit modèle de Herbrand M'_1 du programme similaire à P dans M'_1 contient A, ce qui est absurde.

  - si elle satisfait la propriété du théorème, alors le plus petit modèle de Herbrand M_1 du programme similaire à P dans M_1 ne contient pas A, ce qui est également absurde.

Un ensemble similaire stable est donc bien un modèle similaire. ■

Notons le caractère intuitif de ces propriétés: en effet, mis à part la condition d'effacement qui permet d'éviter l'interdépendance entre un élément négligeable et un élément vrai, il n'y a dans cette démonstration que la "préférence" entre valider la négation d'un élément en prémisses et valider la tête de la formule. Il ne nous reste donc plus qu'à généraliser cette proposition au cas où le programme contient des préférences.

6.3.4 Modèles préférentiels stables

On considère dorénavant un programme défini préférentiel (P, Π) l'ensemble des conséquences de ce programme sachant les connaissances que P détermine. On suppose de plus que chaque règle de P et chaque préférence de Π a été remplacée par toutes ses instantiations, de telle sorte que tous les atomes de P et Π sont liés.

Les modèles préférentiels stables que nous présentons maintenant sont des adaptations de ceux de [Gelfond and Lifschitz 1988]. Pour chaque ensemble d'atomes (M_1, M_2) de P, on définit par $P_{(M_1,M_2)}$ le programme obtenu à partir d'une fonction $Sim(P, \complement\, M_1))$ et de Π:

(1) en effaçant chaque règle C tel qu'il existe un couple (C', Δ') de $Sim(C, (M_1, \complement\, M_1))$ et un couple $(C", \Delta")$ de P issu d'une autre règle, une préférence $p = (a_1, d_1, f_1) \succ (a_2, d_2, f_2)$ tel que $(a_1, f_1, a_2, f_2) \in M_1$ et $d_2 \subseteq \Delta'$ et $d_1 \subseteq \Delta", \Delta' \neq \Delta''$

(2) en effaçant tous les couples $(C', \Delta') \in Sim(P, (M_1, \complement\, M_1))$ tels qu'il existe une suite de couples $(C_i, \Delta_i)_{1 \leq i \leq n}$ de $Sim(P, (M_1, \complement\, M_1))$ avec $(C_n, \Delta_n) \equiv (C', \Delta')$ telle que:

- $\forall 1 \leq i \leq n, \forall A \in C'^{+}_{i}, A \in M_1$ et $C'^{+}_{i} \in M_1, C'^{+}_{i} \in C'^{-}_{i+1}$,
- $C_n^+ \in \Delta_1$

On considère alors le programme constitué par l'ensemble des règles C' non effacées. C'est un programme sans similarité, ni préférences. Il a donc un plus petit modèle de Herbrand $\mathcal{H}_1$, déterminant un ensemble de négligeables $\mathcal{H}_2$ par $\{\Delta'/(C', \Delta') \in Sim(C, (\mathcal{H}_1, \complement\, \mathcal{H}_1))$. Si ce modèle $(\mathcal{H}_1, \mathcal{H}_2)$ coïncide avec M_1, M_2, à l'instar de [Gelfond and Lifschitz 1988], on dit que (M_1, M_2) est un *ensemble préférentiel stable* de (P, Π). De tels ensembles peuvent également être décrits comme étant les points fixes de l'opérateur $T_{(P,\Pi)}$ défini par la condition suivante:

Définition 86. Soit M_1, M_2 un ensemble d'atomes de P, et soit $P_{(M_1,M_2)}$ le programme à partir d'une fonction $Sim(P, \complement\, M_1))$ et de Π par les conditions (1) et (2). On définit alors l'opérateur:

$$\begin{array}{llll} T_{(P,\Pi)}: & \mathcal{P}(At_\Sigma[X]) \times \mathcal{P}(At_\Sigma[X]) & \longmapsto & \mathcal{P}(At_\Sigma[X]) \times \mathcal{P}(At_\Sigma[X]) \\ & (M_1, M_2) & \longmapsto & (M'_1, M'_2) \end{array}$$

tel que M'_1 est le modèle de Herbrand minimal de $P_{(M_1,M_2)}$, M'_2 l'ensemble des éléments négligés pour obtenir cette similarité, i.e. M_2 est l'ensemble $\{\Delta'/(C', \Delta') \in Sim(C, (\mathcal{H}_1, \complement\, \mathcal{H}_1))\}$.

Ces ensembles préférentiels stables ont de plus la propriété suivante:

Théorème 87. *Tout ensemble préférentiel stable de* (P, Π) *est un modèle de Herbrand minimal similaire de* (P, Π).

Preuve. Considérons un ensemble préférentiel stable (M_1, M_2). Montrons tout d'abord que (M_1, M_2) est modèle par similarité de (P, Π).
Soit R une règle de P. Si la similarité appliquée à R détermine un ensemble de négligeables δ tel qu'il existe une autre règle R' de P déterminant un négligeable δ', et qu'il existe une préférence de Π $(a_1, d_1, f_1) \succ (a_2, d_2, f_2)$ telle que $d_1 \subseteq \delta'$ et $d_2 \subseteq \delta$, et $(a_1, f_1, a_2, f_2) \in M_1$ alors R est vraie dans M_1 puisque M_1 valide $\neg\delta$ [24]. On peut donc effacer la règle R. Les règles similaires non effacées sont quant à elles vraies dans M_1 puisque M_1 est leur plus petit modèle de Herbrand. (M_1, M_2) est donc bien un modèle préférentiel similaire de (P, Π), où M_2 est l'ensemble des similarités déterminées par M_1.

Pour montrer que (M_1, M_2) est minimal, considérons maintenant un ensemble plus petit (M'_1, M'_2) modèle préférentiel similaire de (P, Π). Considérons une règle R de P. Par la propriété de monotonie de la fonction de similarité, si un couple (C', Δ') a été effacé dans M_1 à partir de la règle R par la première condition du théorème 87, alors:

- ou bien il existe un couple (C'', Δ'') dans M'_1 qui a été effacé lui aussi,

24. On retrouve là encore une conséquence de notre définition de la satisfaction par similarité (voir définition 72) qui est que la validation de la négation d'une prémisse est "préférable" à la validation d'une tête de clause. Ici, une telle préférence n'est pas déclenchable pour R alors qu'elle l'est pour R'.

- ou bien un tel couple n'existe pas, et la règle est vraie dans M'_1 parce que l'une de ses prémisses est fausse, ce qui d'un point de vue atomes validés revient au même que si la règle avait été effacée.

Les deux modèles (M_1, M_2) et (M'_1, M'_2) sont donc des modèles similaires minimaux du même programme (auquel on a enlevé les règles ayant trait aux préférences). Ils sont donc identiques par le théorème 84. ■

Intuitivement, ces ensembles stables ont la même signification que à la fois les ensembles stables de [Moore 1985] et de [Gelfond and Lifschitz 1988]. Ce sont des ensembles de formules qu'un agent rationnel peut posséder, pour reprendre la formulation de [Moore 1985]. En effet, laissons de côté les interdépendances entre négligeabilité et vérité et concentrons nous sur l'effacement des règles liés aux préférences: il signifie intuitivement que que "*je peux retirer toute règle telle que négliger des éléments en prémisses me conduit à des résultats que je ne préfére pas*". Si les éléments déduits ne sont pas dominés pour les préférences, alors cet ensemble de formules correspond à des assertions qui sont rationnelles vis à vis de mes préférences.

Il s'ensuit donc que l'ensemble des conséquences d'un programme préférentiel peut être calculé par l'opérateur $T_{(P,\Pi)}$, puisqu'il s'agit de l'intersection de ses points fixes.

Proposition 88. *Pour tout programme préférentiel (P, Π), $T_{(P,\Pi)}$ est anti-monotone.*

Il s'ensuit dès lors que $T^2_{(P,\Pi)}$ est monotone. Par le théorème de Tarsky, il a donc un plus petit et un plus grand point fixe. Il s'ensuit alors:

Proposition 89. *([Lifschitz 1996])*
Soit T une fonction anti-monotone, et soit X_0, X_1 les plus grands et plus petits points fixes de T^2. Alors:

- $TX_0 = X_1, TX_1 = X_0$,
- *Pour tout point fixe X de T, $X_0 \subseteq X \subseteq X_1$*

Il s'ensuit bien entendu que, si X est le seul point fixe de T^2, alors X est le seul point fixe de T.

Si l'ensemble des atomes du programme est fini, il s'ensuit que le plus grand et le plus petit point fixe de $T_{(P,\Pi)}$ peuvent être calculés en itérant le calcul de $T_{(P,\Pi)}$ de la manière suivante:

- calculer les ensembles $T^n_{(P,\Pi)}(\emptyset)$ jusqu'à ce qu'il existe une valeur de n telle que $T^n_{(P,\Pi)}(\emptyset) = T^{n+2}_{(P,\Pi)}(\emptyset)$,
- selon la parité de n, l'un de ces deux ensembles est le plus petit point fixe de l'opérateur, l'autre le plus grand.

On a donc ainsi une approximation via cet opérateur des conséquences similaires du programme.

6.3.5 Liens avec le système Soar

On notera que les deux phases de calcul du point fixe de l'opérateur sont non sans analogies avec un système d'intelligence artificielle existant, en l'occurrence le système Soar [Laird et al. 1986] [Newell 1990] L'effacement de règles correspond à la phase de *décision* de Soar, et le calcul de la plus petite algèbre de Herbrand du programme ainsi obtenu correspond à la phase *d'élaboration* de Soar. Cela correspond alors en terme constructif intuitivement à:

1. A partir des faits initialement connus, déterminer quelles sont les règles similaires et appliquer les préférences: dans Soar, cela correspond à une phase de décision *augmentée* des similarités,
2. Déduire tout ce qui est possible en ordre 0 classique de ces algèbres: c'est dans Soar la phase d'élaboration,
3. Retourner à la phase 1.

Le raisonnement s'arrête lorsqu'on a atteint un point fixe. D'autre part, lors de la phase de décision, dans Soar, les règles qui permettent de déduire des éléments qui ne sont pas admissibles au sens des préférences sont *rétractées* [Congdon and Laird 1997], ce qui signifie que leurs conclusions sont retirées de la base de faits. L'effacement de règles correspond également au retrait des règles déduisant des éléments non admissibles au sens des préférences formelles que nous avons défini, et qui, nous l'avons vu, correspondent en partie aux préférences de Soar.

6.4 Conclusion

Nous avons donc établi tant la définition logique que le mode d'implémentation logique du raisonnement par similarité version satisfaction préférentielle forte. Cette logique permet de combiner similarité et préférence, et satisfait comme nous l'avons montré les bonnes propriétés que nous avions au chapitre 4. Cependant, dans le cas où les préférences sont elles aussi incomplètes (et si les connaissances sont incomplètes, pour quelle raison les préférences auraient elles vertu de complétude?), cette logique peut amener à ne rien déduire du tout, puisqu'elle ne permet pas de gérer des hypothèses, comme le montre l'exemple suivant:

$$
\begin{aligned}
&C_1 \equiv Piste(l_1), Taxiway(l_1), TourControle(l_1), Aerogare(l_1) : AeroportCivil(l_1)\\
&C_2 \equiv Piste(l_1), Taxiway(l_1), TourControle(l_1), StationnementNonAligne(l_1)\\
&\qquad\qquad\qquad : AeroportMilitaire(l_1)\\
&C_3 \equiv Cloture(l_1), FormesCirculaires(l_1), BassinRetention(l_1), ReseauRoutier(l_1, l_2),\\
&\qquad\qquad\qquad AeroportCivil(l_2) : SiteCarburant(l_1)\\
&C_4 \equiv Cloture(l_1), FormesCirculaires(l_1), Gardiennage(l_1), ReseauRoutier(l_1, l_2),\\
&\qquad\qquad\qquad AeroportMilitaire(l_2) : SiteMunition(l_1)\\
&C_5 \equiv SiteMunition(l_1), SiteCarburant(l_1) : \perp\\
&C_6 \equiv AeroportMilitaire(l_1), AeroportCivil(l_1) : \perp\\
&Piste(l), Taxiway(l), TourControle(l)\\
&Cloture(l'), FormesCirculaires(l'), ReseauRoutier(l, l')
\end{aligned}
$$

On considère toujours dans l'exemple que la fonction de similarité est une unification partielle, et on considère toujours que la similarité est déterminée par les connaissances de cette théorie. Alors, par similarité, on ne peut rien déduire du tout de cette théorie, puisque la théorie contient un principe d'exclusion empêchant qu'on ne puisse déduire que deux objets se trouvent au même endroit. Ce qui manque donc dans cette logique vis à vis d'un raisonnement par similarité en fonction de préférences dans le cas où celles-ci sont incomplètes, c'est la possibilité de générer des hypothèses de similarité en fonction des négligeabilités, c'est à dire de pouvoir déduire que $AeroportCivil(l), SiteCarburant(l')$ et $AeroportCivil(l), SiteMunition(l')$ sont des conclusions du programme sachant les négligeabilités respectives $Aerogare(l), BassinRetention(l')$, et $StationnementNonAligne(l), Gardiennage(l')$.

7

Similarité - Satisfaction préférentielle faible

Ce chapitre est consacré à poser la définition de la relation de déduction par similarité en supposant que le décideur va faire des hypothèses exclusives les unes des autres d'éléments négligeables. Pour ce faire, nous commençerons donc par poser la définition de la similarité, majoritairement calquée sur celle du chapitre précédent. Puis, nous poserons la définition de la relation de déduction par similarité proprement dîte en prenant en compte les préférences. Nous terminerons en donnant une sémantique de point fixe parallèle permettant l'implantation dans un système d'intelligence artificielle de la logique ainsi définie.

Pour illustrer les intuitions de ce raisonnement, commençons par reprendre l'exemple de la page précédente:

$$
\begin{aligned}
&C_1 \equiv Piste(l_1), Taxiway(l_1), TourControle(l_1), Aerogare(l_1) : AeroportCivil(l_1)\\
&C_2 \equiv Piste(l_1), Taxiway(l_1), TourControle(l_1), StationnementNonAligne(l_1)\\
&\qquad\qquad : AeroportMilitaire(l_1)\\
&C_3 \equiv Cloture(l_1), FormesCirculaires(l_1), BassinRetention(l_1), ReseauRoutier(l_1, l_2),\\
&\qquad\qquad AeroportCivil(l_2) : SiteCarburant(l_1)\\
&C_4 \equiv Cloture(l_1), FormesCirculaires(l_1), Gardiennage(l_1), ReseauRoutier(l_1, l_2),\\
&\qquad\qquad AeroportMilitaire(l_2) : SiteMunition(l_1)\\
&C_5 \equiv SiteMunition(l_1), SiteCarburant(l_1) : \ \perp\\
&C_6 \equiv AeroportMilitaire(l_1), AeroportCivil(l_1) : \ \perp\\
&Piste(l), Taxiway(l), TourControle(l)\\
&Cloture(l'), FormesCirculaires(l'), ReseauRoutier(l, l')
\end{aligned}
$$

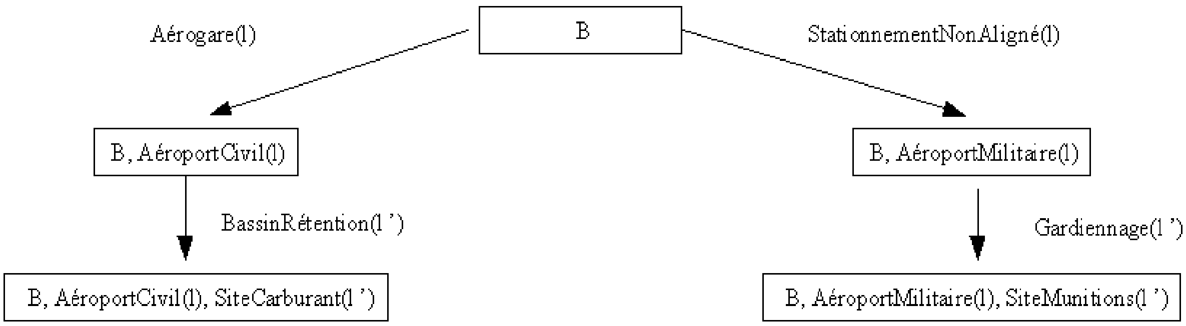


FIG. 7.1: *Le raisonnement par similarité conditionnellement à une hypothèse de négligeabilité*

Le schéma de la figure 7.1 illustre le raisonnement du décideur. N'ayant pas de préférences entre aéroport civil et aéroport militaire, et ces deux notions étant exclusives l'une de l'autre, le décideur pose deux hypothèses de négligeabilité, en l'occurrence $Aerogare(l)$ et $StationnementNonAligne(l')$. Il déduit alors qu'il existe un aéroport militaire et un aéroport civil, par deux "branches" de raisonnement différentes. Il peut alors continuer son raisonnement sur ces deux branches, chaque raisonnement étant conditionnel à une hypothèse de négligeabilité. Les préférences permettent de "couper' certaines branches de raisonnement. Ainsi, dans l'exemple suivant, si l'on rajoute la préférence suivante:

$$P_1 \equiv (AeroportCivil(l_2), BassinRetention(l_1), SiteCarburant(l_1)) \\ \succ \\ (AeroportMilitaire(l_2), Gardiennage(l_1), SiteMunition(l_1))$$

elle induit que le site à munition est préféré au site à carburant, ce qui a pour effet d'arrêter là le raisonnement conditionnellement à l'hypothèse $StationnementNonAligne(l')$. Il s'agit d'un raisonnement intrinsèquement parallèle, chaque hypothèse de négligeabilité étant explorée concurremment aux autres.

7.1 Définition de la notion de similarité

Nous allons maintenant définir sémantiquement un raisonnement par similarité version satisfiabilité faible. Nous allons l'appréhender comme étant une subsomption entre ce que le décideur connaît et ses connaissances représentées par des séquents. Si l'on se réfère au chapitre précédent, les deux modifications principales induites dans la formalisation par l'introduction des éléments négligeables à priori sont:

- en ce qui concerne la fonction de similarité, seuls les éléments appartenant à l'ensemble représentant l'hypothèse de négligeabilité peuvent être négligés,
- en ce qui concerne les préférences, leur application n'est plus déterminée que par l'ensemble des éléments valides dans un monde, les éléments négligés étant par avance connus.

7.1.1 Définition de la fonction de similarité

La théorie que nous proposons ici suppose que le décideur pose des hypothèses de négligeables avant de commencer son raisonnement, auquel cas il n'acceptera plus de raisonnements conduisant à en négliger d'autres que ceux-ci. Nous allons donc raisonner en considérant que le décideur fait des suppositions de négligeables représentés par des ensembles de formules que l'on notera $(D_i)_{1\leq i\leq n}$. On supposera de plus que $\forall\, 1 \leq i \leq n, \nexists\, 1 \leq j \leq n, D_i \subset D_j$. En un mot, les ensembles de négligeables sont supposés "exclusifs" les uns des autres. Dans la suite, on utilisera fréquemment un élément D pour représenter un négligeable parmi les $(D_i)_{1\leq i\leq n}$.

Si l'on compare la définition suivante de la fonction de similarité avec celle donnée au chapitre précédent (voir définition 67), c'est dans le calcul des éléments dissimilaires qu'intervient l'hypothèse de négligibilité, puisqu'elle restreint l'ensemble des éléments négligés à ceux qui sont déductibles de D:

Définition 90. (*Fonction de similarité*)

On définit une fonction $\begin{array}{llll} Sim: & \mathcal{P}(L_\Sigma) \times S_\Sigma \times WD_L & \longmapsto & \mathcal{P}(S_\Sigma \times L_\Sigma) \\ & (D, S, \mathcal{M}) & \longmapsto & \{(S'_k, \Delta_k)_{k\in K}\} \end{array}$

définissant l'ensemble des séquents S'_k similaires à un séquent donné S dans le contexte déterminé par le monde $\mathcal{M}$ sachant les négligeables déterminés par D, fonction qui satisfait les propriétés suivantes :

1. **Calcul des éléments similaires**
   Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

   $$\forall (S', \Delta') \in Sim(D, S, \mathcal{M}), \begin{cases} \bullet\ S'^- \text{ subsume } B \text{ et } S^-, \\ \bullet\ S'^+ \text{ subsume } S^+ \end{cases}$$

   où B est l'ensemble des séquents de la forme suivante $(: \varphi)$ validés par $\mathcal{M}$.

   Intuitivement, cette définition signifie que les prémisses du séquent S' sont plus générales à la fois que ce qui est connu dans le monde et dans le séquent S, et que ses conclusions sont également plus générales que celles de S.

2. **Calcul des éléments dissimilaires**
Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

$$\forall(S', \Delta') \in Sim(D,S,\mathcal{M}), \begin{cases} \bullet\ S^- \models_L \Delta' \text{ et } S'^- \not\models_L \Delta' \\ \bullet\ \forall\varphi \in S^-, S'^- \not\models_L \varphi \Rightarrow \exists\phi \in \Delta', \varphi \models_L \phi \\ \bullet\ \forall\varphi' \in \Delta', \exists\varphi \in S^-, S'^- \not\models_L \varphi \text{ et } \varphi \models_L \varphi' \\ \bullet\ \forall\varphi \in \Delta', D \models_L \varphi \end{cases}$$

Intuitivement, cette condition signifie que ce qui a été négligé ne se déduit pas du nouveau séquent (première condition), que ce qui a été négligé se déduit de l'ancien séquent (seconde condition), et que ce sont les seuls éléments qui sont négligés (troisième condition), et enfin, qu'à une spécialisation près, **tout ce qui a été négligé se déduit des éléments potentiellement négligeables**.

3. **Condition aux limites**
Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

$$\text{Si } \mathcal{M} \models_{h,L} S, \text{alors } Sim(D,S,\mathcal{M}) = \{(S,\emptyset)\}$$

Lorsque le monde satisfait le séquent, sa seule formule similaire est lui-même.

4. **Incertitude des éléments négligés**
Soit $\mathcal{M} \in WD_L, S \in S_\Sigma[X]$

$$\forall(S', \Delta') \in Sim(D,S,\mathcal{M}), \begin{cases} \bullet\ \forall\varphi \in \Delta', V_i^{\mathcal{M}}(\varphi) = U \\ \bullet\ \forall\varphi \in S^-, S'^- \not\models_L \varphi \Rightarrow V_i^{\mathcal{M}}(\varphi) = U \end{cases}$$

On étend la définition à un ensemble de séquents de la manière suivante:

$$\text{Pour } \Gamma \subseteq S_\Sigma[X], \mathcal{M} \in WD_L, Sim(D,\Gamma,\mathcal{M}) = \{Sim(D,S,\mathcal{M}), S \in \Gamma\}$$

La seule différence avec le chapitre précédent étant que la similarité est conduite sous une hypothèse particulière de négligibilité, il est possible de choisir comme fonction de similarité ce qui se réduit informellement à une unification partielle.

7.2 Relation de satisfaction par similarité

Dans cette section, nous commencerons par proposer la définition de la relation de conséquence sémantique par similarité. Après quelques exemples de déductions utilisant cette relation, nous discuterons ensuite des limites et possibilités du formalisme. Comme nous l'avons déjà signalé, la seconde différence entre ce mode de raisonnement et celui présenté à la section précédente tient aux faits que les préférences ne sont plus déterminées que par les éléments que les mondes valident, les éléments négligeables étant connus. On peut aisément reposer la définition de la satisfaction sachant des éléments négligeables (voir définition 68 du chapitre précédent), ce qui permet de reposer la définition de l'admissibilité dans le cas de la satisfiabilité faible de la manière suivante:

Définition 91. Soit WD_L l'ensemble des mondes sur la logique L, soit S un séquent de $S_\Sigma[X]$, soit Γ une théorie de $S_\Sigma[X]$ et soit $P = (\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ une préférence de $P_\Sigma[X]$. On appelle ensemble des admissibles pour S sachant les connaissances déterminées par Γ selon les préférences définies par P l'ensemble suivant, noté $Admiss_{/\Gamma}(S,P) \subseteq WD_L \times \mathcal{P}(L_\Sigma[X])$:

$$\begin{aligned} <\mathcal{M}, D> \in Admiss_{/\Gamma}(S,P) \quad & \Leftrightarrow <\mathcal{M}, D> \models_{\subset, neg/\Gamma\cup S} S \text{ et} \\ & \neg(\exists\mathcal{N} \in WD_L, \exists D' \in (D_i)_{1\leq i\leq n}, \\ & ((\mathcal{N}, D'), (\mathcal{M}, D)) \in Pref(P) \\ & \text{ et } <\mathcal{N}, D'> \models_{\subset, neg/\Gamma\cup S} \Gamma) \end{aligned}$$

où $\models_{\subset, neg/S} \subseteq \mathcal{P}(WD_L) \times \mathcal{P}(S_\Sigma[X]) \times L_\Sigma[X]) \times S_\Sigma[X]$ est la relation de déduction définie par:

$$<\mathcal{M}, D> \models_{\subset, neg/\Gamma} S \Leftrightarrow <\mathcal{M}, D> \models_{neg/\Gamma} S \text{ et } \neg(\exists\mathcal{M}', <\mathcal{M}', D> \models_{neg/\Gamma} S \text{ et } \mathcal{M}' \subset \mathcal{M})$$

L'ensemble des admissibles est défini par les mondes qui satisfont la formule et qui sont maximaux au sens des préférences.

Notons la différence majeure entre cette définition dans la cas de l'admissibilité dans le cas de la satisfiabilité faible et celle de la satisfiabilité forte: le décideur, ayant posé des hypothèses d'ensembles de négligeables, peut alors aboutir à des théories similaires différentes pour chaque ensemble de négligeables, et c'est entre ces théories différentes (qui ont donc des modèles différents) que s'exerce les préférences. De plus, ces préférences sont déterminées par les négligeables à priori D et D', et non pas par les éléments négligés calculés par la fonction de négligeabilité.

En adaptatant alors les définitions de la satisfaction par similarité (voir définition 73 et définition 74), on peut alors poser la définition d'un modèle épistémique similaire dans le cas de la satisfiabilité faible:

Définition 92. (*Modèle épistémique similaire*)
Un modèle *épistémique similaire* est formé par un quadruplet $(WD_L, Sim, (D_i)_{1\leq i\leq n}, \models_S)$ où $\models_S \subseteq (\mathcal{P}(WD_L) \times \mathcal{P}(L_\Sigma[X]) \times \mathcal{P}(S_\Sigma[X])) \times SP_\Sigma[X]$ est définie par:

1. Soit $\mathcal{M} \in WD_L, \Gamma \subseteq S_\Sigma[X], S \in S_\Sigma[X], P \in P_\Sigma[X], D \in (D_i)_{1\leq i\leq n}$:

$$< \mathcal{M}, D > \models_{\mathcal{S}/\Gamma} (P, S) \Leftrightarrow < \mathcal{M}, D > \in AdmissNeg_{/\Gamma}(P, S)$$

2. Soit $\mathcal{M} \in WD_L, \Gamma \subseteq S_\Sigma[X], \Pi \subseteq P_\Sigma[X], D \in (D_i)_{1\leq i\leq n}$:

$$< \mathcal{M}, D > \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Leftrightarrow \forall S \in \Gamma, \forall P \in \Pi, < \mathcal{M}, D > \models_{\mathcal{S}/\Gamma} (S, P)$$

3. Soit $\Gamma \subseteq S_\Sigma[X], \Pi \in P_\Sigma[X], S \in S_\Sigma[X], D \in (D_i)_{1\leq i\leq n}$,

$$< (\Gamma, \Pi), D > \models_{\mathcal{S}/\Gamma} S \Leftrightarrow \forall \mathcal{M} \in WD_L, < \mathcal{M}, D > \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Rightarrow \mathcal{M} \models_{h,L} S$$

Les conséquences d'une théorie sont donc les conséquences valides dans tous les mondes admissibles des théories qui lui sont similaires.

De la même manière que dans le cas de la satisfiabilité forte, la relation de déduction $< \mathcal{M}, D > \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi)$ sera notée $< \mathcal{M}, D > \models_{\mathcal{S}} (\Gamma, \Pi)$ une telle relation pour éviter les surcharges de notations.

7.2.1 Propriétés de la relation de déduction par similarité

A part le fait que l'on restreint ici la négligeabilité à un ensemble de négligeables, voir les propriétés de la section correspondante de la version à satisfiabilité faible.

7.2.2 Exemples de déduction par similarité

Nous considérerons dans cette section deux exemples de raisonnement par similarité en fonction de préférences, un exemple dans un problème de photo-interprétation et un exemple dans un problème de décisions dans le cas d'un forage pétrolier.

7.2.2.1 La photo-interprétation

Considérons la théorie préférentielle suivante:

$$
\begin{array}{rc}
C_1 \equiv & Piste(l_1), Taxiway(l_1), TourControle(l_1) : Aeroport(l_1) \\
C_2 \equiv & Aeroport(l_1), Aerogare(l_1) : AeroportCivil(l_1) \\
C_3 \equiv & Aeroport(l_1), StationnementNonAligne(l_1) : AeroportMilitaire(l_1)
\end{array}
$$

$$
\begin{array}{c}
P_1 \equiv (Piste(l_1), \bot, AeroportMilitaire(l_1)) \\
\succ \\
(Piste(l_1), \bot, AeroportCivil(l_1)) \\
B \equiv Piste(l_1), Taxiway(l_1), TourControle(l)
\end{array}
$$

On considère bien entendu que la fonction de similarité est donnée par une unification partielle, et on s'intéresse aux conséquences préférentielles de cette théorie sachant les connaissances qu'elle détermine. Selon l'ensemble des négligeables considérés, s'il est restreint à:

- $\{Aerogare(l\}$: on peut alors déduire $AeroportCivil(l)$,
- $\{StationnementNonAligne(l)\}$: on peut alors déduire $AeroportMilitaire(l)$,
- $\{\{Aerogare(l\}, \{StationnementNonAligne(l)\}\}$: par application de la préférence, on déduit $AeroportMilitaire(l)$ utilisant la négligeabilité $StationnementNonAligne(l)$,
- $\{Aerogare(l), StationnementNonAligne(l)\}$: par application de la préférence, on déduit $AeroportMilitaire(l)$ utilisant la négligeabilité $StationnementNonAligne(l)$.

7.2.2.2 Les décisions de forage pétrolier

Considérons la théorie préférentielle suivante:

$$
\begin{array}{l}
PremierPuits(L), Hydrocarbures(L), Quantite(L, Faible) : MiseEnProduction(L, StructureLegere) \\
MiseEnProduction(L, StructureLegere) : TauxSoutirage(L, StructureLegere, 200000) \\
PremierPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Faible) : Cout(L, 100) \\
PremierPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Important) : Cout(L, 200) \\
PremierPuits(L), Hydrocarbures(L), Quantite(L, Important) : DeuxPuits(L) \\
SecondPuits(L), Hydrocarbures(L), Quantite(L, Importante) : MiseEnProduction(L, StructureLourde) \\
MiseEnProduction(L, StructureLourde) : TauxSoutirage(L, StructureLourde, 500000) \\
MiseEnProduction(L, StructureLourde), Quantite(L, Important) : Cout(L, 310) \\
SecondPuits(L), Hydrocarbures(L), Quantite(L, Faible) : MiseEnProduction(L, StructureLegere) \\
SecondPuits(L), MiseEnProduction(L, StructureLegere), Quantite(L, Faible) : Cout(L, 10) \\
PremierPuits(l_1), Hydrocarbures(l_1)
\end{array}
$$

$$
\begin{array}{c}
(PremierPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000)) \\
\succ \\
(SecondPuits(L), Quantite(L, Faible), TauxSoutirage(L, StructureLegere, 200000))
\end{array}
$$

$$
\begin{array}{c}
(PremierPuits(L) \land TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Faible), Cout(L, 100)) \\
\succ \\
(SecondPuits(L) \land TauxSoutirage(L, StructureLegere, 200000), Quantite(L, Importante), Cout(L, 200))
\end{array}
$$

On suppose toujours que la fonction de similarité est donnée par une unification partielle, qui de plus ne permet pas de négliger les prédicats $MiseEnProduction, DeuxPuits$. On s'intéresse aux conséquences préférentielles de cette théorie sachant les connaissances qu'elle détermine. Selon l'ensemble des négligeables considérés, s'il est restreint à:

- $\{Quantite(L, Faible)\}$: On choisit alors de faire une mise en production directement à partir d'une structure légère.

- $\{Quantite(L, Importante)\}$: On ne peut rien décider du tout.

- $\{\{Quantite(L, Faible)\}, \{Quantite(L, Importante)\}\}$: On en déduit alors que l'on peut mettre le puits en production avec une structure légère, directement sans second forage, ou le mettre en production avec une structure lourde après un second forage.

7.2.3 Discussion sur la relation de conséquence sémantique

Deux points au moins peuvent être notés sur cette relation de conséquence sémantique sur l'interaction entre ce qui est déductible et les négligeabilités considérées, er sur la manière dont sont agrégées les préférences.

Le premier point concerne l'interaction entre ce qui est déductible et les négligeabilités considérées. En effet, comme nous l'avons vu dans les deux exemples précédents, selon l'ensemble des négligeables considérés, l'ensemble des conclusions déductibles varie complètement. Un point crucial est le partitionnement des négligeables en sous ensembles distincts: en effet, un “mauvais” partitionnement peut entraîner, comme dans le quatrième cas des deux exemples précédents, l'impossibilité d'utiliser les préférences; dans ce cas, le décideur doit reconsidérer l'ensemble des négligeables et surtout la manière dont il a été partitionné. Ce n'est pas le seul point difficile dans le raisonnement, car il peut être parfois difficile de maintenir la cohérence entre l'ensemble des séquents et l'ensemble des règles préférentielles.

À la manière dont nous avons posé les définitions 71 et 73, nous avons supposé que les mondes devaient être maximaux au sens de toutes les préférences du décideur, i.e. de leur union. Néanmoins, on peut penser que pour certains problèmes de décision, une agrégation constituée par l'union des préférences peut s'avérer insuffisante dans certains cas. Ainsi, dans l'exemple précédent, on peut choisir de reconnaître en priorité les aéroports civils plutôt que les aéroports militaires dans un cadre général, mais préférer en cas d'intervention militaire par exemple, préférer reconnaître un aéroport civil plutôt qu'un aéroport militaire (sous entendu préférer prévenir les pilotes qui vont survoler la zone qu'il y a un risque d'attaque plutôt que ne pas les prévenir). Il peut donc être intéressant d'augmenter la théorie proposée en introduisant une fonction d'agrégation des préférences, dans l'exemple une fonction d'agrégation lexicographique.

7.3 La sémantique de point fixe

Le but de cette section est de définir une fonction implantable sur machine permettant de “calculer” l'ensemble des conséquences d'une théorie préférentielle particulière, dans le sens où l'on aurait restreint l'ensemble des formules permettant de former ladite théorie. Pour ce faire, on établit un certain nombre de propriétés formelles qui font l'objet de la section suivante.

7.3.1 Clauses de Horn, préférences de Horn

Tout comme dans le chapitre précédent, on restreint ici le langage considéré aux clauses de Horn, et on restreint la définition de la fonction de similarité (voir définition 90), pour prendre en compte la même propriété de monotonie que dans le chapitre précédent (voir définition 80). On utilise tout comme dans le chapitre précédent les préférences de Horn (voir définition 80), et tout comme dans le chapitre précédent, on utilise dans les définitions les notations $L_\Sigma[X]$ pour désigner l'ensemble des clauses de Horn sur la signature Σ, $P_\Sigma[X]$ pour désigner l'ensemble des formules préférentielles de Horn sur la signature Σ.

7.3.2 Algèbres de Herbrand étendues

Nous réutilisons dans cette section les algèbres de Herbrand étendues introduites dans le cas de la satisfiabilité forte. Ceci permet de poser la définition d'un modèle épistémique similaire de Herbrand dans le cas de la satisfiabibilité faible :

Définition 93. (*Modèle épistémique similaire de Herbrand*)
Un modèle *épistémique similaire* **de Herbrand** est formé par un quadruplet $(At_\Sigma \times At_\Sigma, Sim, (D_i)_{1\leq i\leq n}, \models_S)$ où $\models_S \;\subseteq\; (\mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma) \times \mathcal{P}(L_\Sigma[X]) \times \mathcal{P}(L_\Sigma[X])) \times SP_\Sigma[X]$ est définie par:

1. Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), C \in L_\Sigma[X], P \in P_\Sigma[X], D \in (D_i)_{1\leq i\leq n}, \Gamma \subseteq S_\Sigma[X]$:

$$< (\mathcal{H}_1, \mathcal{H}_2), D > \models_{\mathcal{S}/\Gamma} (P, C) \Leftrightarrow < (\mathcal{H}_1, \mathcal{H}_2), D > \in AdmissNeg_{/\Gamma}(P, C)$$

2. Soit $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), \Gamma \subseteq L_\Sigma[X], \Pi \subseteq P_\Sigma[X], D \in (D_i)_{1\leq i\leq n}$:

$$< (\mathcal{H}_1, \mathcal{H}_2), D > \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Leftrightarrow \forall P \in \Pi, \forall S \in \Gamma, < (\mathcal{H}_1, \mathcal{H}_2), D > \models_{\mathcal{S}/\Gamma} (S, \Pi)$$

3. Soit $\Gamma \subseteq S_\Sigma[X], \Pi \subseteq P_\Sigma[X], C \in L_\Sigma[X], D \in (D_i)_{1\leq i\leq n}$,

$$< (\Gamma, \Pi), D > \models_{\mathcal{S}/\Gamma} C \Leftrightarrow \forall (\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{P}(At_\Sigma) \times \mathcal{P}(At_\Sigma), < (\mathcal{H}_1, \mathcal{H}_2), D > \models_{\mathcal{S}/\Gamma} (\Gamma, \Pi) \Rightarrow \mathcal{H}_1 \models_L C$$

Théorème 94. *Soit* $(WD_L, Sim, (D_i)_{1\leq i\leq n}, \models_S)$ *un modèle épistémique similaire et* $(At_\Sigma \times At_\Sigma, Sim, (D_i)_{1\leq i\leq n}, \models_S)$ *un modèle épistémique similaire de Herbrand. Alors les formules universelles préférentielles valides dans le modèle épistémique similaire sont valides dans le modèle épistémique similaire de Herbrand associé (et réciproquement). Les deux modèles seront alors dits associés.*

Preuve. Immédiat puisque, d'après le théorème de Herbrand, une formule universelle admet un modèle si et seulement si elle admet un modèle de Herbrand, et les clauses de Horn sont des formules universelles et les préférences de Horn constituées de formules universelles. ■

7.3.3 Modèles préférentiels stables

Comparativement à la version préférentielle forte, ces modèles préférentiels stables sont construits de manière parallèle en considérant les négligeabilités à priori, ces derniers éléments étant absents de la version forte. On retrouve dans la définition de ces ensembles l'une des deux propriétés essentielles qui distinguent satisfiabilité faible et forte, à savoir que les préférences utilisées sont uniquement déterminées par les éléments valides, les éléments négligeables étant connus à priori.

On considère dorénavant un programme défini préférentiel (P, Π) l'ensemble des conséquences de ce programme sachant les connaissances que P détermine. On suppose de plus que chaque règle de P, chaque ensemble $(D_i)_{i\in I}$ de négligeables, et chaque préférence de Π a été remplacée par toutes ses instantiations, de telle sorte que tous les atomes de P et Π sont liés.

Les modèles préférentiels stables que nous présentons maintenant sont des adaptations de ceux de [Gelfond and Lifschitz 1988]. Pour chaque ensemble d'ensembles d'atomes $(M_{1_i}, M_{2_i})_{i\in I}$ de P, on définit par $(P_{(M_{1_i}, M_{2_i}), D_i})_{i\in I}$ les programmes obtenus à partir des fonctions $(Sim(D_i, P, (M_{1_i}, \complement M_{1_i}))_{i\in I}$ [25], et de Π pour chaque $i \in I$:

- en effaçant tous les couples (C', Δ') issue d'une règle C de $Sim(C, D_i, (M_{1_i}, \complement M_{1_i}))$ tel qu'il existe un couple $(C", \Delta")$ de $Sim(P, D_j, (M_{1_j}, \complement M_{1_j}))$, une préférence $p = (a_1, d_1, f_1) \succ (a_2, d_2, f_2)$ tel que $(a_1, f_1) \in M_{1_i}$ et $(a_2, f_2) \in M_{1_j}$ et $d_2 \subseteq D_j$ et $d_1 \subseteq D_i$, et tel que $\neg(a_1, f_1) \in M_{1_j}$ et $(a_2, f_2) \in M_{1_i}$ et $d_2 \subseteq D_i$ et $d_1 \subseteq D_j$)

25. La notation $\complement M_{1_i}$ désigne l'ensemble des atomes de D_i qui n'appartiennent pas à M_{1_i}.

- en effaçant tous les couples $(C', \Delta') \in Sim(D_i, P, (M_{1_i}, \complement M_{1_i}))$ tels qu'il existe une suite de couples $(C_i, \Delta_i)_{1 \leq i \leq n}$ de $Sim(D_i, P, (M_{1_i}, \complement M_{1_i}))$ avec $(C_n, \Delta_n) \equiv (C', \Delta')$ telle que:
  - $\forall 1 \leq i \leq n, \forall A \in C'^{-}_{i}, A \in M_1$ et $C'^{+}_{i} \in M_1, C'^{+}_{i} \in C'^{-}_{i+1}$,
  - $C^{+}_{n} \in \Delta_1$

Considérons alors, pour chaque $i \in I$, le programme constitué par l'ensemble des règles similaires de $(P_{(M_{1_i}, M_{2_i}), D_i})_{i \in I}$. C'est un programme sans similarité ni préférences. Il a donc un plus petit modèle de Herbrand. Si, pour tout $i \in I$, ce modèle coïncide avec M_{1_i}, et que l'ensemble des négligeables que détermine M_{1_i} (i.e. $\{\Delta', (S', \Delta') \in Sim(D_i, S, (M_{1_i}, \complement M_{1_i}))\}$ coïncide avec M_{2_i}, à l'instar de [Gelfond and Lifschitz 1988], on dit que $(M_{1_i}, M_{2_i})_{i \in I}$ est un *ensemble préférentiel stable* de $(P, \Pi, (D_i)_{i \in I})$. De tels ensembles peuvent également être décrits comme étant les points fixes de l'opérateur $T_{(P, \Pi, (D_i)_{i \in I})}$ défini par la condition suivante:

Définition 95. Soit M_1, M_2 un ensemble d'atomes de P, et soit $P_{(M_1, M_2)}$ le programme à partir d'une fonction $Sim(P, \complement M_1))$ et de Π par les conditions (1) et (2). On définit alors l'opérateur:

$$\begin{array}{llll} T_{(P, \Pi, (D_i)_{i \in I})}: & \prod_{i \in I} \mathcal{P}(At_{\Sigma}[X]) \times \mathcal{P}(At_{\Sigma}[X]) & \longmapsto & \prod_{i \in I} \mathcal{P}(At_{\Sigma}[X]) \times \mathcal{P}(At_{\Sigma}[X]) \\ & ((M_{1_i}, M_{2_i})_{i \in I}) & \longmapsto & ((M'_{1_i}, M'_{2_i})_{i \in I}) \end{array}$$

tel que pour tout i, M'_{1_i} est le modèle de Herbrand minimal de $(P_{M'_{1_i}, M'_{2_i}, D_i})_{i \in I}$, M'_{2_i} constituant l'ensemble des éléments négligeables $\{\Delta', (S', \Delta') \in Sim(D_i, S, (M_{1_i}, \complement M_{1_i}))\}$.

Ces ensembles préférentiels stables ont de plus la propriété suivante:

Théorème 96. *Tout ensemble préférentiel stable de $(P, \Pi, (D_i)_{i \in I})$ est un modèle de Herbrand minimal similaire de $(P, \Pi, (D_i)_{i \in I})$.*

Preuve. Considérons un ensemble préférentiel stable (M_{1_i}, M_{2_i}). Montrons tout d'abord que ce couple est modèle par similarité de (P, Π).
Soit R une règle de P. S'il existe une couple (C', Δ') issue d'une règle R de $Sim(D_i, C, (M_{1_i}, \complement M_{1_i}))$ tel qu'il existe un couple $(C", \Delta")$ de $Sim(D_j, P, (M_{1_j}, \complement M_{1_j}))$, une préférence $p = (a_1, d_1, f_1) \succ (a_2, d_2, f_2)$ tel que $(a_1, f_1) \in M_{1_i}$ et $(a_2, f_2) \in M_{1_j}$ et $d_2 \subseteq \Delta'$ et $d_1 \subseteq \Delta"$, alors R est vraie dans M_{1_i} puisque M_{1_i} valide $\neg\delta$. On peut donc effacer la règle R. Les règles similaires non effacées sont quant à elles vraies dans M_{1_i} puisque M_{1_i} est leur plus petit modèle de Herbrand. M_{2_i} constituant l'ensemble des négligeabilités nécessaires pour que M_{1_i} soit modèle par similarité du programme, (M_{1_i}, M_{2_i}) est donc bien un modèle préférentiel similaire de (P, Π).

Pour montrer que (M_{1_i}, M_{2_i}) est minimal, considérons maintenant élément plus petit que (M_{1_i}, M_{2_i}), que l'on note (M'_{1_i}, M'_{2_i}), modèle préférentiel similaire de (P, Π). Considérons une règle R de P. Par la propriété de monotonie de la fonction de similarité, si un couple (C', Δ') a été effacé dans M_{1_i} à partir de la règle R, alors:

- ou bien il existe un couple (C'', Δ'') dans M'_{1_i} qui a été effacé lui aussi,
- ou bien un tel couple n'existe pas, et la règle est vraie dans M_{1_i} parce que l'une de ses prémisses est fausse, ce qui d'un point de vue validité des atomes revient au même que si la règle avait été effacée.

Il s'ensuit donc que ce qui les distingue est que (M'_{1_i}, M'_{2_i}) est un modèle similaire du programme sans préférences (puisqu'ils valident les règles effaçées de la même manière. Or nous avons démontré (voir théorème 84) que de tels modèles sont minimaux. Il s'ensuit donc que (M'_{1_i}, M'_{2_i}) et (M_{1_i}, M_{2_i}) sont égaux. ■

Il s'ensuit donc que l'ensemble des conséquences d'un programme préférentiel peut être calculé par l'opérateur $T_{(P, \Pi)}$, puisqu'il s'agit de l'intersection de ses points fixes.

Proposition 97. *Pour tout programme préférentiel (P, Π), $T_{(P, \Pi)}$ est anti-monotone.*

D'après les résultats vus au chapitre précédent, on peut donc calculer les plus petits et plus grands points fixe de $T^2_{(P, \Pi)}$.

7.3.4 Liens avec le système Soar

On notera que les deux phases de calcul du point fixe de l'opérateur sont non sans analogies avec un système d'intelligence artificielle existant, en l'occurrence le système Soar [Newell 1990]. L'effacement de règles correspond à la phase de *décision* de Soar, et le calcul de la plus petite algèbre de Herbrand du programme ainsi obtenu correspond à la phase *d'élaboration* de Soar. Notons de plus que la parallélisme de Soar permet aisément d'implanter ce type de raisonnement intrinsèquement parallèle. Les autres analogies avec le système Soar ont été discutées dans la section correspondante de la logique à satisfiabilité forte.

7.4 Conclusion

Nous avons donc dans ce chapitre établit tant la définition sémantique que l'implantation sous forme de fonctionnelle de point fixe d'un raisonnement par similarité pour une logique à satisfiabilité faible. Cette logique permet de combiner similarité et préférence, et satisfait comme nous l'avons montré les bonnes propriétés que nous avions au chapitre 4. De plus, par rapport à la relation de déduction par similarité (version satisfiabilité forte), elle permet de prendre en compte le cas où les préférences sont incomplètes en gérant des hypothèses de négligeabilité exclusives les unes des autres. Tout comme dans le cas de la logique à satisfiabilité forte, nous avons pu souligner les liens existants entre ce formalisme et le système Soar, système expert parallèle dans lequel nous allons maintenant conduire l'implantation de notre logique. Le choix du système Soar est du, non seulement aux liens étroits que nous avons mentionné entre son mode de fonctionnement et la sémantique point fixe du raisonnement par similarité, mais aussi au fait que Soar va nous permettre d'effectuer automatiquement les hypothèses incompatibles de négligeabilités qui sont à la base du raisonnement par similarité, version satisfiabilité faible. En effet, les éléments négligeables sont pris parmi un sous ensemble de l'ensemble des caractéristiques des cas passés, i.e. le complémentaire des caractéristiques nécessaires pour décrire un problème (par exemple, dans le cas de la photo-interprétation, il ne peut être question de négliger la reconnaissance de la piste pour déduire qu'un objet est un aéroport militaire). Le système Soar a pour caractéristique que deux actions ne peuvent être choisies simultanément: lorsque le choix de deux actions résulte de deux négligeabilités différentes, on a dans Soar un sous problème permettant de choisir entre ces deux actions. Ce sous problème détermine donc les deux hypothèses de négligeabilité incompatibles qui sont à la base de la relation de déduction par similarité (version satisfiabilité faible).

Quatrième partie

Implantation - Application à la photo interprétation

Nous avons présenté dans les chapitres précédents la définition formelle de deux logiques pour raisonner par similarité en fonction de préférences. Nous avons souligné l'analogie existant entre d'une part, la sémantique des préférences dans ces logiques et certaines des préférences du système Soar et d'autre part, la définition des opérateurs de point fixe et le fonctionnement du système Soar.

Nous présentons donc dans un premier temps la manière dont l'implantation de ces opérateurs de point fixe a été réalisée dans le système Soar, et les modifications induites:

- sur la gestion du RETE, en pré-compilant des règles similaires à celles du programme,
- sur la syntaxe de Soar pour prendre en compte la notion de négligeabilité.

Nous présentons ensuite dans un deuxième temps une application utilisant le système Soar ainsi modifié pour modéliser le problème d'interprétation d'images satellites. Nous verrons la manière dont procède le thématicien pour interpréter l'image satellite, le type d'opérations qu'il effectue, le type de connaissances utilisées ainsi que les raisonnements effectués sur l'image. Nous verrons ensuite la manière dont l'implantation a été conduite dans Soar, et la validation qu'elle offre du modèle formel de raisonnement par similarité défini.

8

Implantation dans le système Soar

8.1 Introduction

À partir du schéma de la figure 2.1 de la page 29, nous nous sommes intéressés dans les chapitres précédents, à la formalisation des notions de similarité et de préférence. Nous avons vu qu'il existe une correspondance entre:

- la sémantique des préférences de la logique que nous avons défini et celle des préférences du système Soar,
- l'opérateur de point fixe que nous avons défini pour l'une ou l'autre des deux logiques que nous avons présentés et le fonctionnement du système Soar.

Nous ne nous sommes cependant pas intéressés à l'organisation des connaissances qu'utilise le schéma de la figure 2.1: ni la notion d'état, ni la notion de problème ou celle d'action n'ont pour le moment été modélisées. Le système Soar va nous permettre cette formalisation, de par son architecture que nous commencerons donc par présenter. C'est sur cette architecture que vient se greffer la sémantique de point fixe que nous avons définie. Nous verrons ensuite la manière dont les connaissances et les préférences sont représentées dans le système Soar. Ayant ainsi présenté la partie commune entre Soar et le formalisme que nous avons défini, nous verrons ensuite les modifications que nous avons dû y apporter pour prendre en compte les notions de similarité et de négligeabilité, ce qui informellement revient à mettre en oeuvre un système d'unification partielle dans Soar.

8.2 Le système Soar

Nous présenterons donc dans cette section l'organisation des connaissances et des préférences dans le système Soar, ainsi que la manière dont fonctionne le moteur d'inférence de ce système.

8.2.1 Organisation des connaissances dans le système Soar

L'organisation des connaissances dans le système Soar a son origine dans le rejet par [Laird et al. 1986] de tout système où les méthodes seraient le mode de représentation dominant des connaissances. Ces méthodes (séquences ordonnées d'instructions intégrant entre autres conditionnelles et boucles) partitionnent la connaissance de telle sorte que la connaissance contenue dans une seule méthode (la méthode sélectionnée pour la tâche courante) peut être utilisée. Or, un système intelligent doit sans cesse apprendre de nouvelles connaissances qui doivent être intégrées dans ces méthodes. Un système possédant un ensemble de méthodes plus un mécanisme de sélection d'une méthode en fonction de la tâche donnée ne peut donc exécuter cette forme d'apprentissage.

À partir de ce constat, [Laird et al. 1986][Newell 1990] proposent à la fois une définition de l'intelligence sur laquelle doit reposer tout système d'intelligence artificielle, et un système satisfaisant cette définition, en l'occurrence le système Soar. Les propriétés que doit satisfaire tout système d'intelligence artificielle sont les suivantes [Newell 1990]:

1. Si un système utilise toutes les connaissances qu'il possède, alors il doit être parfaitement intelligent.
2. Si un système ne possède pas certaines connaissances, leur non-utilisation n'est pas un échec de l'intelligence.
3. Si le système a des connaissances et ne réussit pas à les utiliser, il y a alors un échec de l'intelligence.

Pour résumer, un système est intelligent dans la mesure où *il utilise toutes ses connaissances au service de ses buts.*

Pour définir un tel système parfaitement intelligent, l'idée est de sélectionner des méthodes très générales (comme générer-tester, "mini-max search"..., méthodes toutes implantées dans Soar par [Laird et al. 1986]) et de construire un système tel que ces méthodes émergent directement de l'organisation des connaissances dans ce système. Pour ce faire, [Newell and Simon 1972][Laird et al. 1986][Newell 1990] proposent qu'un tel système s'engage dans une recherche, qui s'effectue dans **un espace de problèmes**, en fonction de la proposition suivante [Newell 1990]:

> Si on ne sait comment atteindre X, créer un espace de problèmes qui contient X et chercher X dans cet espace.

On retrouve dans cette notion d'espace de problèmes la notion de problème que nous avions défini (voir figure 2.1 page 29). Au sein de cet espace de problèmes, en fonction de l'organisation des connaissances proposée, le système se trouve dans un certain **état**, possède des **opérateurs** (i.e. des connaissances) pour déterminer un nouvel ensemble **d'états accessibles**, et la résolution de problèmes se ramène alors à identifier **le but** parmi cet ensemble d'états. L'application d'opérateurs mène à un nouvel état, choisi parmi l'ensemble des états accessibles, et la résolution de problèmes continue en reprenant ce nouvel état comme état initial. L'architecture de Soar est bâtie autour des espaces de problèmes (voir figure 8.1), et

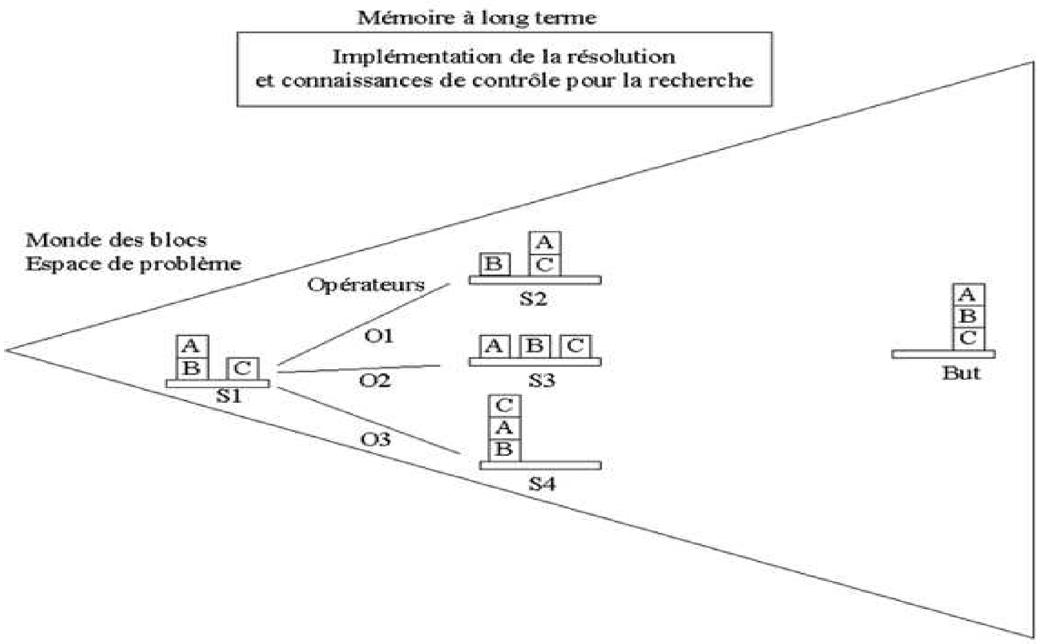


FIG. 8.1: *Architecture de Soar*

les fonctions primaires de l'architecture sont alors:

- choisir un espace de problème.

– choisir un état parmi les états directement accessibles.

– choisir un opérateur.

– appliquer l'opérateur pour obtenir le nouvel état.

La figure 8.1 illustre l'architecture de Soar sur l'exemple désormais classique du monde des blocs (à partir d'un ensemble de blocs se trouvant dans un état initial, trouver la séquence d'opérations permettant d'atteindre l'état final). On retrouve dans ce mode de fonctionnement celui illustré à la figure 2.1, en convenant d'identifier opérateur et action.

D'un point de vue informatique, cette architecture correspond dans l'implémentation à une structure en mémoire. Cette structure consiste en un contexte courant, un certain nombre d'objets et leurs enrichissements (ou augmentations pour reprendre la terminologie de Soar):

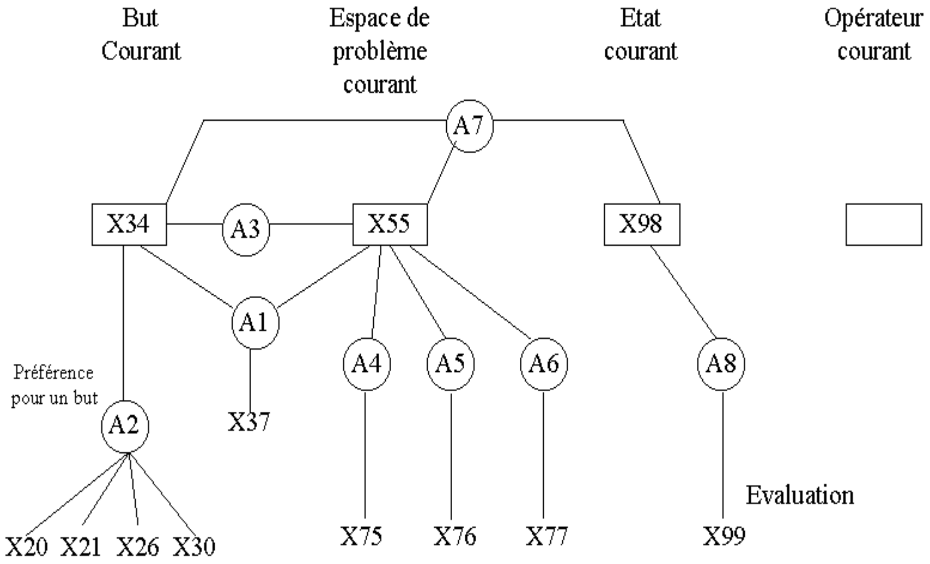


FIG. 8.2: *Architecture de Soar*

le contexte courant contient un emplacement réservé ou *slot* pour chaque type d'objet utilisé par le système: but, espace de problème, état et opérateur (voir exemple 8.2). L'objet occupant le slot but (dans la figure 8.2, $X34$) dans le contexte courant est l'espace de problème courant. Par exemple, dans un problème de photo-interprétation, $X34$ peut représenter le but “interpréter l'image”,

un objet est représenté par un ou des symboles, et chaque slot dans le contexte courant contient un symbole ou est indéfini,

une augmentation est un ensemble de relations entre des objets en mémoire, et peut être vue comme une relation identifiant, attribut, valeur, chaque augmentation d'un objet (identifiant) représentant la valeur de l'un de ses attributs. Ainsi, par exemple, l'état $X98$ dans la figure 8.2 est augmenté par le biais de l'attribut $A8$ de l'évaluation $X99$, l'état $X37$ augmente par l'attribut $A1$ le but $X34$ et l'espace de problème $X55$. La sémantique d'un objet dans Soar est dérivée de l'ensemble de ses augmentations; c'est une représentation inverse de la représentation normale d'un objet, en faisant des augmentations les structures de données qui sont manipulées au lieu que ce soit l'objet lui-même. Par exemple, dans un problème de photo-interprétation, $X98$ peut représenter un objet (par exemple un site missile) sur une image satellite, et A_8 correspond alors à une évaluation de la possibilité de présence d'un tel site sachant qu'aucun site stratégique n'est présent à proximité (la présence d'un site missile indique en général la présence d'un site stratégique à protéger à proximité).

Ces augmentations sont donc représentés par des WME (pour *Working Memory Elements*) qui sont des triplets (identifiant, attribut, valeur)(voir figure 8.3), et tous les WME avec le même identifiant font partie du même objet: Ainsi, par exemple, dans le cas de la photo-interprétation, un ensemble de formes

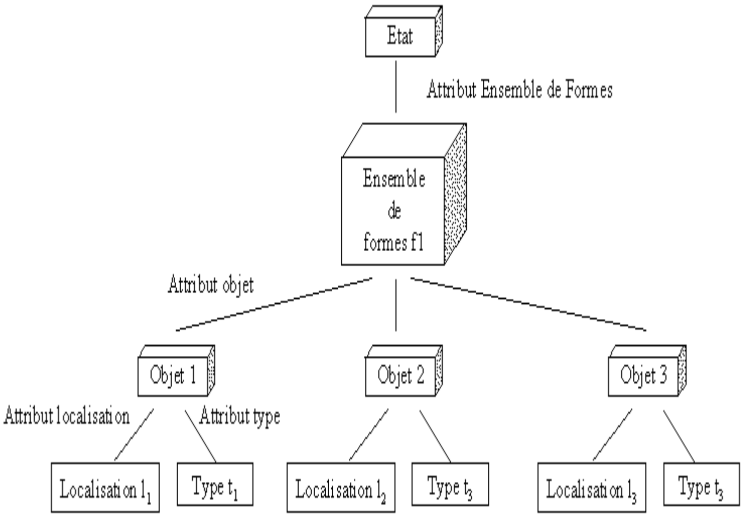


FIG. 8.3: *Les Working Memory Elements*

présents sur l'image se décompose en un ensemble d'objets, chacun d'entre eux étant placé en un certain "endroit" sur l'image et étant d'un certain type (voir figure 8.3).

8.2.2 Le fonctionnement du système Soar

Le traitement informatique réalisé par Soar consiste à remplacer un objet du contexte courant par un nouvel objet, ou à initialiser les slots du contexte courant. Ce traitement est décomposé en un cycle de décision dont le détail de fonctionnement est donné dans la figure 8.4). Ce cycle se décompose suivant plusieurs cycles d'élaborations (fermeture déductive en utilisant les règles du programme), cycles d'élaborations qui ont pour entrées sorties:

Entrées Objets et augmentations en mémoire, accessibles à partir des augmentations du contexte courant.

Résultats Augmentation du contexte courant.

Contrôle Continu jusqu'à la phase de quiescence. La quiescence est une phase où plus aucune connaissance nouvelle n'est accessible, et où le système ne peut donc plus rien déduire.

Un cycle d'élaboration est décomposé (voir 8.4) en une phase de "préférence" (déclenchement de toutes les productions instantiées) et une phase "Working Memory" (utilisation des préférences qui portent sur tous les objets qui ne sont pas des opérateurs). Cette phase aboutit donc à une augmentation des objets du contexte courant (hormis les opérateurs).

La deuxième phase du cycle (voir 8.4) a pour but de déterminer les objets (état, espace de problèmes, but, opérateur) qui vont devenir courant, remplaçant ainsi un objet du contexte. Cette phase a pour entrées sorties

- Entrées: Objets et augmentations en mémoire, accessibles à partir des augmentations du contexte courant.

– Résultats: Vote pour chacune des actions.
 Traduction des augmentations en actions.
 Un nouvel objet est sélectionné dans le contexte courant.

– Contrôle: La procédure de vote est parallèle.

On retrouve donc ici, dans les phases d'élaboration et de décision, une partie de la définition de l'opérateur de point fixe que nous avons défini, puisque l'élaboration consiste bien en une fermeture déductive (voir son mode de contrôle) et que la phase de décision consiste bien à retirer (*rétractation*) de la base de faits les éléments qui ne sont pas admissibles au sens des préférences. La différence avec le système Soar est que celui-ci permet l'écriture de préférences sur d'autres objets que sur les opérateurs, et que ces préférences sont en prises compte au fur et à mesure des élaborations. Dans la sémantique de point fixe que nous avons définie, les élaborations correspondent à un processus d'inférence **monotone**; dans Soar, l'nférence n'est monotone que par rapport aux choix des opérateurs.

8.2.3 Les préférences dans le système Soar

Par rapport à un système classique, le système Soar ne contient pas seulement une base de faits et un ensemble de productions mais de plus, il contient une mémoire préférentielle (voir figure 8.5). Les productions de Soar ne changent pas directement, lorsqu'elle se déclenchent, la base de faits mais créent en mémoire préférentielle des **suggestions** pour effectuer des changements de la base de faits. Le rôle de la phase de décision est alors de choisir parmi ces suggestions en fonction des préférences.

Les préférences ont plusieurs types de valeur dans Soar:

Syntaxe	Sémantique
Acceptable	Candidat pour la sélection
Rejetée	Non candidat pour la sélection
La meilleure	Meilleur que tout autre valeur
La pire	Pire que tout autre valeur
Meilleure que	Meilleur qu'une autre valeur
Pire que	Pire qu'une autre valeur
Indifférent	Choix indifférent
Prohibé	Ne peut être sélectionné si le but doit être atteint
Requis	Doit être sélectionné pour que le but puisse être atteint
Reconsidéré	La préférence de l'opérateur doit être reconsidérée à la prochaine quiesence
Parallel	Deux valeurs ne sont pas mutuellement exclusives

Par rapport aux préférences que nous avons définies, certaines ont la même sémantique que les préférences de Soar:

– une préférence du type $(\perp, \perp, \perp) \succ (\varphi_1, \varphi_2, \phi_2)$, signifie intuitivement: si φ_1 et ϕ_2 sont vraies, si l'on a la possibilité de négliger φ_2, alors préférer ne rien faire plutôt que de faire ϕ_2, ou dit autrement rejeter l'action ϕ_2: c'est la préférence *reject* de Soar,

– les préférences du type $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ correspondent quant à elles intuitivement à un jugement du type: si ϕ_1 résulte de la connaissance de φ_1 et de la négligeabilité de φ_2, et si ϕ_2 résulte de la connaissance de ψ_1 et de la négligeabilité de ψ_2, alors préférer ϕ_1 à ϕ_2: c'est la préférence *better* de Soar,

– Le système Soar définit aussi la préférence *worse* qui est la préférence duale de la précédente, ainsi que les préférences *best* et *worst*, qui correspondent à une quantification sur ϕ_1 dans le premier cas, sur ϕ_2 dans le deuxième cas, donc sont des préférences d'ordre 2.

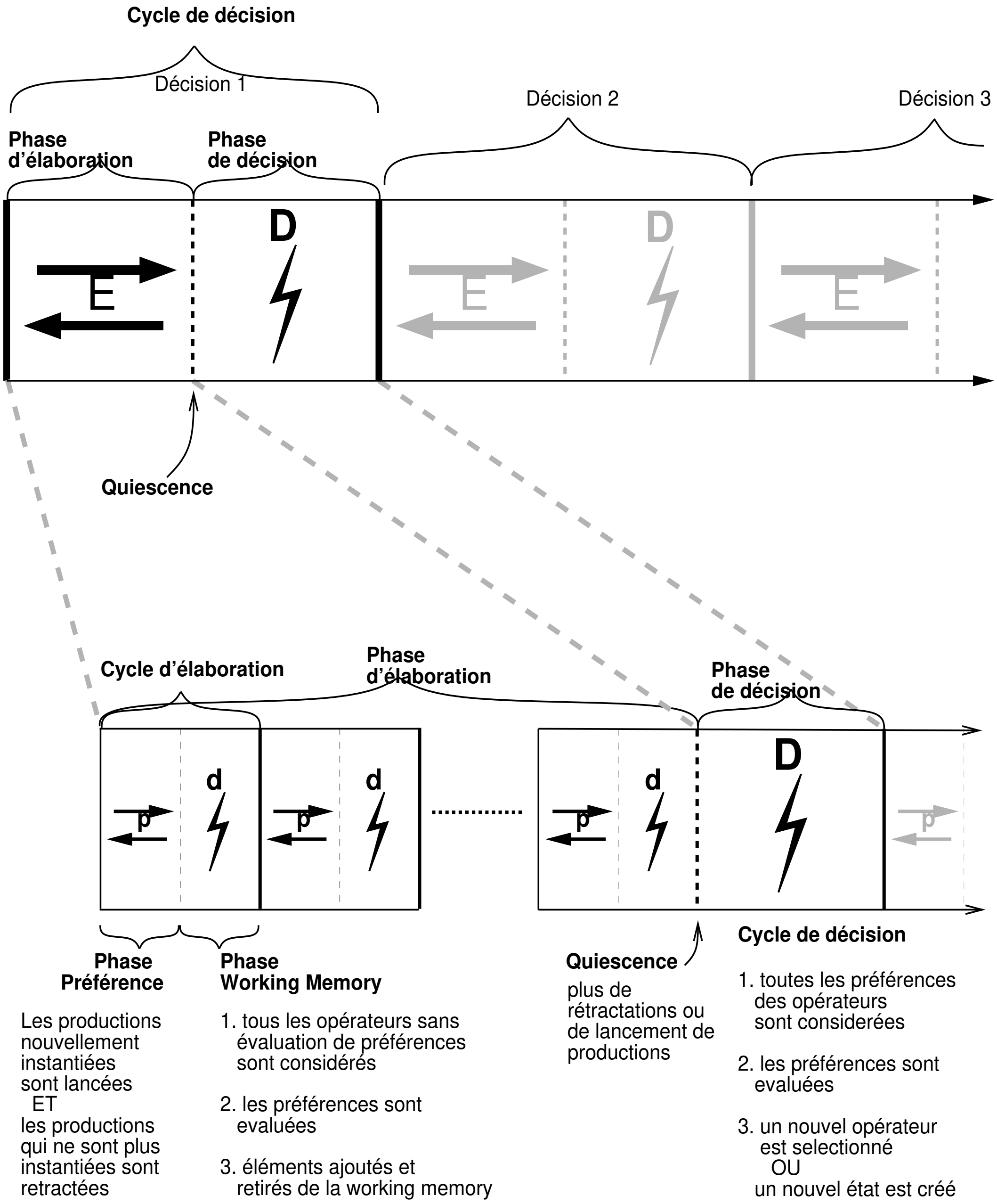


Fig. 8.4: *Architecture de Soar*

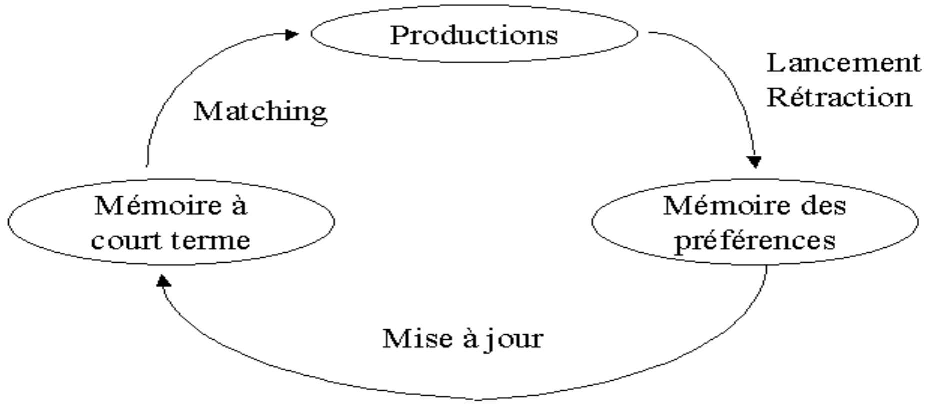


FIG. 8.5: *Architecture de Soar*

Si le système Soar ne possède pas de sémantique formelle justifiant son implémentation, il permet empiriquement d'étendre le formalisme que nous avons défini en étant capable de:

- raisonner avec des préférences contradictoires,
- rechercher la raison pour laquelle la phase de décision peut ne pas systématiquement aboutir à un choix,
- apprendre suite à la résolution de l'un des deux points précédents.

En effet, étant donné la sémantique que nous venons de définir pour les préférences et leur mode de traitement, rien ne permet a priori de supposer que la phase de décision aboutisse systématiquement à un choix pour un but, un espace de problème, un état ou un opérateur. Lorsque ceci n'est pas le cas, on aboutit à une impasse. Le langage des préférences étant donné, la liste des impasses, en nombre fini, est elle aussi donnée [Laird et al. 1986]:

- Il peut y avoir des *impasses d'égalité* (*tie-impasse*): toutes les alternatives sont acceptables, aucune d'entre elle n'est rejetée ou préférée à l'une des autres et elles ne sont pas indifférentes.
- Il peut y avoir des *impasses sans changement* (*no-change impasse*): après que toutes les préférences aient été exprimées, il n'y a plus aucun choix de disponible.
- Il peut y avoir des *impasses rejet* (*reject impasse*): la seule préférence disponible rejette l'une des décisions précédemment prises.
- Il peut y avoir des *impasses conflit* (*conflict impasse*): une production peut fournir une préférence pour un objet o_1 par rapport à un objet o_2 et il existe une préférence contradictoire de l'objet o_2 par rapport à l'objet o_1.
- Il peut y avoir des *impasses violations de contraintes* (*constraint failure impasse*): c'est le cas si il existe plusieurs préférences *requis* sur un attribut, ou s'il existe à la fois une préférence *requis* et *prohibé* sur un attribut.

Grâce à l'architecture de Soar, seules ces cinq difficultés peuvent apparaître. Notons au passage que ces échecs caractérisent le fait que le système ne dispose pas de connaissances directes permettant de les résoudre: il s'agit d'un problème de connaissance et non d'intelligence. Lorsqu'une de ces impasses se produit, Soar crée alors dynamiquement un sous but permettant de la résoudre. La création de ce sous-but est la manifestation de l'intelligence de Soar: un tel sous-but est lui-même soumis à diagnostic, via ses propres augmentations. Ce nouvel objet but est augmenté par les objets ayant entraîné la difficulté. Ceci permet implicitement un passage de paramètres du contexte courant au sous-but, ce dernier déterminant lui même au cours de la résolution quels sont les objets qui servent à la résolution de l'impasse (il est nécessaire de tous les passer car leur utilité n'apparaît qu'au cours de la résolution). Lorsque l'impasse est résolue, la résolution a entraîné des augmentations de certains des objets (passés par paramètres), qui sont repassées par paramètres en retour au super-contexte, réalisant ainsi un passage implicite des paramètres de résolution. Pour illustrer ce processus, considérons l'exemple suivant illustré à la figure 8.6). Au temps

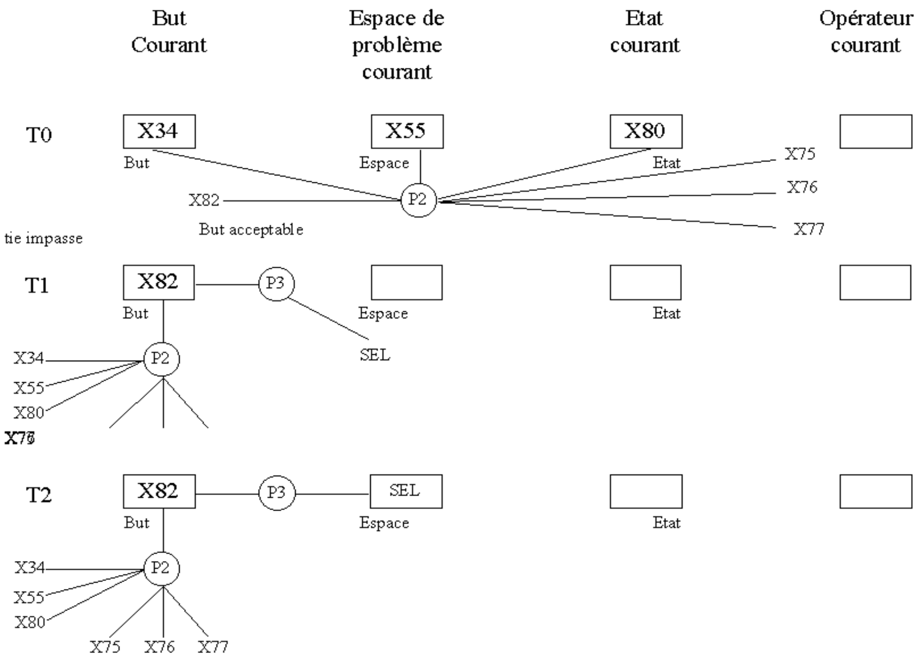


FIG. 8.6: *Passage de paramètres dans Soar (Extrait de [Laird et al. 1986])*

T0, un sous-but a été créé car les opérateurs X75, X76 et X77 étaient en conflit pour occuper le slot de l'opérateur courant dans le contexte courant. Une préférence P2 a donc été construite avec le nouvel objet but X82, cette préférence étant aussi liée aux autres objets du contexte courant. Au temps T1, le sous but X82 remplace le but courant et la résolution de problèmes reprend alors de manière usuelle: ce but est augmenté d'une préférence pour un espace de problèmes de sélection. Au temps T2, le slot du problème courant dans le contexte courant est occupée par SEL etc...

Quand une impasse est résolue, cela signifie que, en effectuant une résolution de problèmes, Soar a eu accès à de la connaissance qui n'était pas directement disponible. Ainsi, quand une impasse est résolue, Soar a eu une opportunité d'apprendre. Le mécanisme d'apprentissage de Soar est appelé *chunking*: quand la résolution de problème, suite à une impasse, a produit un résultat, une nouvelle production est créée, dont les actions sont justement les résultats obtenus par la résolution, et dont les conditions sont les éléments qui existaient en mémoire de travail avant que la résolution de problème ne commence et qui ont effectivement servis à la résolution de problème, comme illustré à la figure 8.7.

Ce mécanisme d'apprentissage (voir [Laird et al. 1986]) permet de rendre compte de la *"Power Law of Practice"*, qui spécifie que le temps nécessaire pour effectuer une tâche décroît en fonction d'une puissance du nombre de fois où elle a été effectuée. Cette adéquation avec la loi est une mise en évidence de la validité cognitive de Soar.

8.3 Implantation de la similarité

Nous avons vu dans les chapitres et sections précédentes que la sémantique des préférences que nous avons définie pouvait être incluse dans une partie des préférences de Soar. Nous nous contenterons donc dans cette section de discuter la manière dont a été modifié l'algorithme du RETE pour implanter une fonction de similarité particulière, qui correspond intuitivement à une unification partielle et qui a été

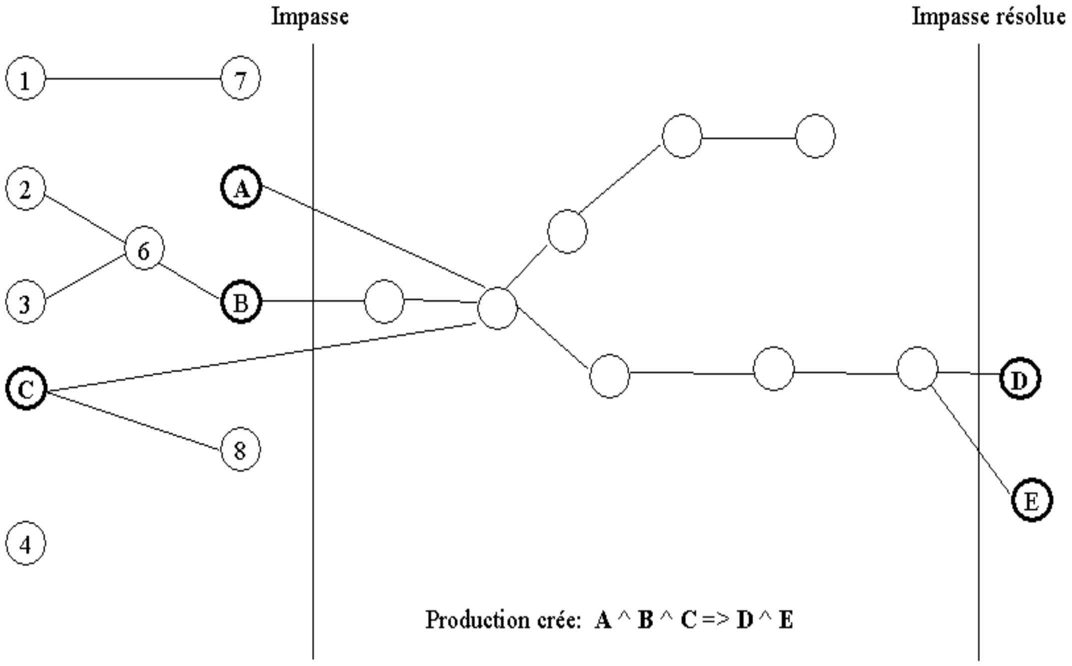


FIG. 8.7: *Le chunking dans Soar*

présentée au chapitre 6. Nous verrons dans un premier temps le fonctionnement de l'algorithme du RETE, puis ses modifications. Nous verrons ensuite les modifications syntaxiques réalisées sur le parseur de Soar pour prendre en compte la notion de négligeabilité, puis nous verrons ensuite comment le mécanisme de sous-buts a été utilisé pour la prise en compte du raisonnement par similarité version satisfiabilité faible.

8.3.1 L'algorithme du réseau RETE

L'algorithme du réseau RETE [Forgy 1982] est un algorithme d'unification qui utilise deux optimisations:

Le partage des conditions (littéraux des prémisses des règles) identiques appartenant à plusieurs règles,

La sauvegarde d'unifications partielles effectuées: même si une unification est incomplète, elle est sauvegardée de telle sorte qu'elle ne doit pas être refaite au cycle suivant.

Pour illustrer le fonctionnement de l'algorithme et la manière dont ces optimisations sont conduites, on prendra l'exemple de la production P_1 suivante:

P_1:	(State <s> cloture <l>)	W_1:	State s_1 cloture l_1
	(cloture <l> endroit <e>)	W_2:	State s_1 cloture l_2
	(cloture <l> aspect <a>)	W_3:	State s_1 cloture l_3
	$\Rightarrow$	W_4:	Cloture l_1 endroit e_1
	qqch	W_5:	Cloture l_2 endroit e_2
		W_6:	Cloture l_1 aspect a_1

qui doit être unifiée avec la base de faits $W_1, W_2.W_3, W_4, W_5, W_6$. La production P_1, écrite dans la syntaxe de Soar, a trois conditions et une action. Les conditions sont des triplets (WME) (identifiant, attribut, valeur); les symboles apparaissant dans les conditions sont ou bien des constantes comme *state, cloture, endroit, aspect* et le test consiste alors à déterminer si des constantes identiques apparaissent dans les WME, ou des variables apparaissant entre <> pour lesquelles des unifications doivent être données qui correspondent à des symboles apparaissant en même position dans les WME.

Le fonctionnement du RETE peut alors aisément être compris en imaginant que chaque condition d'une production fonctionne comme un filtre sur les WME: par exemple, un filtre teste que *State* appa-

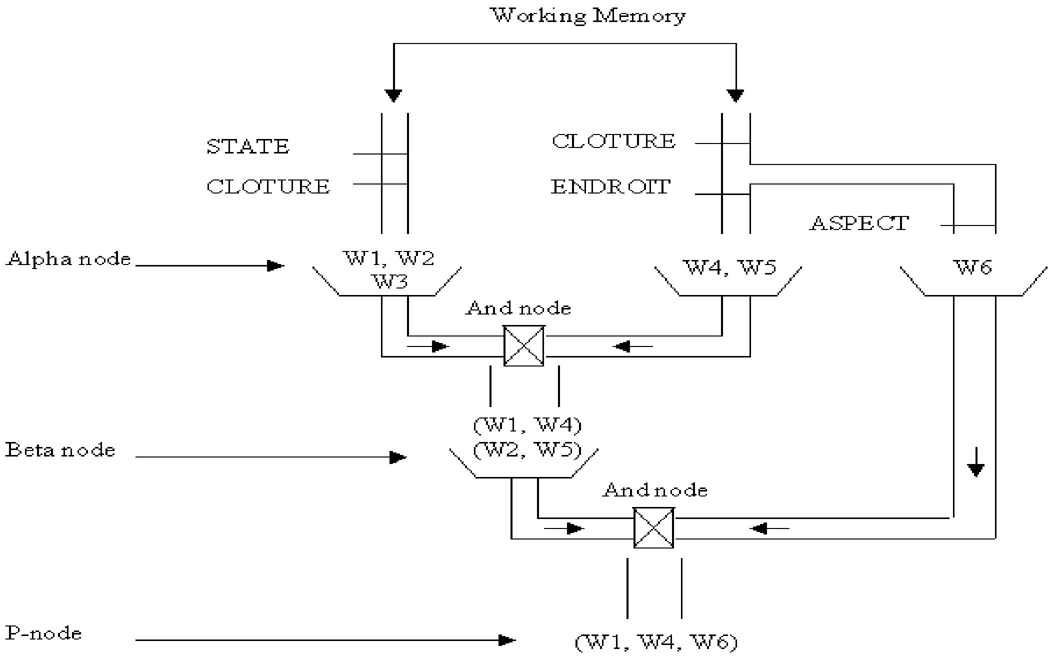


FIG. 8.8: *Le fonctionnement de l'algorithme du Rete [Forgy 1982]*

raît bien dans l'identifiant du WME et que Clôture est bien l'attribut du WME. Seuls certains WME (ceux qui unifient cette condition) passent ce filtre, comme W_1, W_2, W_3 pour la première condition. Ces WME sont stockés dans les **alpha node**.
Outre les WME satisfaisant <u>une</u> condition, les combinaisons de WME qui unifient de manière cohérente plusieurs conditions doivent être générées: c'est le rôle des **And node**. Dans notre exemple (voir figure 8.8), il faut vérifier que c'est bien la même clôture qui est sur l'état et qui se trouve à un endroit donné: les WME (W_1, W_4) et (W_2, W_5) constituent de telles unifications cohérentes. Le résultat d'une unification cohérente est stocké dans un **beta node**. Le nœud auquel on aboutit après avoir unifié toutes les conditions des prémisses des règles s'appelle un **p node**, et contient la partie droite de la règle.

Cet algorithme permet de sauvegarder les unifications partielles: si la base de faits n'avait contenu que W_1, W_2, W_3, W_4, W_5, le contenu du beta node aurait été sauvegardé et si, au cycle suivant, W_6 avait été introduit en base de faits, il aurait été le seul à avoir été unifié. Il permet également le partage de conditions entre productions puisque si l'on ajoute la production:

P_1:	(State <s> cloture <l>)	W_1:	State s_1 cloture l_1
	(cloture <l> endroit <e>)	W_2:	State s_1 cloture l_2
		W_3:	State s_1 cloture l_3
	$\Rightarrow$	W_4:	Cloture l_1 endroit e_1
	qqch	W_5:	Cloture l_2 endroit e_2
		W_6:	Cloture l_1 aspect a_1

il suffit d'ajouter sous le beta node un p node correspondant à cette production. Toutes les unifications réalisées au dessus de ce beta node dans le RETE seront alors communes aux deux productions.

8.3.2 Pré-compilation de règles pour la similarité

Cette pré-compilation nécessite la prise en compte de deux éléments: d'une part, la prise en compte des règles similaires et d'autre part, la prise en compte des éléments négligeables (les éléments non unifiés). La fonction de similarité choisie correspond informellement à une unification partielle (sa formalisation est détaillée à la fois au chapitre 6 et en annexe B).

8.3.2.1 La prise en compte des règles similaires

La prise en compte des règles similaires obéit à deux contraintes: d'une part, limiter le temps de calcul et la complexité du processus d'unification; d'autre part, limiter la taille du RETE en mémoire.

Pour illustrer la manière dont ces deux contraintes ont été prises en compte, considérons la production P_1 suivante associées aux WME suivants:

P_1:	(State <s> cloture <l>)	W_1:	State s_1 cloture l_1
	(cloture <l> endroit <e>)	W_2:	State s_1 cloture l_2
	(cloture <l> aspect <a>)	W_3:	State s_1 cloture l_3
	$\Rightarrow$	W_4:	Cloture l_1 endroit e_1
	qqch	W_5:	Cloture l_2 endroit e_2
		W_6:	Cloture l_1 aspect a_1

et considérons les deux productions suivantes similaires à P_1 et aux WME par la fonction de similarité:

P_2:	(State s_1 cloture l_1)	P_3:	(State s_1 cloture l_2)
	(cloture l_1 aspect a_1>)		(cloture l_2 aspect a_2)
	$\Rightarrow$		$\Rightarrow$
	qqch		qqch

Il est possible d'entrer dans le RETE chacune de ces deux productions, mais c'est une perte de ressources mémoire par rapport à la solution qui consiste à n'entrer que la production suivante, dont les deux productions précédentes ne sont que des instantiations:

P_1':	(State <s> cloture <l>)
	(cloture <l> aspect <a>)
	$\Rightarrow$
	qqch

De ce fait, au lieu d'entrer l'ensemble des productions résultant des unifications partielles avec la base de faits à chaque cycle de décision, on ajoute, au moment de la construction du RETE, à l'ensemble des productions considérés l'ensemble des productions qui leur sont similaires[26], une production similaire étant une production dont les prémisses appartiennent aux prémisses de la règle initiale, la tête de clause restant la même. Une production de cet ensemble se caractérise de plus dans Soar par le fait que ses prémisses doivent **obligatoirement** (contrainte de fonctionnement de Soar) contenir un identifiant d'un état ou d'une impasse, i.e. une condition ($state < s >$ *attribut valeur*) ou ($impasse < i >$ *attribut valeur*) et que, de plus, **chaque** condition des prémisses doit, que ce soit directement ou indirectement, être liée[27] à un état et à une impasse.

Cette pré-compilation de règles similaires induit une modification de l'algorithme du RETE. En effet, toutes les règles qui sont complètement unifiables ne rajoutent pas leur partie action en base de faits: si l'on considère les deux règles similaires suivantes issues de la même production P_1:

P_1':	(State <s> cloture <l>)	P_2':	(State <s> cloture <l>)
	(cloture <l> aspect <a>)		$\Rightarrow$
	$\Rightarrow$		qqch
	qqch		

Alors seule P_1' rajoute sa partie action en base de faits (la fonction de similarité est en effet telle que les prémisses obtenues par application de la fonction de similarité à une règle P_1 et à une base de faits B appartiennent à la borne supérieure au sens du pré-ordre de subsomption de P_1 et de B). De ce fait, lorsque l'algorithme fournit l'ensemble des unifications des règles similaires aux productions initiales, il ne faut retenir que les unifications des règles similaires issues de la même règle initiale et qui sont maximales en terme d'éléments unifiés.

26. Ici, la similarité est appréhendée indépendamment de toute base de faits. C'est l'instantiation de la règle pour cette nouvelle notion de la similarité qui fournit la fonction de similarité que nous avons présenté au chapitre 6.

27. Tous les objets en mémoire à court terme sont liés à d'autres objets (voir figure 8.3), la valeur d'un WME pouvant être l'identifiant d'un autre objet. C'est ainsi que tous les objets en mémoire à court terme doivent, que ce soit directement ou indirectement, être liés à un état.

8.3.2.2 La prise en compte des éléments négligeables

Le mode de gestion des règles similaires étant maintenant détaillé, voyons maintenant la manière dont a été appréhendée la gestion des éléments négligeables associés à chaque règle similaire. D'après la définition que nous avons donnée de chacun des points fixe des deux logiques que nous avons présentées, deux éléments caractérisent la gestion des éléments négligeables :

- chaque instantiation d'une règle similaire détermine des négligeables,
- un élément est ajouté dans la base de faits si et seulement si il n'a pas été déduit à partir de sa propre négligeabilité.

L'association des éléments négligeables à chaque instantiation d'une règle nécessite de distinguer entre les différents types de productions dans Soar [Congdon and Laird 1997]:

- des productions permettant de sélectionner un opérateur:

  Operator Proposal : représente une connaissance permettant de proposer un opérateur comme candidat potentiel,

  Operator Comparison : représente une connaissance permettant de comparer plusieurs opérateurs,

- Des productions permettant d'appliquer un opérateur:

  Operator Application : représente une connaissance sur les modifications induites par un opérateur sur l'état,

  Operator Termination : représente une connaissance permettant de détecter qu'un opérateur a été complètement appliqué,

- Élaborations sur l'état:

  State elaboration : des productions réalisant des inférences sur l'état.

Dans le cas où la proposition ou l'application d'un opérateur est dûe à un ensemble de négligeables, il importe que cet opérateur porte cet ensemble de négligeables comme une de ses augmentations. Ceci permet en effet de comparer des opérateurs en fonction des négligeables qui ont permis de les proposer ou de les appliquer.

Prise en compte des éléments négligeables dans le cas des productions "state elaboration"

Considérons dans un premier temps le dernier type de production, qui réalise des élaborations sur l'état. L'idée est de traduire dans Soar la connaissance qu'une augmentation d'un objet o_1 a été négligée (connaissance d'ordre deux) par une autre augmentation de o_1 créée de la manière suivante:

- un couple (attribut, valeur) ou attribut désigne la constante introduite **négligeable**, et valeur une valeur nouvellement créée désignant la négligeabilité,
- cette variable de négligeabilité est alors augmentée par le couple (attribut, valeur) qui a été négligé

Ainsi, pour prendre l'exemple de la figure 8.9, si une production néglige la présence de *objet* 1, elle néglige que l'ensemble des formes de l'image soit augmenté de l'*objet* 1, ce qui veut dire que l'ensemble des formes de l'image est affecté de la négligeabilité *negligeabilite* 1. Cette négligeabilité est elle même augmentée du fait que c'est l'objet *objet* 1 qu'elle néglige (voir figure 8.9). Lorsque l'élément négligé possède lui-même des augmentations négligées, celles-ci sont affectées elles-mêmes d'une variable représentant leur négligeabilité, comme illustré dans la figure 8.9 avec l'objet *objet* 1 qui est augmenté de sa localisation et de son type.

D'un point de vue syntaxique, ceci correspond donc à l'introduction de l'attribut "*negligible*" dans la syntaxe des programmes Soar, et si l'on considère la production P_1 suivante:

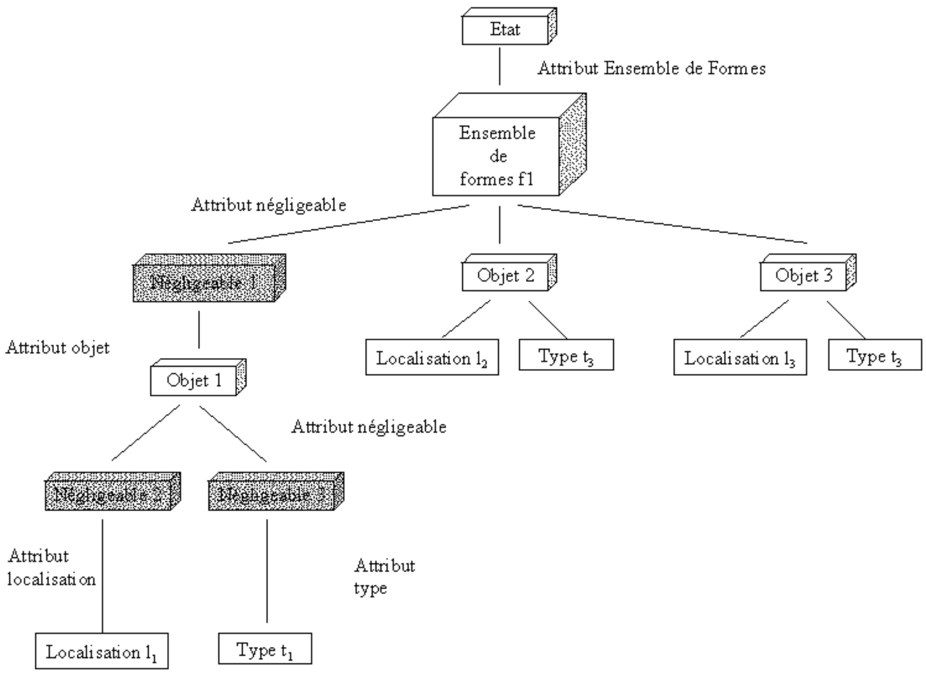


FIG. 8.9: *Architecture de Soar*

```
P1:  (State <s> features <f1>)
     (features <f1> objet <obj1>)
     (objet <obj1> localisation <l1>)
     (objet <obj1> type <t1>)
     (features <f1> objet <obj2>)
     (objet <obj2> localisation <l2>)
     (objet <obj2> type <t2>)
     (features <f1> objet <obj3>)
     (objet <obj3> localisation <l3>)
     (objet <obj3> type <t3>)
     ⇒
     qqch
```

on associe la production similaire P_1' muni de son ensemble de négligeables de la manière suivante, conformément au schéma 8.9:

```
P1':  (State <s> features <f1>)
      (features <f1> objet <obj2>)
      (objet <obj2> localisation <l2>)
      (objet <obj2> type <t2>)
      (features <f1> objet <obj3>)
      (objet <obj3> localisation <l3>)
      (objet <obj3> type <t3>)
      ⇒
      qqch,
      (features <f1> negligible <n1>)
      (negligeable <n1> objet <obj1>)
      (objet <obj1> negligible <n2>)
      (negligeable <n2> localisation <l1>)
      (objet <obj1> negligeable <n3>)
      (negligeable <n3> type <t1>)
```

Prise en compte des éléments négligeables dans le cas des productions "operator proposal" ou "operator application"

Les productions des deux types précédents se distinguent par le fait qu'elles proposent ou appliquent un opérateur. Ce dernier doit donc être proposé ou augmenté avec les négligeables que sa création a nécessité. Aussi, chaque opérateur est augmenté par la méthode qui vient d'être décrite de l'ensemble de négligeables qui a permis de le proposer ou de l'appliquer. L'exemple suivant, extrait du programme de [Rousseau 1998], donne un exemple de règle similaire pour une production "operator proposal" qui reconnaît un aéroport militaire. Dans cet exemple, les négligeabilités $n0$ et $n1$ négligent respectivement un objet aire de stationnement de grande taille et sont associées à l'opérateur $< o1 >$.

```
sp {operateur*reconnaissance*aeroport*militaire*pour*gros*porteurs*similarity-68
    (state <s> ^features <f> ^features <f>)
    (<f>  ^objet <obj5> ^objet <obj3>
          ^objet <obj2> ^objet <obj1>)
    (<obj5> ^type hangars ^location <l>)
    (<obj3> ^location <l> ^rayon-de-courbure grand ^type taxiway)
    (<obj2> ^type aeroport-militaire ^location <l>)
    (<obj1> ^location <l> ^type piste
          ^largeur { > 50 <l*2> } ^largeur <l*2> ^longueur { > 3000 <l*1> }
          ^longueur <l*1>)
    -->
    (<f> ^negligible <n1> + ^negligible <n1> &)
    (<n1> ^objet <obj4> & ^objet <obj4> +)
    (<obj4> ^negligible <n0> + ^negligible <n0> &)
    (<n0> ^location <l> & ^location <l> + ^forme lineaires &
           ^forme lineaires + ^taille grandes & ^taille grandes +
           ^type aires-de-stationnement & ^type aires-de-stationnement +)
    (<s> ^operator <o1> + ^operator <o1> =)
    (<o1> ^negligible <n1> & ^negligible <n1> + ^negligible <n0> &
           ^negligible <n0> +
           ^name reconnaitre-un-aeroport-militaire-pour-gros-porteurs +
           ^strategique tres-important + ^location <l> + ^important hangars +
           ^important hangars & ^important aires-de-stationnement +
           ^important aires-de-stationnement & +)
}
```

Les autres types de productions ne font qu'utiliser les résultats produits par les productions "state elaboration", "operator proposal" et "operator application". Il est donc inutile de leur faire prendre en compte de nouvelles négligeabilités.

Il ne nous reste donc plus que le dernier point de la sémantique de point fixe à prendre en compte, à savoir la prise en compte des cas où un élément est déduit à partir de sa propre négligeabilité. Pour ce faire, il faut pouvoir associer à chaque tête de clause l'ensemble des négligeabilités de ses prémisses.

Considérons l'exemple suivant: une règle similaire permet d'inférer un aéroport à partir, par exemple, d'une aérogare, d'une piste et d'un taxiway en négligeant le fait qu'il existe une tour de contrôle. Une deuxième règle permet d'inférer à partir de la connaissance d'un aéroport et de la taille de sa piste qu'il s'agit d'un aéroport militaire pour bombardiers. Alors intuitivement l'aéroport militaire a lui aussi été inféré en négligeant la tour de contrôle. La prise en compte des cas où un élément est déduit de sa propre négligeabilité est alors immédiate en prenant en compte ce type de transmission.

Dans le cas des productions *state elaboration*, les négligeabilités des prémisses sont des négligeabilités portant sur l'état courant, que la partie action de la production vient augmenter. Il est donc inutile de retransmettre des négligeabilités qui sont déjà sur l'état.

Dans le cas des productions *operator proposal* ou *operator application*, transmettre les négligeabilités revient à transmettre l'union des négligeabilités des prémisses de l'opérateur. Les négligeabilités seront alors transmises comme des augmentations de cet opérateur. Pour réaliser cette transmission, nous avons construit des productions spéciales *transmission production* de la manière suivante:

– les prémisses de ces productions contiennent l'état courant, ainsi que l'opérateur sur lequel va être transmis la négligeabilité, ainsi que les objets instantiés qui se trouvaient en prémisses de la règle *operator proposal* ou *operator application* et dont l'une des augmentations est une négligeabilité,

– la partie action de cette production contient l'opérateur courant augmenté de ces négligeabilités

Le déclenchement de cette production est immédiat au prochain cycle d'appariement (c'est en effet une proposition), permettant ainsi de transmettre les négligeablités.

Finalement, tout fait inféré se voit donc associé les négligeabilités qui ont permis de l'inférer (les jutifications associées du point de vue TMS [DeKleer 1986].

8.3.3 Utilisation du mécanisme de sous-buts pour la similarité

Comme nous l'avons mentionné, l'utilisation du mécanisme de sous-buts de Soar permet de prendre en compte l'implémentation du raisonnement par similarité. En effet, si les préférences du décideur sont complètes sur le problème étudié, lorsque le système doit choisir entre deux opérateurs candidats pour le slot opérateur, il a des préférences lui permettant de choisir l'un des deux. Dans ce cas, le système Soar implémente le raisonnement par similarité version satisfiabilité forte.

Lorsque, par contre, le système Soar n'a pas de préférence lui permettant de choisir entre deux opérateurs qui ont été inférés tous les deux avec des négligeabilités, les deux opérateurs sont en conflit pour le slot opérateur. Ce conflit engendre, comme nous l'avons vu, un sous-but dans lequel ces deux opérateurs sont passés en paramètres: on peut alors mener deux branches de raisonnement parallèles dans ce sous but en choisissant l'un ou l'autre de ces deux opérateurs (c'est un "operator subgoal" au sens de Soar [Laird et al. 1986]). Ces deux opérateurs ayant été inférés avec des négligeabilités, chacune des deux branches de raisonnement correspond à un raisonnement conditionnel à ces négligeabilités: il s'agit donc d'une implémentation du raisonnement par similarité version satisfiabilité faible, le système Soar déterminant **automatiquement** les ensembles de négligeables déterminant les hypothèses de raisonnement.

8.4 Conclusion

Nous avons donc montré dans ce chapitre comment le raisonnement par similarité en fonction de préférences peut être implanté dans le système Soar. Nous avons présenté la manière dont la similarité était prise en compte par la pré-compilation de règles similaires, ainsi que par l'introduction dans Soar des mécanismes permettant de prendre en compte les négligeables. Le système Soar modifié peut alors être décrit par l'algorithme suivant:

```
begin
  while (HALT not true) CycleDécision; end
  CycleDécision
  begin
    PhaseElaboration;
    PhaseDécision;
  end
  PhaseElaboration
  begin
    while (lance ou rétracte des productions) CycleElaboration; end
  end
  CycleElaboration
  begin
    PhasePreference;
    PhaseWorkingMemory;
  end  PhasePreference
  begin
    Lance les productions nouvellement unifiables;
    Lance la production similaire qui néglige le moins
    Rétracte les productions qui ne sont plus unifiables;
    Rétracte les productions similaires qui néglige des faits vrais
  end
  PhaseWorkingMemory
  begin
    EvaluationPréférence;
    Ajout et retrait des éléments dans la WorkingMemory;
    Ajout des productions de transmission des négligeabilités
  end
  PhaseDécision
  begin
    for  (chaque état dans la pile de buts et tant que pas de nouvelle décision)
        EvaluerOpérateurPréférence;
        if NouveauSousÉtat SelectionOpérateurOuImpasse; end
        if (Opérateur précédent terminé)
          SelectionOpérateurOuImpasse;
        end else Impasse;
    end
  end
  SelectionOpérateurOuImpasse
  begin
    if (un seul opérateur après évaluation des préférences)
      SelectionNouvelOpérateur;
    end
  end
  Impasse
  begin
    CréerNouvelEtat
  end
end
```

TAB. 8.1: *Une version simplifiée de l'algorithme de Soar modifié*

9

Application à la photo interprétation

Selon [Brunet et al. 1992], la photo-interprétation est "*un travail d'analyse exécuté à partir de photographies simples ou stéréoscopiques. C'est en grande partie un travail de recherche de formes et même de modèles structurels, qui demande un certain entraînement*". Pour effectuer ce travail, nous allons voir la manière dont procède le thématicien pour interpréter l'image satellite, le type d'opérations qu'il effectue, le type de connaissances utilisées ainsi que les raisonnements effectués sur l'image. Ce mode de travail est dépendant de deux paramètres principaux: la résolution et les buts de l'utilisateur. Avec une grande résolution (par exemple une image aérienne de résolution métrique), le levé de bâtiments est immédiat, puisque les bâtiments sont directement visibles sur l'image. Mais il en est de même avec une image SPOT (résolution 20mx20m), si le but de l'utilisateur est de relever des côtes d'un pays. L'approche que nous présentons est liée à une image de type SPOT (résolution 20mx20m) et le but de l'utilisateur est de reconnaître des objets présents sur l'image. Nous verrons donc dans un premier temps les principes de la télé-détection sur lesquels repose la photo-interprétation, puis la manière dont une image satellite est interprétée. Nous verrons ensuite la manière dont les systèmes existants modélisent la tâche de photo-interprétation. Nous en verrons les limites et la modélisation que nous en proposons.

9.1 L'image satellite: principes de la télé-détection

La télé-détection se définit comme l'ensemble des connaissances et techniques utilisées pour déterminer les caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci (J.O. du 11/12/1980). Il s'agit en fait d'étudier les comportements spectraux des objets par des capteurs (appareils photographiques, radar, ...) qui peuvent être embarqués sur un ensemble de vecteurs (avions, satellites, ...). Le spectre électromagnétique [Le Ber 1993] est divisé en différents domaines:

1. L'ultraviolet: $\lambda < 400nm$, où λ désigne la longueur d'onde,

2. Le visible: $400nm < \lambda < 700nm$,

3. L'infrarouge:

   a Proche $70nm < \lambda < 3\mu m$,

   b Moyen $3\mu m < \lambda < 200\mu m$,

   c Lointain $200nm < \lambda < 1mm$,

4. Les micro-ondes: $\lambda > 1mm$.

Les objets de la surface terrestre ont le pouvoir de réfléchir une partie du rayonnement solaire dans le domaine du visible et du proche infrarouge, et d'émettre un rayonnement dans le moyen infrarouge. C'est ces flux que perçcoivent les récepteurs de télé-détection dits <u>passifs</u>. Les systèmes de télé-détection dits <u>actifs</u> émettent et reçoivent du rayonnement, comme par exemple les radars dans le domaine des micro-ondes. Le flux est reçu par le satellite au moyen de capteurs qui enregistrent le flux transmis par une surface élémentaire (le pixel de l'image) dans différents intervalles de spectre (les canaux du satellite). En ce qui concerne le satellite SPOT, les bandes spectrales sont les suivantes:

	SPOT
Bandes Spectrales	1. 500 - 590 nm 2. 615 - 680 nm 3. 790 - 890 nm Panchromatique 510 - 730 nm

SPOT a une résolution de 20mx20m, et dispose des trois canaux dans le visible et le proche infrarouge; le canal panchromatique n'apporte pas d'information radiométrique supplémentaire malgré un pixel à 10 m. Ayant vu le principe d'obtention des images, voyons maintenant la manière dont un photo-interprète les appréhende.

9.2 Photo-interprétation d'images satellites

La photo-interprétation est un problème de classification, mais s'inscrit dans le cadre d'un problème de décision. C'est un problème de classification puisque, connaissant un certain nombre de formes élémentaires dans une image, le photo-interprète recherche un ensemble de classes d'objets dont ces formes sont parties (par exemple, un aéroport qui se déduit de la présence d'une aérogare, d'une tour de contrôle, de pistes ...). Il s'insère comme première étape dans un problème de décision: ainsi, dans le cadre de la photo-interprétation militaire, des actions militaires peuvent être entreprises en fonction des objets présents dans les images aboutissant à certaines conséquences sur lesquelles le décideur peut avoir des préférences (par exemple, le non bombardement d'un aéroport supposé militaire induit comme conséquence des attaques aériennes possibles et peut être préféré au bombardement d'un aéroport supposé civil induisant comme conséquence un certain nombre de victimes civiles). C'est un problème de décision et de classification en milieu incertain, puisque les connaissances que nous allons maintenant voir et qui permettent de reconnaître un objet dans l'image ne sont pas toujours **toutes** présentes dans l'image. Nous verrons ensuite les étapes principales que suit le processus de photo-interprétation.

9.2.1 Les connaissances du photo-interprète

Deux types de connaissances sont utilisées par le photo-interprète: les connaissances extrinsèques à l'image ou les connaissances exogènes, et les connaissances intrinsèques à l'image.

9.2.1.1 Les connaissances exogènes

On distingue, à l'instar de [Rousseau 1998], trois types de connaissances exogènes liées respectivement:

- au but général dont l'interprétation n'est qu'une étape de résolution. Ce but général peut par exemple être le renseignement militaire, l'étude du ruissellement des eaux. [Rousseau 1998] appelle ce type de connaissance "*contexte général de tâche*".

- au but que l'on veut atteindre à la fin du traitement, et qui est un sous but du but général. Ce but peut par exemple être de détecter les sites stratégiques dans le cas du renseignement militaire, de chercher le sens de la pente, la perméabilité du milieu ou les obstacles naturels pour l'étude du ruissellement des eaux. [Rousseau 1998] appelle ce type de connaissance "*contexte de tâche*".

- à la connaissance du milieu étudié, connaissance notamment du milieu naturel puisqu'en télédétection, tous les phénomènes perceptibles ne le sont qu'au travers de leur impact sur la couche superficielle de la surface terrestre, connaissance également du contexte géopolitique, culturel ou géophysique.

9.2.1.2 Les connaissances intrinsèques à l'image

Ces connaissances sont liées avant tout aux connaissances descriptives sur les objets du monde. Ces connaissances descriptives peuvent se décomposer, pour chaque objet, en la connaissance des attributs suivants:

La forme. La notion de forme regroupe non seulement l'idée d'enveloppe extérieure, mais aussi tout ce qui a trait à la description géométrique de cet objet, par exemple, la hauteur, la géométrie plus ou moins régulière de ses contours, le rayon de courbure pour un détail linéaire ... Cette notion est très utilisée dans l'urbanisme de détail: l'identification des différents bâtiments qui constituent le bâti urbain nécessite souvent le recours à la reconnaissance des formes de ces bâtiments;

La radiométrie. Elle est liée, pour un capteur passif, à la quantité d'énergie émise par un corps en fonction de sa réflectance et d'un filtre $\int reflec(\lambda) filtre(\lambda) d\lambda$. Physiquement, la radiométrie dépend essentiellement de trois facteurs:

- les matériaux composant l'objet, qui induisent des phénomènes d'absorption et de réflexion dans les différentes longueurs d'onde,
- la sensibilité des capteurs à une certaine bande de longueur d'ondes,
- la fonction d'illumination qui dépend de la source lumineuse éclairant la scène.

La structure. Il s'agit du mode d'agencement des formes géométriques. On distingue différents types de structure comme des alignements de points parallèles, des traits parallèles, des structures circulaires concentriques, des quadrillages ...

La texture. C'est l'aspect de surface que prend un objet sur une image. C'est une notion expérimentale qui ne peut être définie avec précision; tout au plus est-on à même de lui accoler un adjectif la représentant: hétérogène, rugueuse ... ou la comparer à un matériau à l'aspect de surface caractéristique (velours, verre ...). C'est une notion relative qui est directement liée à la résolution de l'image. La réduction de la résolution peut entraîner que les limites de certaines structures deviennent des textures.

Le contexte. Il s'agit de l'ensemble des informations extrinsèques à un objet permettant d'influencer ou de guider son identification. On peut le décomposer en:

contexte de perception : il inclue des informations sur la prise de vue,

contexte spatial : il inclue l'ensemble des relations spatiales d'un objet avec d'autres objets.

La vocation. Il s'agit d'une connaissance liée:

- au domaine d'emploi d'un objet. Ainsi, par exemple, un site missile sert à défendre un autre site,
- aux moyens à mettre en oeuvre pour réaliser la fonction d'un objet: par exemple, la présence d'eau près d'une centrale nucléaire.

Ces connaissances sont utilisées dans le processus d'interprétation, processus dont nous allons maintenant présenter les étapes.

9.2.2 Étapes du processus d'interprétation

On ne déchiffre une image qu'en établissant des corrélations entre cette image et les objets observés dans le monde réel connu par le photo-interprète. La photo-interprétation se décompose en plusieurs étapes [Girard 1975][Desachy 1991]:

1. Une première analyse de l'image permet de déterminer les structurations de l'espace engendrée par les changements de radiométrie ou de texture présents sur l'image. Cette phase permet au photo-interprète de choisir dans l'image la zone qu'il juge la plus *intéressante*, sur laquelle il va désormais focaliser son étude. L'intérêt est déterminé par le fait que la reconnaissance d'objets est plus facile dans une zone que dans une autre;

2. Cette étape est une étape d'analyse qui va révéler des formes pour lesquelles des corrélations directes peuvent être effectuées avec le monde connu (ex: routes, ville ...),

3. Dans cette étape, l'observateur a reconnu certains éléments du monde ordinaire, pour rechercher d'autres éléments dont les corrélations avec le monde connu sont plus difficiles à établir, mais pour être sûr de son identification, il lui faut trouver un plus grand nombre de corrélations et que celles-ci convergent,

4. L'observateur a pu identifier plusieurs éléments du monde inconnu dans l'image. Il donne alors un nom à chaque élément: c'est la phase d'identification,

5. L'observateur a identifié son monde avec celui de l'image. Il va tirer des conséquences avec sa logique des nombreuses corrélations qu'il a établies. C'est à partir de cet instant que les erreurs peuvent devenir plus nombreuses car l'observateur ne pourra plus vérifier systématiquement si toutes les relations qui devraient exister entre les objets existent réellement,

6. L'observateur imagine un monde qui est à la rencontre du monde réel et du monde de l'image, c'est à dire qu'il prélève des éléments dans divers ensembles et qu'il réunit dans un nouvel ensemble dont la structure est nouvelle, mais dont les éléments sont anciens. Les éléments viennent de l'image, la structure de la spécialité de l'observateur. C'est à ce point que se situe l'interprétation.

On peut noter que ces trois dernières étapes manipulent des objets qui ont, pour certains d'entre eux, un statut épistémique d'hypothèses plutôt que de connaissances sûres. Les objets identifiés le sont avec incertitude, et la phase d'interprétation combine ces informations incertaines. L'hypothèse que l'on pose est que le raisonnement du photo-interprète est un raisonnement par similarité: le photo-interprète utilise des connaissances en mémoire permettant de reconnaître des objets, et effectue des similarités par rapport à ceux-ci.

On peut résumer cette approche en trois phases principales de travail:

La détection c'est l'association de formes ou de textures à un même objet (elle comprend donc les phases 1, 2, 3),

L'identification C'est la reconnaissance de l'identité d'un objet (elle correspond donc à la phase 4 précédente),

L'interprétation Elle tend à prêter des propriétés aux objets identifiés, propriétés qui ne sont pas visibles dans l'image ou à supposer l'existence d'objets qui ne sont pas directement visibles (par exemple une concentration de routes dans un désert peut laisser supposer la présence d'un puits). Elle comprend les phases 5 et 6 précédentes.

À ces trois phases du travail du photo-interprète correspondent trois phases dans les systèmes d'interprétation informatiques existants [Le Ber 1993][Sandakly 1995]:

- une partie de **bas niveau**, qui contient toutes les implémentations d'algorithmes de traitement d'images (extraction de contours par ex.),

- une partie de **haut niveau**, qui contient les connaissances d'interprétation en fonction d'un but et de l'état du monde, et les connaissances sur les différents objets du monde sous forme symbolique,

- une partie de **niveau intermédiaire** qui contient les connaissances permettant de présider à l'identification, c'est à dire toutes les connaissances permettant de faire un lien entre les primitives images, issues des traitements de bas niveau, et les représentations symboliques qui servent d'entrées à la phase de haut niveau.

C'est selon ces trois niveaux que se structurent les systèmes d'interprétation existants. Issus cependant du domaine de la vision informatique, ces systèmes se focalisent en général sur la phase de bas niveau et la phase de niveau intermédiaire, leur utilisation du haut niveau étant très réduites et se limitant la plupart du temps à la prise en compte de certaines relations spatiales.

9.2.3 Étapes de raisonnement et connaissances utilisées

Comme nous l'avons vu, les étapes de raisonnement 5 et 6 nécessitent, de la part du photo-interprète, des connaissances ayant un statut épistémique d'hypothèses. Les hypothèses doivent vérifier certaines contraintes vis-à-vis des objets auxquels elles sont liées, et vis-à-vis plus particulièrement de leurs caractéristiques contextuelles et de vocation. Deux types de liens peuvent être distingués:

- ceux qui lient usuellement deux types d'objets particuliers,

- ceux qui excluent dans un certain voisinage la présence d'un type d'objet particulier,

Ainsi, par exemple, un site missile est lié à un site stratégique qu'il protège. Un site à munitions exclut dans son voisinage une grande route. On trouvera dans [Rousseau 1998] annexe 1 (pp 37-40) un ensemble de relations contextuelles et de vocation liant les objets utilisés dans le cadre de l'expérimentation menée.

Si un lien d'exclusion n'est pas vérifié avec un objet validé (objet reconnu avec certitude), alors l'hypothèse est rejetée. Par contre, des liens d'exclusion entre hypothèses n'entraînent pas de rejet, mais rend les hypothèses concernées incompatibles ou concurrentes. C'est le cas par exemple des hypothèses émises sur une même région (elle ne peut correspondre qu'à une seule entité), mais c'est aussi le cas des hypothèses en exclusion spatiale (une hypothèse ne peut se trouver dans le voisinage d'une autre) sont aussi incompatibles. Si l'on suppose que le photo-interprète raisonne par similarité, il raisonne alors suivant le mode de raisonnement par similarité version satisfiabilité faible, puisque les hypothèses correspondent alors à des hypothèses de négligeabilité exclusives les unes des autres.

9.3 Les systèmes d'interprétation existants

Le premier type de système intelligent en interprétation d'images concerne des systèmes intelligents en traitement d'images (partie bas niveau de l'interprétation). Dans ce type de système, les connaissances manipulées [Desachy 1991] par le système sont les connaissances que l'on a sur les capacités des divers modules de traitement d'images à résoudre certains problèmes particuliers. En général, le système dispose d'une librairie de modules dont il connaît parfaitement les capacités particulières. Devant un problème de traitement confié au système, celui-ci propose un processus complet d'activation de divers modules pour résoudre le problème posé. On trouvera dans [Desachy 1991] une revue des différents systèmes jusque là proposés pour effectuer cette partie de traitement de bas niveau. Étant donné notre but d'interprétation, ce type de modélisation se situe en amont de la modélisation que nous proposons.

En effet, tenant compte de la résolution du type de scène que nous nous proposons d'interpréter et des buts de l'interprétation, l'objectif à atteindre est la recherche d'objets particuliers (véhicules, routes, bâtiments) et l'analyse va être poussée jusqu'à la reconnaissance d'un type global de scène. Nous nous concentrerons donc sur les parties niveau intermédiaire et niveau haut. Plusieurs systèmes ont été proposés pour réaliser ces différentes tâches, chacun d'entre eux se différençiant des autres par son architecture, la manière dont il représente les objets de l'image et la manière dont est gérée la stratégie d'interprétation et de détection d'objets.

Le système SACAP [Nagao et al. 1979][Nagao and Matsuyama 1980] est dédié à l'interprétation d'images aériennes. L'architecture du système est basée sur un modèle de tableau noir. L'image est d'abord pré-traitée, puis segmentée et chaque segment va être associé à une table de propriétés (ces propriétés étant calculées par des modules particuliers). Les sous-systèmes de détection d'objets (1 module par objet) vont alors examiner chaque segment et ses propriétés. On peut résumer l'architecture de ce système dans le schéma de la figure 9.1. Le blackboard mémorise tous les objets trouvés et devient la nouvelle base de données pour la recherche de nouveaux objets. L'ensemble est piloté et contrôlé par un système qui intègre la résolution de conflits entre sous systèmes de détection d'objets (c'est le rôle du contrôleur). Ce système se cantonne à la modélisation du niveau intermédiaire puisque la stratégie unique de détection et d'identification d'objets est liée à des critères radiométriques. La connaissance sur le contexte et sur les structures des objets n'intervient que très peu.

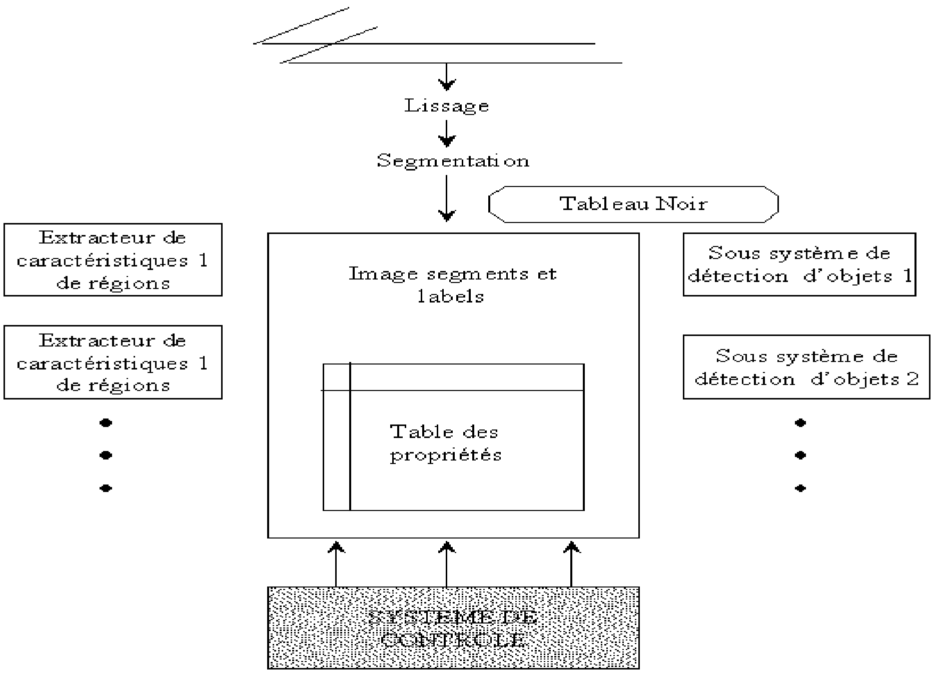


FIG. 9.1: *Le système SACAP (Extrait de [Nagao et al. 1979])*

Le système SPAM (System for Photo-interpretation of Airports Using Maps) a été développé au départ pour l'interprétation des scènes d'aéroport à partir des images aériennes [McKeown et al. 1985], et a été étendu pour traiter des scènes suburbaines [McKeown et al. 1989]. La figure 9.2 montre l'architecture de SPAM. Quatre modules de traitement d'images permettent l'extraction des pri-

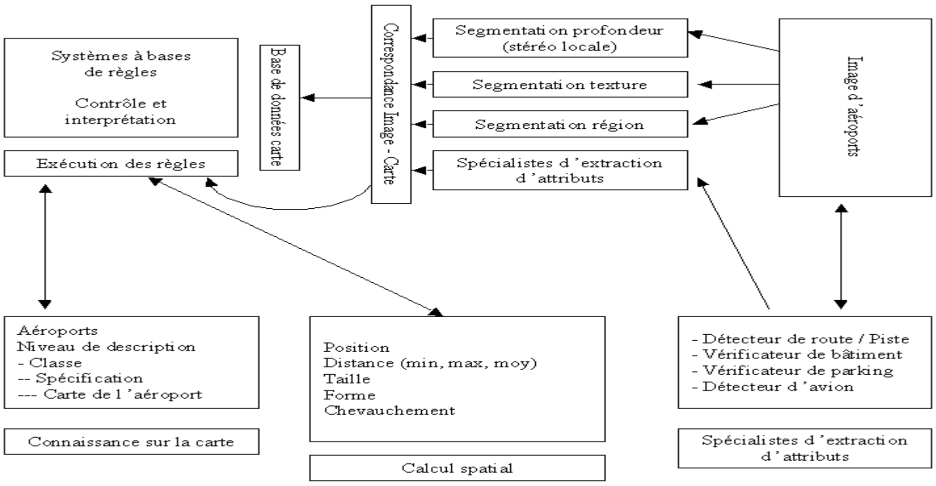


FIG. 9.2: *Le système SPAM (Extrait de [McKeown et al. 1989])*

mitives: un module de segmentation par croissance de régions, un module de détection de formes linéiques, un module d'analyse de cartes de profondeur par stéréo, un module interactif de segmentation d'images. Les primitives sont transformées en une carte (base de données carte) sur laquelle raisonne un système à base de règles contenant la connaissance déclarative sur le contrôle de l'interprétation des scènes d'aéroport. Les classes d'objets sémantiques utilisées pour l'étiquetage des régions dans SPAM sont des sous classes de classes génériques: par exemple, une piste ou une route d'accès sont des sous classes de la classe linéique, un parking est une sous classe de la classe petite tâche. D'autre part, des objets (structures de données) sont définies dans SPAM:

Les régions qui sont les formes extraites des images, contenant des informations comme leur texture, leur forme, ...

Les fragments qui sont des identifications d'une région extraite de l'image, chaque fragment contenant des informations de la classe et des sous classes présidant à son identification,

Les zones fonctionnelles qui sont des sous-parties d'une scène d'aéroport. Elles correspondent à un ensemble d'objets se trouvant généralement à proximité les uns des autres pour des raisons fonctionnelles,

Les modèles qui sont des ensembles cohérents de zones fonctionnelles pouvant correspondre à un aéroport.

Quant à la stratégie d'interprétation, elle est résumée par le schéma suivant: Après avoir classifié

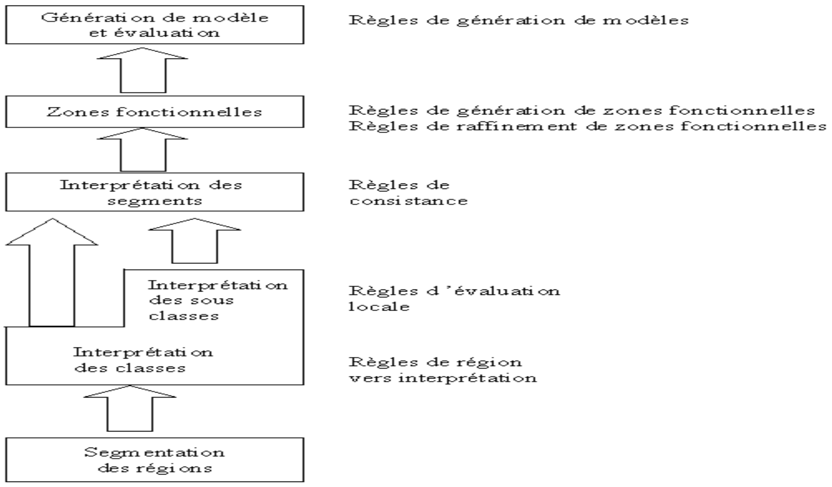


FIG. 9.3: *Le système SPAM (stratégie d'interprétation) (Extrait de [McKeown et al. 1989])*

les régions de l'image, les avoir évalués localement en testant leurs relations spatiales, des zones fonctionnelles sont crées et le modèle global est évalué. Ce système présente l'avantage de prendre en compte en partie le niveau haut par l'utilisation des liens spatiaux et des zones fonctionnelles.

Le système VISIONS présente une architecture en trois niveaux d'abstraction détaillés par le schéma figure 9.4 [Hanson and Riesman 1987]. Le bas niveau correspond aux informations produites par les capteurs; c'est à ce niveau que se situent l'extraction des primitives images. Au niveau intermédiaire correspond un ensemble de primitives symboliques telles que des fragments ou des régions. Ces primitives sont stockées dans la base de données ISR, qui permet de stocker les primitives sous forme de structures de données qui ressemblent aux frames. Dans le haut niveau, la connaissance est encapsulée dans des schémas [Draper et al. 1989] qui représentent chacun un concept de haut niveau. Chaque concept contient toutes les connaissances nécessaires à la reconnaissance des objets de la classe. La communication entre les différents schémas passe par l'intermédiaire d'un blackboard global. La stratégie d'interprétation commence par l'instantiation du schéma de plus haut niveau correspondant à la scène (le schéma scène routière, scène d'aéroport ...). Chaque schéma instantié met en oeuvre ses différentes stratégies pour détecter les objets de sa classe en localisant leurs régions dans l'image ou en activant des schémas des sous parties qui, à leur tour peuvent activer d'autres schémas. La figure 9.5 donne l'exemple d'une instance d'un schéma: à chaque schéma est associé une stratégie OHM (Objet Hypothesis Maintenance). Lorsqu'un schéma est instancié, la stratégie OHM est activée pour surveiller le tableau noir global et pour déclencher les différentes stratégies en fonction des données sur le tableau noir. La figure 9.5 présente les sept différentes stratégies associées à la vie d'une hypothèse.

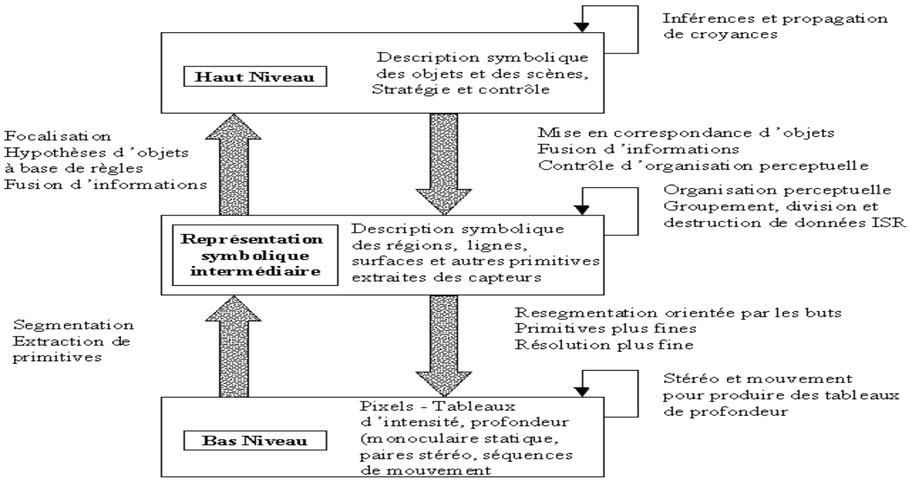


FIG. 9.4: *Le système VISIONS (Extrait de [Draper et al. 1989])*

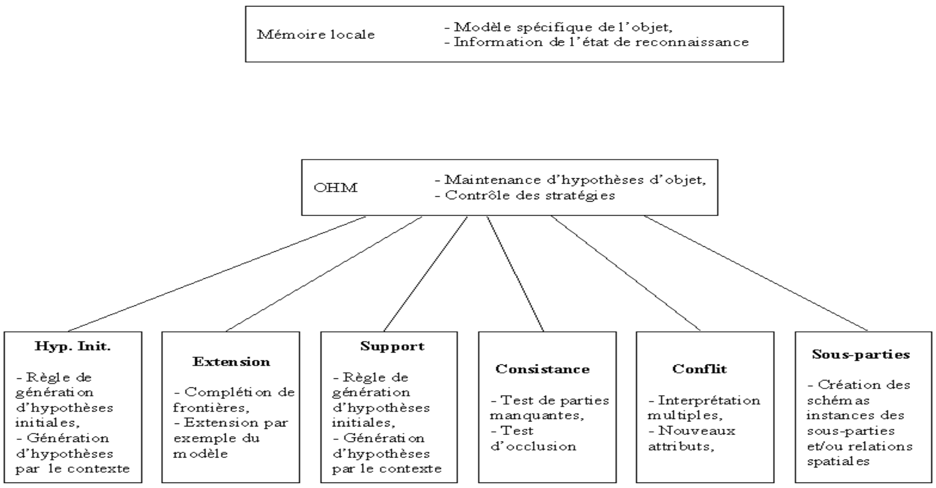


FIG. 9.5: *Le système VISIONS (Extrait de [Hanson and Riesman 1987])*

Le système SIGMA présente une architecture représentée par trois experts selon le schéma suivant (voir figure 9.6) [Matsuyama 1990].

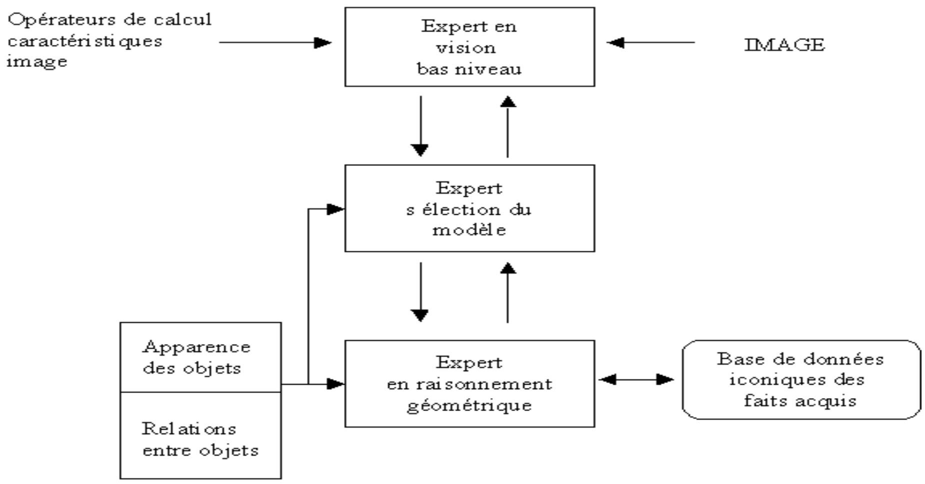


FIG. 9.6: *Le système SIGMA (Extrait de [Matsuyama 1990])*

- L'expert sur le raisonnement géométrique correspond à la connaissance nécessaire pour le raisonnement spatial entre les objets au niveau de la scène,
- L'expert de sélection de modèle correspond à la connaissance nécessaire pour le raisonnement sur l'apparence des objets dans la scène, et sur la transformation des données images en objets de haut niveau dans la scène,
- L'expert sur le bas niveau correspond à la connaissance nécessaire pour le raisonnement au niveau de l'extraction de primitives.

La structure principale de la représentation des connaissances dans SIGMA est la classe d'objets. Une classe SIGMA est inspirée à la fois du formalisme des frames et de celui des langages orientés objet. Une classe d'objets est constituée d'un ensemble de champs (slots), de liens et de règles. La stratégie d'interprétation est une boucle sur un cycle d'interprétation en trois phases. Une étape de segmentation précède cette boucle, et le cycle d'interprétation comprend:

- une phase de génération d'hypothèses,
- une phase d'accumulation d'évidence, où les hypothèses sont testées, où leur consistance est évaluée,
- la résolution de situations, qui permet:
  - si l'objet de l'hypothèse a déjà été trouvé, de valider celle-ci,
  - si l'objet de l'hypothèse n'a pas été trouvé, alors de déclencher un module de traitement d'images permettant de le rechercher

Le système ICARE a été développé pour l'interprétation des images satellites [Desachy 1991]. La figure 9.7 présente l'architecture de ICARE. Ce système raisonne tout d'abord avec des connaissances dites de niveau 1 (analyse du contexte particulier d'un pixel, considéré isolement des pixels voisins). Cette première étape permet donc de déterminer la meilleure classification possible pour chaque pixel. Cette étape étant effectuée, on peut alors réaliser la deuxième étape qui consiste à analyser chaque pixel en intégrant la connaissance des pixels voisins: forme, relation topologique avec une autre classe ... Suit alors un deuxième processus déductif qui permet de classer chaque pixel en fonction des connaissances et des données exogènes possibles. Le système ICARE est un système expert et intègre donc les connaissances sous forme de règles de production.

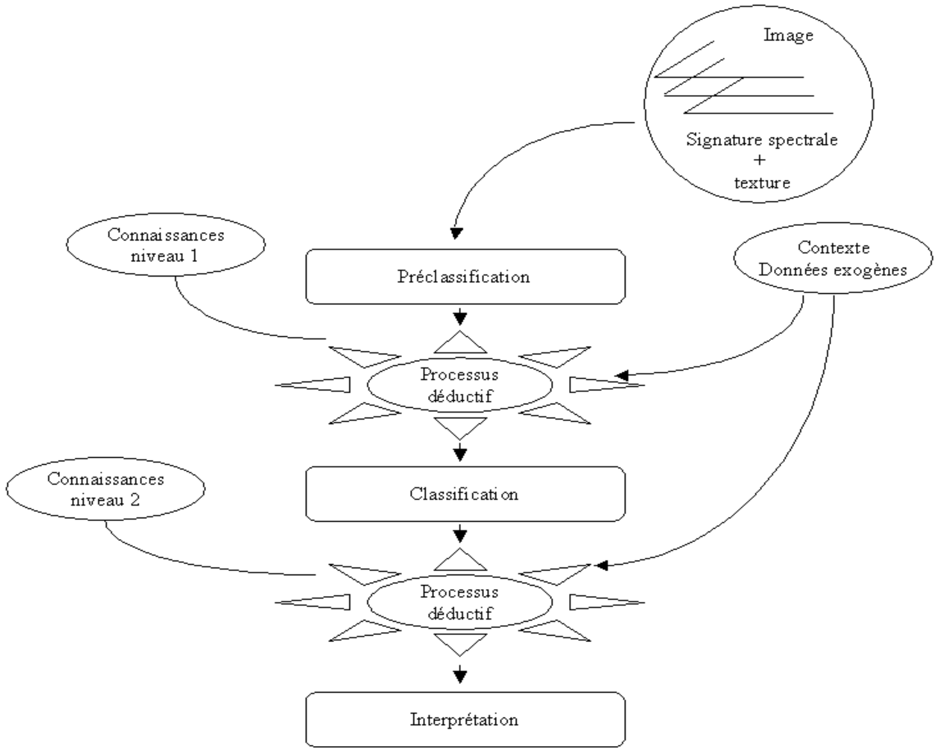


FIG. 9.7: *Le système ICARE (Extrait de [Desachy 1991])*

Le système MESSIE présente une architecture orientée suivant trois niveaux selon le schéma 9.8). L'architecture de MESSIE est un blackboard hiérarchique à contrôle par requêtes et événements. Les entités de traitement du blackboard coopèrent pour interpréter une scène. La connaissance de

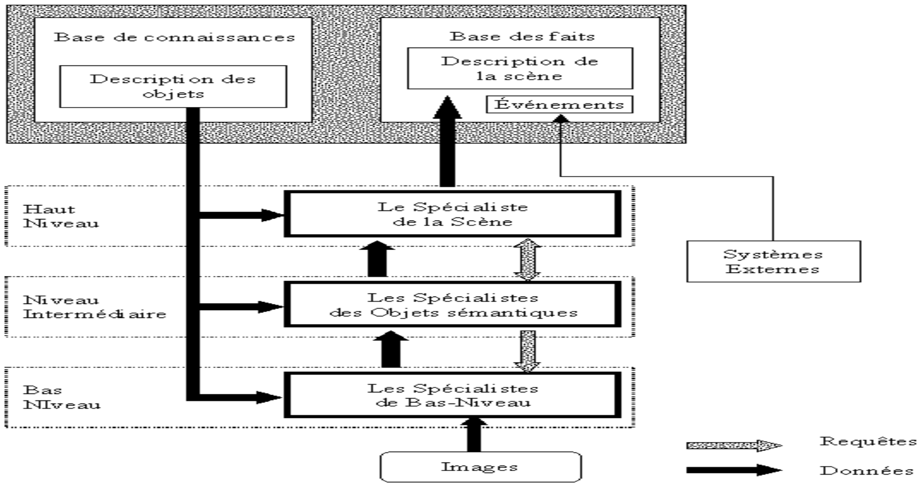


FIG. 9.8: *Le système MESSIE (Extrait de [Sandakly 1995])*

MESSIE est hiérarchique et constituée de trois niveaux d'abstraction: le bas niveau, le niveau intermédiaire et le haut niveau. A chaque niveau d'abstraction correspond des tâches et des spécialistes particuliers. Un spécialiste générique, le spécialiste de la scène, est responsable de la gestion de la stratégie d'interprétation, de la validation des objets détectés et de la résolution des conflits de localisation. Les connaissances sur les objets sémantiques peuvent être décrites selon quatre points de vue:

- la géométrie: elle est décrite par un ensemble de sous-parties reliées par des opérateurs de localisation,
- la radiométrie: elle est déduite à partir de la composition des objets en matériaux et de la sensibilité des capteurs aux aspects de ces matériaux,
- le contexte spatial: il est décrit d'une manière déclarative par un ensemble de propositions exprimant des relations spatiales entre les objets,
- La vocation permet la catégorisation des objets.

Les connaissances sont exprimées dans un langage d'héritage équivalent aux frames permettant le prototypage, l'expression de contraintes ainsi que des mécanismes de déclenchement automatique de traitements et de vérification de cohérence. Les stratégies des différents spécialistes sont implantées sous forme de règles. Ils communiquent par l'intermédiaire d'un tableau noir, un agent jouant le rôle de contrôleur.

Plusieurs idées clefs ressortent de la présentation de ces systèmes mettant en évidence quelques règles pour la construction d'un système d'interprétation. Comme la modélisation en étapes du raisonnement du photo-interprète avait permis de le montrer, l'architecture d'un système de photo interprétation doit être hiérarchique en intégrant au moins les trois niveaux d'abstraction reflétant les transformations nécessaires dans un système d'interprétation. Le système Soar, par son architecture sous forme d'espaces de problèmes, correspond bien à ce critère de modélisation. Au niveau des formalismes de représentation des connaissances, les systèmes que nous venons de décrire utilisent, soit les frames, soit des systèmes à règles de production. C'est cette deuxième approche que l'on utilisera dans le système Soar.

Il aurait également été possible d'utiliser des ATMS possibilistes [Monai and Chehire 1992][Benferhat et al. 1994] [Benferhat et al. 1997a] pour modéliser la démarche de photo-interprétation. Ces auteurs ont en effet présenté une extension des ATMS, appelée ATMS possibiliste dans le cadre de laquelle la prise en compte de l'incertitude est directement integrée dans l'ATMS. L'incertitude liée aux justifications ou aux suppositions graduées est prise en compte par la théorie des possibilités. Elle est propagée par l'intermédiaire d'une extension du principe de résolution. Ce mécanisme de raisonnement a été utilisé par [Monai and Chehire 1992] pour déterminer sur un champ de bataille le type de formation militaire en fonction de la nature, du nombre et des positions respectives des unités militaires élémentaires. Un problème de photo-interprétation étant aussi un problème de reconstruction de scénarios en fonction d'informations partielles, un ATMS possibiliste aurait pu être utilisé pour modéliser le problème de photo-interprétation. Notre approche plus orientée vers une sémantique de point fixe nous à conduit à privilégier l'emploi de Soar, d'autant qu'à l'analyse, il s'est avéré que les préférences de Soar pouvaient être très naturellement associées à celle de notre formalisme logique.

9.4 Modélisation dans le système Soar

Une modélisation des connaissances et des étapes du raisonnement dans le système Soar a été proposée par [Rousseau 1998]. Nous présenterons ci-après la modélisation et l'implantation effectuées.

9.4.1 Modélisation des connaissances dans Soar

Comme nous l'avons vu, les systèmes existants en photo-interprétation se concentrent plus en général sur la partie traitement d'images (étape de bas niveau du problème de photo-interprétation) plutôt que sur la partie interprétation. C'est au niveau intermédiaire et au haut niveau que se situe la modélisation que nous proposons, dans lequel le raisonnement est identifié à un raisonnement par similarité version satisfiabilité faible. Les connaissances intrinsèques à l'image sont modélisées par:

La forme Elle peut être exprimée sous forme de combinaisons d'objets géométriques simples, comme par exemple des rectangles, des cercles ...

La radiométrie Elle peut être exprimée sous forme d'intervalles de niveaux de gris,

La structure Elle peut être exprimée sous forme de combinaisons de structures élémentaires, comme par exemple alignement de points parallèle, structures circulaires ...

Ces connaissances intrinsèques ont été obtenues par interrogation d'un photo-interprète militaire expert. On trouvera ci-après deux exemples de descriptions d'objets (extraites de [Rousseau 1998] fournies par le photo-interprète interrogé, avec pour chaque élément de description un critère qualitatif d'importance dont la manipulation sera décrite ci-après:

- Site à munitions

Description du site	**Importance du critère dans la description**
bâtis circulaires espacés de x mètres entre eux	peu important
bâti organisé selon un schéma polygonal	Important
Clôture polygonale	Important
Voies de communication avec l'extérieur	Important
Bâti administratif de type gardiennage	Très important
Relations avec d'autres sites	**Importance du critère relation**
Proximité de sites militaires utilisant ces munitions	Important
Loin de sites civils	Peu important

- Sites à carburant

Description du site	**Importance du critère dans la description**
bâtis circulaires espacés de $x - dx$ [32] mètres entre eux	Peu Important
bâti organisé selon un schéma polygonal	Important
Clôture polygonale	Important
Voies de communication avec l'extérieur	Important
Relations avec d'autres sites	**Importance du critère relation**
Proximité d'une raffinerie	Tres Important
Site de consommation	Capital
Site de transport	Capital

Dans le système Soar, les connaissances relatives à la l'identification d'un objet sont implantées sous forme de règles de production. On trouvera dans [Rousseau 1998] et dans les sections suivantes des exemples de telles règles de production. Le critère d'importance fourni sur l'échelle qualitative $PeuImportant \prec Important \prec TresImportant \prec Capital$ permet d'ordonner facilement les ensembles de négligeables les uns par rapport aux autres en fonction de cet ordre. Dans le cas où plusieurs négligeabilités sont associées à la reconnaissance d'un objet, la fonction d'agrégation des importances utilisées est le min.

Les connaissances contextuelles et de vocation sont modélisées par le biais de leurs interactions sur le raisonnement. En effet, comme nous l'avons vu, ces connaissances permettent d'établir des liens entre les objets, par le biais de relations qui lient usuellement deux types d'objets particuliers, ou par le biais de relations qui excluent dans un certain voisinage la présence d'un type d'objet particulier. Ces

32. Mesure obtenue par comparaison avec sites à munitions.

connaissances ont été modélisées dans Soar sous forme de priorités (préférences), permettant de déterminer si une interprétation est meilleure qu'une autre:

- un objet o_1 est meilleur qu'un objet o_2 pour une identification ou une interprétation ssi o_1 possède un lien avec un objet requis dans une caractéristique spatiale ou de vocation et o_2 ne possède pas un tel lien,
- un objet o_1 qui exclue dans le voisinage la présence d'un objet particulier o_2:
  - est rejeté si un tel objet o_1 est déduit sans négligeabilités,
  - conduit à la création deux hypothèses concurrentes ssi o_1 et o_2 sont déduits avec négligeabilités (on retrouve alors les ensembles exclusifs les uns des autres de négligeables avec ceux qui ont permis de déduire respectivement o_1 et o_2).

9.4.2 Modélisation des étapes du raisonnement dans Soar

Nous avons présenté dans les sections précédentes les différentes étapes que suit le raisonnement du photo-interprète. Nous présentons dans cette section la correspondance dans Soar de ces différentes étapes sous forme d'organisation des connaissances dans différents espaces de problèmes [Rousseau 1998]:

Interpréter l'image Il correspond à la prise en compte par le photo-interprète du contexte général de tâche, du contexte de tâche (photo-interprétation militaire, ruissellement des eaux).

Focalisation sur une région de l'image Il permet de définir la zone de l'image qui le photo-interprète juge la plus intéressante pour commencer l'interprétation. On trouve dans cet espace de problème des connaissances permettant de sélectionner une zone comme intéressante en fonction de l'image (aérienne, radar, satellite ...), de la résolution ...

Détection des objets C'est l'étape qui va permettre de révéler des formes pour lesquels des corrélations peuvent être établies avec le monde réel: on associe donc un ensemble de formes ou de textures à un objet,

Identification des objets L'observateur donne un nom à chacun des objets qu'il a identifié dans l'image,

Interprétation C'est à ce niveau que sont intégrées les connaissances qui président aux liens entre hypothèses, qu'elles soient liées aux contraintes de vocation ou contextuelles. C'est à ce niveau que les différentes hypothèses exclusives les unes des autres sont crées, et à ce niveau également que les hypothèses non compatibles avec les connaissances validées sont utilisées.

Dans l'implantation, seuls les deux derniers espaces de problème ont été utilisés. Le raisonnement se déroule alors de la manière suivante: une étape d'identification (les descriptions permettant de reconnaître un objet sont encodées dans des règles de production) permet de reconnaître les objets potentiellement identifiables en fonction des négligeabilités. Les connaissances contextuelles et de vocation sont ensuite déclenchées pour prendre en compte ces relations.

9.5 Implantation

Les tests du système ont été effectués sur des images SPOT de la région de Sébastopol (Ukraine) et Mourmelon (France) par [Rousseau 1998]. Ces images présentent de nombreux sites militaires, et sont suffisamment bien connues des photo-interprètes pour que certains objets soient reconnus directement. Deux types de limitations ont été considérées dans cette implantation:

- d'une part, des limitations des connaissances utilisées. Nous supposerons en effet que la partie de bas niveau a été effectuée et nous utiliserons directement les résultats (connaissances symboliques) que doit donner une phase de détection. Ce résultat a été obtenu en interrogeant un photo interprète-expert. Par rapport aux connaissances descriptives que nous avons présentées, les connaissances relatives aux formes, aux structures, aux contextes spatiaux et aux vocations sont utilisées. Ces connaissances sont formalisées sous la forme de descriptions symboliques, qui peuvent être par exemple "ensemble de formes circulaires en l_1". Les connaissances liées à la radiométrie, aux textures et au contexte de prise de vue n'ont pas été utilisées,

- d'autre part, une restriction au niveau des étapes de raisonnement considérées. Seuls les espaces de problèmes "Identification des objets" et "Interprétation" ont été utilisés.

9.5.1 Description du programme

La structure du programme respecte les différentes étapes de raisonnement du photo-interprète. Ce programme, écrit par [Rousseau 1998], est décomposé suivant les différentes étapes suivantes:

- La première partie du programme consiste en la description de la base de faits (l'ensemble des connaissances relatives à l'image courante). Cette description comprend donc l'état, l'espace de problème, l'ensemble des objets qui ont été détectés dans l'image qui sont décrits sous la forme des caractéristiques forme, structure, contexte spatial (caractéristiques qui sont censées avoir été perçues par le module de bas niveau) et sous la forme d'une localisation qui est un attribut symbolique ("location") représentant l'endroit où les formes ont été perçues. Ainsi, par exemple, la production suivante décrit une image comportant un ensemble de formes *features*, où l'on trouve en l_3 un ensemble de formes circulaires en forme de fer à cheval, une clôture polygonale au même endroit ...

```
sp {description*image
    (state <s> ^superstate nil)
    (<s> ^name recognize
         ^problem-space <p>)
    (<p> ^name interpretation-image)
     -->
     (<s> ^features <f> + &)
     (<obj19> ^type ensemble-de-formes-circulaires
             ^forme fer-a-cheval
             ^location l3)
    (<f> ^objet <obj20> + &)
    (<obj20> ^type cloture
             ^forme polygonale
             ^location l3)
    (<f> ^objet <obj21> + &)
    (<obj21> ^type ensemble-de-formes-circulaires
             ^forme polygonale
             ^position-relative regulierement-espacees
             ^eloignement-relatif peu-proche)
```

- La deuxième partie du programme comporte les règles d'identification, c'est à dire les règles permettant de déduire de la coexistence d'un ensemble de formes au sein d'un objet la nature de celui-ci. Chaque identification d'un objet est implantée sous la forme d'un opérateur: ces règles sont celles qui donnent lieu à la création de règles similaires, et sont des opérateurs de type "*operator proposal*" dans Soar. Ainsi, par exemple, la production suivante permet d'identifier un site à munitions comme étant composé d'un ensemble de formes circulaires régulièrement espacées, proches les unes des autres, d'une clôture polygonale, d'un bâtiment de gardiennage et d'une voie de communication situés à un endroit donné ici désigné par l.

```
sp {operateur*reconnaissance*site*a*munition
    %{(state <s> ^features <f>)
      (<f> ^objet <obj1>
```

```
          ^hypothese <h>)
      (<obj1> ^type ensemble-de-formes-circulaires
              ^position-relative regulierement-espacees
              ^eloignement-relatif peu-proche
              ^location <l>)}
    %{(<obj1> ^forme polygonale)}
    %{(<f> ^objet <obj2>)
      (<obj2> ^type cloture
              ^forme polygonale
              ^location <l>)}
    %{(<f> ^objet <obj3>)
      (<obj3> ^type batiment
              ^fonction gardiennage
              ^location <l>)}
    %{(<f> ^objet <obj4>)
      (<obj4> ^type voie-de-communication
              ^extremite1 <l>
              ^extremite2 exterieur
              ^location <l>)}
    -->
    (<s> ^operator <o> + =)
    (<o> ^name reconnaitre-un-site-a-munition
         ^strategique tres-important
         ^location <l>
         ^hypothese <h>)}
```

– La troisième partie consiste à définir l'influence des relations de proximité ou de vocations sur la confiance accordée à la reconnaissance d'un objet. Ainsi, par exemple, la production suivante permet d'affirmer que l'action qui consiste à reconnaître une raffinerie loin d'un port pétrolier (i.e. loin d'une étendue d'eau) doit être rejetée.

```
sp {operateur*relations*raffinerie*1
    (state <s> ^features <f>
               ^operator <o> +)
    (<o> ^name reconnaitre-une-raffinerie
         ^confiance moyen
         ^location <l>
         ^hypothese <h>)
    -{(<f> ^objet <obj>
          ^hypothese <h>)
      (<obj> ^location <l1>
             ^matiere eau)
      (<f> ^proche-de <p>)
      (<p> ^location <l>
          ^location <l1>)}
    -->
    (<s> ^operator <o> -)}
```

– La quatrième partie sert à définir les conditions qui nécessitent ou non la création d'une hypothèse. Ainsi, par exemple, la production suivante affirme que la reconnaissance au même endroit d'un aéroport militaire et d'un aéroport civil nécessite la création d'une hypothèse.

```
sp {operateur*conflit*aeroport*civil*aeroport*militaire2
    :no-similarity
```

```
(state <s> ^features <f>
           ^operator <o1> +
           ^operator <o2> +)
(<o1> ^location <l>)
(<o2> ^location <l>)
(<o1> ^name <> reconnaitre-un-aeroport-civil)
(<o2> ^name reconnaitre-un-aeroport-militaire)
-->
(<s> ^operator <o2> = <o1>)}
```

– La cinquième partie permet d'écrire le niveau de confiance symbolique associé à chaque opérateur en fonction de l'évaluation de l'importance des éléments négligés dans sa description. Ainsi, par exemple, la production suivante permet d'écrire que la confiance accordée à la reconnaissance d'un objet sans négligeabilités est totale.

```
sp {operateur*ecrire*le*niveau*de*confiance*totale*
    :no-similarity
    (state <s>  ^operator <o> +)
   -(<o> ^negligible <neg>)
    -->
    (<o> ^confiance totale + &)}
```

– La sixième partie concerne les actions des opérateurs, c'est à dire l'identification des objets reconnus par l'opérateur qui a été proposé. Cette partie permet de fournir pour chaque objet reconnu l'endroit où il se trouve ainsi que le niveau de confiance et l'ensemble des négligeabilités qu'il a été nécessaire de créer pour l'identifier. Cette production permet d'affirmer que l'opération de reconnaître un site à munitions engendre la création d'un objet site à munitions muni des attributs (location, confiance) de l'opérateur.

```
sp {operateur*reconnaissance*site*a*munition*
    :no-similarity
    (state <s>  ^operator <o>)
    (<o> ^name reconnaitre-un-site-a-munition
         ^location <l>
         ^confiance <conf>
         ^hypothese <h>)
    (state <ss> ^features <f>)
    (<f> ^hypothese <h>)
    -->
    (<f> ^objet <obj> + &
         ^confiance <conf> + &)
    (<obj> ^type site-a-munition
           ^location <l>
         ^confiance <conf>
         ^negligible <neg>
           ^hypothese <h>)
    (<o> ^objet-reconnu <obj> + &)}
```

– La septième partie est celle concernant les créations d'hypothèses. Un opérateur devant créer une hypothèse crée celle-ci en lui assignant un ensemble de faits déduits sans négligeabilités (faits-averes) et un ensemble de faits déduits à partir de l'hypothèse.

```
sp {operateur*hypothese*creer
```

```
:no-similarity
(state <s> ^operator <o>)
(<o> ^hypothese <h>)
-->
(<s> ^features <f1> + &)
(<f1> ^hypothese <h>)
(<h> ^faits-averes <fa>
     ^faits-deduits-dans-l-hypothese <fdh>
     ^negligible <neg>)}
```

A la fin de l'exécution, le programme fournit plusieurs schémas d'interprétations incompatibles entre eux correspondant à des ensembles de négligeables incompatibles entre eux. La sémantique formelle permet d'assurer que les hypothèses fournies sont maximales au sens des priorités.

La mise en oeuvre informatique [Rousseau 1998] a été effectuée sur deux types de programmes différents: un programme fonctionnant avec une description assez complète des images afin de valider les règles par rapport aux connaissances fournies par le photo-interprète. Un deuxième programme inclut une base de faits minimale où l'on ne décrit que peu d'objets (le système ayant plus de négligeabilités et de similarités à effectuer). En ce qui concerne le premier programme, le système fournit en sortie les hypothèses correspondant aux différents scénarios d'interprétation, munies de la confiance associée aux éléments qui ont été négligés. Il a permis de montrer la validité du mode de raisonnement proposé puisque ces scénarios correspondent à un jeu d'essai effectué avec un photo-interprète. Le deuxième programme avait pour but de tester le traitement des connaissances dans le cas d'une base de faits plus réduite. Les résultats [Rousseau 1998] montrent que la réduction du nombre d'objets présents dans l'image intervient surtout sur les objets qui étaient auparavant de confiance faible, certains d'entre eux disparaissant même lors de la réduction de la base de faits. Nous présentons ici un exemple de fonctionnement et de conclusions du système extrait de [Rousseau 1998] (les commentaires sont précédés de //)

```
     0: ==>S: S1 // Le programme cree l'etat dans lequel l'image est decrite.

     1:    O: O24 (reconnaitre-un-aeroport-militaire-pour-helicopteres). //  L'operateur est choisi.
 l operateur O24 neglige cloture sur la localisation l1.  // Le programme annonce les elements
negliges

     2:    O: O36 (reconnaitre-un-site-a-munition)

     3:    O: O54 (reconnaitre-un-site-missile)

     4:    O: O40 (reconnaitre-un-aeroport-militaire)

     5:    O: O43 (reconnaitre-un-aeroport-militaire)

     6:    ==>S: S2 (operator tie)   // Deux operateurs ont le meme niveau de priorite et
                                     // ne sont pas indifferents, le programme cree donc un sous but.
     // On cree ici une hypothese: sur la localisation l2, il y a soit
     // un aeroport militaire pour gros porteurs, soit un aeroport
     // militaire pour avion de chasse.
     7:    O: O62 (reconnaitre-un-aeroport-militaire-pour-gros-porteurs)

     8:    O: O63 (reconnaitre-un-aeroport-militaire-pour-avions-de-chasse)

     9:    O: O67 (reconnaitre-un-aeroport-militaire-pour-gros-porteurs-bombardiers-tactiques)
    10:    O: O48 (reconnaitre-un-camp-militaire)
 l operateur O48 neglige cloture sur la localisation l1
```

```
    11:    ==>S: S3 (state no-change)
    12:       O: O84 (terminer-la-reconnaissance)
la reconnaissance est terminee
System halted.
```

Citons quelques résultats du système extraits du travail de [Rousseau 1998]

- Les connaissances déduites sans faire d'hypothèses:

```
(F1 ^confiance moyen ^negligible N36
       ^negligible N38 ^objet O83 ^objet O57 ^objet O56 ^objet O55 ^objet O50
       ^objet O49 ^objet O2 ^objet O3 ^objet O4 ^objet O5 ^objet O6 ^objet O7
       ^objet O8 ^objet O9 ^objet O10 ^objet O11 ^objet O12 ^objet O13
       ^objet O14 ^objet O15 ^objet O16 ^objet O17 ^objet O18 ^objet O19
       ^objet O20 ^objet O21 ^objet O22 ^proche-de P2 ^proche-de P3)
```

  Les objets $o2$ à $o22$ sont ceux qui décrivent l'image au départ. Les autres objets sont ceux qui ont étés reconnus sans faire appel à une hypothèse.

- Les hypothèses:

```
(H1 ^conf moyen ^confiance moyen ^faits-averes F6
       ^faits-deduits-dans-l-hypothese F7 ^negligible N44
       ^strategique tres-important)

(H2 ^conf moyen ^confiance moyen ^faits-averes F3
       ^faits-deduits-dans-l-hypothese F4 ^negligible N40
       ^strategique tres-important)
```

- Les faits déduits dans l'hypothèse:

```
(F4 ^objet O81 ^objet O66)

O81 ^confiance moyen ^hypothese H2 ^localisation l2 ^negligible N52
       ^propriete consommation-de-munitions ^propriete militaire
       ^propriete consommation-de-carburant ^strategique tres-important
       ^type aeroport-militaire-pour-gros-porteurs-bombardiers-tactiques)

(O66 ^confiance moyen ^hypothese H2 ^localisation l2 ^negligible N41
       ^propriete militaire ^propriete consomation-de-carburant
       ^propriete transport ^strategique tres-important
       ^type aeroport-militaire-pour-gros-porteurs).
```

9.6 Conclusion

Le propos de cette application à la photo-interprétation était d'avoir une confirmation de la validité de l'approche et de l'implantation réalisée à l'aide de Soar. En conséquence, nous ne prétendons ni avoir fourni un état de l'art complet, ni une modélisation parfaitement aboutie du raisonnement du photo-interprète. En revanche, cette expérience a bien démontré que:

- l'application fonctionne correctement. En effet, la réalisation d'une maquette complète a permis de réaliser une application sans bogues, et la modification de l'algorithme du RETE que nous avons proposée permet d'offrir des temps de calcul compatibles avec l'utilisation du raisonnement par similarité dans une application,

- la logique de raisonnement par similarité version satisfiabilité faible associée à une modélisation dans Soar du problème de photo-interprétation constitue bien une approche du raisonnement qui fournit des résultats qui rendent bien compte du comportement humain. En effet, cette étude a permis de montrer que le raisonnement par similarité permet d'obtenir des résultats stables et fidèles aux résultats obtenus par un photo-interprète, par la génération d'hypothèses concurrentes (scénarios d'interprétations différents).

Conclusion et perspectives

Une des difficultés majeures pour représenter le raisonnement humain dans la décision réside dans la prise en compte des interactions entre des problèmes de décision mémorisés par le décideur, le problème courant et les préférences du décideur. Chaque problème courant contient un état courant que le décideur tente d'identifier par rapport à son expérience: la première phase de la décision est donc de reconnaître un ou plusieurs états mémorisés similaires à l'état courant. Cet ensemble E d'états similaires à l'état courant permet alors de déterminer l'ensemble des états accessibles C à partir de l'état courant, qui sont l'ensemble des états atteints en appliquant l'action entreprise dans un des état de E. Le décideur a des préférences sur les éléments de C, qui déterminent alors rétroactivement l'action choisie. Deux points clefs apparaissent donc dans la formalisation du raisonnement humain dans la décision:

- d'une part, la formalisation de la notion de similarité,
- d'autre part, la formalisation de la notion de préférence, en tenant compte des similarités (les préférences s'expriment sur les éléments de C atteints à partir des états *similaires* à l'état courant).

Ces deux éléments doivent donc être intégrés dans un formalisme de déduction qui soit bien fondé, i.e. telle que la relation de déduction qui exploite similarité et préférences satisfasse de *"bonnes"* propriétés formelles. L'idée est que l'ensemble de la démarche du décideur est partagée entre une partie reconnaissance de l'état courant et une partie plus déductive concernant le futur.

La démarche proposée est alors la suivante. Nous commençons par déterminer la représentation des connaissances et des préférences. La représentation des connaissances est donnée par des séquents, symboles de la forme $\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m$, exprimant que de $\varphi_1, \ldots, \varphi_n$, on peut déduire ψ_1 ou ψ_2 ou ... ou ψ_m. Dans le cas d'une décision, ces formules correspondent respectivement à l'une des deux représentations suivantes:

- A partir de la description de problème $\varphi_1, \ldots, \varphi_n$, on peut entreprendre les actions ψ_1 ou ψ_2 ou ... ou ψ_m;
- A partir de l'action ψ_1, et de la description de problème $\varphi_2, \ldots, \varphi_n$, on obtient les conséquences ψ_1 ou ψ_2 ou ... ou ψ_m;

La similarité entre un état courant constitué par un ensemble de séquents $(: \psi'_1), (: \psi'_2), \ldots, (: \psi'_k)$ et un couple (état, action) mémorisé représenté par le séquent $\varphi_1, \ldots, \varphi_n : \psi_1, \ldots, \psi_m$ est définie par le séquent $\varphi'_1, \ldots, \varphi'_k : \psi''_1, \ldots, \psi''_m$, où:

- $\varphi'_1, \ldots, \varphi'_k$ contient l'ensemble des caractéristiques communes à l'état courant et à l'état mémorisé (i.e. $\varphi'_1, \ldots, \varphi'_k$ subsume $\varphi_1, \ldots, \varphi_n$ et $(: \psi'_1), (: \psi'_2), \ldots, (: \psi'_k)$). Les parties distinguant état courant et état mémorisé sont appelées négligeables: elles correspondent aux parties non subsumées de $\varphi_1, \ldots, \varphi_n$. Ces parties sont incertaines dans le problème courant : on ne sait pas si elles sont vraies ou fausses,
- $\psi''_1, \ldots, \psi''_m$ est l'adaptation de l'action du problème passé au problème courant ($\psi''_1, \ldots, \psi''_m$ subsume $\psi_1, \ldots, \psi_m$)

Par rapport aux approches usuelles du raisonnement à partir de cas [Schank 1982][Kolodner 1993], où un cas est une description complète d'une partie de la connaissance, un cas est ici décrit par un ensemble de formules et la similarité entre un problème de décision mémorisé et un problème de décision courant est évaluée au fur et à mesure des inférences effectuées: une première similarité concerne la reconnaissance de l'état courant et permet d'inférer l'action à effectuer, une deuxième similarité concerne la reconnaissance des anticipations et permet de déduire les conséquences attachées au choix de cette action. De plus, par rapport aux formalisations du raisonnement à partir de cas [Napoli 1992] [Koehler 1994], cette approche a l'avantage de prendre en compte les éléments négligés qui sont ensuite intégrés dans les préférences. Une préférence est un objet syntaxique $(p_1 \wedge a_1, \varphi_1, \Phi_1) \succ (p_2 \wedge a_2, \varphi_2, \Phi_2)$ tel que:

- lorsque l'action a_1 peut être entreprise en sachant au moins la description de problème p_1, résultant en au moins la conséquence Φ_1, et en négligeant φ_1,

- et si l'action a_2 peut être entreprise en sachant au moins la description de problème p_2, résultant en au moins la conséquence Φ_2, et en négligeant φ_2,

- alors la conséquence Φ_1 est préférée à la conséquence Φ_2.

Ce mode de formalisation des préférences se distingue des approches existantes de par son caractère descriptif et de par sa représentation syntaxique explicite. Nous avons montré que les formalisations existantes des préférences, qu'elles soient qualitatives comme dans l'approche de [Lehmann 1996] qui encode implicitement une monotonie des raisonnements, ou qu'elles soient quantitatives comme dans l'approche de [Gilboa and Schmeidler 1995], où la similarité n'est pas syntaxiquement exprimable, cherchent une modélisation plus normative de la décision, sans y introduire les raisonnements.

Cette nécessité d'explorer l'introduction des raisonnements dans la décision nous a alors amené à considérer les propriétés que doit satisfaire une *"bonne"* logique pour raisonner par similarité en fonction de préférences. Nous avons donc introduit une relation $\vdash_S$, lue relation de déduction par similarité sachant des préférences, et étudié le système syntaxique minimal que doit satisfaire une telle relation de déduction. Ces méta propriétés sont les suivantes, si Γ désigne un ensemble de séquents, et Π un ensemble de préférences:

- Méta propriétés concernant la similarité:

Supra Classique	$\dfrac{\Gamma \vdash a}{(\Gamma, \emptyset) \vdash_S a}$
Inclusion	$(\Gamma \cup a, \emptyset) \vdash_S a$
Monotonie prudente	$\dfrac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma, \Pi) \vdash_S a}{(\Gamma \cup y, \Pi) \vdash_S a}$
Coupure	$\dfrac{(\Gamma, \Pi) \vdash_S y \quad (\Gamma \cup y, \Pi) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$
$\wedge$	$\dfrac{(\Gamma, \Pi) \vdash_S a \quad (\Gamma, \Pi) \vdash_S b}{(\Gamma, \Pi) \vdash_S a \wedge b}$
$\vee$	$\dfrac{(\Gamma \cup y, \Pi) \vdash_S a \quad (\Gamma \cup z, \Pi) \vdash_S a}{(\Gamma \cup (y \vee z), \Pi) \vdash_S a}$

- Méta propriétés concernant les préférences:

Anti-Monotonie	$\dfrac{(\Gamma, (\Pi \cup p)) \vdash_S a}{(\Gamma, \Pi) \vdash_S a}$
Monotonie conditionnelle aux préférences	$\dfrac{(\Gamma, \Pi) \vdash_S a \quad \Gamma \nvdash \varphi_1, \phi_1}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$

Ces méta axiomes ne sont pas satisfaits par les formalismes non standards existants, comme par exemple la logique des défauts avec priorités de [Brewka 1994]. Nous avons proposé deux relations de déduction par similarité avec des préférences satisfaisant ces méta axiomes. Ces deux logiques intègrent les préférences en plus de la similarité, en choisissant comme conséquences d'une théorie (un ensemble de problèmes de décision mémorisés) les conséquences admissibles pour les préférences du décideur. La première de ces deux logiques, ou relation de déduction forte par similarité, suppose que tous les éléments de description des cas passés sont potentiellement négligeables, sans à priori d'incompatibilités. Ceci induit une relation de déduction par similarité qui peut s'avérer trop faible lorsque les préférences ne sont pas complètes. La seconde logique, ou relation de déduction faible par similarité, lève ce pré-requis en permettant de poser des ensembles de négligeables <u>exclusifs</u> les uns des autres, ce qui permet d'obtenir des conclusions relativement à des hypothèses de négligeabilités. Pour chacune de ces deux logiques, nous avons défini une sémantique de point fixe, permettant une implantation dans un système d'intelligence artificielle.

En résumé, nous avons donc formalisé un raisonnement par similarité en fonction de préférences, établit les propriétés qu'un tel système de raisonnement devait satisfaire (système syntaxique), et fournit deux relations de déduction par similarité dont nous avons établi la complétude par rapport à ces méta-axiomes et donné un mécanisme d'implantation sous forme d'une sémantique point fixe.

L'implantation qui en a été effectuée dans le système Soar repose en partie sur l'adéquation entre la sémantique formelle des préférences que nous avons donnée et une partie des préférences de Soar. En effet,

- une préférence, du type que nous venons de voir, $(\perp, \perp, \perp) \succ (\varphi_1, \varphi_2, \phi_2)$, signifie intuitivement: si φ_1 et ϕ_2 sont vraies, si l'on a la possibilité de négliger φ_2, alors préférer ne rien faire plutôt que de faire ϕ_2, ou dit autrement rejeter l'action ϕ_2: c'est la préférence *reject* de Soar,

- les préférences du type $(\varphi_1, \varphi_2, \phi_1) \succ (\psi_1, \psi_2, \phi_2)$ correspondent quant à elles intuitivement à un jugement du type: si ϕ_1 résulte de la connaissance de φ_1 et de la négligeabilité de φ_2, et si ϕ_2 résulte de la connaissance de ψ_1 et de la négligeabilité de ψ_2, alors préférer ϕ_1 à ϕ_2: c'est la préférence *better* de Soar,

- Le système Soar définit aussi la préférence *worse* qui est la préférence duale de la précédente, ainsi que les préférences *best* et *worst*, qui correspondent à une quantification sur ϕ_1 dans le premier cas, sur ϕ_2 dans le deuxième cas, donc sont des préférences d'ordre 2.

D'autre part, le système Soar définit de manière structurelle les notions de problèmes, d'actions et d'états. Grâce à ces bonnes propriétés du système Soar, nous avons pu l'adapter en y introduisant une fonction de similarité particulière, correspondant intuitivement à un processus d'unification partielle. L'implantation s'appuie sur une pré-compilation des règles pour prendre en compte la similarité avant insertion dans le RETE, et sur l'ajout dans la syntaxe des programmes Soar de la notion de négligeable. Cette notion est introduite par le biais d'un prédicat d'ordre 2 "Negligeable", qui contient les éléments négligés. Introduit alors dans la base de faits lors d'une unification partielle, une préférence peut alors prendre en compte les éléments négligés.

Enfin, la validité de ce modèle a été testée dans le cadre de la modélisation de la tâche d'un photo-interprète [Rousseau 1998]. En effet, lorsqu'un photo-interprète travaille sur une image satellitale, il n'a accès qu'à une description partielle des objets capturés dans la scène. S'appuyant alors sur son expérience (des cas similaires rencontrés dans le passé), il peut, en tenant compte de ses objectifs de photo-interprétation, établir une interprétation de la scène. Le travail a donc consisté à:

- interroger les experts de photo-interprétation (ainsi qu'à consulter la littérature du domaine) pour établir les règles et les préférences applicables (la mémoire des cas passés à considérer),
- une fois ces éléments encodés dans le formalisme syntaxique de Soar adapté, analyser une scène SPOT comportant plusieurs sites de manière à disposer (en choisissent différentes options d'omission de connaissances) d'un ensemble de cas courants.

Les résultats de ces expériences ont montré que l'approche retenue apporte effectivement une formalisation pertinente du raisonnement préférentiel à partir d'informations incomplètes.

En ce qui concerne les perspectives associées à ce travail, on peut noter que le formalisme présenté ne fournit aucune hypothèse de continuité pour les préférences comme c'est classiquement le cas en théorie de la décision: le modèle ne peut donc pas prendre en compte des préférences qui n'auraient pas été mémorisées. On peut par contre aisément étendre le modèle pour prendre en compte des préférences *"similaires"* à des préférences mémorisées. Il s'agit en effet de la traduction symbolique de l'approche de [Rubinstein 1988] que nous avons présentée à la section 1.4. En effet, ce que [Rubinstein 1988] propose, pour la comparaison des loteries de la forme $(p, x;\ 1-p, 0)$ [28] est une heuristique de choix de la forme suivante:

1. Etape 1:
   - Si $x_1 > x_2$ et $p_1 > p_2$ alors $L_1 > L_2$.
   - Sinon, aller à l'étape 2.
2. Etape 2:
   - Si $x_1 \sim_x x_2$ et $\neg(p_1 \sim_p p_2)$ et $p_1 > p_2$ alors $L_1 > L_2$.
   - Si $p_1 \sim_p p_2$ et $\neg(x_1 \sim_x x_2)$ et $x_1 > x_2$ alors $L_1 > L_2$.
   - Sinon, aller à l'étape 3.
3. Etape 3: Non spécifiée.

Si l'on pose que x_1, x_2, p_1, p_2 sont des pré-requis sur lesquels s'appuient les préférences, l'approche revient à appliquer une préférence lorsque des pré-requis similaires sont vrais. Traduite dans le domaine symbolique, cette approche revient donc à permettre la prise en compte de préférences similaires aux préférences mémorisées, c'est à dire d'appliquer la préférence $(p_1 \wedge a_1, \varphi_1, \Phi_1) \succ (p_2 \wedge a_2, \varphi_2, \Phi_2)$ lorsque:

- l'action a_1 peut être entreprise en sachant *au plus* la description de problème p_1, résultant en *au plus* la conséquence Φ_1, et en négligeant *au plus* φ_1,
- et l'action a_2 peut être entreprise en sachant *au plus* la description de problème p_2, résultant en *au plus* la conséquence Φ_2, et en négligeant *au plus* φ_2,
- alors la conséquence Φ_1 est préférée à la conséquence Φ_2.

Une deuxième perspective est d'étudier la validité cognitive du système Soar ainsi modifié par l'étude du temps de prise de décision en fonction du nombre de fois où une alternative a été proposée. En effet, le mécanisme d'apprentissage de Soar [Laird et al. 1986], le chunking, a pour origine la loi puissance sur l'apprentissage (*"Power Law of Practice"*), qui spécifie que le temps mis pour effectuer une tâche est une fonction puissance décroissante du nombre de fois où elle a été effectuée. Dans Soar, la formalisation de l'apprentissage permet de rendre compte de cette loi. Pour tester la validité cognitive du modèle que nous proposons, il pourrait être intéressant d'étudier le temps de prise de décision en fonction et du nombre d'alternatives et du nombre de négligeables pour déterminer si le lien entre similarité et chunking satisfait ou non une telle loi. Pour ce faire, on pourrait commencer par étudier le lien entre temps de décision

28. Lire probabilité de gagner x avec la probabilité p, probabilité de gagner 0 avec la probabilité $1-p$.

et négligeables, puisque le lien entre apprentissage et temps de décision est déjà connu. À cet égard, il pourrait être intéressant de se référer à [Newell 1990], qui cite un autre principe que doit satisfaire un système universel de la cognition, à savoir le "principe d'incertitude" (Principe 7 p32), ou loi de Hick qui s'exprime comme suit:

> Le temps de décision T s'accroît avec l'incertitude sur le jugement ou la décision à prendre par $T = ICH$, où IC est une constante ($IC = 150 ms/bit$) et H est l'entropie de l'information. Pour n alternatives équiprobables, $H = \log_2(n+1)$, et pour n alternatives de probabilités différentes d'occurrence p_i, $H = \sum_i p_i \log(1/p_i + 1)$.

L'idée, pour valider le raisonnement par similarité, serait de chercher s'il existe une loi donnant le temps de décision en fonction du nombre de négligeables et du nombre d'alternatives utilisées dans le raisonnement. Ayant déterminé cette loi, il faudrait alors vérifier si le système Soar muni du raisonnement par similarité la satisfait ou pas. Il sera alors nécessaire d'établir les bases expérimentales pour la validation, c'est à dire de constituer un ensemble d'exemples offrant un suffisamment grand nombre de possibilités de raisonnement pour obtenir une confirmation statistiquement valide.

Annexes

A

Démonstration des théorèmes

A.1 Théorème de [Neumann and Morgenstern 1947]

Axiomatique 98. *[Neumann and Morgenstern 1947]*

Axiome 1 *(Classement des loteries)*
L'ensemble de choix $\mathcal{P}$ est constitué de toutes les lois de probabilité sur $(C, \mathcal{G})$, où C est l'ensemble des conséquences et $\mathcal{G}$ une tribu contenant les singletons $\{c\}, c \in C$. La relation de préférence $\succeq$[29] *est un pré-ordre large*[30] *dans $\mathcal{P}$.*

Axiome 2 *(Continuité des préférences)*
Pour toutes les lois $P', P'', Q \in \mathcal{P}$, vérifiant $P' \succ Q \succ P''$, on peut trouver $\lambda, \mu \in]0,1[$ tel que $\lambda P' + (1-\lambda)P'' \succ Q$. et $Q \succ \mu P' + (1-\mu)P''$.

Axiome 3 *(Axiome d'indépendance)*
Pour toutes lois $P', P'', Q \in \mathcal{P}$ et tout $\lambda \in]0,1]$, $P' \succ P'' \Leftrightarrow \lambda P' + (1-\lambda)Q \succ \lambda P'' + (1-\lambda)Q$.

Théoreme E. *[Neumann and Morgenstern 1947] Sous les axiomes 1, 2, 3, il existe sur $(\mathcal{P}, \succ)$ une fonction d'utilité linéaire, unique à une transformation affine strictement croissante près.*

Preuve. **i)** Soit $P, Q \in \mathcal{P}$ et $\mu \in]0,1]$. Alors $P \succ Q \Rightarrow P \succ \mu Q + (1-\mu)P$ par l'axiome d'indépendance et par $P = \mu P + (1-\mu)P$.

ii) Pour tous $P, Q \in \mathcal{P}$ et $\lambda, \mu \in [0,1]$ $P \succ Q, \lambda > \mu \Rightarrow \lambda P + (1-\lambda)Q \succ \mu P + (1-\mu)Q$
Soit $\alpha = \frac{\mu}{\lambda} \in [0,1[$. En appliquant l'axiome d'indépendance à $P, \alpha P + (1-\alpha)Q, Q$

$$\begin{aligned} \lambda P + (1-\lambda)Q &\succ \lambda(\alpha P + (1-\alpha)Q) + (1-\lambda)Q \\ &\succ \mu P + (1-\mu)Q \end{aligned}$$

iii) Pour tous $P', P'', Q \in \mathcal{P}$, tels que $P' \succ Q \succ P''$, il existe un unique $\lambda_0 \in]0,1[$ tel que $Q \sim \lambda_0 P' + (1-\lambda_0)P''$.
Soit $E = \{\mu \in]0,1[;\ \mu P + (1-\mu)P'' \succ Q\}$. Cet ensemble n'est pas vide, c'est un intervalle d'après *ii)* puisque, si μ lui appartient, tout $\lambda > \mu$ lui appartient aussi. Si λ_0 désigne sa borne inférieure, E est donc soit l'intervalle $]\lambda_0, 1[$, soit l'intervalle $[\lambda_0, 1[$.
Montrons alors que $\lambda_0 P' + (1-\lambda_0)P'' \sim Q$.

- Si l'on avait $\lambda_0 P' + (1-\lambda_0)P'' \succ Q \succ P''$, il existerait d'après l'axiome de continuité $\alpha \in]0,1[$ tel que $\alpha(\lambda_0 P' + (1-\lambda_0)P'') + (1-\alpha)P'' \succ Q$, soit $\alpha\lambda_0 P' + (1-\alpha\lambda_0)P'' \succ Q$, ce qui contredit la définition de λ_0 puisque $\alpha\lambda_0 < \lambda_0$.

29. On notera respectivement $\succ$ et $\sim$ les parties asymétriques et symétriques de $\succeq$.
30. On rappelle qu'un pré-ordre large est une relation asymétrique et négativement transitive.

- Si l'on avait $P' \succ Q \succ \lambda_0 P' + (1-\lambda_0)P''$, il existerait d'après l'axiome de continuité $\beta \in]0,1[$ tel que: $Q \succ \beta P' + (1-\beta)(\lambda_0 P' + (1-\lambda_0)P'')$, soit $Q \succ (\beta + (1-\beta)\lambda_0)P' + (1-\lambda_0)(1-\beta)P''$, ce qui est impossible puisque $\beta + (1-\beta)\lambda_0 \in]\lambda_0, 1]$.

La seule possibilité est donc que $\lambda_0 P' + (1-\lambda_0)P'' \sim Q$, λ_0 étant unique d'après *ii*).

iv) Étant donné $P_0, P_1 \in \mathcal{P}$ tels que $P_0 \succ P_1$, il existe une utilité linéaire sur $\mathcal{L} = \{P \in \mathcal{P}, P_0 \succeq P \succeq P_1\}$. Posons $U(P_0) = 1$ et $U(P_1) = 0$. Pour tout P tel que $P_0 \succeq P \succeq P_1$, soit λ_P l'unique valeur telle que $\lambda_P P_0 + (1-\lambda_P)P_1 \sim P$ (axiome de continuité) et posons $U(P) = \lambda_P$.

- Alors U est une utilité représentant $\succeq$ dans $\mathcal{L}$. En effet, $P \succeq Q \Leftrightarrow \lambda_P P_0 + (1-\lambda_P)P_1 \succeq \lambda_Q P_0 + (1-\lambda_Q)P_1$ d'après ii).
- U est une fonction linéaire sur $\mathcal{L}$. Soit $P', P'' \in \mathcal{L}$ et $P = \lambda P' + (1-\lambda)P''$.
  Soit

$$\begin{array}{lll} U(P') = \lambda', & \text{où } \lambda' \text{ est tel que} & P' = \lambda' P_0 + (1-\lambda')P_1 \\ U(P'') = \lambda'', & \text{où } \lambda'' \text{ est tel que} & P'' = \lambda'' P_0 + (1-\lambda'')P_1 \end{array}$$

  $P = \alpha P' + (1-\alpha)P'' = [\alpha\lambda' + (1-\alpha)\lambda'']P_0 + [\alpha(1-\lambda') + (1-\alpha)(1-\lambda'')]P_1$, soit
  $U(P) = \alpha\lambda' + (1-\alpha)\lambda'' = \alpha U(P') + (1-\alpha)U(P'')$.

v) Cette utilité est linéaire à une transformation affine strictement croissante près.
Soit V une fonction d'utilité linéaire sur $\mathcal{L}$. Si $P \in \mathcal{L}$ est tel que $P \sim \lambda_P P_0 + (1-\lambda_P)P_1$, alors par linéarité de V:

$$\begin{aligned} V(P) = \lambda_P V(P_0) + (1-\lambda_P)V(P_1) &= \lambda_P(V(P_0) - V(P_1)) + V(P_1) \\ &= a\lambda_P + b \\ &= aU(P) + b \end{aligned}$$

Réciproquement, U étant une fonction d'utilité linéaire, a, b des nombres réels et $a > 0$, $V = aU + b$ est encore une utilité linéaire:

- $V(P) \geq V(Q) \Leftrightarrow U(P) \geq U(Q)$ car $a > 0$,
-

$$\begin{aligned} V(\alpha P' + (1-\alpha)P) &= aU(\alpha P' + (1-\alpha)P) + b \\ &= \alpha(aU(P') + b) + (1-\alpha)(aU(P) + b) \\ &= \alpha V(P') + (1-\alpha)V(P) \end{aligned}$$

vi) Existence et unicité à une transformation affine strictement croissante sur tout $\mathcal{P}$.
Si tous les éléments de $\mathcal{P}$ sont indifférents, c'est évident. Soit donc $P_0, P_1 \in \mathcal{P}$ tels que $P_0 \succ P_1$. Pour tout $P \in \mathcal{P}$, il existe $Q'_0, Q'_1 \in \mathcal{P}$ tels que P, P_0, P_1 appartiennent à $\mathcal{L}' = \{Q \in \mathcal{P}, Q'_0 \succeq Q \succeq Q'_1\}$ et d'après ce qui précède, il existe sur $\mathcal{L}'$ une utilité linéaire vérifiant $U(P_0) = 1$ et $U(P_1) = 0$. $U(P)$ ne dépend pas du choix de Q'_0 et Q'_1. Soit $\mathcal{L}'' = \{Q \in \mathcal{P}, Q''_0 \succeq Q \succeq Q''_1\}$ un autre ensemble contenant P, P_0, P_1, U'' une utilité linéaire sur $\mathcal{L}''$ vérifiant $U''(P_0) = 1$ et $U''(P_1) = 0$. Les fonctions U et U'' coïncidant sur $\mathcal{L}' \cup \mathcal{L}''$, d'où $U''(P) = U(P)$.

Le procédé précédent associe donc pour tout $P \in \mathcal{P}$ un nombre $U(P)$ déterminé. Cette application est une utilité linéaire, unique à une transformation affine strictement croissante commune à toutes les utilités linéaires introduites.

■

A.2 Théorème de [Savage 1954]

On ajoute ici deux axiomes à l'axiomatique de [Savage 1954] dans le but de simplifier les démonstrations, en l'occurrence les axiomes 8 et 9.

Axiomatique 99. *[Savage 1954][Jaffray 1985]*

Axiome 1 *(Existence d'un ordre complet)*

- *$\mathcal{E}$ est une σ algèbre,*
- *A contient les actes en escalier, les actes en escalier généralisé et est fermé par greffe.*
- *$\succeq$ est un pré-ordre large total.*

Axiome 2 *(Axiome d'indépendance)*
Si des actes a, a', b, b' et un événement E vérifient

$$\begin{array}{ll} a_{|_E} = a'_{|_E} & a_{|_{E^C}} = b_{|_{E^C}} \\ b_{|_E} = b'_{|_E} & a'_{|_{E^C}} = b'_{|_{E^C}} \end{array}$$

Alors $a \succeq b \Leftrightarrow a' \succeq b'$. Les préférences entre deux actes ne dépendent donc que des parties qui les distinguent. Définissons donc pour chaque événement E une relation de pré-ordre large $\succeq_E$, que l'on interprète comme les préférences du décideur sachant E réalisé, par:

$$a \succeq_E b \Leftrightarrow [\forall a', b' \in At.q, \quad \begin{array}{c} a_{|_E} = a'_{|_E} \\ b_{|_E} = b'_{|_E}, \\ a'_{|_{E^C}} = b'_{|_{E^C}} \end{array} \quad a' \succeq b'$$

Axiome 3 *(Ordonnancement des conséquences)*
Pour que cette interprétation de la relation de préférence sachant un événement E réalisé soit compatible avec les préférences dans le certain du décideur, il faut que pour tous résultats c', c'' et tout événement E non nul,

$$a_{c'} \succeq_E a_{c''} \Leftrightarrow c' \succeq_C c''$$

Axiome 4 *(Non trivialité de $\succeq_C$)*
Il existe $c, c' \in C$ tels que $c' \succ_C c$.

Axiome 5 *(Comparaison des événements)*
Pour tous résultats $c' \succ_C c''$ et $k' \succ_C k''$, étant donné les actes a_E et $a_{E'}$ tels que:

$$\begin{array}{ll} a_E(w) = c' \text{ pour } w \in E, & a_E(w) = c'' \text{ pour } w \in E^C \\ a_{E'}(w) = c' \text{ pour } w \in E, & a_{E'}(w) = c'' \text{ pour } w \in {E'}^C \end{array}$$

et les actes b_E et $b_{E'}$ tels que:

$$\begin{array}{ll} b_E(w) = k' \text{ pour } w \in E, & b_E(w) = k'' \text{ pour } w \in E^C \\ b_{E'}(w) = k' \text{ pour } w \in E, & b_{E'}(w) = k'' \text{ pour } w \in {E'}^C \end{array}$$

on a alors:

$$a_E \succeq a_{E'} \Leftrightarrow b_E \succeq b_{E'}$$

On interprète intuitivement cet axiome en introduisant la notion de prime. En effet, on dira qu'un acte a_E offre une prime (c'/c'') lorsque E est réalisé si a_E est définit par:

$$a_E(w) = c' \text{ pour } w \in E, a_E(w) = c'' \text{ pour } w \in E^C$$

L'axiome traduit l'intuition que le décideur qui préfère une prime (c'/c'') lorsque l'événement E est réalisé à la même prime si E' est réalisé est un décideur qui juge E plus probable que E'. Encore faut-il pour cela que cette préférence ne dépend pas des montants de la prime, ce que traduit cet axiome.
Cet axiome, combiné à la non-trivialité de $\succeq$, permet maintenant de définir une relation d'ordre $\succeq_{\mathcal{E}}$ dans $\mathcal{E}$ par:

$$E \succeq_{\mathcal{E}} E' \Leftrightarrow \forall c, c' c \succeq_C c', a_E \succeq a_{E'}$$

avec a_E (resp $a_{E'}$) acte offrant une prime (c/c') si E (resp si B).

Axiome 6 *(Divisibilité et continuité)*
Pour tout couple d'actes a, b tels que $a \succ b$ et tout résultat c, on peut trouver une partition finie de l'événement certain $\{E_i \in \mathcal{E}, i \in I\}$ telle que pour tout $i \in I$:

1. $$a_i \succ b, \qquad \text{où } a_i \text{est l'acte défini par}$$
$$a_i(w) = a(w) \text{ pour } w \notin E_i, a_i(w) = c \text{ pour } w \in E_i$$

2. $$a \succ b_i, \qquad \text{où } b_i \text{est l'acte défini par}$$
$$b_i(w) = b(w) \text{ pour } w \notin E_i, b_i(w) = c \text{ pour } w \in E_i$$

Axiome 7 *(Axiome de dominance)*
Pour tout événement E.

1. *Si $a \succ_E a_c$ pour tout $c \in b(E)$, alors $a \succeq_E b$,*
2. *Si $a \prec_E a_c$ pour tout $c \in b(E)$, alors $a \preceq_E b$,*

Axiome 8 *(Axiome de mesurabilité)*
Pour tout acte a et tout résultat c, $\{w \in E, a(w) \preceq_C c\}$ et $\{w \in E, a(w) \succeq_C c\}$ sont des événements de E.

Axiome 9 *(Axiome de continuité monotone de $\succeq_{\mathcal{E}}$)*
Pour tous événements E, E', et toute suite d'événements $(E_n)_{n\in\mathbb{N}}$ telle que $E_n \uparrow E$:

$$\forall n \in \mathbb{N}, E_n \preceq_{\mathcal{E}} E' \Rightarrow E \preceq_{\mathcal{E}} E'$$

Théoreme F. *[Savage 1954] Sous les axiomes 1 à 9, il existe une fonction d'utilité représentant $\succeq$, donnée par une expression de la forme*

$$a \in A \longmapsto U(a) = E_P U(a(.))$$

où P est une loi de probabilité sur $(E, \mathcal{E})$. $P(.)$ est alors nécessairement l'unique loi de probabilité compatible avec $\succeq_{\mathcal{E}}$. La fonction U est unique à une transformation affine strictement croissante près.

Preuve du théorème F. On démontre en deux temps ce théorème de [Savage 1954], d'abord en montrant l'existence de la loi de probabilité compatible avec $\succeq_{\mathcal{E}}$, puis en démontrant que les actions peuvent être représentées par une hypothèse d'espérance d'utilité.

Existence et unicité de la loi de probabilité compatible avec $\succeq_{\mathcal{E}}$ Commençons par remarquer que par l'axiome d'indépendance (Axiome 2) et par l'ordonnancement des conséquences (Axiome 3):

Proposition 100. *Si $\{E_i \in \mathcal{E}, i \in I\}$ est une partition finie de $E \in \mathcal{E}$, si f et g sont des actes constants sur chaque $(E_i)_{i \in I}$, i.e. tels que $f(E_i) = \{c'_i\}, g(E_i) = \{c''_i\}, i \in I$ alors*

$$\forall i \in I, c'_i \succeq_C c''_i \Rightarrow f \succeq_E g$$

et si de plus, il existe i_0 tel que E_{i0} soit non nul et $c'_{i_0} \succ_C c''_{i_0}$, alors $f \succ g$.

Cette propriété permet de démontrer que $\succeq_{\mathcal{E}}$ vérifie **Q1** de la définition 12 (il suffit de remarquer que (A, A^C) est une partition de Ω, et de prendre des constants $f_\Omega(A) f_\Omega(A^C) = c'$ et $f_A(A) = c'', f_A(A^C) = c', c' \succ_C c''$ et de noter qu'alors $f_\Omega \succ f_A$, soit $\Omega \succeq_{\mathcal{E}} A$ d'après l'axiome 5. La comparaison $A \succeq_{\mathcal{E}} \emptyset$ est identique.).

Remarquons tout d'abord que par la proposition 100, on a $E \supset F \Rightarrow E \succeq_{\mathcal{E}} F$ ainsi que $[E \succeq_{\mathcal{E}} F, G \cap E = \emptyset, G \cap F = \emptyset] \Rightarrow E \cup G \succeq_{\mathcal{E}} F \cup G$ (Même mode démonstration que pour la propriété Q1 ci-dessus.). La combinaison de ces deux résultats permet alors de démontrer que $[E \succeq_{\mathcal{E}} F, G \cap E = \emptyset] \Rightarrow E \cup G \succeq_{\mathcal{E}} F \cup G$. (il suffit de prendre $G' = G \cap F^C$, alors $E \cup G \succeq_{\mathcal{E}} E \cup G' \succeq_{\mathcal{E}} F \cup G' = F \cup G$). Ceci permet de montrer que la relation $\succeq_{\mathcal{E}}$ vérifie également **Q2** de la définition 12. Supposons donc que $A_1 \succeq_{\mathcal{E}} B_1, A_2 \succeq_{\mathcal{E}} B_2$ et $A_1 \cap A_2 = \emptyset$. Si l'on pose $A'_1 = A_1 \cap {B_2}^C$ et $B'_2 = B_2 \cap {A_1}^C$, il vient:

$$A'_1 \cup A_2 \succeq_{\mathcal{E}} A'_1 \cup B_2 = A_1 \cup B_2 = A_1 \cup B'_2 \succeq_{\mathcal{E}} B_1 \cup B'_2$$

puis

$$A_1 \cup A_2 = (A'_1 \cup A_2) \cup (A_1 \cap B_2) \succeq_{\mathcal{E}} (B_1 \cup B'_2) \cup (A_1 \cap B_2) = B_1 \cup B_2$$

La démonstration de ii de **Q2** se fait de la même manière.

Comme l'axiome **Q3** a été posé en axiome (Axiome 9), il en résulte que $(\Omega, \mathcal{E}, \succeq_{\mathcal{E}})$ est un espace de probabilité qualitative. Il est de plus sans atome si l'on fait la remarque suivante.
Soit $A \in \mathcal{E}, A \succeq_{\mathcal{E}} \emptyset$. Soit f_A (resp. $f_\emptyset$) un acte offrant une prime (c'/c'') si A $(\emptyset)$. D'après l'axiome 6, étant donné les actes f_A et $f_\emptyset$, qui vérifient $f_A \succ f_\emptyset$, il existe une partition finie $\{E_i \in \mathcal{E}, i \in I\}$ telle que pour tout $i \in I, f_A \succ f_{E_i}$, f_{E_i} acte offrant la prime (c'/c'') si E_i, donc $A \succ_{\mathcal{E}} E_i$. De plus, il existe i_0 telle que $A \cap E_i \succ_{\mathcal{E}} \emptyset$, sinon $[\forall i \in I, A \cap E_i \sim_{\mathcal{E}} \emptyset] \Rightarrow A \cup_{i \in I} E_i \sim_{\mathcal{E}} \emptyset, i.e. A \sim_{\mathcal{E}} \emptyset$. L'espace de probabilité qualitative est donc sans atome.

Il existe donc sur l'espace probabilisable $(\Omega, \mathcal{E})$ une loi de probabilité unique $P(.)$ compatible avec $\succeq_{\mathcal{E}}$.

Préférence sur les loteries Les préférences du décideur vont engendrer des préférences sur les loteries. En effet, si $P(.)$ est la loi de probabilité subjective du décideur, on peut distinguer parmi les actes en escalier les actes caractérisés par la partition finie $\{E_i, i \in I\}$ et $f(E_i) = \{c_i\}, i \in I$ et tels que $P(E_i) \neq 0$. Un tel acte est dit simple. On appelle alors loterie engendrée par l'acte en escalier simple f la loterie $Q = \Sigma_{i \in I} P(E_i) \mathcal{C}_{c_i}$, où $\mathcal{C}_{c_i}$ est une loterie certaine de support $\{c_i\}$. Pour montrer qu'une loterie est engendrée par au moins un acte en escalier simple, il suffit de noter la proposition suivante:

Proposition 101. *Soit $(\Omega, \mathcal{E}, \succeq_{\mathcal{E}})$ un espace de probabilité qualitative sans atome, $P(.)$ la loi de probabilité compatible. Alors, pour tout $E \in \mathcal{E}$ et $\rho in [0, 1]$, on peut trouver $A \in \mathcal{E}$ tel que $A \subset E$ et $P(A) = \rho P(E)$.*

On peut alors montrer que toute loterie est engendrée par un acte en escalier simple, car si $Q = \Sigma_{i \in I} \lambda_i \mathcal{C}_{c_i}$, on peut trouver par la proposition précédente une famille $B_i \in \mathcal{E}$ $t.q$ $P(B_i) = \lambda_i$. L'acte ainsi défini par les familles $(B_i)_{i \in I}, (c_i)_{i \in I}$ engendre Q.
Inversement, on peut démontrer que si deux actes en escalier simple engendrent la même loterie, ils sont équivalents (voir [Jaffray 1985], Prop 6, p15-17).

Il suffit maintenant de retrouver sur l'ensemble des loteries ainsi définis, totalement pré-ordonné les axiomes du théorème E. C'est le cas du premier axiome, puisque les actes en escalier simples sont totalement ordonnés. On note $\succeq_{\mathcal{L}}$ l'ordre induit par $\succeq$ sur les loteries. Cet ordre est continu (Deuxième axiome du théorème E car:

Proposition 102. *Pour toutes les lois $P', P'', Q \in \mathcal{L}$, vérifiant $P' \succ_{\mathcal{L}} Q \succ_{\mathcal{L}} P''$, on peut trouver $\lambda, \mu \in]0,1[$ tel que $\lambda P' + (1-\lambda)P'' \succ_{\mathcal{L}} Q$. et $Q \succ_{\mathcal{L}} \mu P' + (1-\mu)P''$.*

Soit $P' = \Sigma_{i\in I}\alpha'_i\mathcal{C}_{c'_i}, P'' = \Sigma_{i\in I}\alpha''_i\mathcal{C}_{c''_i}$. Soit k un acte en escalier simple engendrant Q, f acte en escalier simple engendrant R'. On opère sur f des transformations successives de manière à obtenir un acte engendrant une loterie $\lambda P' + (1-\lambda)P''$ qui soit strictement préféré à k.
Soit $\overline{c} = Max\{c'_i, i \in I\}$ et $\underline{c} = Min\{c'_i, c''_i, i \in I\}$.

1. Par l'axiome 6, il existe une partition fine $\{F_k, k \in K\}$ de Ω telle que l'acte f_k obtenu en modifiant f sur l'un quelconque des F_k pour qu'il y donne le résultat $\underline{c}$ vérifie encore $f_k \succ k$. Comme $P(A_1) = \alpha'_1 > 0$, $A_1 \succ_{\mathcal{E}} \emptyset$; il existe donc k_0 tel que $F_{k_0} \cap A_1 succ_{\mathcal{E}} \emptyset$. L'acte g obtenu en modifiant f sur $F_{k_0} \cap A_1$ pour y donner le résultat $\overline{c}$ vérifie alors $g \succ k$.
2. Soit $P(F_{k_0} \cap A_1) = 1 - \lambda > 0$. Pour tout $i \in I$, soit $B_i \in \mathcal{E}$ tel que $B_i \subset A_i$ et $P(B_i) = (1-\lambda)P(A_i)$. L'acte h obtenu en modifiant g sur chaque B_i pour qu'il y donne le résultat $\overline{c}$ vérifie encore $h \succeq g$, soit $h \succeq k$. Cet acte donne les résultats $(c'_i)_{i\in I}$ avec des probabilités $\lambda P(A_i)$ et le résultat $\underline{c}$ avec la probabilité $1-\lambda$.
3. Enfin, on peut trouver une partition $\{E_j, j \in J\}$ de $F_{k_0} \cap A_1$ par l'axiome 6 telle que $P(E_j) = (1-\lambda)\alpha''_i$. L'acte l obtenu en modifiant h sur chacun des E_j de manière qu'il y donne le résultat c''_j vérifie encore $l \succ h$, donc $l \succ k$. Or l engendre la loterie $\lambda P' + (1-\lambda)P''$, d'où le résultat.

La deuxième proposition se démontre de manière identique.

Il ne reste alors plus qu'a démontrer que l'on peut obtenir le troisième axiome de l'axiomatique 98, ce que la proposition suivante permet d'établir:

Proposition 103. *Pour toutes lois $P', P'', Q \in \mathcal{L}$ et tout $\lambda \in]0,1], P' \succ P'' \Leftrightarrow \lambda P' + (1-\lambda)Q \succ \lambda P'' + (1-\lambda)Q$.*

Soit donc $P' = \sum_{i\in I}\lambda'_i\ \mathcal{C}_{c'_i}, P'' = \sum_{j\in J}\lambda''_j\ \mathcal{C}_{c''_i}, Q = \sum_{k\in K}\mu_k\ \mathcal{C}_{c_k}$.
Soit donc a' un acte en escalier simple engendrant P' et soit $\{B_i, i \in I\}$ la partition de Ω associée. De même, soit a'' l'acte en escalier simple engendrant P'' de partition associée $\{C_j, j \in J\}$. Par la proposition 101, pour tout B_i et C_j, il existe $D_{ij} \in \mathcal{E}$ tel que $D_{ij} \subset B_i \cap C_j$ et $P(D_{ij}) = \lambda P(B_i \cap C_j)$. Soit alors $D = \cup_{(i,j)\in I\times J} D_{ij}, P(D) = \lambda, P(D^C) = 1-\lambda$; on peut alors construire une partition $\{E_k, k \in K\}$ de D^C telle que $P(E_k) = (1-\lambda)\mu_k$ pour tout $k \in K$. Si l'on prend alors les restrictions à D de a' et a'', et qu'on les prolonge de manière commune de manière à obtenir sur chaque E_k le résultat c_k, les actes obtenus engendrent respectivement les loteries $\lambda P' + (1-\lambda)Q$ et $\lambda P'' + (1-\lambda)Q$. Tout revient donc à démontrer que $a' \succ a'' \Leftrightarrow a' \succ_D a''$.
Remarquons que si le choix des D_{ij} n'est pas unique, les actes qui les prolonges sur les intervalles D associées sont eux équivalents. Ce qui permet d'introduire une relation $\succ_\lambda$ sur les actes en escalier simples tels que:

$$a' \succ_\lambda a'' \Leftrightarrow \forall \Delta = \cup \Delta_{ij}\ t.q.\ \Delta_{ij} \subset B_i \cap C_j \text{ et} P(\Delta_{ij}) = \lambda P(B_i \cap C_j), a' \succ_\Delta a''$$

On peut est donc amené à montrer que $a \succ a'' \Leftrightarrow a' \succ_\lambda a''$. Pour cela, on remarque que l'on a

$$\forall \lambda > 0, \gamma > 0, \lambda + \gamma \prec 1, [a' \succ_\lambda a'', a' \succ_\gamma a''] \Rightarrow a' \succ_{\lambda+\alpha} a''$$

De cette propriété, on peut déduire que $a' \succ a'' \Leftrightarrow a' \succ_{\frac{1}{2}} a''$, et de la déduire directement

$$\forall \beta \in [0,1], a' \succ_\beta a'' \Leftrightarrow a' \succ_{\frac{k\beta}{2^p}} a''$$

Montrons que l'on peut alors passer à la limite:

$a' \succ_\lambda a'' \Rightarrow a' \succ a''$ Soit $D \in \mathcal{E}$, et soit b', b'' actes obtenus en prolongeant les restrictions respectives de a' et a'' à D par une partie commune. En appliquant deux fois l'axiome 6, on montre l'existence de $\Delta \succ_{\mathcal{E}} \emptyset, \Delta \cap D \neq \emptyset$ tel que f', identique à l'acte b' sauf sur Δ où il donne le résultat $Min\{c'_i, i \in I\}$ et f'', l'acte identique à b'' sauf sur Δ où il donne le résultat $Max\{c''_j, j \in J\}$ vérifient encore $f' \succ f''$. Des majorations de f' sur Δ et des minorations de f'' sur Δ les transforment en actes g' et g'' engendrant $\beta R' + (1-\beta)Q$ et $\beta R'' + (1-\beta)Q$, où $\beta = P(D \cup \Delta) > \lambda$; comme $g' \succeq f'$ et $f'' \succeq g''$, il vient $g' \succ g''$ et donc $a' \succ_b etaa''$. Plus généralement, les mêmes transformations opérées seulement sur les sous ensembles de Δ permettent de montrer que $a' \succ_\gamma a''$ pour tout $\gamma \in [\alpha, beta]$.. Or ont peut toujours trouver p et k tels que $\frac{k}{2^p} \in [\alpha, \beta]$, donc $a' \succ_{\frac{k}{2^p}} a''$ et $a' \succ a''$.

$a' \succ a'' \Rightarrow a' \succ_\lambda a''$ Un raisonnement analogue au précédent permet de montrer (par l'axiome 6) que pour $\beta \succ \beta_0$, où $\beta_0 < 1$ $a' \succ_\beta a''$; d'où de nouveau, pour un p et un k tels que $\frac{k\beta}{2^p} \succ \beta_0$, et donc de nouveau $a' \succ_\beta a''$.

Ayant ainsi retrouvé l'axiomatique 98, on peut alors en retrouver la conséquence, à savoir **qu'il existe sur** $(\mathcal{L}, \succeq_{\mathcal{L}})$ **une fonction d'utilité linéaire**. On peut alors montrer que ce résultat peut être étendu à tout acte à partir des actes en escalier simple.

■

B

Généralisations du modèle de l'utilité espérée

B.1 Introduction

L'échec du modèle de l'utilité espérée à décrire les comportements effectifs des individus a entraîné tout un courant de recherche dans deux directions:

- d'une part, la recherche systématique par les psychologues de tout comportement constituant une déviation (d'où le terme employé de *biais cognitif*) par rapport aux axiomes ou aux implications des axiomes de l'utilité espérée,
- d'autre part, la construction de nouveaux modèles de décision permettant de prendre en compte le manque de validité descriptive du modèle de l'utilité espérée.

Après une revue rapide des études psychologiques menées dans le cadre du manque de validité descriptive des axiomes de l'utilité espérée, nous verrons quelques uns des modèles utilisés pour généraliser le modèle de l'utilité espérée. L'objet du présent chapitre n'est pas de faire le point sur tous les progrès effectués récemment sur l'ensemble des modèles de décision, mais de montrer, à la lumière de ces généralisations et de leur comparaison, que le manque de validité descriptive de ces nouveaux modèles conduit à poser une hypothèse de représentation du raisonnement dans la décision, et non plus seulement des résultats. De ce fait, on se concentrera sur la décision dans le risque où l'on constate, depuis le début des années 1980, une importante remise en cause des axiomes de l'utilité espérée. On ne se concentrera donc pas sur les modèles de décision en incertitude pour lesquels on renverra le lecteur à [Munier 1995] pour un panorama de ces modèles. Nous verrons que ces remises en cause, dans la mesure où elle ne sont que partielles en retenant en général les deux premiers axiomes de l'axiomatique 4[31], ne peuvent prédire l'ensemble des effets observés dans le seul triangle de Marschak-Machina. La prise en compte de ces effets nécessite la prise en compte d'une nouvelle hypothèse, à savoir la représentation des raisonnements.

B.2 Le triangle de Marschak-Machina

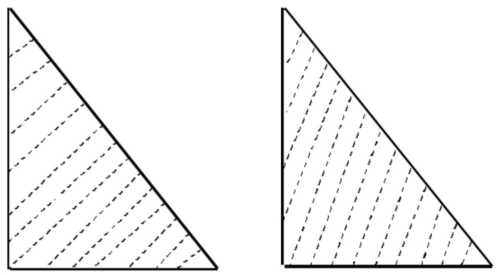
Pour comparer les unes par rapport aux autres ces généralisations du paradigme de l'utilité espérée, nous utiliserons le fait que chacun de ces modèles définissent des profils caractéristiques de préférence et d'indifférence. Pour visualiser ces lieux d'indifférence entre loteries, on considère un ensemble de loteries n'impliquant que trois conséquences distinctes x_1, x_2, x_3, conséquences identiques pour l'ensemble des loteries considérées, classées dans l'ordre $x_1 \prec x_2 \prec x_3$, les probabilités p_1, p_2, p_3 étant elles variables. On représente ces courbes d'indifférence dans un triangle unité, l'axe des abscisses mesurant la probabilité p_1

31. Ces axiomes permettent en effet de prouver l'existence d'une fonction représentatrice de l'ordre défini par les préférences, et l'hypothèse d'indépendance permet elle de lui donner une forme particulière: l'espérance d'utilité.

et l'axe des ordonnées la probabilité p_3 [32]. Ainsi, par exemple, dans le cas de l'utilité espérée, les courbes d'indifférence sont des droites parallèles.

$$\begin{aligned}
E(L) = U(x_1) * p_1 + U(x_2) * p_2 + U(x_3) * p_3 &= k \\
\text{Soit } [U(x_1) - U(x_2)] * p_1 + [U(x_3) - U(x_2)] * p_3 &= k - U(x_2) \\
\text{soit } \frac{dp_3}{dp_1}_{E=k} &= \frac{U(x_2) - U(x_1)}{U(x_2) - U(x_3)},
\end{aligned}$$

Cette équation caractérise la pente de cette droite. Ces courbes se présentent alors de la manière suivante dans ce triangle unité, ou triangle de Marschak - Machina:



(a) Inclinaison au risque

(b) Aversion au risque

FIG. B.1: *Le triangle de Marschak - Machina*

La pente des droites dans le triangle représente l'aversion ou l'inclinaison au risque des décideurs: Plus la pente des droites est élevée, plus l'individu aura d'aversion pour le risque.

De la même manière, on peut définir des courbes d'indifférence pour chacun des modèles que nous allons présenter.

B.3 Autres invalidités empiriques du modèle de l'utilité espérée

Nous ne reviendrons pas dans cette partie sur l'ensemble des expériences qui démontrent que le raisonnement humain ne se fait pas sur la base du calcul mathématique des probabilités (le lecteur est pour ce faire renvoyé à [Tversky and Kahneman 1982][Hogarth 1988]). Nous ne détaillerons ici que les expériences qui sont nécessaires pour mener à bien notre démonstration, donc celles qui contredisent les conséquences des axiomes de l'utilité espérée puisque nous avons déjà présenté les expériences qui invalident les axiomes:

Effet de proportionalité Il consiste à comparer des paires de loteries proportionnelles en probabilité. Par exemple, la comparaison entre les deux paires de loteries suivantes:

$$\begin{array}{lcllcl}
L_1(3000) & = & 1 & L_2(4000) & = & 0.8 \\
 & & & L_2(0) & = & 0.2 \\
L_3(3000) & = & 0.25 & L_4(4000) & = & 0.2 \\
L_3(0) & = & 0.75 & L_4(0) & = & 0.8
\end{array}$$

La théorie de l'utilité espérée impose $L_1 \succ L_2$ et $L_3 \succ L_4$, alors que la majorité des individus choisit $L_1 \succ L_2$ et $L_3 \prec L_4$. Cet effet de proportionalité tend à donner une structure en éventail aux courbes d'indifférence dans le triangle de Marschak-Machina (voir [Kahneman and Tversky 1979]).

32. On obtient la troisième composante de la loterie p_2 en mesurant la distance horizontale qui sépare le point de coordonnées (p_1, p_3) de l'hypothénuse car $p_2 = 1 - p_1 - p_3$.

Effet de réflexion Cet effet considère que les comportements face à un risque diffèrent selon que les loteries engendrent des gains, des pertes ou les deux. De ce fait, l'allure des courbes dans le triangle de Marschak-Machina est symétrique pour les gains et pour les pertes (voir [Kahneman and Tversky 1979]).

Surévaluation des petites probabilités Cet effet se fonde sur le fait que les individus ont tendance à surévaluer les loteries où les gains importants ont une probabilité très faible. Cet effet se traduit par des courbes d'indifférence ayant des pentes plus élevées à la base comme sur le côté gauche du triangle.

B.4 Les généralisations du paradigme de l'utilité espérée

Ces généralisations, toutes effectuées dans le cas du risque, visent à remplacer ou à substituer aux axiomes de la théorie de la décision de nouveaux axiomes qui doivent permettre de rendre compte des comportements observés. Le manque de validité descriptive de la théorie étant dû à la transitivité de la relation de préférence, à l'axiome d'indépendance, c'est autour de trois idées que se sont organisées les généralisations du paradigme de l'utilité espérée: relaxation de la transitivité de la relation de préférence, affaiblissement de l'axiome d'indépendance (mais en gardant des courbes d'indifférence linéaire dans le triangle de Marschak-Machina), rejet de l'axiome d'indépendance (les courbes d'indifférence ne sont alors plus forcément linéaire). Nous organiserons ces présentations en donnant l'intuition à la base de la théorie, qui est en général une intuition de raisonnement, la forme de la fonctionnelle de préférence ainsi que la forme des courbes d'indifférence qu'elle induit dans le triangle de Marschak-Machina, ceci dans le but d'ensuite comparer ces différents modèles entre eux. Chacune de ces théories posant une hypothèse de raisonnement particulière, leur comparaison engendre une hypothèse de représentation des raisonnements dans la décision.

B.4.1 Les modèles intransitifs

Ces modèles sont fondés sur l'hypothèse que c'est l'intransitivité de la relation de préférence qui est à la base des déboires de la théorie de l'utilité espérée. Les deux théories suivantes, théorie du regret et du désappointement, sont deux théories visant à prendre en compte une réaction psychologique devant le risque, réaction face à la possibilité de prendre une mauvaise décision. La différence entre les deux est que, dans le cas du regret, cette réaction est déterminée avant la décision (donc fondée sur les conséquences) alors que dans le cas du désappointement, cette réaction est déterminée par les informations disponibles au moment de la décision (donc fondées sur les utilités espérées). On mentionnera aussi la théorie SSB [Fishburn 1982] qui résulte quand à elle de la volonté de rejeter axiome de transitivité de la relation de préférence et axiome d'indépendance de la définition de la rationalité.

La théorie du regret [Loomes and Sugden 1982] [Sugden 1993] Elle se fonde sur une intuition relative à la psychologie des décideurs: *lorsqu'ils ont effectué un choix entre deux alternatives risquées, les individus expriment un sentiment de réjouissance ou de regret face à leur choix, selon la conséquence issue de l'état de la nature effectivement réalisé.* Ainsi, si c_{ij} désigne la conséquence de l'acte i lorsque l'état de la nature est j,

- l'individu a du regret si $u(x_{ij}) < u(x_{kj})$,
- l'individu se réjouit si $u(x_{ij}) > u(x_{kj})$

Le comportement de l'individu est alors conforme à la maximisation d'une fonctionnelle d'espérance d'utilité modifiée, soit

$$U(L) = \sum_{j=1}^{n} p_j m_{ij}^k \text{ où } m_{ij}^k = u(x_{ij}) + R(u(x_{ij}) - u(x_{kj}))$$

où R est une fonction strictement croissante, nulle en zéro. Le lieu des courbes d'indifférence pour ce modèle est donné à la figure B.2.

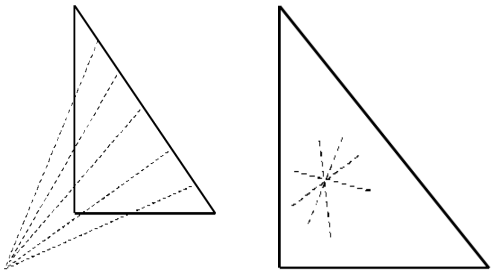
Le modèle du désappointement [Bell 1985] Cette théorie est basée sur une explication du renversement des préférences dans le paradoxe de Allais, explication qui est la suivante:

- dans la comparaison entre les deux premières loteries, si l'on choisit de parier et que l'on n'obtient rien, le désappointement ressenti lors du passage de 10% de chances de gagner 2500000 à 0 est très important, et rend, en regard, le gain sûr de 500000 très attractif,
- dans le second cas, une grande disparité n'est pas attendue entre le désappointement de perdre à l'une ou l'autre de ces deux loteries.

Fondant son argumentation sur le fait qu'il existe une réaction psychologique causée par la comparaison du résultat actuel avec celui que l'individu avait anticipé, [Bell 1985] montre que cette hypothèse conduit l'individu à maximiser une espérance d'utilité modifiée de la forme suivante:

$$U(L) = \sum_{i=1}^{n} p_i * (u(x_i) + C[u(x_i) - \sum_{i=1}^{n} p_i * u(x_i)])$$

La figure B.2 illustre la forme des cartes d'indifférence pour la théorie du désappointement.



(a) (b)

FIG. B.2: *Cartes d'indifférence dans le cas des théories du regret et du désappointement*

On signalera également la théorie de l'utilité espérée SSB [Fishburn 1982] (pour la représentation des loteries par une fonctionnelle d'utilité ayant des propriétés de symétrie oblique (*skew symmetry*, qui donne des courbes d'indifférence similaires à celles des modèles du regret et du désappointement lorsque les loteries sont indépendantes.

B.4.2 Les modèles affaiblissant l'axiome d'indépendance

Ces modèles ont été élaborés les auteurs qui ont posé comme hypothèse que l'axiome d'indépendance est la cause principale des déboires de la théorie de l'utilité espérée.

L'utilité pondérée Cette théorie repose aussi sur une interprétation de pourquoi l'axiome d'indépendance est violé. En effet, celui-ci impose que pour toutes lois $P', P'', Q \in \mathcal{P}$ et tout $\lambda \in]0,1]$, $P' \succ P'' \Leftrightarrow \lambda P' + (1-\lambda)Q \succ \lambda P'' + (1-\lambda)Q$. Cette égalité reste valable lorsque $Q = P'$. Or, si l'on interprète le mixage comme nous l'avons fait par des choix de loteries en deux temps, l'incertitude additionnelle induite par la première étape rend le mixage $\lambda P' + (1-\lambda)Q$ suffisamment distinct de la loterie simple P' ou P'' pour que, contrairement aux prescriptions de l'axiome d'indépendance, $\lambda P' + (1-\lambda)Q$ n'est pas indifférent à P'. D'où l'idée de supposer qu'il existe un nombre β, tel que si le choix de λ est fait, alors $\lambda P' + (1-\lambda)Q$ et $\beta P' + (1-\beta)Q$, qui implique tous les deux une incertitude dans la première étape, peuvent être indifférents pour une certaine valeur de β. On remplace donc l'axiome d'indépendance par:

Indépendance faible: $\forall P, Q, R, si P \sim Q, \forall \beta \in [0,1], \exists \alpha \in [0,1], \beta P + (1-\beta)R \sim \alpha Q + (1-\alpha)R$

Il existe alors une fonction ω, linéaire en probabilité telle que:

$$U(L) = \sum_{i=1}^{n} \frac{p_i * \omega(x_i) * u(x_i)}{\sum_{k=1}^{n} p_k * \omega(x_k)}$$

et les courbes d'indifférence ont alors la forme évoquée à la figure B.3.

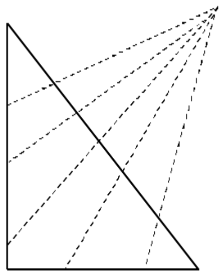


FIG. B.3: *Cartes d'indifférences dans le cas de la théorie l'utilité pondérée*

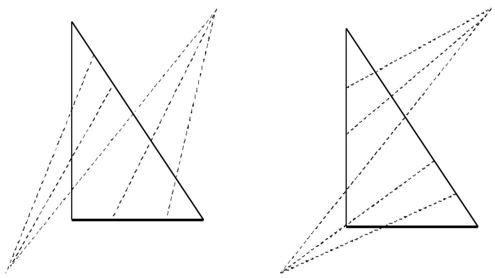
La théorie de l'aversion au désappointement [Gul 1991] Le fondement de cette théorie est un affaiblissement de l'axiome d'indépendance le plus faible possible pour prendre en compte le paradoxe de Allais. Si l'on reprend ce fameux paradoxe, notons que la loterie Q n'a aucune chance d'offrir un résultat décevant, alors que la loterie P offre une probabilité 0.01 d'offrir un résultat décevant. Une explication possible de l'invalidité de l'axiome d'indépendance dans cet exemple particulier est que la loterie avec la probabilité la plus faible de désappointement souffre le plus lorsqu'elle est mixée avec une loterie inférieure (probabilité 1 de gagner 0); c'est à dire que si les loteries étaient à peu près indifférentes initialement, la loterie avec la probabilité la plus forte de désappointement (S) peut être préférée après avoir été mixée avec une loterie inférieure.

À partir de cette analyse, [Gul 1991] définit une fonctionnelle d'utilité espérée modifiée qui se décompose en utilité espérée avec réjouissance et utilité espérée avec désappointement:

$$U(L) = \gamma(\alpha) \sum_{c \geq \sum_j p_j c_{ij}} [u(c) * \frac{p_j}{\alpha}] + (1 - \gamma(\alpha)) \sum_{c \leq \sum_j p_j c_{ij}} [u(c) * \frac{p_j}{1-\alpha}]$$

où $\gamma(\alpha) = \frac{\alpha}{1-(1-\alpha)\beta}$ représente l'importance accordée à la réjouissance, et $\alpha = \sum_{c \geq \sum_j p_j c_{ij}} p_j$ représente la condition de réjouissance.

On trouvera alors à la figure B.4 les cartes d'indifférence obtenues avec ce modèle.



(a) (b)

FIG. B.4: *Cartes d'indifférence dans le cas de la théorie de l'aversion au désappointement*

La théorie de la perspective de référence [Viscusi 1989] Cette théorie a été développée à partir de l'idée que, pour chaque loterie, la perception des probabilités diffère dès le début de l'analyse des probabilités réelles selon l'expérience individuelle: l'individu se forge une idée précise sur la distribution de probabilité en fonction de distributions déjà rencontrées. On définit donc une fonction de déformation des probabilités par la moyenne pondérée de la représentation q de l'individu de la probabilité d'origine p et de cette dernière, soit $\Pi(p) = \frac{\lambda q+\xi p}{\lambda+\xi}$ où λ est constant, et où ξ varie avec les conséquences.

On peut alors définir la fonctionnelle de préférence par:

$$U(L) = \sum_{j=1}^{n} \Pi(p_j)u(c_{ij})$$

Dans le modèle relaxant ou diminuant l'axiome d'indépendance, on trouve également les théories de l'utilité espérée implicite et utilité pondérée implicite.

B.4.3 Les modèles rejetant l'axiome d'indépendance

Ces modèles ont été élaborés par ceux qui pensent qui posent comme hypothèse que la fonctionnelle de préférence n'est pas linéaire en probabilités. Certains ont bâti leur théorie à partir de fonctions d'utilité locales, d'autres intègrent des fonctions de pondération non linéaire des probabilités.

L'utilité espérée généralisée [Machina 1982] Son fondement est un rejet total de l'axiome d'indépendance. Ce rejet est motivé par une interprétation de la manière dont cet axiome est violé, à savoir l'hypothèse que les individus sont très sensibles aux changements de probabilités d'événements isolés. Reprenons encore une fois l'exemple du paradoxe de Allais. Notons que le changement de Q à P et celui de R à S peut être pensé comme un mouvement de 0.10 masses de probabilité du résultat 500000 au 2500000 et un déplacement d'une masse de 0.01 unités de probabilité du résultat 500000 à 0. Quand la première perspective est Q, le mouvement ascendant de 0.10 masses de probabilités n'est pas suffisant pour compenser le mouvement descendant de 0.01 masses de probabilité. Néanmoins, lorsque la perspective initiale est R, le résultat 0 n'est plus un résultat si isolé de la distribution initiale puisque, relativement à Q, sa probabilité a augmenté de 0 à 0.89. L'individu n'est plus alors aussi réactif à l'augmentation de 0.01 en probabilité et ce mouvement est dès lors compensé par un mouvement ascendant de 0.10 masses de probabilité, aussi S est-il préféré.

Dit autrement, un mouvement différentiel d'une distribution $F(.)$ à une distribution $F'(.)$ change la valeur d'utilité espérée par la différence en valeur de $U(x, F)$ en fonction des distributions F et F': un mouvement différentiel de F à F' change alors la fonctionnelle d'utilité par $\int_D u(c, Q)[dF_P(c) - dF_Q(c)]$. Des préférences lisses (i.e. locales) sont introduites en intégrant des fonctions d'utilité locales dans la fonctionnelle d'utilité. On a donc, pour deux loteries P et Q,

$$U(P) - U(Q) = \int_D u(c, Q)[dF_P(c) - dF_Q(c)] + o(||P - Q||)$$

où $F_P(.)$ représente les distributions cumulatives de probabilité des loteries P et Q et D est l'ensemble des conséquences envisagées pour les loteries P et Q. On ajoute une hypothèse de structure en éventail de telle sorte que, si P domine stochastiquement Q au premier ordre, $\frac{-u''(c,P)}{u'(c,P)} < \frac{-u''(c,Q)}{u'(c,Q)}$ qui signifie que l'aversion au risque augmente au fur et à mesure que les loteries sont plus favorables.

Les cartes d'indifférence pour cette théorie sont à la figure B.5.

L'utilité quadratique [Chew et al. 1991] C'est l'observation du comportement concret des individus, en contradiction avec l'axiome d'indépendance, qui a conduit [Chew et al. 1991] à poser cette théorie. L'interprétation est un affaiblissement de celle évoquée dans le cas de l'utilité pondérée.

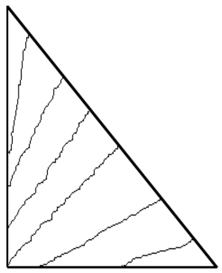


FIG. B.5: *Cartes d'indifférence dans le cas de l'utilité espérée généralisée*

L'idée est que tout mixage qui place moins de poids sur P doit être indifférent à un autre mixage qui place plus de poids sur P. Plus formellement,

$$\forall P, Q, R, \forall \beta \in [0, \frac{1}{2}], \exists \alpha \in [\frac{1}{2}, 1], \beta P + (1-\beta)R \sim \alpha Q + (1-\alpha)R$$

On peut alors définir une fonctionnelle d'utilité modifiée de la forme

$$U(L) = \sum_i \sum_j \Phi(c_i, c_j) p_i p_j \text{ où } \Phi(c_i, c_j) = \Phi(c_j, c_i)$$

On trouvera à la figure B.6 les courbes d'indifférence données par ce modèle.

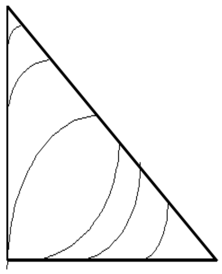


FIG. B.6: *Cartes d'indifférence dans le cas de l'utilité quadratique*

La théorie des perspectives aléatoires [Kahneman and Tversky 1979][Tversky and Kahneman 1992]

Cette théorie, conçue par des psychologues, résulte plus d'une somme de constatations empiriques dans le triangle unité, où elle reste d'ailleurs seule valable, que d'une axiomatique fondée sur des intuitions normatives. Elle se décompose en deux phases: une phase d'édition et une phase d'évaluation.

La phase d'édition est une réorganisation et une reformulation des perspectives afin de simplifier la phase d'évaluation, et se décompose comme suit:

- une première étape de codage des conséquences en terme de gains et de pertes par rapport à un point de référence,
- la combinaison des conséquences communes d'une même loterie (on fait la somme des probabilités des mêmes conséquences),
- l'élimination des caractéristiques communes des loteries à comparer,
- l'approximation des probabilités et utilités à la première décimale.

La phase suivante est une phase d'évaluation. Elle permet de choisir parmi l'ensemble des perspectives proposées celle qui maximise la fonctionnelle d'utilité proposée, construite à partir d'une combinaison d'une fonction de déformation des probabilités π et d'une fonction d'utilité u.

- la fonction π possède les caractéristiques suivantes:
  - Elle est monotone, croissante et continue sur $[0, 1]$ de telle sorte que $\pi(0) = 0$ et $\pi(1) = 1$,

- Elle est sous-additive: $\pi(p) + \pi(1-p) < 1$,
- Elle satisfait une propriété de sous proportionalité $\frac{\pi(p.q)}{\pi(p)} \leq \frac{\pi(p.q.r)}{\pi(p.r)}$,
- Elle surévalue les petites probabilités $\pi(p) > p$,
- Elle sous-évalue les autres probabilités $\pi(p) < p$.

- la fonction u évalue les conséquences en fonction d'un point de référence, en faisant une distinction entre gains et pertes.

La fonctionnelle de préférence s'écrit alors, dans le cadre du triangle de Marschak-Machina, selon le classement des conséquences par rapport au point de référence:

- $x_1 < 0 < x_3$ et $p_1 + p_3 < 1$ alors $U(L) = \pi(p_1)x_1 + \pi(p_3)x_3$,
- $x_1 < x_2 < 0$ ou $0 < x_2 < x_3$ et $p_1+p_2+p_3 = 1$ alors $U(L) = u(x_2)+\pi(p_i)[u(x_1)-u(x_2)]$, $i = 1, 3$

Le modèle dichotomique [Quiggin 1982][Yaari 1987][Segal 1989] Les modèles dichotomiques, développés indépendamment par ces trois auteurs, se fondent sur une fonction de pondération des probabilités des conséquences, définie en fonction de l'ordre sur les conséquences. Elle est fondée sur une intuition de Allais, que la fonctionnelle de préférence s'exprime comme produit scalaire d'un vecteur de fonctions de probabilités par un vecteur des incréments d'utilité des conséquences, conséquences ordonnées par ordre croissant, soit:

$$U(L) = \theta_1 u_1 + \theta_2(u_2 - u_1) + \ldots + \theta_n(u_n - u_{n-1})$$

Allais démontre alors que tout se passe comme si l'on avait:

$$\theta_1 = 1; \quad \theta_2 = \theta(1-p_1) = \theta(p_2 + p_3 + \ldots + p_n); \quad \theta_n = \theta(1 - p_1 - p_2 - p_3 - \ldots - p_{n-1}) = \theta(p_n)$$

Ceci correspond à un raisonnement du type suivant: "... je suis sûr d'obtenir la conséquence que me procure l'utilité u_1, mais au delà de cette utilité, j'aurais avec une probabilité $\theta(1-p_1)$ l'incrément $u_1 - u_2$, avec une probabilité $\theta(1-p_1-p_2)$ l'incrément $u_3 - u_2$ ....

L'axiomatisation suivant ce raisonnement donne alors la représentation suivante pour la fonctionnelle de préférence par, si $\forall i, x_i \preceq x_{i+1}$,

$$U(L) = \sum_{i=1}^{n} u(x_i)[\theta(\sum_{j=1}^{i} p_j) - \theta(\sum_{j=1}^{i} p_j)]$$

où la fonction g est monotone croissante sur $[0,1]$, avec $g(0) = 0$ et $g(1) = 1$. On trouvera à la figure B.7 les cartes d'indifférence de ce modèle, ainsi que la forme de la fonction de pondération g, à la fois concave et convexe, qui rend au mieux compte de comportements individuels.

La principale caractéristique des modèles dichotomiques est de représenter l'inclinaison au risque par le biais d'une fonction de déformation des probabilités et non par la concavité de la fonction neumanienne d'utilité. C'est en effet plus intuitif, dans la mesure où l'aversion ou l'inclinaison au risque a plus à voir avec la perception des probabilités qu'avec celle des conséquences. Ainsi, si la fonction déformation est concave, l'individu a un comportement d'aversion au risque, et si elle est convexe, il a un comportement d'inclinaison au risque.

On notera également parmi la classe de ces modèles rejetant l'axiome d'indépendance l'utilité dépendant des loteries [Becker and Sarin 1987]]. Cette théorie est fondée sur la possibilité de faire varier la fonction d'utilité en fonction de la loterie considérée. On notera également la théorie du choix dans le risque prenant en compte un facteur de sécurité [Jaffray 1988], où un niveau de sécurité est intégré dans le processus de décision, et permet de prendre en compte les plus mauvais résultats des décisions aussi bien que les résultats obtenus avec peu de risques.

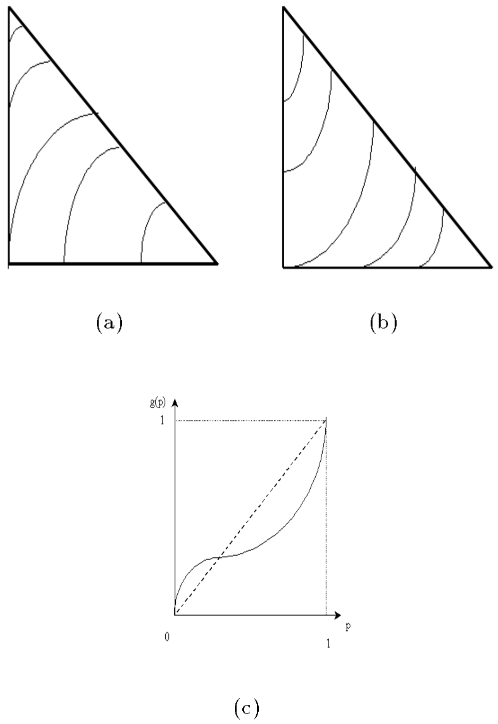


FIG. B.7: *Cartes d'indifférence dans le cas du modèle dichotomique*

B.5 Comparaison des généralisations de l'utilité espérée

On peut se demander quelle est la validité de ses différents modèles théoriques, i.e. leur capacité à représenter des décisions individuelles, qui reste le meilleur critère de validité d'une théorie à vocation descriptive, comme le souligne Maurice Allais, cité par [Munier 1995]:

> "Il est bien sûr facile de développer des théories, de formuler des axiomes et de prouver des théorèmes si, comme trop d'"économistes mathématiques" le font, on se confine au seul champ délicieux de la pure abstraction mathématique .... Mon point de vue personnel est tout à fait différent. *J'ai une profonde aversion pour les raffinements mathématiques inutiles et l'expérience me semble être le seul critère*. Pour moi, empiler axiome sur axiome ne rend pas une théorie vraie, et tout construction mathématique est inutile si elle est jetée à bas par les faits".

Le triangle de Marschak-Machina va nous servir à tester la validité empirique de ces théories en permettant de comparer entre elles ces généralisations de l'utilité espérée. En effet, face à cette multiplication des critères de décision, le problème devient dès lors de choisir le bon. Pour comparer ces modèles entre eux, on peut, pour chaque modèle, produire les courbes d'indifférence qui lui sont associées et les comparer expérimentalement avec celles de décideur. Deux méthodes ont été proposées et implémentées pour ce faire:

Les méthodes de proportion de choix cohérents Il s'agit d'estimer la vraisemblance des choix proposés par un modèle en mesurant la distance entre le choix de ce modèle et celui réalisé de manière aléatoire, les deux par rapport aux choix du sujet. Ce critère n'aboutit cependant pas à des conclusions bien arrêtées pour le choix de référence, sans pour autant qu'une fonctionnelle généralisation le modèle de l'utilité espérée face mieux que celle-ci (voir [Chew and Waller 1986][Camerer 1989][Battalio et al. 1990][Harless and Camerer 1994]).

Les méthodes économétriques Il s'agit d'estimer la pertinence d'un modèle en supposant que le choix d'un décideur par rapport à celui du modèle est représenté par une variable aléatoire y qui prend une valeur positive ou négative selon la préférence révélée. Supposant de plus que le choix du décideur est entachée d'un bruit blanc ϵ[33], si P et Q sont deux loteries, on a $y = V(P) - V(Q) + \epsilon$, où V est la fonctionnelle de préféérence à tester. On peut alors faire un test statisque (de Show) permettant de tester l'adéquation de y aux données fournies par le décideur. Cependant, cette approche ne permet pas plus de discriminer les modèles puisque, sur les dix testés, huit se classent premier ex-aqueo (dont le modèle de l'utilité espérée) (voir [Hey and Orme 1994]).

L'idée, avancée par [Abdellaoui and Munier 1994][Munier 1995], est que ces comparaisons cherchent à départager les modèles sur l'ensemble du triangle de Marschak-Machina, alors que si l'on considère par exemple le paradoxe de Allais qui implique que les préférences sont déformées au voisinage de la certitude, il faudrait plutôt chercher à qualifier les performances des modèles sur des sous-zones appartenant au triangle de Marschak-Machina. [Abdellaoui and Munier 1994] propose donc de définir une notion de structure de risque comme partageant le triangle en zones dans lequel chaque fonctionnelle de préférence peut être comparée aux autres. En découpant de cette manière le triangle de Marschak - Machina en ensembles bornés distincts, et en évaluant les performances descriptives de chaque modèle sur chacun de ces sous-ensembles, [Abdellaoui and Munier 1994][Munier 1995][Dernis 1995] parviennent à définir cinq zones dans le triangle correspondant à quatre fonctionnelles de préférence et une zone de turbulence où aucune fonctionnelle n'apparaît clairement:

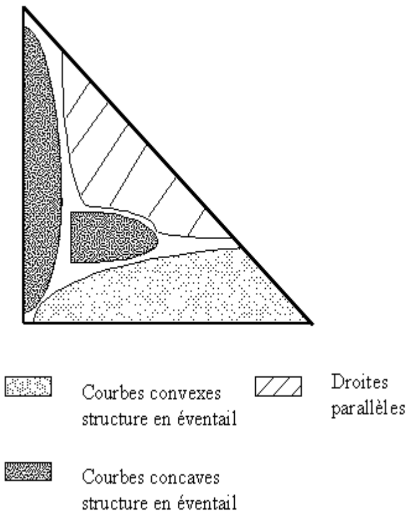


FIG. B.8: *Partage en structures de risque du triangle de Marschak-Machina*

Le parallélisme des courbes d'indifférence le long de l'hypothénuse et au dessus de la première bissectrice permet d'affirmer que le critère de l'utilité espérée y est valide.

la structure en éventail et la convexité le long des cotés principaux du triangle et dans ses parties inférieures permet d'affirmer que le modèle de l'utilité quadratique et du modèle dichotomique y sont valides,

la structure en éventail et la concavité le long du coté gauche et dans la partie milieu du triangle permet d'affirmer que le modèle dichotomique ou la théorie des perspectives aléatoires y sont valides.

L'idée derrière cette notion de structure de risque est que les individus utilisent une heuristique différente selon la zone du triangle dans laquelle ils se trouvent pour évaluer les perspectives. Cette

33. Bruit blanc, car en économie, même les imperfections humaines ont leurs limites.

idée de structure de risque aborde au moins implicitement l'idée de représentation de résultat d'un raisonnement en changeant la fonctionnelle de préférence $f(x)$, x représentant les différentes alternatives, en $f[y](x)$ où y est maintenant un paramètre prenant en compte la structure de risque: la fonctionnelle est dépendante de la structure de risque. Pour généraliser ce propos, il est alors nécessaire, non plus d'avoir une axiomatique, mais une mécanique contextuelle permettant de manipuler les axiomes représentant le raisonnement dans la décision; on a donc besoin d'une logique.

B.6 Conclusions

Les limites expérimentales de la théorie de l'utilité espérée et de ses différentes généralisations permettent de supposer que le manque de vertu descriptive de ce modèle tient sa source dans les présupposés de base du modèle, à savoir la rationalité substantive et non procédurale, et la représentation de la structure des choix et non pas de leur contenu. Entamer une approche descriptive de la décision revient donc à remplacer l'hypothèse de représentation de la décision (à l'origine de ces présupposés) par l'hypothèse de représentation du raisonnement humain dans la décision.

C

Borne supérieure pour la subsomption de n formules du premier ordre

C.1 Notations

En plus des notations dj́à introduites, on définit pour toute formule φ (voir aussi [Lalement 1990]):

- L'ensemble $\mathcal{O}(\varphi)$ des occurrences de φ,
- La sous formule φ/u de φ en u pour $u \in \mathcal{O}(\varphi)$
- Le symbole $\varphi(u)$ en u pour $u \in \mathcal{O}(\varphi)$

i) Si $\varphi \in At_\Sigma[X]$ ou $\varphi = \perp$, alors $\mathcal{O}(\varphi) = \{\varepsilon\}, \varphi(\varepsilon) = \varphi$ et $\varphi/\varepsilon = \varphi$.

ii) si $\varphi = R\varphi_1 \ldots \varphi_n$ où R est un connecteur ou un $\forall \vec{x}$ ou un $\exists \vec{x}$, alors:

$$\mathcal{O}(\varphi) = \varepsilon \cup \bigcup_{i=1}^{n} i.\mathcal{O}(\varphi_i)$$

$$\begin{array}{cc} \varphi(\varepsilon) = R & \varphi/\varepsilon = \varphi \\ \varphi(i.u) = \varphi_i(u) & \varphi/i.u = \varphi_i/u \end{array}$$

On définit l'ensemble des atomes contenus dans une formule φ par:

$$A_\varphi = \{R \in \Sigma_R, \exists u \in \mathcal{O}(\varphi), \varphi(u) = R\}$$

C.2 Existence d'une borne supérieure

Comme (ψ "subsume" φ) $\Leftrightarrow$ ($\varphi \models \psi$) $\Leftrightarrow$ $\mathbf{Mod}\{\varphi\} \subseteq \mathbf{Mod}\{\psi\}$, **S** est une relation de pré-ordre sur L_Σ, par "héritage" du pré-ordre $\subseteq$. On peut d'autre part montrer:

Lemme 104. *Soit $\mathcal{E}$ une classe d'algèbres finiment axiomatisable. Alors il existe $\varphi \in L_\Sigma$ telle que $\mathcal{E} = \mathbf{Mod}\{\varphi\}$.*

Preuve. Soit $\mathcal{E}$ une classe d'algèbres finiment axiomatisable. Alors il existe $\Gamma \subseteq L_\Sigma$, Γ fini tel que $\mathcal{E} = \mathbf{Mod}\ \Gamma$. Γ étant fini, soit $\Gamma = (\varphi_i)_{1 \leq i \leq n}$, et soit $\varphi = \bigwedge_{i=1}^{n} \varphi_i$.

Soit $\Delta = \{\psi \in L_\Sigma;\ \Gamma \vdash \psi\}$.

Alors $\varphi \in \Delta$ (par n-1 introductions du $\wedge$) et comme **Mod** $\Gamma=$ **Mod** Δ, on a **Mod** $\Delta \subseteq$ **Mod**$\{\ \}\varphi$, donc $\mathcal{E} \subseteq \mathbf{Mod}\{\varphi\}$.

Réciproquement, montrons $\mathbf{Mod}\{\varphi\} \subseteq \mathcal{E}$.
Soit $\mathcal{A} \in \mathbf{Mod}\{\varphi\}$.
Alors $\forall \xi \in A^X,\ \varphi^{\mathcal{A}}(\xi) \ = \ 1 \ = \ min_{1\leq i\leq n}\varphi_i^{\mathcal{A}}$
donc $\forall i, \varphi_i^{\mathcal{A}} = 1$, i.e. $\mathcal{A} \in \mathcal{E}$. Donc $\mathbf{Mod}\{\varphi\} \subseteq \mathcal{E}$.

De ces deux inclusions, on déduit $\mathcal{E} = \mathbf{Mod}\{\varphi\}$. ■

Lemme 105. *L'ensemble des classes de modèles finiment axiomatisables est un treillis pour l'inclusion.*

Preuve (sup-demi treillis). Soient $(\mathcal{E}_i)_{1\leq i\leq n}$ une famille quelconque de classes d'algèbres finiment axiomatisables telles que $\bigcap_{i=1}^n \mathcal{E}_i \neq \emptyset$.
D'après le lemme 104, ces classes d'algèbres étant finiment axiomatisables, il existe une famille de formules φ_i telles que :

$$\forall i, 1 \leq i \leq n, \mathcal{E}_i = \mathbf{Mod}\{\varphi_i\}$$

Par définition, toute algèbre de $\bigcap_{i=1}^n \mathcal{E}_i$ est modèle de $\bigwedge_{i=1}^n \varphi_i$, i.e.
$\bigcap_{i=1}^n \mathcal{E}_i \subseteq \mathbf{Mod}\{\bigwedge_{i=1}^n \varphi_i\}$.

Réciproquement, montrons $\mathbf{Mod}\{\bigwedge_{i=1}^n \varphi_i\} \subseteq \bigcap_{i=1}^n \mathcal{E}_i$.
Soit $\mathcal{A} \in \mathbf{Mod}\{\bigwedge_{i=1}^n \varphi_i\}$.
Alors $\forall \xi \in A^X,\ (\bigwedge_{i=1}^n \varphi_i)^{\mathcal{A}}(\xi) \ = \ 1 \ = \ min_{1\leq i\leq n}\varphi_i^{\mathcal{A}}$
donc $\forall i, \varphi_i^{\mathcal{A}} = 1$, i.e. $\mathcal{A} \in \bigcap_{i=1}^n \mathcal{E}_i$. Donc $\mathbf{Mod}\{\bigwedge_{i=1}^n \varphi_i\} \subseteq \bigcap_{i=1}^n \mathcal{E}_i$.

La réunion d'un nombre fini de parties finies de L_Σ étant elle-même finie, il en résulte que $\bigcap_{i=1}^n \mathcal{E}_i$ est finiment axiomatisable. Par les propriétés classiques sur les ensembles, il résulte donc que l'ensemble des classes d'algèbres finiment axiomatisables est un sup-demi treillis. ■

Preuve (inf-demi treillis). Soient $(\mathcal{E}_i)_{1\leq i\leq n}$ une famille quelconque de classes d'algèbres finiment axiomatisables. Soit $(\varphi_i)_{1\leq i\leq n}$ telle que $\forall i,\ 1 \leq i \leq n, \mathcal{E}_i = \mathbf{Mod}\{\varphi_i\}$.

Montrons que $\bigcup_{i=1}^n \mathcal{E}_i$ est une classe d'algèbres finiment axiomatisable. On va plus précisement montrer que $\psi \equiv \bigvee_{i=1}^n \varphi_i$ est une axiomatisation de cette classe d'algèbres.

Il est immédiat que ψ subsume chacune des formules $(\varphi_i)_{1\leq i\leq n}$ par la règle d'introduction du $\vee$ en logique minimale. Ceci établit donc que $\bigcup_{i=1}^n \mathcal{E}_i \subseteq \mathbf{Mod}\{\psi\}$.

D'autre part, soit $\mathcal{A}$ un élément de $\mathbf{Mod}\{\psi\}$.
Alors,
$\forall \xi \in A^X, (\psi^{\mathcal{A}})(\xi) \ = \ 1 \ = \ max_{1\leq i\leq n}\varphi_i^{\mathcal{A}}$
Donc $\exists i, 1 \leq i \leq n, \mathcal{A} \subseteq \mathcal{E}_i$. Ceci établit donc que $\mathbf{Mod}\{\psi\} \subseteq \bigcup_{i=1}^n \mathcal{E}_i$
Ce qui termine la démonstration. ■

Théorème 106. $(L_\Sigma, \mathbf{S})$ *est un treillis.*

Preuve. Ceci résulte de l'équivalence entre ψ **S** φ et $\mathbf{Mod}\{\varphi\} \subseteq \mathbf{Mod}\{\psi\}$. L'ensemble des classes de modèles finiment axiomatisables étant un treillis pour l'inclusion, l'équivalence permet de transporter la structure de treillis sur L_Σ. ■

Dans cette section, on considérera une famille $(\varphi_i)_{1\leq i\leq n}$ de formules de L_Σ universellement quantifiées. On établira certaines propriétés de la borne supérieure ψ (pour la relation de subsomption) de cet ensemble de formules, propriétés visant à établir des conditions "syntaxiques" sur l'ensemble des formules ψ solutions de notre problème. On considérera, sauf indication contraire, que les formules $(\varphi_i)_{1\leq i\leq n}$ sont en forme normale conjonctive:

$$\forall i, 1 \leq i \leq n, \varphi_i = \forall \vec{x} \bigwedge_{j=1}^{p} \bigvee_{k=1}^{p_j} \varphi_{i,j}^k.$$

Proposition 107. *Soit φ une formule universelle de L_Σ, écrite en forme prénexe normale disjonctive $\varphi \equiv \forall \vec{x} \bigvee_{i=1}^{n} \bigwedge_{j=1}^{p_i} \varphi_i^j$. Alors ψ subsume φ ssi ψ subsume $\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j$ pour tout i.*

Preuve (condition suffisante). La condition est suffisante en écrivant chaque formule $\vdash \exists \vec{x}(\psi \vee \neg\varphi_{i_0})$, puis en utilisant la règle d'élimination du $\exists$ de la logique minimale, puis en utilisant la règle d'introduction du $\vee$ de la logique minimale pour tout $i \neq i_0$ de $\varphi_i(\ \vec{t}\ )$ avec $t \in T_\Sigma[X]$, puis en ré-introduisant le quantificateur $\exists$, on obtient $\vdash \forall \vec{x} \bigvee_{i=1}^{n} \bigwedge_{j=1}^{p_i} \varphi_i^j \Rightarrow \psi$. ∎

Preuve (condition nécessaire). On raisonne par l'absurde.
Soit ψ telle que ψ subsume φ, i.e. telle que $\vdash \forall \vec{x} \bigvee_{i=1}^{n} \bigwedge_{j=1}^{p_i} \varphi_i^j \Rightarrow \psi$.
On considère l'ensemble $I_0 = \{i, 1 \leq i \leq n, \vdash \forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j \not\Rightarrow \psi\}$. Par hypothèse, $I_0 \neq \emptyset$.
L'ensemble des classes de modèles des formules de I_0 est donc tel que
$\forall i \in I_0, \mathbf{Mod}\{\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j\} \not\subseteq \mathbf{Mod}\{\psi\}$.
Or

$$\forall i, 1 \leq i \leq n, \mathbf{Mod}\{\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j\} \subseteq \mathbf{Mod}\{\forall \vec{x} \bigvee_{i=1}^{n} \bigwedge_{j=1}^{p_i} \varphi_i^j\} \subseteq \mathbf{Mod}\{\psi\}$$

par les règles d'élimination du quantificateur universel, d'introduction du $\vee$ de la logique minimale pour la première inclusion, puis de ré-introduction du quantificateur universel, ce qui entraîne:

$$\forall i, 1 \leq i \leq n, \mathbf{Mod}\{\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j\} \subseteq \mathbf{Mod}\{\psi\}.$$

On a donc

$$\begin{array}{lccc} \forall i \in I_0, & \mathbf{Mod}\{\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j\} & \not\subseteq & \mathbf{Mod}\{\psi\} \\ & & \text{et} & \\ & \mathbf{Mod}\{\forall \vec{x} \bigwedge_{j=1}^{p_i} \varphi_i^j\} & \subseteq & \mathbf{Mod}\{\psi\}, \end{array}$$

ce qui est absurde. ∎

En jouant sur la dualité entre les connecteurs $\wedge$ et $\vee$ vis à vis de la négation, ainsi que sur la dualité borne inférieure et supérieure au sens du pré-ordre de subsomption, on peut montrer:

Remarque 108. Soit φ une formule universelle de L_Σ, écrite en forme prénexe normale conjonctive $\varphi \equiv \forall \vec{x} \bigwedge_{i=1}^{n} \bigvee_{j=1}^{p_i} \varphi_i^j$. Alors ψ est subsumée par φ ssi ψ est subsumée par $\forall \vec{x} \bigvee_{j=1}^{p_i} \varphi_i^j$ pour tout i.

Proposition 109. *Soient ψ et ψ' respectivement borne supérieure et borne inférieure au sens de la relation de subsomption des formules universelles $(\varphi_i)_{1\leq i\leq n}$ de L_Σ. Alors ψ et ψ' sont des formules universelles.*

Preuve. Montrons la proposition par récurrence sur n, nombre de formules subsumées.

n=1 ψ est borne supérieure (resp. ψ' est borne inférieure) au sens du pré-ordre de subsomption de $\varphi = \forall \vec{x} \bigwedge_{j=1}^{m} \varphi^j$. Montrons par induction sur φ que ψ et ψ' sont universelles, en utilisant la base de connecteurs $\vee, \neg, \forall$ pour la borne supérieure et $\wedge, \neg, \forall$ pour la borne inférieure:

- Si φ est un littéral, alors $\psi = \psi' = \varphi$ et la proposition est vraie.

- *(Borne supérieure)* Si $\varphi = \phi \vee \phi'$, alors ψ est borne supérieure de ϕ et ϕ' (voir 107) et est donc universelle par la première phase d'induction.
  (Borne inférieure) Si $\varphi = \phi \wedge \phi'$, alors ψ' est borne inférieure de ϕ et ϕ' (voir 108) et est donc universelle par la première phase d'induction.

La proposition est vraie au rang $n = 1$.

n=2 ψ étant borne supérieure de φ_1 et φ_2 est borne supérieure de $\varphi_1 \vee \varphi_2$, et par application de l'hypothèse de récurrence au rang $n = 1$, est universelle.
ψ' étant borne inférieure de φ_1 et φ_2 est borne inférieure de $\varphi_1 \wedge \varphi_2$, et par application de l'hypothèse de récurrence au rang $n = 1$, est universelle.

n-1, n Supposons la propriété vraie au rang n-1. Soit ψ la borne supérieure au sens du pré-ordre de subsomption de $(\varphi_i)_{1 \leq i \leq n}$. Alors ψ subsume $(\varphi_i)_{1 \leq i \leq n-1}$. Deux cas se présentent alors:

- Si ψ est la borne supérieure de $(\varphi_i)_{1 \leq i \leq n-1}$ alors par hypothèse de récurrence, ψ est universelle, et la proposition est démontrée,
- Sinon ψ subsume ψ_{n-1}, formule universelle par hypothèse de récurrence, borne supérieure de $(\varphi_i)_{1 \leq i \leq n-1}$, et φ_n. ψ est donc borne supérieure de ψ_{n-1} et de φ_n, et par hypothèse de récurrence au rang 2, ψ est donc une formule universelle.

La proposition est donc vraie pour tout n. ■

Remarque 110. Soit φ une formule universellement quantifiée de L_Σ. Alors, pour toute substitution σ, $\sigma\varphi$ est conséquence de φ.

Preuve. Soit donc une formule $\varphi \in L_\Sigma[x_1, \dots, x_m]$, et soit $t_1, \dots, t_m \in T_\Sigma[y_1, \dots, y_n]$. Alors, par les règles d'élimination du $\forall$ appliquée aux $\forall x_i, 1 \leq i \leq m$, puis les n introductions du $\forall y_i, 1 \leq i \leq n$, on obtient une règle dérivée:

$$\forall\forall: \frac{\dfrac{\forall \vec{x}\varphi}{\varphi[x_1 = t_1, \dots, x_n = t_m]}}{\forall \vec{y}(\varphi[x_1 = t_1, \dots, x_n = t_m])}$$

correspondant à une règle notée $\forall\forall$ permettant une dérivation de $\sigma\varphi$ à partir de φ, où σ est une substitution (voir [Lalement 1990]). ■

La forme syntaxique de la borne inférieure au sens du pré-ordre de subsomption de n formules universelles a été étudiée par [Plotkin 1970][Plotkin 1971]. Nous ne nous intéresserons plus par la suite qu'au cas de la borne supérieure de n formules universelles.

Le but étant de calculer cette borne supérieure, on va d'abord établir un certain nombre de propriétés de cette borne supérieure qui vont permettre d'énoncer un algorithme de calcul de ψ énumérant toutes les solutions dans le cas où les deux formules φ_1 et φ_2 sont universellement quantifiées.

C.3 Propriétés de la borne supérieure

On appelle formule conjonctive pure une formule ne contenant comme connecteur logique binaire que des conjonctions. La proposition 107 va nous permettre dans la suite du raisonnement de ne plus considérer que des formules conjonctives pures. On l'utilisera ensuite pour généraliser le raisonnement établi aux formules disjonctives.

Dans la suite, on ne considérera plus comme ensemble de formules de L_Σ qu'un ensemble $(\varphi_i)_{1 \leq i \leq n}$ de formules purement conjonctives quantifiées universellement. On écrira chacune de ces formules en forme prénexe:

$$\forall i, 1 \leq i \leq n, \varphi_i \equiv \forall \vec{x} \bigwedge_{j=1}^{m} \varphi_i^j$$

Proposition 111. *Soit $(\varphi_i)_{1\leq i\leq n}$ une famille de formules universelles conjonctives de L_Σ. Alors, si ψ est une borne supérieure de cette famille au sens de la relation de subsomption, tous les littéraux de ψ appartiennent aux $(\varphi_i)_{1\leq i\leq n}$ à une substitution σ près.*

Preuve. $\forall i, 1 \leq i \leq n, \varphi_i \models_L \psi$. Comme $(\varphi_i)_{1\leq i\leq n}$ et ψ sont des formules universelles, le problème de leur satisfiabilité peut être restreint aux algèbres de Herbrand. Donc, pour tout i, tout modèle de Herbrand de (φ_i) est modèle de ψ. Tous les littéraux de ψ appartiennent donc aux $(\varphi_i)_{1\leq i\leq n}$ à une substitution σ près. ■

Proposition 112. *Soit ψ une borne supérieure au sens de la relation de subsomption des formules $(\varphi_i)_{1\leq i\leq n}$ de L_Σ. Alors, si parmi l'ensemble des $(\varphi_i)_{1\leq i\leq n}$:*

- *une formule φ_{i_0} est telle qu'aucun de ses littéraux n'admet d'unificateur principal avec les autres formules $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$, alors ψ s'écrit $\psi \equiv \psi' \vee \varphi_{i_0}$.*
- *un littéral χ présent dans une formule φ_{i_0} n'admet aucun unificateur principal avec les autres formules $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$ et, pour chaque formule $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$, d'autres littéraux de φ_{i_0} admettent des unificateurs principaux, alors ψ ne contient pas le littéral χ.*

Preuve (Première partie de la proposition). Soient φ_{i_0} une formule ayant la propriété annoncée, et soit ψ une formule subsumante des $(\varphi_i)_{1\leq i\leq n}$. Ecrivons ψ en forme normale disjonctive $\psi \equiv \forall\vec{x} \bigvee_{k=1}^{m} \bigwedge_{j=1}^{p_k} \psi_k^j$.

ψ contient au moins un littéral de φ_{i_0}, sinon la satisfiabilité de ψ dans les modèles de Herbrand de φ_{i_0} ne serait pas assurée.

Soit k_0 la conjonction contenant ce littéral. Supposons que cette conjonction contienne un littéral χ des $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$. Comme $\varphi_{i_0} \models \psi$, tout modèle de Herbrand de φ_{i_0} est modèle de ψ. Si, pour toute algèbre de Herbrand $\mathcal{H}$ modèle de φ_{i_0}, $\forall \xi \in H^X, \bigvee_{j=1}^{p_{k_0}} (\psi_{k_0}^j)^{\mathcal{H}}(\xi) = 0$, alors on définit une formule ψ' de la manière suivante:

$$\begin{array}{lll} \psi' \equiv & \forall\vec{x} \bigvee_{k=1, k\neq k_0}^{m} & \bigwedge_{j=1}^{p_k} {\psi'}_k^j \\ & \forall k \neq k_0, & {\psi'}_k \equiv \psi_k \end{array}$$

Alors $\forall i, 1 \leq i \leq n, \psi'$ subsume ψ, ce qui est absurde puisque ψ est une borne supérieure.
Donc, pour toute algèbre de Herbrand $\mathcal{H}$ modèle de φ_{i_0}, $\forall \xi \in H^X, \bigvee_{j=1}^{p_{k_0}} (\psi_{k_0}^j)^{\mathcal{H}}(\xi) = 1$. Aucun littéral de φ_{i_0} n'admettant d'unificateur principal avec un autre littéral des $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$, χ n'admet pas d'unificateur principal avec les atomes de φ_{i_0}. Donc la conjonction k_0 ne contient aucun littéral qui n'appartienne pas à φ_{i_0}. Supposons qu'elle ne contienne que certains des littéraux de φ_{i_0}. Soit alors ψ' identique à ψ si ce n'est que la k_0 i-ème conjonction a été modifiée comme nous allons l'indiquer.
On écrit $\varphi_{i_0} \equiv \forall\vec{x}(\bigwedge_{k=1}^{k_0} \psi_k^j) \wedge (\bigwedge_{k=k_0}^{m} \psi_k^j)$ avec $k_0 \neq m$ (l'indiçage du $\wedge$ ayant été au besoin ré-effectuée), et on construit une formule ψ' en remplaçant la sous formule correspondant à la conjonction k_0 par φ_{i_0}. ψ' vérifie donc les propriétés suivantes:

$$\begin{array}{lll} \psi' \equiv & \forall\vec{x} \bigvee_{k=1}^{m} & \bigwedge_{j=1}^{p_k} {\psi'}_k^j \\ & \exists k_0, & \bigwedge_{j=1}^{p_{k_0}} {\psi'}_k^j \equiv \varphi_{i_0} \\ & \forall k \neq k_0, & {\psi'}_k \equiv \psi_k \end{array}$$

Alors $\mathbf{Mod}\{\psi'\} \subseteq \mathbf{Mod}\{\psi\}$ et $\forall i, 1 \leq i \leq n, \varphi_i \models \psi'$. Donc ψ' est plus petit que la borne supérieure. Ce qui est absurde. ■

Preuve (Deuxième partie de la proposition). On raisonne par l'absurde.
On suppose que ψ contienne le littéral χ. On écrit ψ en forme normale prénexe disjonctive, i.e. $\psi \equiv \forall\vec{x} \bigvee_{k=1}^{m} \bigwedge_{j=1}^{p_k} \psi_k^j$.
$\forall i, 1 \leq i \leq n, \varphi_i \models \psi$, et les formules ψ et $(\varphi_i)_{1\leq i\leq n}$ étant universelles, le problème de la satisfiabilité peut, d'après le théorème de Herbrand, être restreinte aux algèbres de Herbrand. Pour $i \neq i_0$, la conjonction contenant χ n'est jamais valuée à 1 dans les algèbres de Herbrand de φ_i.

S'il existe une conjonction la $k_0^{\text{i-ème}}$ valuée à 1 dans les algèbres de Herbrand modèle de φ_{i_0} qui n'est pas celle contenant χ, alors définissons ψ' par:

$$\begin{array}{lll} \psi' \equiv & \forall\vec{x} \bigvee_{k=1, k\neq k_0}^{m} & \bigwedge_{j=1}^{p_k} {\psi'}_k^j \\ & \forall k \neq k_0, & {\psi'}_k \equiv \psi_k \end{array}$$

Alors, ψ' subsume chacune des formules $(\varphi_i)_{1\leq i\leq n}$ et subsume ψ, ce qui est absurde. Donc la conjonction

contenant χ est valuée à 1 dans les modèles de Herbrand de φ_{i_0}. D'autre part, ψ étant la borne supérieure des $(\varphi_i)_{1\leq i\leq n}$, la conjonction contenant χ est la formule φ_{i_0}.
On définit alors une formule ψ' de la manière suivante:

1. On choisit un littéral de φ_{i_0} pour lequel il existe un unificateur principal σ pour tous les $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$ et on l'intègre à ϕ, ϕ initialement vide.

2. Si il existe un autre littéral ϕ' de $(\varphi_i)_{1\leq i\leq n, i\neq i_0}$ tel qu'il soit unifiable pour les $(\varphi_i)_{1\leq i\leq n}$ par une substitution θ et que $\sigma\circ\theta$ ne contienne ni conflit ni cycle, alors $\phi \equiv \forall\ \vec{x}\ (\sigma\ \circ\ \theta)\ [\phi\ \wedge\ \phi'\ ]$.

On repète l'étape 2 jusqu'à ce qu'il n'existe plus de substitutions ou de littéraux permettant de l'appliquer.
On définit alors ψ' par:

$$\begin{array}{lll} \psi' \equiv & \forall\vec{x}\bigvee_{k=1,k\neq k_0}^{m} & \bigwedge_{j=1}^{p_k}{\psi'}_k^j \\ & \forall k\neq k_0, & {\psi'}_k \equiv \psi_k \\ & k=k_0 & {\psi'}_{k_0}\equiv\phi \end{array}$$

La conjonction ϕ ainsi obtenue étant déductible de $(\varphi_i)_{1\leq i\leq n}$ par la remarque 110 et par la règle d'introduction du $\wedge$ de la logique minimale, et déductible de ψ, $\mathbf{Mod}\{\psi'\}\subseteq\mathbf{Mod}\{\psi\}$. ψ' est donc plus petit que la borne supérieure, ce qui est absurde.
Ce qui termine la démonstration. ∎

Corollaire 113. *Soit ψ la borne supérieure au sens de la relation de subsomption des formules $(\varphi_i)_{1\leq i\leq n}$ de L_Σ. Alors, si parmi les littéraux des $(\varphi_i)_{1\leq i\leq n}$, aucun n'est unifiable pour une famille quelconque $(\varphi_i)_{1\leq i\leq k, k<n}$*

$$\psi\equiv\forall\vec{x}\bigvee_{i=1}^{n}\varphi_i$$

Preuve. D'après la première partie de la proposition précédente, toute formule ψ s'écrit $\psi'\vee\varphi_i$ pour tout i. Donc ψ s'écrit $\psi'\vee\forall\vec{x}\bigvee_{i=1}^{n}\varphi_i$. Or il est immédiat que $\forall\vec{x}\bigvee_{i=1}^{n}\varphi_i$ subsume φ_i pour tout i. ψ étant la borne supérieure nécessairement, ψ' est vrai. Ce qui établit le résultat annoncé. ∎

Pour la dernière proposition, on se restreindra à deux formules, φ_1 et φ_2.

Corollaire 114. *Soit ψ une borne supérieure au sens de la relation de subsomption de φ_1 et φ_2. Si parmi l'ensemble des littéraux de φ_1, l'un deux eux admet un unificateur principal avec un littéral de φ_2, ψ est une formule purement conjonctive et tous les littéraux présents dans ψ sont communs à φ_1 et φ_2.*

Preuve. D'après la deuxième partie de la proposition 112, ψ ne contient que des littéraux communs à φ_1 et φ_2, à une substitution σ près.
Le fait que ψ soit purement conjonctive résulte intuitivement de la même proposition. Raisonnons par l'absurde et supposons que ψ contienne une disjonction. Ecrivons ψ en forme normale disjonctive, i.e. $\forall\vec{x}\bigvee_{k=1}^{m}\bigwedge_{j=1}^{p_k}\psi_k^j$.

Considérons une conjonction de ψ, la k_0-ième. Alors, elle contient au moins un littéral unifiable de φ_1 et φ_2, que l'on notera $\sigma\varphi_{1,j_0}$. Si il existe un autre littéral de φ_1 et φ_2 tel qu'il soit unifiable par une substitution θ et que $\sigma\circ\theta$ ne contienne ni conflit ni cycle, alors $(\sigma\circ\theta)[\varphi_{1,j_0}\wedge\varphi_{1,j_1}]$ (où φ_{1,j_1} est ce second littéral) est nécessairement contenu dans cette conjonction. On peut ainsi continuer le raisonnement jusqu'à ce qu'il n'existe plus d'unificateur pour la sous-formule ainsi obtenue.

Considérons alors la formule ψ' constituée par la fermeture universelle de cette conjonction. Alors ψ' subsume φ_1 et ψ' subsume φ_2. ψ' est maximale au sens du pré-ordre partiel de subsomption, puisqu'on ne peut ajouter de nouveau littéral commun à φ_1 et φ_2 qui seuls composent ψ.
Par la règle d'introduction du $\vee$ de la logique minimale, il est immédiat que ψ' subsume ψ. Ce qui rend absurde l'affirmation ψ est disjonctive. ∎

C.4 Calcul de la solution

Soient $(\varphi_i)_{1\leq i\leq n}$ une famille de formules universelles de L_Σ. On va chercher à construire la formule ψ telle que ψ soit la borne inférieure de l'ensemble $(\varphi_i)_{1\leq i\leq n}$ pour la relation de subsomption. On considérera tout d'abord le cas des formules purement conjonctives, avant de nous intéresser aux autres formules universelles.
On définit l'ensemble des littéraux communs à toutes les formules $(\varphi_i)_{1\leq i\leq n}$ par

$$L_{(\varphi_i)_{1\leq i\leq n}} = \bigcap_{i=1}^{n} L\varphi_i$$

On considère également l'ensemble Θ_L des permutations de L.
Soit $E(\psi)$ une liste de formules. On désigne par la formule $E(\psi, k)$ la k i-ème formule de la liste.

C.4.1 Cas des formules universelles purement conjonctives

C.4.1.1 Cas de deux formules φ_1 et φ_2

Théorème 115. *La liste des formules ψ telles que ψ soit la borne supérieure au sens de la relation de subsomption des φ_1 et φ_2 peut être énumérée par l'algorithme suivant, appelé dans la suite $Calcul\_Subsomption_2(\varphi_1, \varphi_2)$.*

DEBUT
$E(\psi) = \emptyset$
$k = 1$
  SI $L = \emptyset$ **ALORS**
   $E(\psi, 1) = \forall \vec{x}[\varphi_1 \vee \varphi_2]$
  FINSI
  POUR *tout* $\vartheta_i \in \Theta_L$ **FAIRE**
   SI $\sigma \neq Id$ **ALORS**
    $k = k + 1$
   FINSI
   $\sigma = Id$
   POUR *tout* χ *de* L **FAIRE**
    SI *il existe un unificateur* θ *de* $\chi_{(\varphi_1,\varphi_2)}$ **ALORS**
     SI $\sigma \circ \theta$ *ne contient ni conflit ni cycle* **ALORS**
      $\sigma = \sigma \circ \theta$
      $E(\psi, k) = \forall \vec{x} [(\sigma \circ \theta) (\psi \wedge \chi_{\varphi_1})]$
      Pour ce faire, ré-élargissons le langage considéré à toutes les formules sur la signature Σ, noté rappelons le, L_Σ.
     FINSI
    FINSI
   FINPOUR
  FINPOUR
  SI $k = 1$ **ALORS**
   $E(\psi, 1) = \forall \vec{x}[\varphi_1 \vee \varphi_2]$
   RETOURNE$E(\psi)$
  FINSI
 FIN

Preuve (Maximalité des solutions obtenues). D'après la remarque 110, l'algorithme ci-dessus consiste à appliquer aux formules φ_1 et φ_2 m éliminations du $\forall$, puis autant d'éliminations du $\wedge$ que nécessaire pour obtenir l'ensemble A des atomes communs à toutes les formules φ, puis à appliquer ensuite des introductions du $\wedge$ et des introductions du $\forall$ pour obtenir une formule ψ. Par simple examen de l'algorithme, l'ensemble des formules ψ obtenues subsume donc φ_1 et φ_2 et est maximale par le corollaire 113. ■

Preuve (Complétude de l'algorithme). D'après les propositions 112 et 109, toutes les solutions sont des formules universelles purement conjonctives. La proposition 113 montre que les solutions sont maximales. ■

C.4.1.2 Cas d'un ensemble de n formules $n > 2$

En plus des notations précédentes, on définit l'ensemble $\Theta_{(\varphi_i)_{1\leq i\leq n}}$ des permutations d'une liste de formules $(\varphi_i)_{1\leq i\leq n}$. On désigne d'autre part par $E(\varphi_1, \varphi_2)$ la liste de formules ψ susbsumant deux formules φ_1 et φ_2, obtenue par application de l'algorithme présenté dans le cadre du théorème 115.

Théorème 116. *La liste des formules ψ telles que ψ soit la borne supérieure au sens de la relation de subsomption des $(\varphi_i)_{1\leq i\leq n}$ peut être énumérée par l'algorithme suivant, appelé dans la suite $Calcul\_Subsomption_N((\varphi_i)_{1\leq i\leq n})$.*

$k = 1$
$Calcul\_Subsomption_N((\phi_i)_{1\leq i\leq n})$
DEBUT
  SI $(\phi_i)_{1\leq i\leq n} = (v, v')$ **ALORS**
   $\psi \equiv Calcul\_Subsomption_2(v, v')$
  FINSI
  SI $L_{(\phi_i)_{1\leq i\leq n}} = \emptyset$ **ALORS**
   $\psi \equiv \forall\vec{x} \bigvee_{i=1}^{n} \phi_i$
  FINSI
  POUR *tout* $\vartheta_i \in \Theta_{(\phi_i)_{1\leq i\leq n}}$ **FAIRE**
   $j = 1$
   TANT QUE $j \leq n-1$ **FAIRE**
    $l = j+1$
    TANT QUE $l \leq n$ **FAIRE**
     $E(\phi_j, \phi_l) = Calcul\_Subsomption_2(\phi_j, \phi_l)$
      $E(\psi, k) = Calcul\_Subsomption_N((\phi_i)_{1\leq i\leq n, i\neq j, i\neq l} \cup E(\phi_j, \phi_l))$
      $k = k + 1$
      $l = l + 1$
     FIN TANT QUE
    $j = j + 1$
   FIN TANT QUE
  FINPOUR
FIN

Preuve (Maximalité des solutions obtenues). On raisonne par l'absurde.

i) Si il existe un littéral χ unifiable pour tous les $(\varphi_i)_{1\leq i\leq n}$, on sait d'après la proposition 112 que l'ensemble des $(\varphi_i)_{1\leq i\leq n}$ ne contient que des littéraux communs à ces formules. Par simple inspection de l'algorithme, les formules obtenues sont maximales.

ii) Sinon, soit ψ une formule obtenue par l'algorithme précédent. Supposons qu'elle ne soit pas maximale. Alors il existe une formule ψ' telle que ψ' subsume ψ et φ_i pour tout $i \in \{1, \ldots, n\}$. D'après cette dernière propriété, ψ' subsume donc tout couple de formules φ_j, φ_l pour $j, l \in \{1, \ldots, n\}$. Intuitivement, l'absurdité de l'existence de ψ' résulte de la maximalité des solutions du théorème 115. Pour le prouver, il suffit de:

   – Choisir un couple j, l correspondant au choix réalisé dans l'algorithme du théorème 116 pour calculer la formule ψ.
   – De réitèrer comme dans l'algorithme. ψ' subsume φ_j, φ_l et subsume $\forall i, 1 \leq i \leq n, i \neq j, i \neq l, \varphi_i$, donc ψ' subsume $\forall i, 1 \leq i \leq n, i \neq j, i \neq l, \varphi_i$ et la borne supérieure de φ_j, φ_l, maximale d'après le théorème 115.

   En réitérant ainsi ces deux étapes, ψ' doit subsumer (conditions de terminaison de l'algorithme), ou la borne supérieure de deux formules, ce qui est absurde d'après le théoreme 115, ou bien subsumer la borne supérieure de p formules n'ayant aucun littéral unifiable, ce qui est absurde d'après la proposition 113.

Donc les solutions énumérées par cet algorithme sont maximales. ■

Preuve (Complétude de l'algorithme). La complétude de l'algorithme est assurée par simple inspection d'après le théorème 115 et la proposition 113 dans les cas où il existe des littéraux unifiables pour tous les $(\varphi_i)_{1\leq i\leq n}$ ou dans le cas où il n'existe aucun littéral unifiable pour une famille quelconque $(\varphi_i)_{1\leq i\leq k}$ pour $k<n$.
Dans les autres cas, la complétude de l'algorithme résulte de la complétude de l'algorithme dans le cas $n=2$ et du fait que, comme montré dans la première partie de la preuve, toute formule ψ subsumant $(\varphi_i)_{1\leq i\leq n}$ doit subsumer tout couple de formules φ_j, φ_k, et que l'algorithme dans le cas de deux formules est complet. ■

C.4.2 Cas des formules universelles

C.4.2.1 Cas d'un ensemble de n formules $n \geq 2$

Le théorème 116 de la section précédente peut facilement être généralisé à n formules $(\varphi_i)_{1\leq i\leq n}$ universellement quantifiées que l'on écrira en forme normale disjonctive:

$$\forall i, 1 \leq i \leq n, \varphi_i = \forall \vec{x} \bigvee_{j=1}^{m} \bigwedge_{k=1}^{p_j} (\varphi_i)_j^k .$$

On utilisera la proposition 107 pour généraliser la solution de la section précédente.
On appelle Ω l'ensemble des subsitutions bijectives (renommages). On appelle $A(\{\psi\})$ un ensemble d'ensemble de formules. On reprendra les notations de la section précédente dans l'énoncé du théorème.

Théorème 117. *La liste des formules ψ telles que ψ soit la borne supérieure au sens de la relation de subsomption des $(\varphi_i)_{1\leq i\leq n}$ peut être énumérée par l'algorithme suivant:*

DEBUT
$E(\psi) = \emptyset$
$A(\{\psi\}) = \{\emptyset\}$
$k = 1$
  POUR *tout* $i \in \{1,\ldots,n\}$ **FAIRE**
   $A(\psi, i) = Calcul\_Subsomption_N((\bigwedge_{k=1}^{p_j}(\varphi_i)_j^k)_{1\leq j\leq m})$
  FINPOUR
  POUR *tout* $i \in \{1,\ldots,n\}$ **FAIRE**
   Chercher un élément ω *de* Ω *tel que* $\exists j, \exists k, \exists l, A(\psi,i,k) = \omega A(\psi,j,l)$
   $E(\psi,k) = A(\psi,j,l)$
   $k = k + 1$
  FINPOUR
 FIN

Preuve. D'après la proposition 107, la liste des ψ solutions doit subsumer chacune des propositions $\bigwedge_{j=1}^{p_i}\varphi_i^j$. Ce qui signifie que la liste des ψ solutions est constituée des formules communes aux renommages près des listes de propositions subsumant les $\bigwedge_{j=1}^{p_i}\varphi_i^j$, que calcule l'algorithme précédent. ■

Dorénavant,pour toute clause $C \equiv A:-A_1\ldots A_p$, on note, pour toute formule $C' \equiv A'_1 \ldots A'_q$, si C'' est la borne supérieure au sens du pré-ordre de subsomption de C et de C':

- σ la substitution résultat du calcul de C'' à partir de C et C'.
- $C_{C''}$ la partie subsumée de C.
- $C_{\sigma C-C''}$ la fermeture universelle des parties non subsumées de C.

L'algorithme précédent nous donne donc une représentation syntaxique de la borne supérieure au sens du pré-ordre de subsomption de deux formules, comme "unification partielle" de celles-ci.

D

Autre méta propriétés de la relation de déduction liées aux préférences

Cette annexe est destinée à donner les règles donnant les propriétés de la relation de dérivation par similarité avec des préférences en fonction des symboles logiques usuels. Il fait donc suite aux méta propriétés de la relation de déduction liées à la notion de préférence et présentées dans le chapitre 3:

$$\vee_{1,0,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1 \vee \varphi'_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi'_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\vee_{2,0,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1 \vee \phi'_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi'_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\vee_{0,0,1} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1 \vee \psi'_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi'_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\vee_{0,0,2} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2 \vee \phi'_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi'_2))) \vdash_S a}$$

$$\vee_{0,1,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1, \varphi'_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2 \vee \varphi'_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\vee_{0,2,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi'_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2 \vee \psi'_2, \phi_2))) \vdash_S a}$$

Intuitivement, les quatre premières règles signifient que plus l'ensemble des préconditions des préférences est restreint, plus l'ensemble des conclusions est large, étant bien sûr entendu que la disjonction de deux préconditions est moins restreinte que l'une d'entre elles. Les deux dernières règles signifient que si deux éléments sont par hypothèse négligeables dans des préférences par ailleurs identiques, alors les déductions que l'on peut faire sachant ces préférences sont des déductions que l'on pourra faire utilisant la disjonction de ces deux négligeables.

Voyons maintenant le cas de la règle $\wedge$:

$$\wedge_{1,0,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1', \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1 \wedge \varphi_1', \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\wedge_{2,0,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1') \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1 \wedge \phi_1') \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\wedge_{0,0,1} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1', \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1 \wedge \psi_1', \psi_2, \phi_2))) \vdash_S a}$$

$$\wedge_{0,0,2} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2'))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2 \wedge \phi_2'))) \vdash_S a}$$

$$\wedge_{0,1,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2 \wedge \varphi_2', \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi_1, \varphi_2', \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a}$$

$$\wedge_{0,2,0} \quad \frac{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2 \vee \psi_2', \phi_2))) \vdash_S a}{(\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2, \phi_2))) \vdash_S a \quad \text{et } (\Gamma, \Pi \cup ((\varphi_1, \varphi_2, \phi_1) \prec (\psi_1, \psi_2', \phi_2))) \vdash_S a}$$

Intuitivement, les quatre premières règles signifient que plus l'ensemble des préconditions des préférences est restreint, plus l'ensemble des conclusions est large, étant bien sûr entendu que la conjonction de deux préconditions est plus restreinte que l'une d'entre elles. Les deux dernières règles signifient que si deux éléments sont par hypothèse négligeables dans des préférences par ailleurs identiques, alors les déductions que l'on peut faire utilisant la conjonction de ces deux négligeables seront des déductions que l'on pourra faire utilisant chacun de ces deux négligeables.

Bibliographie

[Abdellaoui and Munier 1994] M. Abdellaoui and B. Munier. The closing in method: An experimental tool to investigate individual choice patterns under risk. *Models and Experiments on Risk and Rationality*, 1994.

[Adams 1975] E. Adams. *The Logic of Conditionals*. D. Reidel, Dordrecht, 1975.

[Allais 1953] M. Allais. Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine. *Econometrie*, 40:257–332, 1953.

[Baader and Hollunder 1991] F. Baader and B. Hollunder. Kris: Knowledge representation and inference system. *SIGART Bulletin*, 2(3):8–14, 1991.

[Battalio et al. 1990] R. Battalio, J. Kagel, and K. Jiranyakul. Testing between models of choice under uncertainty: some initial results. *Journal of Risk and Uncertainty*, 3:25–50, 1990.

[Becker and Sarin 1987] J. Becker and R. Sarin. Lottery dependent utility. *Management Science*, 33(11): 1367–1382, November 1987.

[Bell 1985] D. Bell. Disappointment in decision making under uncertainty. *Operations research*, 33(1): 1–27, January-February 1985.

[Bell et al. 1991] D. Bell, H. Raiffa, and A. Tversky. *Decision Making: descriptive, normative and prescriptive interactions*, chapter Savage revisited. Bell D.E., Raiffa H., Tversky A, 1991.

[Benferhat et al. 1997a] S. Benferhat, T. Chehire, and F. Monai. *Fuzzy information engineering*, chapter Possibilistic ATMS in a Data Fusion Problem, pages 417–434. John Wiley and Sons, INC, 1997a.

[Benferhat et al. 1994] S. Benferhat, D. Dubois, J. Lang, and H. Prade. *Between Mind and Computer: Fuzzy Science and Engineering*, chapter Hypothetical Reasoning in possibilistic logic: basic notions, applications and implementation issues, pages 1–29. World Scientific, 1994.

[Benferhat et al. 1992] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In B. Nebel, C. Rich, and W. Swartout, editors, *Principles of Knowledge Representation and Reasoning*, pages 673–684, 1992.

[Benferhat et al. 1997b] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92:259–176, 1997b.

[Bernoulli 1738] D. Bernoulli. Specimen theoriae novae de mensura sortis. In *Academiae Scientiarum Imperialis Petropolitanae*, volume 5, pages 175–192, 1738. Translated by L.Sommer, Econometrica, 22, (1954), 23-36.

[Besnard 1989] P. Besnard. *An introduction to default logic*. Springer Verlag, 1989.

[Boutilier 1990] C. Boutilier. Conditional logics of normality as modal systems. In *National Conference of the American Association for Artificial Intelligence*, pages 594–599. MIT Press, 1990.

[Boutilier 1994] C. Boutilier. Toward a logic for qualitative decision theory. In *International Conference on Knowledge Representation and Reasoning*, pages 75–86, 1994.

[Brachman et al. 1985] R. Brachman, V. Gilbert, and H. Levesque. An essential hybrid reasoning system: Knowledge and symbol level accounts in krypton. In *9th International Joint Conference in Artificial Intelligence*, pages 532–539, Los Angeles, CA, 1985.

[Brachman and Levesque 1984] R. Brachman and H. Levesque. The tractability of subsumption in frame-based description languages. In *4th National Conference of the American Association for Artificial Intelligence*, pages 34–37, Austin, Texas, 1984.

[Brachman and Schmolze 1985] R. Brachman and J. Schmolze. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.

[Brewka 1989] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proc. International Joint Conference in Artificial Intelligence*, pages 1043–1048, 1989.

[Brewka 1994] G. Brewka. Reasoning about priorities in default logic. In *Proc. American Association for Artificial Intelligence*, pages 940–945, 1994.

[Brewka 1996] G. Brewka. Well-founded semantics for extended logic programs with dynamic preferences. *Journal of Artificial Intelligence Research*, 4:19–36, 1996.

[Brewka et al. 1997] G. Brewka, J. Dix, and K. Konolige. *Nonmonotonic reasoning: an overview*. Number 73 in CSLI Lecture Notes. CSLI Publications, 1997.

[Brunet et al. 1992] Brunet, P. Georges, and F. Verger. *Dictionnaire de la géographie*. Presses Universitaires de France, 1992.

[Camerer 1989] C. Camerer. An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 8:167–196, 1989.

[Charreton and Bourdaire 1985] R. Charreton and J. Bourdaire. *La décision économique*. Number 2235. Que sais-je?, 1985.

[Chew et al. 1991] S. Chew, L. Epstein, and U. Segal. Mixture symmetry and quadratic utility. *Journal of Mathematical Psychology*, 59(1):139–163, 1991.

[Chew and Waller 1986] S. Chew and W. Waller. Empirical tests of weighted utility theory. *Journal of mathematical psychology*, 30:55–72, 1986.

[Clark 1978] K. Clark. *Logic and Data-Bases*, chapter Negation as Failure, pages 293–322. H. Gallaire and J. Minker, 1978.

[Congdon and Laird 1997] C. Congdon and J. Laird. *The Soar User's Manual, Version 7.0.4*. Electrical Engineering and Computer Science Department, University of Michigan, draft edition, May 1997.

[Coupey and Fouqueré 1997] P. Coupey and C. Fouqueré. Extending conceptual definitions with default knowledge. *Computational Intelligence Journal*, 13(2):258–299, 1997.

[Coupey et al. 1998] P. Coupey, C. Fouqueré, and S. Salotti. Formalizing partial matching and similarity in case-based reasoning with a description logic. *Journal of Applied Artificial Intelligence*, 12: 71–112, 1998.

[Coupey and Salotti 1997] P. Coupey and S. Salotti. Une logique de descriptions comme cadre formel d'un système de raisonnement à partir de cas. *Revue d'intelligence artificielle*, 11:127–177, 1997.

[De Finetti 1937] B. De Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré*, 7:1–68, 1937.

[DeKleer 1986] J. DeKleer. An assumption-based tms. *Artificial Intelligence*, 28:163–196, 1986.

[Dernis 1995] H. Dernis. Etude des comportements individuels face au risque: pour un partage en structures de risque. Mémoire de DEA, GRID: ENS-CACHAN, 1995.

[Desachy 1991] J. Desachy. *Interprétation automatique d'images satellite: le système ICARE*. Thèse de doctorat, Université Paul Sabatier, 1991.

[Dionne et al. 1993a] R. Dionne, E. Mays, and F. Oles. The equivalence of model theoretic and structural subsumption in description logics. In *13th International Joint Conference in Artificial Intelligence*, pages 710–716, Chambéry, France, 1993a.

[Dionne et al. 1993b] R. Dionne, R. Mays, and F. Oles. Concept algebras, objects and k-rep. In *13th International Joint Conference in Artificial Intelligence: workshop on object based representation system*, pages 61–66, Chambéry, France, 1993b.

[Dix 1992] J. Dix. Default theories of poole-type and a method for constructing cumulative versions of default logic. In B. Neumann, editor, *Proc. of 10th European Conference on Artificial Intelligence ECAI'92*, pages 289–293. John Wiley & Sons, 1992.

[Donimi et al. 1991] F. Donimi, M. Lenzerini, D. Nardi, and W. Nutt. The complexity of concept languages. In *Principles of Knowledge Representation and Reasoning*, pages 151–162, Cambridge, Mass., 1991. J.A.R. Fikes and E. Sandewall.

[Doyle 1985] J. Doyle. Reasoned assumptions and pareto optimality. In C. n. J. C. o. A. I. Menlo Park, editor, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 88–90, 1985.

[Draper et al. 1989] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riesman. The schema system. *International Journal of Computer Vision*, 2:209–250, January 1989.

[Dubois et al. 1998a] D. Dubois, L. Godo, H. Prade, and A. Zapico. Possibilistic representation of qualitative utility: an improved characterization. In . P. EDK, 10 Villa d'Orléans, editor, *IPMU'98: Internationl Conference on Information Processing and Management of Uncertainty*, pages 180–187, Juillet 1998a.

[Dubois et al. 1998b] D. Dubois, D. Le Berre, H. Prade, and R. Sabbadin. Logical representation and computation of optimal decisions in a qualitative setting. In *AAAI'98: Fifteenth National Conference on Artificial Intelligence*, pages 588–593, Juillet 1998b.

[Dubois and Prade 1989] D. Dubois and H. Prade. Order-of-magnitude reasoning with fuzzy relations. *Revue d'Intelligence Artificielle*, 3(4):1989, 1989.

[Dubois and Prade 1998] D. Dubois and H. Prade. An introductory survey of possibility theory and its recent developments. *Journal of Japan Society for Fuzzy Theory and Systems*, 10(1):21–42, February 1998.

[Dubois et al. 1997] D. Dubois, H. Prade, and R. Sabbadin. A possibilistic logic machinery for qualitative decision. In *AAAI'1997 Spring Symposium Series: Qualitative Preferences in Deliberation and Practical Reasoning*, pages 47–54, 1997.

[Etherington 1989] D. Etherington. A semantics for default logic. In *Proc. International Joint Conference in Artificial Intelligence*, pages 495–498, 1989.

[Fishburn 1970] P. Fishburn. *Utility theory for decision making*. New-York: Wiley, 1970.

[Fishburn 1982] P. Fishburn. Nontransitive measurable utility. *Journal of Mathematical Psychology*, 26: 31–67, 1982.

[Forgy 1982] C. Forgy. Rete: A fast algorithm for the many patterns/many object pattern match problem. *Artificial Intelligence*, 19:17–37, 1982.

[French 1988] S. French. *Decision theory: an introduction to the mathematics of rationality.* Ellis Horwood Series in Mathematics and its applications, 1988.

[Gallier 1986] J. Gallier. *Logic for Computer Science.* Harper and Row, 1986.

[Geffner and Pearl 1992] H. Geffner and J. Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, pages 209–244, 1992.

[Gelfond and Lifschitz 1988] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. *Logic Programming: Proc. of the Fifth International Conference and Symposium*, pages 1070–1080, 1988.

[Gilboa and Schmeidler 1995] I. Gilboa and D. Schmeidler. Case-based decision theory. *Quaterly Journal of Economics*, CX:605–639, 1995.

[Girard 1975] M. Girard. *Applications de la télédétection à la biosphère.* Masson col Sciences agronomiques, 1975.

[Goldszmidt and Pearl 1992] M. Goldszmidt and J. Pearl. Rank-based systems: a simple approach to to belief revision belief update, and reasoning about evidence and action. In M. Kaufmann, editor, *Proc. of the 3rd International Conference on Knowledge Representation and Reasoning*, pages 661–672, San Mateo, CA, October 1992.

[Gul 1991] F. Gul. A theory of disappointment aversion. *Econometrica*, 59:667–686, 1991.

[Hadj-Ali et al. 1998] A. Hadj-Ali, D. Dubois, and H. Prade. Raisonnement sur les ordres de grandeurs relatifs avec des relations floues, appliqué aux probabilités qualitatives. In *LFA'98: Rencontres Francophones sur la logiques floue et ses applications*, Rennes, Novembre 1998 1998.

[Hanson and Riesman 1987] A. Hanson and E. Riesman. *Advances in Computer Vision*, chapter The VISIONS Image-understanding system, pages 1–114. Erlbaum Assoc, 1987.

[Harless and Camerer 1994] D. Harless and C. Camerer. The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289, November 1994.

[Heinsohn 1991] J. Heinsohn. A hybrid approach for modeling uncertainty in terminological logics. *Lecture Notes in Computer Science*, 548:198–205, 1991.

[Hey and Orme 1994] J. Hey and C. Orme. Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326, November 1994.

[Hogarth 1988] R. Hogarth. *Judgement and choice : the psychology of decision.* John Wiley and Sons, Second Edition, 1988.

[Jaffray 1985] J. Jaffray. *Cours de DEA.* Université Pierre et Marie Curie, 1985.

[Jaffray 1988] J. Jaffray. Choice under risk and the security factor: an axiomatic model. *Theory and Decision*, 24:169–200, 1988.

[Kahneman and Tversky 1979] D. Kahneman and A. Tversky. Prospect theory : an analysis of decisions under risk. *Econometrica*, 47:263–291, 1979.

[Klein 1993] G. Klein. *A Recognition-Primed Decision (RPD) Model of Rapid Decision Making*, pages 138–147. Ablex Publishing Corporation, 1993.

[Klein et al. 1993] G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok. *Decision Making in Action: Models and Methods.* Ablex Publishing Corporation, 1993.

[Koehler 1994] J. Koehler. An application of terminological logics to case-based reasoning. In *Proc. of 5th International Conference on Principles of Knowledge representation and reasoning*, pages 351–362, 1994.

[Koehler 1996] J. Koehler. Planning from second principles. *Artificial Intelligence*, 87:145–186, 1996.

[Kolodner 1993] J. Kolodner. *Case-based reasoning*. Morgan Kaufmann Publishers, San Mateo CA, 1993.

[Konolige 1988] K. Konolige. On the relation between default and autoepistemic reasoning. *Artificial Intelligence*, 35:343–382, 1988.

[Kowalski and van Emden 1974] R. Kowalski and M. van Emden. The semantics of predicate logic as a programming language. *Journal of the ACM*, 23(4):733–742, 1974.

[Krauss et al. 1990] S. Krauss, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Laird et al. 1986] J. Laird, P. Rosenbloom, and A. Newell. *Universal subgoaling and chunking: the automatic generation of goal hierarchies*. Kluwer Academic Publishers, 1986.

[Lalement 1990] R. Lalement. *Logique, Réduction, Résolution*. Masson, 1990.

[Lamarre 1991] P. Lamarre. S4 as the conditionnal logic of nonmonotonicity. In *Proceedings of the International Conference on Knowledge Representation and Reasoning*, pages 357–367. Morgan Kaufmann, 1991.

[Langlotz and Shortliffe 1989] C. Langlotz and E. Shortliffe. Logical and decision-theoretic methods for planning under uncertainty. *AI Magazine*, 10(1):39–47, 1989.

[Le Ber 1993] F. Le Ber. *Modélisation des connaissances et raisonnements pour l'analyse des paysages agraires à partir de données satellitaires*. Thèse de doctorat, Univeersité de Nancy 1, 1993.

[Lehmann 1996] D. Lehmann. Generalized qualitative probability: Savage revisited. In *Proc. Uncertainty in Artificial Intelligence*, pages 381–388, 1996.

[Leland 1994] J. Leland. Generalized similarity judgments: an alternative explanation for choices anomalies. *Journal of Risk and Uncertainty*, 9:151–172, 1994.

[Levesque 1990] H. Levesque. All i know: A study in autoepistemic logic. *Artificial Intelligence*, 42: 263–309, 1990.

[Lieber and Napoli 1997] J. Lieber and A. Napoli. Planification à partir de cas et classification. In *Proc. JICAA'97*, pages 617–628, 1997.

[Lieber and Napoli 1998] J. Lieber and A. Napoli. Correct and complete retrieval for case-based problem-solving. In *13th European Conference on Artificial Intelligence ECAI'98*, pages 68–72, 1998.

[Lifschitz 1985] V. Lifschitz. Computing circumscription. In *International Joint Conference on Artificial Intelligence*, pages 121–127, 1985.

[Lifschitz 1996] V. Lifschitz. *Principles of Knowledge representation*, chapter Foundations of declarative logic programming. CSLI, 1996.

[Lin and Shoham 1992] F. Lin and Y. Shoham. A logic of knowledge and justified assumptions. *Artificial Intelligence*, 57:271–289, 1992.

[Lipshitz 1993] R. Lipshitz. *Converging Themes in the Study of Decision Making in Realistic Settings*, pages 103–137. Ablex Publishing Corporation, 1993.

[Loomes and Sugden 1982] G. Loomes and R. Sugden. Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal*, 92:805–824, 1982.

[Luce and Raiffa 1957] R. Luce and H. Raiffa. *Games and decisions: introduction and critical survey.* John Wiley and Sons, 1957.

[MacGregor and Bates 1987] R. MacGregor and R. Bates. *The LOOM Knowledge representation language Technical Report RS-87-188.* Information Science Institute, University of Southern California, Maria del Ray, CA, 1987.

[Machina 1982] M. Machina. Expected utility analysis without the independance axiom. *Econometrica*, 50(2):277–323, 1982.

[Makinson 1989] D. Makinson. *General theory of cumlative inference*, chapter Non monotonic Reasoning, LNAI 346, pages 1–18. Reinfrank, de Kleer, Ginsberg and Sandewall, 1989.

[Marek 1991] W. Marek. Computing intersection of autoepistemic expansions. In M. Press, editor, *Logic Programming and Non-Monotonic Programming, Proceedings of the First International Workshop*, pages 276–288, Cambridge Mass, 1991.

[Matsuyama 1990] T. Matsuyama. *SIGMA, a Knowledge-based Aerial Image Understanding System.* Advances in Computer Vision and Machine Intelligence. Plenum, New York, 1990.

[McCarthy 1980] J. McCarthy. Circumscription: A form of nonmonotonic reasoning. *Artificial Intelligence*, 13(2):27–39, 1980.

[McCarthy 1986] J. McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[McDermott and Doyle 1980] D. McDermott and J. Doyle. Non-monotonic logic i. *Artificial Intelligence*, 13:41–72, 1980.

[McKeown et al. 1985] D. McKeown, W. Harvey, and J. McDermott. Rule-based interpretation of aerial imagery. *PAMI*, 7(5):570–585, September 1985.

[McKeown et al. 1989] D. McKeown, J. Wilson, W. Harvey, and L. Wixson. Automating knowledge acquisition for aerial image interpretation. *Computer Vision Graphics and Image Processing*, 46: 37–81, 1989.

[Milnor 1954] J. Milnor. Games against nature in decision processes. Wiley, 1954.

[Minksy 1975] M. Minksy. *The psychology of computer vision*, chapter A framework for representing knoweldge, pages 211–277. P.H. Winston, 1975.

[Monai and Chehire 1992] F. Monai and T. Chehire. Possibilistic assumption based truth maintenance system, validation in a data fusion application. In D. Dubois, M. Wellman, B. D'Ambrosio, and P. Smets, editors, *Proc. of the Conference on Uncertainty in Artificial Intelligence*, pages 83–91, 1992.

[Montgomery 1983] H. Montgomery. *Analysing and aiding Decision Processes*, chapter Decision rules and the search for a dominance structure: towards a process model of decision making, pages 343–369. 1983.

[Moore 1985] R. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25: 75–94, 1985.

[Munier 1988] B. Munier. A guide to decision making under uncertainty. In B. Munier, editor, *Risk, Decision, Rationalité*, 1988.

[Munier 1995] B. Munier. Entre rationalités instrumentales et cognitive: contributions de la dernière décennie à la modélisation du risque. *Revue d'économie politique*, 105(1):6–70, 1995.

[Nagao and Matsuyama 1980] M. Nagao and T. Matsuyama. *A structural analysis of complex aerial photographs.* Plenum, New York, 1980.

[Nagao et al. 1979] M. Nagao, T. Matsuyama, and H. Mori. Structured analysis of complex aerial photographs. In *Proceedings International Joint Conference of Artificial Intelligence*, pages 610–616, Tokyo, 1979.

[Napoli 1992] A. Napoli. Subsumption and classification-based reasoning in object-based representation. In *10 th European Conference on Artificial Intelligence*, pages 425–429, Vienna, Austria, 1992.

[Napoli et al. 1996] A. Napoli, J. Lieber, and R. Curien. Classification based problem solving in case-based reasoning. In *Advances in case-based reasoning - Third European Workshop EWCBR'96*, number 1168 in LNAI, pages 295–308, 1996.

[Nebel 1990] B. Nebel. Reasoning and revision in hybrid representation systems. *Lecture Notes in Computer Science*, 422, 1990.

[Neumann and Morgenstern 1947] J. V. Neumann and O. Morgenstern. *Theory of games and economic behavior.* Princeton University Press, 1947.

[Newell 1990] A. Newell. *Unified theories of cognition.* Harvard University Pres, 1990.

[Newell and Simon 1972] A. Newell and H. Simon. *Human problem solving.* Prentice-Hall: Englewood Cliffs, 1972.

[Nute 1984] D. Nute. *Topics in Conditionnal Logic.* Reidel, Dordrecht, 1984.

[Padgham and Zhang 1993] L. Padgham and T. Zhang. A terminological logic with defaults: A definition and an application. In *13th International Joint Conference in Artificial Intelligence*, pages 663–668, Chambéry, France, 1993.

[Patel-Schneider 1984] P. Patel-Schneider. Small can be beautiful in knowledge representation. In *IEEE Workshop on principles of knowledge-based systems*, pages 11–16, 1984.

[Pearl 1993] J. Pearl. From conditional oughts to qualitative decision theory. In *Proc. of the Conference on Uncertainty in Artificial Intelligence*, pages 12–20, 1993.

[Plotkin 1970] G. Plotkin. A note on inductive generalisation. *Machine Intelligence*, 5:153–163, 1970.

[Plotkin 1971] G. Plotkin. A further note on inductive generalisation. *Machine Intelligence*, 6:101–124, 1971.

[Pomerol 1997] J. Pomerol. Artificial intelligence and human decision making. *European Journal of Operational Research*, 99:3–25, 1997.

[Poole 1988] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[Quantz and Royer 1992] J. Quantz and V. Royer. A preference semantics for default in terminologic logics. In C. R. B. Bebel and W. Swartout, editors, *Principles of Knowledge Representation and Reasoning: 3rd International Conference*, pages 294–305, Cambridge MA, 1992.

[Quiggin 1982] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:323–343, 1982.

[Reiter 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Reiter 1982] R. Reiter. Circumscription implies predicat completionv (sometimes). In *American Association form Artificial Intelligence*, pages 418–420, 1982.

[Rousseau 1998] L. Rousseau. Raisonnement par similarité sur des espaces de problèmes: application à la modélisation du raisonnement du photo-interprète. Mémoire de DEA, DEA IARFA Université Paris 6, 1998.

[Rubinstein 1988] A. Rubinstein. Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of Economic Theory*, 46:145–153, 1988.

[Sandakly 1995] F. Sandakly. *Contribution à la mise en oeuvre d'une architecture à base de connaissances pour l'interprétation de scènes 2D et 3D.* Thèse de doctorat, Université de Sophia-Antipolis, 1995.

[Savage 1954] L. Savage. *The foundations of statistics.* Dover Publications Incorporated: New York, 1954.

[Schank 1982] R. Schank. *Dynamic memory: a theory of reminding and learning in computers and people.* Cambridge University Press, New York, 1982.

[Segal 1989] U. Segal. Anticipated utility: a measure representation approach. *Theory and decision*, 19: 359–373, 1989.

[Shafer 1991] G. Shafer. *Decision Making: descriptive, normative and prescriptive interactions*, chapter Savage revisited. Bell D.E., Raiffa H., Tversky A, 1991.

[Shoham 1988] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, 1988.

[Slade 1989] S. Slade. Case-based reasoning: A research paradigm. *The AI Magazine*, 12(1):43–55, 1989.

[Sugden 1993] R. Sugden. An axiomatic foundation for regret theory. *Journal of Economic Theory*, 60: 159–180, 1993.

[Tversky 1969] A. Tversky. Intransitivity of preferences. *Psychological Review*, 76:31–48, 1969.

[Tversky and Kahneman 1981] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[Tversky and Kahneman 1982] A. Tversky and D. Kahneman. *Judgment under uncertainty : heuristics and biases.* Cambridge University Press, 1982.

[Tversky and Kahneman 1992] A. Tversky and D. Kahneman. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992.

[Viscusi 1989] K. Viscusi. Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, 2:235–264, 1989.

[Voorbraak 1993] F. Voorbraak. Preference-based semantics for nonmonotonic logics. In *Proc. International Joint Conference in Artificial Intelligence*, pages 584–589, 1993.

[Woods 1975] W. Woods. *Representation and understanding: studies in cognitive science*, chapter What's a link? Foundations for semantic networks, pages 35–82. Academic Press, New York, 1975.

[Yaari 1987] M. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.