Synthèse des travaux

Andreas Herzig

juin 1998

Table des matières

1 Introduction

1.1 Thème de recherche

Mon travail porte sur la *formalisation du raisonnement en logique*, allant de la représentation des connaissances jusqu'aux méthodes de démonstration automatique. Dans cette perspective, j'étudie *l'évolution des bases de connaissances*, qui est un problème essentiel en intelligence artificielle et plus généralement en informatique. Ce thème comprend aussi bien la mise-à-jour des bases de données que la planification en robotique et l'évolution des croyances dans le dialogue homme-machine. Ces applications étant d'un intérêt primordial, ma démarche consiste à les modéliser formellement dans un premier temps, et dans un second temps à leur associer des procédures de démonstration automatique.

Dans ces applications, apparaissent de façon naturelle les concepts de croyance, de connaissance, d'intention et d'obligation. ainsi que des problèmes d'incertitude. Dans un langage logique, à chacun de ces concepts peut être associé un opérateur intensionnel (opérateur modal). On obtient alors une famille d'opérateurs modaux paramétrés par des concepts. Ainsi, la formule modale

$$[p]A$$

peut être lue « l'agent p sait que A », « l'agent p croit que A », « pour l'agent p il est obligatoire que A », etc.

Mon but général est de placer dans un cadre unifié les croyances et leur dynamique en définissant des formalismes de type logique non-classique suffisamment expressifs pour pouvoir les représentater, et d'y associer des méthodes de raisonnement automatique. La démarche classique consiste d'abord à étudier séparément la logique non-classique associée à chaque concept, puis à intégrer plusieurs de ces concepts en une seule logique.

L'étude d'opérateurs de connaissance, croyance et obligation dans des logiques modales faisant partie des travaux classiques du domaine (voir par exemple les travaux de J. Hintikka sur les logiques epistémiques), les opérateurs modaux peuvent également rendre compte de la dynamique des croyances. Ainsi ils peuvent être associés à des actions : la formule $[p]A$ peut être lue « après exécution de l'action p, A sera vrai ». De telles actions peuvent être itérées, et un plan correspond alors à une suite d'actions $p_1, \ldots, p_n$. Des auteurs en logique philosophique comme Segerberg ont rapproché l'exécution d'une action à celle d'un programme, permettant ainsi d'utiliser la logique dynamique (qui est une logique modale particulière où le paramètre p est vue comme un programme) comme une logique d'actions et de plans.

Au lieu de considérer que les actions sont abstraites on peut aussi les étudier à travers leurs résultats. Une action est donc confondue avec les formules rendues vraies par son exécution. Ainsi le paramètre p peut être analysé comme $faire(B)$, où B est une formule. La formule modale $[faire(B)]A$ est lue « après avoir rendu B vrai, A sera vrai », ou « après mise-à-jour par B, A sera vrai », ou encore « sous l'hypothèse B, A est vrai ». La dernière formule a la forme d'un conditionnel, et si l'on utilise une flèche (habituellement associée aux conditionnels), la formule d'en haut s'écrit

$$A \Rightarrow B$$

où $\Rightarrow$ (aussi noté $>$) est un symbole d'implication différent de l'implication matérielle de la logique classique $\rightarrow$. De telles logiques des conditionnels ont été étudiées en philosophie formelle.

Les logiques des conditionnels – de par leur lien avec le raisonnement hypothétique – permettent de traiter naturellement l'aspect dynamique des bases de connaissances : si $C = m\grave{a}j(B, A)$ est le résultat de la mise-à-jour de la base de croyances B par une nouvelle information A, alors nous avons

$$m\grave{a}j(B, A) \rightarrow C \quad \text{si et seulement si} \quad B \rightarrow (A \Rightarrow C)$$

Ainsi, l'affirmation d'un conditionnel $A \Rightarrow C$ signifie que C résultera d'une mise-à-jour (hypothétique) par A. Cette équivalence est connue sous le nom « Test de Ramsey » dans la littérature. À travers ce test, les logiques des conditionnels s'appliquent à l'analyse et à la formalisation des opérations de mise-à-jour.

Les formalismes logiques que j'ai mentionnés, et plus généralement tous les formalismes qui ont été proposés dans la littérature pour prendre en compte l'évolution des bases de connaissances (comme les formalismes de raisonnement non-monotone, de mise-à-jour ou de raisonnement sur les actions) se heurtent à plusieurs difficultés, dont le fameux problème de la persistance des informations lors d'un changement (problème du décor, « frame problem »), qui correspond au problème de l'héritage dans le raisonnement non-monotone. Les approches pour résoudre ces problèmes, qui est d'abord un problème de représentation des connaissances, ont rencontré des difficultés autant sur le plan sémantique (par exemple l'absence de sémantique des approches basées sur la syntaxe) que sur le plan syntaxique (l'absence de résultats de d'axiomatisabilité ou de décidabilité).

J'ai fondé mes travaux sur l'hypothèse que ces difficultés viennent du fait que traditionnellement en logique, on suppose que l'ensemble des atomes d'un langage logique n'est muni d'aucune structure : en logique, tout ce qui n'est pas déclaré dépendant est supposé indépendant (tandis que le cas contraire prévaut en théorie des probabilités). Comme beaucoup d'auteurs dans la littérature l'ont constaté, l'implication matérielle $P \rightarrow Q$ de la logique classique n'est pas suffisante pour représenter ce type de dépendance.

Dans mes travaux j'ai introduit une structure de dépendance à la logique, et je me suis proposé

1. d'étudier en détail ce concept de dépendance dans des cadres appropriés (théorie des probabilités, théorie des possibilités, théories de la causalité),

2. d'intégrer les structures de dépendance à la logique,

3. d'étudier la notion de déduction pour les logiques non-classiques en général et les logiques basées sur la dépendance en particulier,

4. de définir des procédures générales de démonstration automatique,

5. de mettre en œuvre ces procédures dans les applications comme la mise-à-jour des bases de croyances, la génération de plans d'actions ou la reconstruction des croyances dans le dialogue homme-machine.

1.2 Historique des travaux

Les deux côtés de mes travaux, à savoir la modélisation de la dynamique des croyances et la démonstration automatique ont été développés parfois en alternance et parfois en parallèle.

Tandis que mon DEA et les débuts de ma thèse portaient plutôt sur l'aspect évolutif, ma thèse et mes débuts de chercheur au CNRS ont été surtout consacrés à la définition d'une méthode très générale de démonstration automatique pour les logiques non-classiques. Ce sont mes travaux récents qui ont abouti à une synthèse de ces deux activités, en s'appuyant sur le concept-clef de la dépendance : depuis 1994 j'ai consacré presque toutes mes publications à des méthodes de démonstration automatique pour des systèmes logiques de type non-classique basés sur la dépendance et dédiés à la modélisation de l'évolutif. Enfin, dans des travaux très récents j'ai appliqué cette méthodologie aux dialogues homme-machine coopératifs.

Les publications que j'ai faites s'inscrivent dans le contexte du groupe Logique appliquée de l'IRIT dirigé par Luis Fariñas del Cerro, et beaucoup d'entre elles sont cosignées par lui et d'autres membres de cette équipe.

Dès mes débuts dans la recherche je me suis intéressé au rapport entre révision, opérations de mise-à-jour des bases de données et conditionnels. Dans mon rapport de DEA de 1986 j'ai étudié une logique des conditionnels dans laquelle la sémantique du connecteur $\Rightarrow$ est celle de la logique intuitionniste. Les résultats ont été repris et publiés en 1996 [Fariñas del Cerro et Herzig, 1996b].

Au début de ma thèse j'ai également esquissé en 1988 une sémantique pour les mises-à-jour dans le cadre d'une logique des conditionnels très proche du « Possible Models Approach » (PMA) de Winslett (qui date également de 1988). J'ai continué le développement de cette approche à partir de 1992. L'intérêt de la communauté scientifique pour le thème de la mise-à-jour était alors (en 1993) à l'origine du projet inter-PRC « Gestion de l'évolutif et de l'incertain dans une base de connaissances », qui regroupait des chercheurs venant de l'intelligence artificielle, des bases de données et de l'automatique et dont j'étais co-responsable (avec Ch. Froidevaux). Les résultats de ces travaux sont synthétisés dans [Collectif, 1995].

À cette époque il m'a semblé important de faire le point sur l'état des recherches en logiques des conditionnels, et j'ai entamé en 1993 avec G. Crocco et L. Fariñas del Cerro l'édition du livre « Conditionals: From Philosophy to Computer Science » paru chez Oxford University Press [Crocco *et al.*, 1995].

Aussi bien intuitivement que formellement, il y a un lien fort entre un conditionnel $A \Rightarrow C$ et une inférence non-monotone $A \mid\!\sim C$. J'ai étudié les liens entre les logiques des conditionnels et les propriétés générales des relations d'inférence non-monotone en collaboration avec L. Fariñas del Cerro ainsi qu'avec J. Lang de l'équipe « Raisonnement plausible, décision et méthodes de preuve ». En étendant un résultat de Gärdenfors et Makinson, nous avons montré que c'est la même théorie d'incertitude qui est à la base des deux formalismes, à savoir la théorie des possibilités. Les résultats ont été publiés d'abord à l'ECAI [Fariñas del Cerro *et al.*, 1992] et ensuite dans la revue « Artificial Intelligence » [Fariñas del Cerro *et al.*, 1994]. Ce dernier article constitue la partie I de ce document.

Dans la suite j'ai proposé dans [Fariñas del Cerro et Herzig, 1994b, Fariñas del Cerro et Herzig, 1995a] un cadre général pour la mise-à-jour basé sur une combinaison des conditionnels avec une relation de dépendance ; dans celui-ci, à l'axiomatique des logiques des conditionnels on rajoute un « axiome de décor » générique paramétrisé par la dépendance. Cet axiome prend la forme

$$C \rightarrow (A \Rightarrow C) \quad \text{si } A \text{ et } C \text{ sont indépendants}$$

et stipule que si C est vrai et n'a « rien à voir » avec A, alors C continue à être vrai après la mise-à-jour par A.

Dans ce même cadre, j'ai associé à la PMA une axiomatique [Fariñas del Cerro et Herzig, 1994a, Fariñas del Cerro et Herzig, 1995d, Fariñas del Cerro et Herzig, 1997] et une méthode de démonstration automatique [Herzig, 1996]. L'approche PMA ayant été critiquée pour certaines de ses propriétés, dans des travaux récents j'ai passé en revue les corrections de la PMA publiées dans la littérature, et j'ai proposé une variante basée sur la dépendance qui a une complexité théorique particulièrement intéressante (dans la classe NP) et qui est munie d'une procédure de démonstration automatique par réécriture [Herzig et Rifi, 1998].

La mise-à-jour et son rapport avec les conditionnels est le thème d'un article de référence qui doit paraître dans le « Handbook of Defeasible Reasoning and Uncertainty Management » [Herzig, to appear]. Cet article constitue la partie II de ce document. Dans cet article, je montre en particulier comment, en partant du Test de Ramsey, des résultats existants sur les conditionnels peuvent être transférés au cadre formel général pour les opérations de mise-à-jour proposées par Katsuno-Mendelzon, Satoh et Grahne, et comment ce cadre peut être généralisé en s'inspirant des conditionnels. Ce résultat m'a semblé par ailleurs avoir des conséquences philosophiques intéressantes en ce qui concerne la distinction entre mise-à-jour et révision des croyances habituellement faite dans la littérature, et j'ai rédigé un article sur ce sujet en collaboration avec G. Crocco du département de philosophie de l'Université de Constance, Allemagne [Crocco et Herzig, 1997].

Mon approche étant basée sur le concept de dépendance, depuis 1993 je me suis posé la question des fondements et des propriétés de cette notion. Ces réflexions ont aussi été menées dans le cadre d'un groupe de travail du PRC Intelligence artificielle « Lois, structures et dépendances » (LSD) dont j'étais le responsable et dont les discussions ont été publiées dans [Collectif, 1997].

Au moins trois possibilités existent en ce qui concernent les fondements. La première et la plus simple est de supposer que la dépendance est un concept primitif, à partir duquel par exemple la causalité peut ensuite être définie. Cette option avait été prise dans [Fariñas del Cerro et Herzig, 1994b, Fariñas del Cerro et Herzig, 1995a], et c'est aussi le choix (en général implicite) dans la plupart des approches causales de la littérature récente en raisonnement sur les actions. Un inconvénient de ces approches est que leur notion de dépendance, de par son caractère *ad hoc*, ne nous renseigne que peu sur les propriétés de la dépendance, d'autant plus que nos intuitions premières à ce sujet semblent être peu claires. Par exemple des propriétés comme le « critère conjonctif pour la dépendance » CCD

$$\text{si } C \text{ dépend de } A \text{ et } C \text{ dépend de } B \text{ alors } C \text{ dépend de } A \wedge B$$

postulées par J.M. Keynes en 1921 et étudiées par Gärdenfors doivent-elles être acceptées ? En effet, toutes ces approches font l'impasse sur ces questions pourtant fondamentales, en se restreignant à des dépendances entre formules atomiques.

Ce constat m'a amené à étudier des notion de dépendance basées dans des théories de l'incertitude. En effet, en théorie des probabilités il existe une notion bien connue d'indépendance définie à travers la loi de multiplication :

$$A \text{ et } B \text{ sont indépendants si et seulement si } P(A \wedge B) = P(A) \cdot P(B)$$

et il existe des résultats formels (en particulier une axiomatique) dans la littérature. Cependant, ces formalisations sont assez complexes et délicates à manipuler. De plus, les mesures de probabilité constituant un cadre numérique, il s'est avéré difficile de transférer la notion de dépendance associée dans un cadre logique. Je me suis intéressé alors à la théorie des possibilités, qui est une théorie ordinale de l'incertitude (Gärdenfors et Makinson [Gärdenfors et Makinson, 1994] utilisent l'expression «a 'lazy man's approximation' to a probability districution») et dont les « pères fondateurs » sont D. Dubois et H. Prade de l'équipe « Raisonnement plausible, décision et méthodes de preuve » de l'IRIT. Cette théorie peut être identifiée à la théorie de l'enracinement épistémique de Gärdenfors, qui elle est la base de deux formalismes importants en dynamique des croyances, à savoir les révisions des croyances dites AGM et les inférences non-monotones dites KLM (j'avais étudié ces derniers liens dans [Fariñas del Cerro *et al.*, 1994]). En collaboration avec L. Fariñas del Cerro ainsi qu'avec D. Dubois et H. Prade j'ai donc entrepris une étude de fond de cette option, ce qui nous a permis d'obtenir des axiomatiques simples et des résulats de complétude. Nous nous sommes d'abord basé directement sur les opérations de révision de croyances AGM, en définissant A et C comme dépendants par rapport à une base de croyances B si C est conséquence de B mais non de $B - A$, où $B - A$ est la contraction de B par A [Fariñas del Cerro et Herzig, 1994c, Fariñas del Cerro et Herzig, 1995c, Fariñas del Cerro et Herzig, 1996a]. Dans la suite nous avons obtenu des résultats similaires en abordant directement la notion de dépendance possibiliste à travers la conditionalisation, c.-à.-d. en étudiant le rapport entre le dégré de possibilité $\Pi(C)$ d'une formule C et le dégré de possibilité conditionnel $\Pi(C|A)$ [Dubois *et al.*, 1994, Dubois *et al.*, 1997, Dubois *et al.*, 1998]. La partie IV de ce document contient [Fariñas del Cerro et Herzig, 1996a] et [Dubois *et al.*, 1998].

Une troisième voie pour fonder la dépendance passe par la notion de thème d'intérêt. Ici on considère qu'à chaque énoncé du langage on peut associer un ensemble de thèmes, en s'inspirant de travaux en linguistique informatique et en philosophie formelle. Par exemple si A signifie « l'interrupteur est en position haute » alors A a comme thème la position de l'interrupteur, mais aussi l'éclairage de la pièce. Deux formules sont alors indépendantes si elles n'ont aucun thème en commun. Une telle notion est *a priori* seulement syntaxique. Malgré cet inconvénient, elle a des propriétés très intéressantes au niveau de la représentation, et par exemple le critère conjonctif pour la dépendance CCD se trouve naturellement vérifié. C'est à cause de cette simplicité que la dépendance thématique m'a parue la plus appropriée pour être à la base d'une théorie de la dynamique des croyances à la fois suffisamment riche d'un point de vue théorique et suffisamment simple d'un point de vue pratique, c'est-à-dire ayant des propriétés de calcul « raisonnables », et je l'ai retenue pour la suite.

En effet, la variante de la PMA que j'ai proposée récemment dans [Herzig et Rifi, 1998] est basée sur les thèmes (tandis que les formalisations de la PMA que j'ai étudiées précédemment étaient fondées sur une dépendance thématique particulière). Il en est de même en ce qui concerne une approche pour la révision des bases de croyance qui étudie sa compatibilité avec l'approche AGM [Herzig, 1997].

J'ai également retenu cette approche à partir de 1995 dans mes recherches en raisonnement sur les actions. Ici, des thèmes sont associés non seulement aux formules, mais aussi aux actions. Par exemple à l'action de basculer un interrupteur peuvent être associés le thème de la position de l'interrupteur ainsi que l'éclairage de la pièce. Une action et une formule sont alors indépendantes si elles n'ont aucun thème en commun, et nous pouvons alors formuler un « axiome de décor » générique similaire à celui pour les conditionnels :

$$C \to [p]C \quad \text{si } p \text{ et } C \text{ sont indépendants}$$

En collaboration avec M. A. Castilho et O. Gasquet, j'ai rajouté à la logique dynamique une notion de dépendance entre actions et formules pour résoudre le problème du décor. Afin de mécaniser cette logique j'ai adapté les méthodes des tableaux sémantiques standards des logiques modales, en les paramétrant par des relations de dépendance. Cette méthode permet également de générer des plans [Castilho *et al.*, 1997b]. Un article qui étend ce travail a été soumis au « Journal of Logic and Computation » (après un «accepted subject to modifications », la version corrigée vient d'être envoyée). Il constitue la partie III de ce document.

Il s'est avéré que la méthode des tableaux pour les logiques des actions constitue aussi une généralisation très intéressante des méthodes connues. Ce constat nous a permis d'aller bien au-delà de ces méthodes et d'intégrer des logiques modales pour lesquelles aucune méthode de démonstration automatique n'était connue à ce jour, comme celles associées à la propriété de densité ou de confluence. Ces travaux ont été menés en collaboration avec M. A. Castilho, L. Fariñas del Cerro et O. Gasquet et ont été publiés dans [Castilho *et al.*, 1997a], qui apparaît comme la partie V de ce document. Ils sont à l'origine d'un projet ambitieux de mon équipe qui a démarré en 1997 et dans lequel un démonstrateur automatique accessible par Internet vient d'être implanté.

Ces tableaux ont l'avantage d'être décidables, et constituent ainsi une alternative intéressante à la méthode de traduction en logique classique du premier ordre proposée dans ma thèse que j'ai continué à étudier par ailleurs dans une série d'articles [Gasquet et Herzig, 1993, Gasquet et Herzig, 1994b, Balbiani et Herzig, 1994, Gasquet et Herzig, 1994a, Gasquet et Herzig, 1995, Gasquet et Herzig, 1996, Gasquet et Herzig, 1996]. Ces travaux ont également fait l'objet d'un article de référence paru dans le « Handbook of Logic in Artificial Intelligence and Logic Programming » [Fariñas del Cerro et Herzig, 1995b]. L'extension de la méthode de traduction aux logiques d'actions augmentées par la dépendance ne devrait pas poser de problèmes majeurs, mais n'a pas encore été entreprise à ce jour. J'ai choisi de ne pas inclure ces travaux dans le présent document.

2 L'architecture de base d'un agent

L'objectif de l'informatique est la conception et la construction de machines capables de mémoriser et de traiter l'information (et l'objectif déclaré de l'intelligence artificielle est la construction de telles machines dont on peut prétendre qu'elles sont intelligentes). En ce qui concerne ces tâches, l'analogie humaine a été évoquée dès les débuts. Ainsi l'architecture de base d'un système cognitif en termes d'état interne, perception et action s'applique à ces agents artificiels que sont les machines. En IA on va plus loin dans cette analogie, et on analyse l'état mental en termes de plusieurs attitudes mentales : les croyances, intentions, etc.

Comment peut-on être sûr qu'une machine « fait » en effet ce qu'on attend d'elle ? La difficulté de cette question grandit avec la complexité des machines, et le recours à des procédures formelles de spécification, conception et vérification s'est imposé. L'outil formel par excellence étant la logique, je me suis intéressé à la formalisation d'un aspect essentiel des agents artificiels, à savoir la dynamique de leurs croyances, c-à-d. leur capacité d'évoluer en fonction de la perception de leur environnement, mais aussi la capacité d'agir sur cet environnement afin d'accomplir certaines tâches.

Les formalismes logiques permettent non seulement la modélisation de systèmes (c'est leur côté sémantique) mais aussi le calcul : c'est ce qui est étudié en théorie de la démonstration. Cette dernière a des aspects à la fois théoriques (décidabilité, complexité) et pratiques (procédures de décision, de démonstration automatique). En informatique, cette mise en œuvre à travers des procédures de calcul constitue le but et l'aboutissement obligatoire de la démarche logique.

Chaque agent qu'il soit humain ou artificiel, et plus généralement un système cognitif, possède deux capacités fondamentales. Il s'agit d'une part de la possibilité de percevoir le monde extérieur, et d'autre part de la faculté d'agir sur celui-ci. La perception s'effectue à travers des capteurs (« sensors »), tandis qu'une action est accomplie à travers le contrôle moteur (les effecteurs d'un robot). Dans le langage informatique et dans la théorie du contrôle, ce sont ces capteurs qui fournissent les entrées («input »), tandis que les effecteurs fournissent les sorties («output »). Une instance de ce schema perception/action est par exemple la communication entre agents. A partir de cette analyse, on obtient l'architecture de base d'un agent si on ajoute comme troisième composante l'état interne dans lequel l'agent se trouve à un instant donné. Les actions entreprises par l'agent sont alors fonction de cet état interne. D'autre part, cet état doit prendre en compte les perceptions, qui entraînent en général son évolution : l'état interne doit être modifié en fonction des perceptions.

L'architecture que je viens de décrire est aussi simple que générale, et on la retrouve sous des étiquettes différentes en philosophie (système cognitif) et en informatique (automate). Elle apparaît plus particulièrement en intelligence artificielle et en sciences cognitives, mais elle s'applique aussi en informatique théorique (perception et action sont alors réspectivement entrées et sorties d'un programme, dont l'état interne correspond aux valeurs des variables).

Si on considère que le monde est constitué de faits (c'est la position par exemple de Wittgenstein dans [Wittgenstein, 1927]), que nous pouvons identifions à des propositions, il est alors naturel de supposer qu'une composante essentielle de l'état interne d'un agent est une représentation symbolique de ces propositions, qui rend compte du monde extérieur. Nous appellons ces « propositions internes » les croyances de l'agent. Bien sûr, ces croyances sont en général imparfaites (incomplètes, erronées). Notons que la croyance est un état mental particulier, et que d'autres états mentaux existent comme les intentions, désirs et préférences, qui font également partie de l'état interne d'un agent.

Quand l'état interne d'un agent est modifié suite à une perception, étant donné que cette dernière est censée refléter le monde extérieur, il s'agit d'abord et en premier lieu d'un changement de croyances (qui sera repercuté ensuite sur les autres états mentaux comme les intentions). C'est pour cette raison qu'on s'est surtout intéressé dans la littérature à l'effet d'un changement sur les croyances. J'ai fait la même chose dans ce travail, et je me suis intéressé surtout aux interactions entre perception, croyances et actions (cependant, les intentions jouent un rôle à travers les buts dans la génération de plans d'actions).

Notons ici que le temps n'est pas une notion de base dans la déscription d'un agent et de son évolution. Cette dernière est rythmée par la perception d'objets, évènements et propositions, et par l'execution d'actions. Nous souscrivons donc à la vue relationnelle du temps, dans l'esprit de [Whitehead, 1919, Russell, 1914, Russell, 1956], et qui s'oppose à une vue absolue où les instants ou intervalles temporels sont primitifs.

2.1 Croyances

Qu'est-ce que les croyances? Beaucoup de modèles concurrents ont été proposés, tous plus ou moins fondés sur et motivés par des considérations rélatives à la cognition humaine. Ainsi, ont été proposés les réseaux probabilistes (bayesiens), les réseaux de neurones, ainsi que des formalismes plus inspirés par l'architecture des ordinateurs tels les bases de données relationelles et déductives, les réseaux sémantiques ou les systèmes de maintien de la cohérence.

Une approche alternative qui est plus prudente et plus modeste permet d'éviter de faire ces hypothèses fortes sur la structure des croyances. Elle prend comme point de départ leurs critères d'acceptation : quand est-il possible de dire qu'un agent entretient une certaine croyance? Ceci correspond au « knowledge level » postulé par Newell [Newell, 1982]. Si on identifie l'état mental doxastique d'un agent (son état de croyance) à l'ensemble des croyances acceptées par lui, et si on introduit des connecteurs entre croyances acceptées comme conjonction, implication et négation, on arrive de cette manière à un des modèles les plus simples des croyances, à savoir le modèle logique.

Ce choix a été fait souvent en intelligence artificielle, et j'ai fait de même. On y gagne en clarté et en généralité; par contre une telle représentation ne fournit que peu de possibilités pour structurer les croyances.

En effet, ce manque de structure dans le modèle logique est à l'origine des difficultés majeures qu'on rencontre dans ce qu'il m'intéresse ici, à savoir la prise en compte des perceptions et la gestion des actions. La solution que je préconise et que j'ai développé revient d'ailleurs à l'introduction dans le metalangage d'une structuration des croyances, à savoir une structure de dépendance, qui constitue un outil relativement simple (comparé par exemple à la théorie des probabilités) mais suffisamment puissant pour resoudre les problèmes principaux posés par le changement des croyances.

2.2 Perception

Au niveau de la perception, on distingue généralement deux étapes, la première correspondant à la perception d'objets et d'évènements, et la seconde à la perception de faits sur ces objets (propositions) : nous voyons d'abord un avion en train de décoller et nous percevons ensuite (du fait) qu'un avion est en train de décoller ; nous entendons un autre agent dire « la machine à café est en panne » et nous entendons qu'il dit que la machine à café est en panne. La première étape est non-épistémique : elle ne fait pas appel à des concepts et à des connaissances ou croyances de l'agent. C'est le contraire lors de la seconde étape : ce sont les croyances qui interviennent lorsque l'agent passe de la perception d'un objet à la perception d'une proposition. Dans la suite je ferai abstraction de la première étape et je supposerai que la perception est « de haut niveau » : les agents perçoivent des propositions.

Dans le cas de ce qu'on pourrait appeller un fonctionnement normal, la perception d'une proposition résulte en l'acceptation de cette proposition par l'agent. Cette proposition peut être en conflit avec les autres croyances acceptées :

dans l'exemple de la perception du décollage d'un avion, il se peut que l'agent acceptait jusqu'alors que l'avion était en stationnement ; or un objet ne peut être à plusieurs endroits à la fois. Les croyances doivent alors subir ce qu'on appelle une révision en philosophie formelle et en intelligence artificielle, et une mise à jour en bases de données, pour rétablir son adéquation avec l'état (supposé) du monde extérieur à l'agent. Les détails du déroulement de ce processus dépendent bien sûr fortement du modèle des croyances.

2.3 Action

Une division en deux niveaux peut être faite quant aux actions : à un niveau élémentaire un agent exécute des actions, qui font partie de la classe des évènements qui peuvent être produits par des agents et qui entraînent des changements. Ces actions s'inscrivent en général dans une suite d'actions (un plan) que l'agent est en train d'executer. À un niveau plus haut il s'agit alors d'atteindre un but que l'agent s'est fixé (qui est en particulier fonction de ses intentions), et auquel il croit parvenir en exécutant un plan généré préalablement. Ce sont les croyances qui interviennent alors pour permettre de générer un plan d'actions élémentaires à executer. Nous pouvons identifier un but à une proposition ou à un ensemble de propositions, dont l'agent croit en général qu'elles ne sont pas le cas.

Notons qu'un plan peut aussi faire appel aux capacités perceptives à travers des actions de test (ce qui rend possible une réduction de la perception, chose que nous n'allons pas faire ici).

3 Le langage de la logique

Dans quel formalisme pouvons-nous travailler ? Comme je viens de dire, j'ai choisi d'utiliser un langage logique. À partir de ce choix, j'ai continué à développer l'architecture des agents.

3.1 Le langage des croyances

Chaque croyance est une formule du langage. J'ai considéré que les croyances d'un agent sont organisées dans une base de croyances, qui est un ensemble fini de propositions.

Une alternative serait de considérer des ensembles satisfaisant certaines propriétés de fermeture, telle que la fermeture par modus ponens. Cependant, de tels ensembles sont infinis, alors que pour représenter des croyances sur un ordinateur nous sommes toujours amenés à considérer des représentations finis (qui peuvent cependant représenter des ensembles infinis).

D'autres arguments pour le choix de bases de croyances finies ont été débatues dans la littérature, à savoir qu'elles véhiculent des informations sur la structure justificationnelle des croyances. En philosophie, la position « fondationaliste » affirme que les différentes croyances n'ont pas toutes le même statut : quelques unes sont justifiées par d'autres croyances, tandis que d'autres sont justifiés par elles-mêmes (« self-justifying »). Ainsi, la base

$$\{p, p \to q\}$$

correspondrait à une structure justificationnelle différente de celle de

$$\{p, q\}$$

(alors que les deux bases sont logiquement équivalentes au moins en logique classique et aussi intuitionniste). L'évolution des croyances modifie alors en premier lieu les croyances auto-justifiées, et les croyances déduites sont ensuite modifiées en conséquence. A l'opposé la position cohérentiste nie une telle différence et donne le même statut à toutes les croyances.

Tout en ayant une représentation finie, j'ai considéré dans mon travail que des bases logiquement équivalentes doivent être modifiées de la même manière. La raison principale de ce choix est que sinon des différences apparemment insignifiantes pourraient devenir importantes, telles que celles entre les bases

$$\{p, q\}, \{p \wedge q\}, \{q \wedge p\}, \{p \wedge (p \wedge q)\}$$

etc. Cependant, l'approche que j'ai proposé est orthogonale à cette distinction et pourrait également être appliquée à partir de la position fondationaliste. Notons cependant que mon concept central de dépendance permet déjà une structuration des bases de croyances. Ainsi, des raisons de simplicité (rasoir d'Occam) m'ont amené à renoncer dans mon étude à un second outil de structuration que constituent les justifications.

Notons ici que depuis les travaux de Hintikka [Hintikka, 1962], les croyances ainsi que les connaissances des agents sont surtout étudiées à l'aide de logiques dites doxastiques et épistémiques. Dans ces logiques, on rajoute des opérateurs dits non-classiques au langage de la logique classique. Ainsi, la formule $B_i A$ est lue « l'agent i croit que A ». Comme l'a remarqué Gärdenfors [Gärdenfors, 1988, p. 29], l'intérêt principal de ces opérateurs est de pouvoir être imbriqués afin d'étudier des principes épistémiques tels que « si un agent croit quelque chose alors il croit qu'il le croit », tandis qu'on peut s'en passer au moins dans un premier temps lors de l'analyse de l'évolution des croyances. Cependant, rien ne nous empêche de les rajouter à l'approche ultérieurement, et de tels opérateurs auront certainement leur utilité en particulier dans le contexte de dialogue avec d'autres agents, où il s'agit de faire évoluer les croyances d'un agent au sujet des croyances (et aussi intentions, etc.) d'un autre agent qui vient de communiquer avec lui.

3.2 Le langage des perceptions

Comme je l'ai dit, j'ai suppose que ce qui est perçu par un agent est une proposition.

J'ai considéré que la prise en compte d'un perception par un agent peut se faire moyennant une fonction ayant comme arguments la proposition perçue et la base de croyances antérieure à la perception, et fournissant la base de croyances résultante. J'appelle cette fonction une opération de mise à jour.

Se pose alors la question à quel niveau linguistique cette opération se situe : dans la plupart des approches de la littérature, il est considéré qu'elle ne fait pas partie du langage mais du métalangage. En conséquence cette notion est analysé non pas en termes d'axiomes et règles d'inférence, mais en termes de postulats, qui contrairement aux précédents sont en général non-constructives (dans le sens qu'ils ne permettent pas de définir la notion de déduction). Au contraire, les systèmes logiques que j'ai étudiés contiennent des opérateurs du langage objet.

Comme opérateur je n'ai pas seulement considéré celui de la mise à jour, mais aussi un opérateur qui est en quelque sorte l'inverse de celui-ci, à savoir l'opérateur conditionnel. Tandis que le premier permet de parler des mises à jour effectives, le second correspond à une mise à jour hypothétique. Ceci m'a permi d'établir un lien avec des logiques des conditionnels, qui sont des systèmes logiques pour des expressions du type « si ... alors ... » bien étudiées en philosophie analytique. Ce lien se manifeste par la validité de ce qui est convenu d'appeller le test de Ramsey.

3.3 Le langage des actions et plans

J'ai dit que les actions sont une classe d'évènements bien particulière : ce sont des évènements qui peuvent être produits par des agents et qui entraînent des changements. Comme par ailleurs on peut considérer que les évènements sont des propositions, il est possible de réduire les actions à des propositions. Ceci revient à identifier une action à ses conséquences (ce qui a été proposé par des philosophes comme von Wright ou Segerberg [Segerberg, 1992], et par des informaticiens comme Winslett [Winslett, 1988]). Ainsi, l'exécution d'une action peut être connectée à une mise à jour.

Je n'ai pas poursuivi cette approche pourtant séduisante par sa simplicité, et j'ai considéré que les actions sont une entité à part bien différente des propositions. Je n'ai pas de justification philosophique particulière à offrir pour ce choix, qui a été motivé plutôt par des aspects algorithmiques : il s'est avéré que la variante de la logique dynamique que nous avons étudié pour traiter des actions peut être mécanisée d'une manière relativement simple et efficace, ce qui m'a semblé être moins le cas des procédures associées aux mises à jour.

J'ai considéré un langage avec un ensemble d'actions atomiques. À partir de ces actions atomiques des formules peuvent être construites exprimant qu'une proposition est vraie après l'exécution d'une action. Ainsi, ces formules correspondent aux conditionnels (tandis que les mises à jour correspondraient ici à l'inverse d'un opérateur d'action, exprimant qu'une certaine proposition était vraie avant l'exécution de l'action).

Dans la suite je vais examiner les inconventients de mon approche logique pour l'évolution des croyances, en montrant qu'ils sont d'ordre plus général et pèsent sur potientiellement toutes les formalismes logiques lors qu'on les applique à des problèmes de perception et d'action, ou plus généralement à des problèmes d'évolution des croyances.

3.4 Le manque de structure

Quand on adopte une représentation des connaissances en logique, des difficultés sérieuses apparaissent lorsqu'on essaie d'aller au-delà du cas assez limité d'une seule base de croyances sûres et invariantes, c'est-à-dire de croyances qui n'admettent pas d'exceptions et qui ne peuvent évoluer. L'apparition de ces difficultés est un phénomène général qu'on rencontre autant en mise à jour des bases de données logiques qu'en raisonnement sur les actions, révision des croyances, raisonnement à partir de règles ayant des exceptions, raisonnement temporel, fusion de plusieurs bases de croyances (raisonnement multi-sources) etc., en somme tout ce qui a trait à la dynamique des bases de croyances. Ce sont J. McCarthy et P. Hayes [McCarthy et Hayes, 1969] (voir aussi [McCarthy, 1990]) qui ont les premiers fait état de ces problèmes dans le cadre de la logique classique, et plus tard ce sont S. Hanks et D. McDermott [Hanks et McDermott, 1986] qui ont affirmé qu'elles sont inhérentes à toutes les approches logiques (classiques ou non- classiques), et que même les formalismes qui ont été développés explicitement pour les surmonter devront toujours capituler devant un contre-exemple ou un autre pour lequel ils ne réussissent pas à donner la solution intuitive. Constatant l'impossibilité de présenter une théorie qui résout tous les exemples, Sandewall [Sandewall, 1993, Sandewall, 1994] a ainsi proposé une typologie d'exemples qui conduit à une typologie de théories.

Mon hypothèse générale de départ est que les difficultés viennent du fait que traditionnellement en logique, on suppose que l'ensemble des atomes d'un langage logique n'est muni d'aucune structure de dépendance : en logique, tout ce qui n'est pas déclaré comme dépendant est supposé indépendant.

Ceci se manifeste par exemple par le fait que deux atomes logiques qui ont un nom différent, sont considérés comme (logiquement) indépendants et n'ayant « rien à voir» l'un avec l'autre. Concrètement, ceci veut dire que pour les deux constantes propositionnelles, toute combinaison booléenne est satisfaisable, ce qui est effectivement le cas dans toutes les formalismes logiques connus. Que ceci soit contre-intuitif est assez clair, il suffit de considérer « il pleut » et « la route r est mouillée », ou « Jean est professeur (d'école publique) » et « Jean est fonctionnaire ».

Bien sûr, dans les systèmes logiques on trouve la notion de théorie (propositionnelle ou du premier ordre) qui permet de formuler les lois d'un domaine d'application. Ainsi, dans l'exemple ci-dessus la structure de dépendance entre p et q peut être reflétée par la théorie

$$T = \{p \to q\}$$

où $\to$ est l'implication matérielle, et p et q ne seront plus alors indépendants dans la théorie T (car $p \wedge \neg q$ est insatisfaisable avec T). Ainsi une théorie logique est un moyen pour rendre compte d'un ensemble de lois du domaine ou contraintes d'intégrité en intelligence artificielle et bases de données logiques. Etant donné une théorie T, on veut alors savoir si une formule F est conséquence logique d'une base de croyances K, ce qui peut s'écrire symboliquement

$$K \vdash_T F$$

où $\vdash_T$ représente une relation de conséquence non-classique.

Il se trouve que la notion de théorie s'adapte bien au cas évoqué ci-dessus d'une base de croyances sûres et invariantes : étant donné une théorie T, savoir si on a $F \vdash_T G$ veut dire établir la déduction $T \vdash F \rightarrow G$ en logique classique. Des exemples standards de raisonnements sur des croyances sûres et invariantes dans une théorie particulière se rencontrent en mathématiques (théorie des groupes, etc.). Dès qu'on va au-delà de ce cadre, on se heurte à des problèmes majeurs. D'abord et comme l'ont constaté McCarthy et Hayes [McCarthy et Hayes, 1969], l'inadaptation de la logique classique au raisonnement sur des règles avec exceptions ou sur des systèmes évolutifs, est due à sa propriété de monotonie :

$$\text{si } F \vdash_T G \text{alors } F, F' \vdash_T G$$

où F' est une formule quelconque. Intuitivement, cette monotonie n'est pas souhaitable inconditionnellement, mais seulement lorsque l'ajout de F' ne met pas en question la déduction de G : il faut et il suffit que F' soit indépendant de G (en un sens à préciser) pour que l'ajout puisse être effectué.

Malgré une multitude de recherches très variées qui sont parties de ce constat, aboutissant à diverses propositions de formalismes logiques, on semble loin d'un consensus même après 30 ans de recherche, et le constat dressé par Hanks et McDermott en 1986 est encore valable : Il n'existe actuellement aucun système logique pour raisonner sur les actions qui soit à la fois simple, naturel et muni de bonnes propriétés logiques et qui permette de donner la solution intuitive à tous les exemples académiques de la littérature (en particulier le fameux « Yale Shooting Problem »). On trouve une situation similaire dans le cas du raisonnement à partir de croyances génériques (règles avec exceptions), où l'on étudie des relations d'inférence non-monotones $\mid\!\sim$ qui sont plus faibles que la relation d'inférence classique. Dans ce cadre, on semble considérer que G est indépendant de F si on n'a pas la relation de conséquence $F \vdash_T \neg G$, comme le suggère la propriété de monotonie rationnelle

$$\text{si } F \vdash_T H \text{ et } F \not\vdash_T \neg G \text{ alors } F \wedge G \vdash_T H$$

Cependant, comme l'a constaté Delgrande [Delgrande et Pelletier, 1994], ceci ne résoud pas les problèmes de spécificité, d'irrelevance et d'héritage. La situation est légèrement différente dans le domaine des bases de données logiques, où le besoin de solutions opérationnelles a justifié de fortes restrictions sur la forme syntaxique des lois du domaine et des formules logiques qui peuvent être représentées dans la base de croyances.

Notons ici que le type de loi que j'ai donné en exemple correspond au cas simple d'une loi statique, c'est-à-dire ne se référant qu'à un seul état de la base de croyances. Pendant longtemps, ce sont uniquement des lois statiques qui ont été considérées. Des lois dynamiques (qui apparaissent par exemple dans le « Yale Shooting Problem ») traitent d'au moins deux états de la base et posent des problèmes strictement plus complexes que les lois statiques.

En résumé, utiliser des théories logiques pour rendre compte, à l'intérieur même de la logique, de la structure de dépendance du domaine, s'avère d'une mise en œuvre délicate, sans qu'il n'y ait de démarche générale pour le faire.

4 Raisonnement et incertitude

En intelligence artificielle, il y a beaucoup d'approches dont la démarche peut être vue comme un ajout de structure à l'ensemble des formules du langage. Ces structures sont des ordres (ou des pré-ordres) généralement totaux qui pour deux formules données indiquent laquelle des deux est la plus préférée (la moins susceptible de présenter des exceptions, la plus certaine, la plus enracinée, etc.). Ce type de structure permet de rendre compte des lois du domaine d'une manière plus nuancée qu'avec des théories : il suffit d'attribuer aux formules une préférence maximale afin de les préserver lors d'un changement, ou de les appliquer en priorité lors d'un raisonnement avec exceptions.

Ces ordres (et les formalismes non-monotones associés) ont été étudiés tant sémantiquement qu'axiomatiquement (relations d'inférence). Ce type d'approches, en particulier le système P, fournit une sorte de consensus minimal sur les principes généraux du raisonnement non-monotone [Kraus *et al.*, 1990, Gärdenfors et Makinson, 1988]. Ayant montré comment ces ordres peuvent être identifiés aux modèles des logiques des conditionnels [Fariñas del Cerro et Herzig, 1991], le lien entre relations d'inférence non-monotones et logiques des conditionnels a été établi à travers ces ordres dans [Fariñas del Cerro *et al.*, 1994].

De tels formalismes peuvent potentiellement résoudre les problèmes posés. Cependant, il n'est pas concevable de représenter de tels ordres totaux dans des applications concrètes, car leur taille est exponentielle avec le nombre d'atomes du langage. On a alors recours à des ordres spécifiés partiellement, et on applique des principes de fermeture méta-linguistiques (qui ne peuvent être écrits sous forme d'axiomes) tel la fermeture rationnelle [Lehmann et Magidor, 1992], le minimum de spécificité [Benferhat *et al.*, 1992], ou le système Z [Pearl, 1990]. Un point faible des approches utilisant des ordres est qu'elles ne permettent pas d'obtenir toutes les conséquences non-monotones souhaitées. Ainsi, des problèmes d'irrelevance[1] ou de « noyade »[2] [Benferhat *et al.*, 1993,

1. Par exemple, pouvoir déduire que les pingouins pondent des œufs à partir de l'ensemble des connaissances par défaut « les oiseaux volent », « les pingouins sont des oiseaux », « les pingouins ne volent pas », « les oiseaux pondent des œufs », n'est pas possible avec les approches citées plus haut : une sous-classe (ici, les pingouins) exceptionnelle d'une classe (ici, les oiseaux) par rapport à une propriété (ici, voler) est implicitement considérée, par ces approches, comme potentiellement exceptionnelle par rapport à n'importe quelle autre propriété (ici, pondre des œufs). Pourtant on souhaiterait pouvoir exprimer et exploiter le fait que pondre des œufs n'a rien à voir avec le fait de voler (dans le contexte des oiseaux).

2. Le problème de la noyade correspond au fait que dans beaucoup d'approches basées

Delgrande et Pelletier, 1994] posent un problème à l'outil des ordres.

D'une manière générale, le problème est que dans ces formalismes on a abandonné la propriété de monotonie, et donc toutes les hypothèses d'indépendance sous-jacentes à la logique classique, ce qui rend l'inférence dans ces formalismes très prudente. L'enjeu est donc de se rapprocher le plus possible de la monotonie en gérant les relations d'indépendance. Les règles de cumulativité

$$\text{Si } F \mid\!\sim G \text{ alors } (F \mid\!\sim H \text{ si et seulement si } F \wedge G \mid\!\sim H)$$

et de monotonie rationnelle peuvent être interprétées dans ce sens. Dans cette perspective, c'est ce problème de gestion explicite des indépendances qui est au cœur des limitations des logiques non-monotones.

En résumé, même si ces ordres permettent effectivement de munir le langage d'une structure (qui reste à un niveau méta-linguistique), on constate que celle-ci ne rend pas compte de la notion d'indépendance. De plus, se posent des problèmes représentationnels et même dans le cas d'un ordre spécifié partiellement, on n'atteint pas le pouvoir déductif souhaité.

Il est intéressant de noter que – alors que j'ai souligné que l'indépendance des atomes est une hypothèse implicite en logique (qu'il est possible d'annuler par la déclaration explicite d'une contrainte reliant deux atomes) – la situation inverse prévaut en théorie des probabilités : deux variables aléatoires sont par défaut dépendantes, et le cas échéant il faut déclarer explicitement qu'elles sont indépendantes : étant donné les probabilités $P(F)$ et $P(G)$, on ne connaît en général qu'un encadrement de la probabilité $P(F \wedge G)$ (qui ne vaut $P(F) \cdot P(G)$ qu'en cas d'indépendance stochastique).

Dès qu'on passe à la modélisation d'un domaine par une méthode probabiliste, on adopte toujours la démarche suivante (que ce soit par les réseaux bayésiens ou markoviens) : les connaissances du domaine s'expriment sous la forme de contraintes probabilistes, et des hypothèses d'indépendance conditionnelle sont formulées qui permettent de décomposer la distribution de probabilités sous-jacente. Pearl [Pearl, 1988] a proposé une caractérisation axiomatique de la notion de l'indépendance $I(X, Y|Z)$ (X est indépendant de Y dans le contexte Z) entre des variables (et non des évènements ou propositions). Il conjecture que pour toute relation d'indépendance qui satisfait cette axiomatique (ou encore qui a une structure de graphoide définie dans [Pearl, 1988]), il existe une distribution de probabilités P telle que :

$$P(X|Z) = P(X|Z \cup Y) \text{ si et seulement si } I(X, Y|Z)$$

En revanche, on ne sait pas encore très bien faire ce travail sur les représentations ordinales des croyances typiques des formalismes non-monotones telles que celle offerte par la théorie des possibilités. De telles notions d'indépendance sont présentées dans la section suivante.

sur l'idée d'ordre, les formules moins « enracinées » que celles invalidées directement par la situation exceptionnelle rencontrée, se trouvent elle-mêmes invalidées, même si elles ne sont pas parti prenantes dans l'incohérence créée par cette situation exceptionnelle.

5 Les fondements de la dépendance

L'indépendance étant a priori un concept non-logique, son utilisation dans des formalismes logiques m'a amené à la question de ses propriétés formelles.

La notion d'indépendance en probabilités s'exprime multiplicativement, par

$$F \text{ et } G \text{ sont indépendants si et seulement si } P(F \wedge G) = P(F) \cdot P(G).$$

Il s'ensuit alors de l'axiomatique des probabilités que l'indépendance est une relation symétrique qui est insensible à la négation, dans le sens où l'indépendance de F et G implique l'indépendance de F et $\neg G$. Cependant, l'indépendance probabiliste a été critiquée dans la littérature dès les années 20 par des chercheurs comme J.M. Keynes, critiques qui ont été reprises dans les travaux de Gärdenfors [Gärdenfors, 1978, Gärdenfors, 1990]. Ces critiques ont porté en particulier sur l'absence de régularités par rapport à la conjonction. Ainsi, des définitions différentes de la loi de multiplication ont été proposées afin de satisfaire au critère conjonctif pour l'indépendance :

(CCI) F indépendant de G et F indépendant de H implique F indépendant de $G \wedge H$,

et au critère conjonctif pour la dépendance :

(CCD) F dépendant de G et F dépendant de H implique F dépendant de $G \wedge H$.

Il existe donc (au moins chez les philosophes) plusieurs notions d'indépendance probabiliste. Il semble qu'elles s'appliquent à des problèmes différents tel que la mise à jour, le raisonnement avec des exceptions, la causalité, les thèmes d'intérêt, etc., et qu'elles leur soient spécifiques. Sur un plan plus général, il se pose la question de savoir si l'indépendance peut être construite sur d'autres bases que la théorie des probabilités, en particulier sur d'autres théories de l'incertain. J'ai en particulier étudié des notions fondées sur l'idée de thème et sur la théorie des possibilités. Comme certaines de ces relations ne sont pas symétriques, j'ai préféré la formulation « F est indépendant de G » à « F et G sont indépendants ».

5.1 La dépendance basée sur les théories de l'incertitude

La théorie des possibilités [Dubois et Prade, 1988] propose une modélisation de l'incertain différente de la théorie des probabilités. Elle offre en particulier un cadre ordinal, plus qualitatif que les probabilités. Par ailleurs, le cadre possibiliste permet de raffiner, par l'introduction de dégrés, les trois états épistémiques qui existent pour une proposition en logique classique (certainement vraie, certainement fausse, possiblement vraie et possiblement fausse). C'est sa nature qualitative, et sa compatibilité avec le caractère ternaire des états de connaissance, qui rendent la théorie des possibilités intéressante pour les applications au raisonnement de sens commun.

A quoi correspond la loi de la multiplication ici ? La probabilité d'une conjonction de formules indépendantes étant égal au produit des probabilités des conjoints, la contrepartie naturelle dans le cadre des possibilités est l'opérateur min. Ainsi, on peut penser définir que F et G sont indépendants, en théorie des possibilités, si le dégré de possibilité $\Pi(F \wedge G)$ est égal à $\min(\Pi(F), \Pi(G))$. Dans la terminologie de [Nahmias, 1978], les formules F et G sont dites non-reliées (« unrelated »).

Cependant, d'autres définitions peuvent être obtenues si on se réfère à une formulation plus intuitive de l'indépendance probabiliste, à savoir que F et G sont indépendants si et seulement si la probabilité de F est insensible au conditionnement par G, c.à.d. $P(F|G) = P(F)$. Tandis que les axiomes pour la fonction Prob font que cette formulation est équivalente à la définition multiplicative, ceci n'est pas le cas dans la théorie des possibilités. Ainsi, si on stipule que F et G sont fortement indépendants si et seulement si $N(F|G) = N(F)$, [3] on obtient une notion différente de celle de la relation ci-avant exprimant l'absence de liens.

Dans [Fariñas del Cerro et Herzig, 1995c, Fariñas del Cerro et Herzig, 1996a, Dubois *et al.*, 1997], j'ai établi la carte des différentes notions d'indépendance, qui correspondent aux différents changements d'état envisageables pour F à l'arrivée d'une nouvelle information G.

J'ai étudié leurs propriétés et axiomatisations. En particulier, j'ai identifié les notions qui satisfont les critères conjonctifs de dépendance et d'indépendance (CCD) et (CCI), ainsi que les critères correspondants pour la disjonction (DCI) et (DCD), qui sont plausibles sous certaines interprétations. Il est à souligner qu'aucune de ces notions n'est sensible à la négation, et que seule l'absence de lien possède la propriété de symétrie. Dans [Dubois *et al.*, 1994, Dubois *et al.*, 1997, Dubois *et al.*, 1998], des liens formels ont été établis entre ces notions et les postulats de Kraus, Lehmann et Magidor [Kraus *et al.*, 1990] pour les relations d'inférence non-monotones et les postulats AGM pour la révision des croyances de Alchourón, Gärdenfors et Makinson (voir [Gärdenfors, 1988]).

Les différents types d'indépendance possibiliste se généralisent aisément au cas général d'une relation d'indépendance est une relation ternaire (F est indépendant de G étant donné un certain contexte K).

5.2 La dépendance basée sur la notion de thème

Supposons une base de croyances K telle que F peut être déduit de K. Si la nouvelle information G traite de tout à fait autre chose que F, nous pouvons considérer que F est indépendant de G et doit donc survivre à la révision. L'analyse de la signification de « traiter de quelque chose » amène naturellement à la

3. N dénote une mesure de nécessité associée par dualité à une mesure de possibilité Π, c'est-à-dire $N(F) = 1 - \Pi(\neg F)$ et $N(F|G) = 1 - \Pi(\neg F|G)$, où la mesure de possibilité conditionnelle satisfait la relation $\Pi(F \wedge G) = \min(\Pi(F|G), \Pi(G))$. La relation d'indépendance est plutôt exprimée en terme de nécessité $N(F|G) = N(F)$, car $\Pi(F) < 1$ a un sens « négatif », à savoir que F n'est pas complètement possible, tandis que $N(F) > 0$ exprime que F est quelque peu certain.

notion de thème d'intérêt, qui a été étudié en particulier dans le contexte des bases de données [Cuppens et Demolombe, 1989a, Cuppens et Demolombe, 1989b, Cuppens et Demolombe, 1991, Cholvy et Demolombe, 1994, Fariñas del Cerro et Lugardon, 1991, Lugardon, 1996]. On en trouve d'autres applications dans un contexte d'apprentissage supervisé ou de communication entre agents, où il est important de ne prendre en considération que les croyances qui apportent de l'information sur le thème sur lequel on se focalise. Par exemple, le lecteur de cette synthèse s'attend sans doute à y trouver des informations nouvelles (c'est-à-dire qu'il ne connait pas déjà), à condition que celles-ci soient relatives à la notion de dépendance. Ainsi, des informations sur les loisirs de l'auteur le laissera sûrement de marbre ...

Formellement, cette analyse présuppose l'existence d'un ensemble de thèmes T et d'une fonction t associant à chaque formule F un ensemble de thèmes $t(F)$. F et G sont alors indépendantes (par rapport à T et t) si et seulement si elles n'ont pas de thèmes en commun, c'est-à-dire si

$$t(F) \cap t(G) = \emptyset$$

La question des propriétés de cette notion d'indépendance peut être réduite à celle des propriétés naturelles de la fonction thématique t. Intuitivement, il est plausible de supposer que l'ensemble des thèmes associé à $F \wedge G$ soit contenu dans l'union des thèmes associés à F et à G :

$$t(F \wedge G) \subseteq t(F) \cup t(G)$$

Au niveau de l'indépendance induite, ceci correspond exactement au critère conjonctif pour l'indépendance (CCI). De même, le critère (CCD) est validé dès que $t(F) \cup t(G) \subseteq t(F \wedge G)$, ce qui peut paraître moins plausible ici : si l'on instancie F par p et G par $p \vee q$, alors $F \wedge G$ est équivalent à p, et (si l'on admet un principe d'irrelevance de la syntaxe[4]) on a $t(F \wedge G) = t(p)$. Ceci contraindrait $t(p \vee q)$ à être contenu dans $t(p)$, ce qui ne peut être le cas en général. Une solution (adopté dans la plupart des approches basées sur les thèmes) consiste alors à rejeter le principe d'indépendance de la syntaxe. Parallèlement, le critère disjonctif pour l'indépendance (DCI) correspond à

$$t(F \vee G) \subseteq t(F) \cup t(G)$$

et le critère disjonctif pour la dépendance (DCD) à la propriété (moins plausible) que

$$t(F) \cup t(G) \subseteq t(F \vee G)$$

Admettre ces quatre propriétés revient donc à admettre les quatre critères (CCI), (CCD), (DCI) et (DCD). Si l'on rajoute à cette hypothèse que $t(F) = t(\neg F)$, alors on obtient un représentation des croyances particulièrement économique : en effet, la valeur de t pour une formule quelconque F est déterminée par les valeurs des atomes de F. Ceci a été fait dans le raisonnement multi-sources [Cholvy,

4. Ici, ce principe nous amène à postuler que des formules logiquement équivalentes ont les mêmes thèmes : si $\vdash F \leftrightarrow G$ alors $t(F) = t(G)$.

1993, Cholvy, 1994b, Cholvy, 1994a] (voir aussi [Cazalens et Demolombe, 1992, Cazalens *et al.*, 1992]). Comme nous l'avons vu, le prix de cette économie est l'abandon du principe d'irrelevance de la syntaxe.

Une autre hypothèse est que l'unique thème d'une formule atomique est la formule elle-même : $t(p) = \{p\}$. Si on la rajoute aux précédentes on obtient que le thème d'une formule est l'ensemble des formules atomiques qu'elle contient. Dans ce cas la représentation explicite de la fonction t est superflue. Il a été montré dans [Herzig, 1996, Fariñas del Cerro et Herzig, 1994a, Fariñas del Cerro et Herzig, 1997] que ceci correspond à la notion d'indépendance sous- jacente dans l'approche des mondes possibles de [Winslett, 1988].

Fondamentalement, nous disposons maintenant de trois types de relations pour structurer l'ensemble des formules :

- les indépendances ;
- les dépendances (qui peuvent être de différentes natures) ;
- les lois du domaine : Elles établissent des liens entre formules, dont la nature est exprimée au moyen de l'implication matérielle (ou plus généralement, d'une formule logique quelconque).

Dans la suite ces outils formels vont être appliqués à l'analyse des perceptions et des actions.

6 L'utilisation des dépendances dans l'évolution des bases de croyances

L'information sur l'indépendance de deux formules F et G peut être appliquée directement dans le cas de l'évolution d'une base de croyances K ; voir [Fariñas del Cerro et Herzig, 1994b, Fariñas del Cerro et Herzig, 1995d]. En effet, le postulat suivant paraît raisonnable et naturel :

(I$\diamond$) Si G est indépendant de F et $G \in K$, alors $G \in K \diamond F$

où $\diamond$ est un opérateur de changement qui étant donné une base de croyances K et une nouvelle information F renvoie une nouvelle base $K \diamond F$ incorporant la nouvelle information. Dans le cas contraire où F et G sont dépendants, nous ne pouvons *a priori* rien dire au sujet de G dans $K \diamond F$: nous savons bien que F et G sont reliés, mais nous ignorons de quelle manière. Ce sont les lois du domaine qui doivent permettre de conclure ici.

Depuis les travaux de [Katsuno et Mendelzon, 1992] il est devenu habituel de classifier les évolutions des bases de croyances en révisions et en mises à jour ; voir aussi [Sombé, 1994]. Pour ces dernières on suppose que le monde (dont la base est censée rendre compte) a évolué, tandis que pour les premières le monde n'évolue pas : ce ne sont que nos croyances sur le monde qui évoluent. Ainsi, quand on met à jour, on s'imagine qu'une action a été exécutée dont les effets doivent être répercutés sur la base ; quand on révise, on ajuste ses croyances sur un monde statique à la lumière de nouvelles informations, à l'image d'un détective qui enquête sur un meurtre en partant d'indices et d'alibis. Réviser un document veut dire corriger des erreurs et éventuellement le compléter, et quand on le met à jour on modifie ce qui a « bougé dans le monde », par exemple des adresses et numéros de téléphone. Le postulat (I$\diamond$) peut être appliqué aux deux types d'évolution.

Au niveau de la mise à jour, dans [Herzig, 1996, Fariñas del Cerro et Herzig, 1997] il a été montré que si l'opérateur de changement $\diamond$ est celui de l'approche des mondes possibles de [Winslett, 1988], alors cet opérateur peut être axiomatisé avec (I$\diamond$), où la relation d'indépendance est basée sur la fonction de thème t « extrême » tel que $t(F)$ est l'ensemble des atomes de F. Cette axiomatique fournit également une méthode de démonstration automatique basée sur la mise en forme normale. Cependant, cette approche ne permet pas de caractériser les formules préservées lors de la mise à jour d'une base (elle met seulement en évidence une condition suffisante). Dans [Marquis, 1994] une caractérisation exacte des formules indépendantes d'une mise à jour donnée est fournie. Cette caractérisation a permis de montrer que lorsque la base de croyances est compilée sous forme de ses impliqués premiers, la mise à jour avec certains types de formules syntaxiquement contraintes (conjonction ou disjonction de littéraux) peut être réalisée en temps polynomial (même quasi-linéaire) par rapport à la taille des données, tandis que dans le cas général, le problème de l'interrogation de la base mise à jour est Π_2^p-complet [Eiter et Gottlob, 1992].

Dans [Herzig, 1997], les lois du domaine on été intégrées aux postulats AGM pour la révision de Alchourròn, Gärdenfors et Makinson, et il a été étudié comment pour une relation d'indépendance basée sur les thèmes le postulat (I$\diamond$) peut être rajouté aux postulats AGM étendus par les lois. Etant donné que l'indépendance basé sur les thèmes est en général sensible à la syntaxe, le point central ici est de garantir l'irrelevance de la syntaxe qui est requis par les postulats AGM.

7 L'utilisation des dépendances dans le raisonnement sur les actions

Comme je l'ai évoqué plus haut, une mise à jour d'une base de croyances correspond souvent à l'exécution d'une action dans le monde physique. A quoi correspond le postulat (I$\diamond$) dans une logique des actions ? J'ai choisi comme logique d'actions la logique dynamique, qui augmente la logique classique avec des opérateurs modaux de la forme $[a]$, où a est une action. La formule $[a]F$ peut être lue « F est vrai après exécution de l'action a ».

L'approche des mondes possibles se voulant une logique des actions, ceci est possible si on identifie une action à l'ensemble de ses résultats (qui lui est décrit par une formule logique). Dans la logique dynamique, les actions sont une catégorie linguistique à part entière et différente de celle des formules. Ceci nous amène à modifier le domaine de la relation de dépendance : ici, elle lie des actions à des formules.

La persistance des formules indépendantes peut alors être décrite par

(I[a]) Si F est indépendant de l'action a, alors $F \to [a]F$.

Si on rajoute cet axiome à la logique dynamique (ou à une version plus faible sans l'opérateur d'itération, car celui-ci augmente la complexité alors qu'il n'est pas nécessaire pour traiter les problèmes qu'on se pose traditionnellement dans le domaine du raisonnement sur les actions), on obtient alors une famille de logiques paramétrées par une relation de dépendance.

Dans [Castilho *et al.*, 1997b] il a été montré comment les problèmes du décor et de la ramification peuvent être résolus dans ce cadre, et une méthode de démonstration automatique par tableaux a été exhibée.

8 Perspectives

Ayant bien établi une structuration des croyances à travers la dépendance, j'ai dans mon travail intégré cette notion dans deux cadres formels de type logique non-classique importants, à savoir les logiques des conditionnels et la logique dynamique. À ces deux formalismes j'ai pu associer des procédures mécaniques de décision, en tirant profit de mes travaux antérieurs sur des méthodes générales de démonstration automatique pour les logiques non-classiques.

Ceci constitue une avancée importante dans le programme de recherche que je me suis fixé. Il s'agit maintenant d'une part d'étendre cette démarche à des formalismes plus riches, et d'autre part de passer à l'étape suivante et de mettre en œuvre cette méthodologie dans les applications que j'ai énumérées plus haut. Ces deux objectifs doivent être poursuivis en parallèle et même interagir.

En effet, les cadres logiques que j'ai établi sont focalisés sur la dynamique et n'intègrent pas encore des opérateurs modaux de croyance ou d'intention. De tels travaux viennent d'être entamés en collaboration avec L. Fariñas del Cerro, D. Longin et O. Rifi, et les premiers résultats sur ce sujet seront publiés dans [Fariñas del Cerro *et al.*, 1998].

Cette publication vise également une application qui me paraît particulièrement prometteuse, à savoir le dialogue homme-machine coopératif, où il s'agit de reconstruire les croyances des locuteurs et auditeurs après un acte de discours.

À l'origine de ces dernières recherches il y a un projet industriel avec le CNET (depuis 1997) dans le domaine de l'audiotel, et il est permis d'espérer que des applications plus ambitieuses de type « agent personnel » seront à la portée de mon approche.

Références

[Balbiani et Herzig, 1994] Philippe Balbiani et Andreas Herzig. A translation from the modal logic of provability into K4. *J. of Applied Non-Classical Logics (JANCL)*, 4(1):73–78, 1994.

[Benferhat *et al.*, 1992] Salem Benferhat, Didier Dubois, et Henri Prade. Representing default rules in possibilistic logic. In *Proc. 4th Int. Conf. on Knowledge Representation and Reasoning (KR'92)*, pages 673–684. Morgan Kaufmann Publishers, 1992.

[Benferhat *et al.*, 1993] Salem Benferhat, Claudette Cayrol, Didier Dubois, Jerôme Lang, et Henri Prade. Inconsistency management and prioritized syntax-based entailment. In Ruzena Bajcsy, editor, *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*, pages 640–645. Morgan Kaufmann Publishers, 1993.

[Castilho *et al.*, 1997a] Marcos A. Castilho, Luis Fariñas del Cerro, Olivier Gasquet, et Andreas Herzig. Modal tableaux with propagation rules and structural rules. *Fundamenta Informaticae*, 32(3/4):281–297, 1997.

[Castilho *et al.*, 1997b] Marcos A. Castilho, Olivier Gasquet, et Andreas Herzig. Modal tableaux for reasoning about actions and plans. In Sam Steel et Rachid Alami, editors, *European Conference on Planning (ECP'97)*, number 1348 in LNAI, pages 104–116. Springer Verlag, 1997.

[Cazalens *et al.*, 1992] Sylvie Cazalens, Robert Demolombe, et Andrew Jones. A logic for reasoning about is about. Technical report, ONERA-CERT, Toulouse, 1992.

[Cazalens et Demolombe, 1992] Sylvie Cazalens et Robert Demolombe. Intelligent access to date and knowldge bases via user's topics of interest. In *Proc. IFIP'92*, 1992.

[Cholvy et Demolombe, 1994] Laurence Cholvy et Robert Demolombe. Reasoning with information sources ordered by topics. In *Proc. AIMSA'94*, 1994.

[Cholvy, 1993] Laurence Cholvy. Proving theorems in a multisources environment. In *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*, pages 66–71, 1993.

[Cholvy, 1994a] Laurence Cholvy. Fusion de sources d'informations contradictoires ordonnées en fonction des thèmes. *Revue d'Intelligence Artificielle (RIA)*, 8(2):187–204, 1994.

[Cholvy, 1994b] Laurence Cholvy. A logical approach to multisources reasoning. 1994.

[Collectif, 1995] Collectif. Gestion de l'evolutif et de l'incertain dans les bases de connaissances. In *Actes des 5èmes journées nationales du PRC-GDR Intelligence Artificielle*, pages 77–122, Nancy, France, February 1995. (Collectif du projet Inter-PRC : N. Bidoit, S. Cerrito, L. Cholvy, M.-O. Cordier, P. Dague, D. Dubois, F. Dupin de Saint-Cyr, D. Fontaine, C. Froidevaux, M. Ghallab, J.-L. Golmard, A. Herzig, J. Lang, F. Lévy, Y. Moinard, O. Papini, H. Prade, C. Sayettat, C. Schwind, L. Travé-Massuyès, R. Valette).

[Collectif, 1997] Collectif. Lois, structures et dépendances. In *Actes des sixièmes journées nationales du PRC-GDR Intelligence Artificielle*. Hermès, Paris, March 1997. (Groupe PRC "LSD" : S. Benferhat, Ph. Besnard, L. Cholvy, R. Demolombe, D. Dubois, O. Gasquet, J.-L. Golmard, A. Herzig, P. Marquis, Y. Moinard, H. Prade, C. Schwind).

[Crocco *et al.*, 1995] Gabriella Crocco, Luis Fariñas del Cerro, et Andreas Herzig, editors. *Conditionals: From Philosophy to Computer Science*. Studies in Logic and Computation. Oxford University Press, 1995.

[Crocco et Herzig, 1997] Gabriella Crocco et Andreas Herzig. The Ramsey Test as an inference rule. In Paul Weingartner, Gerhard Schurz, et Georg Dorn, editors, *The role of pragmatics in contemporary philosophy: Contributions of the Austrian Ludwig Wittgenstein Society*, volume 5, Kirchberg, 1997. Austrian Ludwig Wittgenstein Society.

[Cuppens et Demolombe, 1989a] Frédéric Cuppens et Robert Demolombe. Cooperative answering: a methodology to provide intelligent access to databases. In *Proc. 2nd Int. Conf. on Expert Database Systems*, 1989.

[Cuppens et Demolombe, 1989b] Frédéric Cuppens et Robert Demolombe. How to recognize interesting topics to provide cooperative answering. *Information Systems*, 14(2), 1989.

[Cuppens et Demolombe, 1991] Frédéric Cuppens et Robert Demolombe. Extending answers to neighbour entities in a cooperative answering context. *Decision Support Systems*, 7(1), 1991.

[Delgrande et Pelletier, 1994] James P. Delgrande et Francis J. Pelletier. A formal approach to relevance. In *Proc. AAAI Fall Symposium on Relevance*, New Orleans, November 1994.

[Dubois *et al.*, 1994] Didier Dubois, Luis Fariñas del Cerro, Andreas Herzig, et Henri Prade. An ordinal view of independence with applications to nonmonotonic reasoning. In Ramon Lopez de Mantaras et David Poole, editors, *Proc. Int. Conf. on Uncertainty in AI (UAI'94)*, pages 1855–203, Seattle, 1994.

[Dubois *et al.*, 1997] Didier Dubois, Luis Fariñas del Cerro, Andreas Herzig, et Henri Prade. Qualitative relevance and independence: a roadmap. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI-97)*, pages 62–67, Nagoya, 1997.

[Dubois *et al.*, 1998] Didier Dubois, Luis Fariñas del Cerro, Andreas Herzig, et Henri Prade. Qualitative relevance and independence: a roadmap. In Didier Dubois, Erich P. Klement, et Henri Prade, editors, *Fuzzy Sets, Logics and Knowledge Bases*, Trends in Logic. Kluwer Academic Publishers, 1998. to appear.

[Dubois et Prade, 1988] Didier Dubois et Henri Prade. *Possibility Theory – An Approach to Computerized Processing of Uncertainty.* Plenum Press, New York, 1988. (with the collaboration of H. Farreny, R. Martin-Clouaire, C. Testemale).

[Eiter et Gottlob, 1992] Thomas Eiter et Georg Gottlob. On the complexity of propositional knowledge base revision, updates, and counterfactuals. *Artificial Intelligence J.*, 57:227–270, 1992.

[Fariñas del Cerro *et al.*, 1992] Luis Fariñas del Cerro, Andreas Herzig, et Jerôme Lang. From expectation-based nonmonotonic reasoning to conditional logics. In Bernd Neumann, editor, *10th Eur. Conf. on Artificial Intelligence (ECAI'92), Vienna*, pages 314–318. John Wiley & Sons, Ltd., 1992. (final version published in the AI Journal 66:375–393).

[Fariñas del Cerro *et al.*, 1994] Luis Fariñas del Cerro, Andreas Herzig, et Jerôme Lang. Ordering-based nonmonotonic reasoning. *Artificial Intelligence Journal (AIJ)*, 66:375–393, April 1994.

[Fariñas del Cerro *et al.*, 1998] Luis Fariñas del Cerro, Andreas Herzig, Dominique Longin, et Omar Rifi. Belief reconstruction in cooperative dialogues. In Fausto Giunchiglia, editor, *Proc. 8th Int. Conf. on Artificial Intelligence: Methods, Systems, Applications (AIMSA'98)*, Lecture Notes in Artificial Intelligence. Springer Verlag, September 1998.

[Fariñas del Cerro et Herzig, 1991] Luis Fariñas del Cerro et Andreas Herzig. A modal analysis of possibility theory (invited paper). In Philippe Jorrand et Josef Kelemen, editors, *Proc. of the Int. Workshop on Foundations of AI Research (FAIR 91)*, number 535 in LNAI, pages 11–18. Springer Verlag, September 1991. (short version published in ECSQAU'91).

[Fariñas del Cerro et Herzig, 1994a] Luis Fariñas del Cerro et Andreas Herzig. A conditional logic for updating in the possible models approach. In Bernhard Nebel et Leonie Dreschler-Fischer, editors, *Proc. 18th German Conf. on Artificial Intelligence (KI'94)*, number 861 in LNAI, pages 237–247. Springer Verlag, September 1994.

[Fariñas del Cerro et Herzig, 1994b] Luis Fariñas del Cerro et Andreas Herzig. Interference logic = conditional logic + frame axiom. *Int. J. of Intelligent Systems*, 9(1):119–130, 1994.

[Fariñas del Cerro et Herzig, 1994c] Luis Fariñas del Cerro et Andreas Herzig. Possibility theory and independence. In Bernadette Bouchon-Meunier, Ronald R. Yager, et Lotfi A. Zadeh, editors, *Proc. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'94)*, pages 512–517, Paris, june 1994.

[Fariñas del Cerro et Herzig, 1995a] Luis Fariñas del Cerro et Andreas Herzig. Changements minimaux constructifs. In Jacques Dubucs et François Lepage, editors, *Méthodes logiques pour les sciences cognitives*, pages 195–204. Hermès, Paris, 1995.

[Fariñas del Cerro et Herzig, 1995b] Luis Fariñas del Cerro et Andreas Herzig. Modal deduction with applications in epistemic and temporal logic. In Dov Gabbay, Chris J. Hogger, et J. A. Robinson, editors, *Handbook of Logic and Artificial Intelligence*, volume 4 - Epistemic and Temporal Reasoning, pages 499–594. Oxford University Press, 1995.

[Fariñas del Cerro et Herzig, 1995c] Luis Fariñas del Cerro et Andreas Herzig. Possibility theory and independence. In Bernadette Bouchon-Meunier, Ronald R. Yager, et Lotfi A. Zadeh, editors, *Advances in Intelligent Computing - IPMU'94, Selected Papers*, number 945 in LNCS, pages 292–301. Springer Verlag, 1995.

[Fariñas del Cerro et Herzig, 1995d] Luis Fariñas del Cerro et Andreas Herzig. Revisions, updates, and interference. In André Fuhrmann et Hans Rott, editors, *Logic, Action, and Information*, pages 189–207. DeGruyter Publishers, Berlin, 1995.

[Fariñas del Cerro et Herzig, 1996a] Luis Fariñas del Cerro et Andreas Herzig. Belief change and dependence. In Yoav Shoham, editor, *Proc. 6th Conf. on Theoretical Aspects of Rationality and Knowledge (TARK'96)*, pages 147–162. Morgan Kaufmann Publishers, 1996.

[Fariñas del Cerro et Herzig, 1996b] Luis Fariñas del Cerro et Andreas Herzig. Combining classical and intuitionistic logic, or: intuitionistic implication as a conditional. In Franz Baader et Klaus U. Schulz, editors, *Frontiers in Combining Systems (Proc. Int. Workshop FroCoS'96)*, number 3 in Applied Logic Series, pages 93–102. Kluwer Academic Publishers, 1996.

[Fariñas del Cerro et Herzig, 1997] Luis Fariñas del Cerro et Andreas Herzig. All other things being equal: on a notion of inertia in conditional logic. In Eva Ejerhed et Sten Lindström, editors, *Logic, Action and Cognition - Selection of papers of the Umeå colloquium on dynamic approaches in Logic, Language and Information (UmLLI-93)*, Trends in Logic, pages 137–148. Kluwer Academic Publishers, 1997.

[Fariñas del Cerro et Lugardon, 1991] Luis Fariñas del Cerro et Valérie Lugardon. Sequents for dependence logics. *Logique et Analyse*, 133-134:57–71, 1991.

[Gärdenfors et Makinson, 1988] Peter Gärdenfors et David Makinson. Revisions of knowledge systems using epistemic entrenchment. In *Proc. 2nd Conf. on Theoretical Aspects of Reasoning about Knowledge*, pages 83–95, 1988.

[Gärdenfors et Makinson, 1994] Peter Gärdenfors et David Makinson. Nonmonotonic inference based on expectation ordering. *Artificial Intelligence J.*, 65:197–245, 1994.

[Gärdenfors, 1978] Peter Gärdenfors. Conditionals and changes of belief. In I. Niiniluoto et R. Tuomela, editors, *The Logic and Epistemology of Scientific Change*, volume 30, pages 381–404. Acta Philosophica Fennica, 1978.

[Gärdenfors, 1988] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, 1988.

[Gärdenfors, 1990] Peter Gärdenfors. Belief revision and irrelevance. *PSA*, 2:349–356, 1990.

[Gasquet et Herzig, 1993] Olivier Gasquet et Andreas Herzig. Translating inaccessible worlds logic into bimodal logic. In Mike Clarke, Rudolf Kruse, et Serafin Moral, editors, *2nd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 93), (Granada, Spain)*, volume 747 of *Lectures Notes in Computer Science*, pages 145–150. Springer Verlag, 1993.

[Gasquet et Herzig, 1994a] Olivier Gasquet et Andreas Herzig. Automated deduction for non-normal modal logics. In Andrews I. Jones et Marek Sergot, editors, *2nd International Workshop on Deontic Logic in Computer Science (DEON'94)*, pages 136–144. TANO, Oslo, January 1994.

[Gasquet et Herzig, 1994b] Olivier Gasquet et Andreas Herzig. La méthode de traduction fonctionnelle pour les logiques modales. *Revue d'Intelligence Artificielle (RIA)*, 8(2):119–144, 1994.

[Gasquet et Herzig, 1995] Olivier Gasquet et Andreas Herzig. Translation-based deduction methods for modal logics. In B. Bouchon-Meunier, R. R. Yager, et L. Zadeh, editors, *Advances in Intelligent Computing - IPMU'94, Selected Papers*, number 945 in LNCS, pages 399–408. Springer Verlag, 1995.

[Gasquet et Herzig, 1996] Olivier Gasquet et Andreas Herzig. From classical to normal logics. In Heinrich Wansing, editor, *Proof Theory of Modal Logics*, number 2 in Applied Logic Series, pages 293–311. Kluwer Academic Publishers, 1996.

[Hanks et McDermott, 1986] S. Hanks et D. McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proc. 5th Conf. on Artificial Intelligence (AAAI'86)*, pages 328–333, 1986.

[Herzig et Rifi, 1998] Andreas Herzig et Omar Rifi. Update operations: a review. In *Proc. Eur. Conf. on Artificial Intelligence (ECAI'98)*, pages 13–17. John Wiley & Sons, Ltd., August 1998.

[Herzig, 1996] Andreas Herzig. The PMA revisited. In Luigia Carlucci Aiello et Stuart Shapiro, editors, *Proc. Int. Conf. on Knowledge Representation and Reasoning (KR'96)*, pages 40–50. Morgan Kaufmann Publishers, November 1996.

[Herzig, 1997] Andreas Herzig. How to change factual beliefs using laws and independence information. In Dov M. Gabbay, Rudolf Kruse, Andreas Nonnengart, et Hans Juergen Ohlbach, editors, *Qualitative and Quantitative Practical Reasoning - First Int. Joint Conf. on Qual. and Quant. Practical Reasoning; ECSQARU-FAPR'97*, number 1244 in LNCS, pages 311–321. Springer Verlag, June 1997.

[Herzig, to appear] Andreas Herzig. Logics for belief base updating. In Didier Dubois, Dov Gabbay, Henri Prade, et Philippe Smets, editors, *Handbook of defeasible reasoning and uncertainty management, volume 'Belief change'*. Kluwer Academic Publishers, to appear.

[Hintikka, 1962] Jaakko K. K. Hintikka. *Knowledge and belief.* Cornell University Press, Ithaca, N.Y., 1962.

[Katsuno et Mendelzon, 1992] Hirofumi Katsuno et Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, editor, *Belief revision*, pages 183–203. Cambridge University Press, 1992. (preliminary version in Allen, J.A., Fikes, R., and Sandewall, E., eds., Principles of Knowledge Representation and Reasoning: Proc. 2nd Int. Conf., pages 387–394. Morgan Kaufmann Publishers, 1991).

[Kraus *et al.*, 1990] Sarit Kraus, Daniel Lehmann, et Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence J.*, 44:167–207, 1990.

[Lehmann et Magidor, 1992] D. Lehmann et M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence J.*, 55:1–60, 1992.

[Lugardon, 1996] V. Lugardon. *Sur les fondements de la notion de dépendance et de son application à la théorie de l'action.* PhD thesis, Institut de recherche en informatique de Toulouse (IRIT), Université Paul Sabatier, 1996.

[Marquis, 1994] Pierre Marquis. Possible models approach via independency. In *Proc. 11th Eur. Conf. on Artificial Intelligence*, pages 336–340, Amsterdam, 1994.

[McCarthy et Hayes, 1969] J. McCarthy et P.J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer et D. Mitchie, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.

[McCarthy, 1990] John McCarthy. *Formalizing common sense: papers by J. McCarthy.* Ablex, Norwood, N.J., 1990.

[Nahmias, 1978] S. Nahmias. Fuzzy variables. *Fuzzy Sets and Systems*, 1(2):97–110, 1978.

[Newell, 1982] A. Newell. The knowledge level. *Artificial Intelligence J.*, 18:87–127, 1982.

[Pearl, 1988] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann Publishers, 1988.

[Pearl, 1990] J. Pearl. System Z : a natural ordering of defaults with tractable applications for default reasoning. In M. Vardi, editor, *Proc. 3rd Conf. on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, San Mateo, 1990. Morgan Kaufman.

[Russell, 1914] Bertrand Russell. *Our knowledge of the external world: as a field for scientific method in philosophy.* The Open Court Publishing Company, London, 1914.

[Russell, 1956] Bertrand Russell. On order in time. In R.C. Marsh, editor, *Russell, Bertrand: Logic and Knowledge (essays 1901-1950).* London, 1956.

[Sandewall, 1993] Erik Sandewall. The range of applicability of nonmonotonic logics for the inertia problem. In *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*, 1993.

[Sandewall, 1994] Erik Sandewall. *Features and Fluents.* Oxford University Press, 1994.

[Segerberg, 1992] Krister Segerberg. Getting started: beginnings in the logic of action. *Studia Logica*, 51:347–378, 1992.

[Sombé, 1994] Léa Sombé. A glance at revision and updating in knowledge bases. *International J. of Intelligent Systems*, 9(1):1–27, January 1994. (P. Besnard, L.Cholvy, M.O. Cordier, D. Dubois, L. Fariñas del Cerro, C. Froidevaux, F. Lévy, Y. Moinard, H. Prade, C. Schwind, P. Siegel).

[Whitehead, 1919] Alfred N. Whitehead. *An enquiry concerning the principles of natural knowledge.* Cambridge University Press, 1919.

[Winslett, 1988] Mary-Anne Winslett. Reasoning about action using a possible models approach. In *Proc. 7th Conf. on Artificial Intelligence (AAAI'88)*, pages 89–93, St. Paul, 1988.

[Wittgenstein, 1927] Ludwig Wittgenstein. *Tractatus logico-philosophicus.* 1927.