Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Résolution d'un paradoxe en logique de croyance et de connaissance

Simon, Laurent

Award date:
1998

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 19. Apr. 2024

Facultés universitaires Notre Dame de la Paix
Institut d'informatique

Résolution d'un paradoxe en logique de croyance et de connaissance

Promoteur : Pierre-Yves SCHOBBENS
Auteur : Laurent SIMON

Mémoire réalisé en vue de l'obtention du grade de maître en informatique
Année académique 1997-1998

Merci à monsieur Pierre-Yves Schobbens sans qui ce travail n'aurait jamais vu le jour et ne se serait jamais terminé. Les nombreuses heures qu'il m'a consacrées ont été aussi agréables qu'indispensables.

Merci à monsieur Eric Gillet pour ses conseils.

Merci à mes parents pour leur soutien matériel et moral.

Merci à Cécile pour ses corrections et sa présence apaisante.

Introduction

Depuis l'arrivée de l'informatique et plus particulièrement des techniques et des méthodes de développement en intelligence artificielle, la logique a pu quitter le cadre strictement théorique auquel elle était cantonnée.
En effet depuis sa naissance au cours de l'Antiquité, cette discipline s'occupait principalement de modéliser la structure du langage et du raisonnement humain afin de mieux les appréhender.
L'intervention de l'informatique a permis d'ajouter tout un pan d'analyse, inconnu jusqu'alors. Les systèmes logiques pouvaient désormais servir à modéliser le comportement d'agents non-humains tels que des robots ou des agents logiciels. La différence essentielle dans ce cas réside dans le fait qu'il ne s'agit plus seulement de modéliser pour comprendre, mais également pour obtenir un comportement défini de la part des agents spécifiés.
Ce travail vise à analyser certaines conséquences étranges d'un système logique développé dans le but de modéliser le comportement *d'agents rationnels* du point de vue de leurs connaissances et de leurs croyances[1]. Le moyen employé pour ce faire fut de combiner différentes logiques modales, la *logique épistémique* (ou logique de la connaissance) et la *logique doxastique* (ou logique de la croyance).
Nous avons d'abord commencé par un panorama historique et thématique concernant la connaissance, la croyance et les nombreux systèmes logiques développés dans le but d'analyser les différentes notions s'y rapportant. Ceci nous a paru important pour fixer le contexte de notre travail et pour donner au lecteur différentes pistes bibliographiques sur les diverses notions que nous n'avons pas développées et qui sont apparentées aux concepts clés de notre analyse.
Puis nous avons développé la première partie de l'étude proprement dite qui consiste en un exposé systématique de la structure développée par Van Linder[2] dans sa thèse.
Le système comprend la définition formelle d'un langage et une interprétation sémantique de ce langage basée sur des modèles de Kripke en terme de *mondes possibles,* ainsi qu'une série d'axiomes de base qui correspondent à la sémantique du système.
Un des théorèmes qui dérive de ce système nous a semblé être paradoxal lorsqu'on le réinterprète en langage naturel. Il nous a donc paru pertinent, dans une seconde partie, d'analyser les prémisses dont proviennent ce paradoxe et de voir comment on pourrait les supprimer ou les modifier afin que le système soit philosophiquement plus justifiable. Puis nous avons tenté de redéfinir le système en supprimant l(es) axiome(s) qui entraînent le paradoxe.
Ceci nous conduit enfin à la partie la plus personnelle du travail, c'est-à-dire la reconstruction d'un système, apparenté au système de Van Linder d'où nous sommes partis.

---

[1] Le système est développé dans : VAN LINDER, B. ; « *Modal logics for rational agents* », Proefschrift Universiteit Utrecht, Faculteit Wiskunde en Informatica, Utrecht, 1996.
[2] *ibid.*

Dans notre nouveau système, nous avons démontré que le paradoxe considéré n'apparaît plus.
Enfin il nous a semblé bon de vérifier toute une série de conséquences assez classiques dans un système de logique de croyance et de connaissance, afin d'expliciter le comportement de ce que nous avons construit.

Chapitre 1 :
Panorama historique et thématique du domaine

Les problèmes de la connaissance et de la croyance ont traversé la culture et le questionnement de l'homme depuis l'aube de l'histoire scientifique.
Les grecs antiques présocratiques, et particulièrement Parménide d'Elée[3], distinguaient deux voies dans le raisonnement humain : la voie de la connaissance vraie (qu'ils appelaient *episteme*) et la voie de la croyance, de l'opinion trompeuse (qu'ils appelaient *doxa*).
Platon et Aristote[4], à leur suite, s'interrogèrent sur les possibilités qu'a l'homme de connaître et sur la façon dont il peut connaître en vérité ou se tromper.
Le questionnement sur la connaissance et la croyance s'est poursuivi durant l'antiquité tardive et tout le moyen âge à travers des auteurs comme Plotin[5], Saint Bonaventure[6] et Saint Thomas d'Aquin[7].
L'époque moderne continua la recherche et tous les plus grands penseurs (Descartes[8], Spinoza[9], Locke[10], Leibniz[11], Hume[12]...) abordèrent la question de la connaissance et de l'erreur.
Kant fit du problème de la connaissance humaine la clef de voûte de son oeuvre[13] et le dix-neuvième siècle poursuivit ou critiqua son analyse.

Cependant c'est seulement à partir de la deuxième moitié du vingtième siècle que l'on assista à cette étape fondamentale : la formalisation des notions de connaissance et de croyance dans un système logique.
C'est le philosophe finlandais Jaakko Hintikka qui fonda véritablement le domaine de la logique de connaissance et de croyance en publiant un livre comprenant la première formalisation de la connaissance et de la croyance[14].

---

[3] PARMENIDE ; « *De la nature* », traduit par J.P.Dumont, in *les présocratiques*, Bibliothèque de la Pléiade, Gallimard, 1988.
[4] ARISTOTE ; « *La métaphysique* », traduit par J. Tricot, Vrin, Paris, 1986.
[5] PLOTIN ; « *Ennéades* », traduit par E. Bréhier, Les Belles Lettres, 1989.
[6] SAINT BONAVENTURE ; « *Itinéraire de l'esprit vers Dieu* », Vrin, Paris, 1924.
[7] SAINT THOMAS D'AQUIN ; « *Somme contre les gentils* », traduit par R. Bernier, M. Corvez, L. J. Moreau, 4 vol., Lethielleux, 1951-1961.
[8] DESCARTES, R. ; « *Oeuvres et lettres* », Bibliothèque de la Pléiade, Gallimard, 1953.
[9] SPINOZA, B. ; « *L'Ethique* », traduit par R. Misrahi, Presses Universitaires de France, 1990.
[10] LOCKE, J. ; « *An essay concerning human understanding* », édité par P. Nidditch, Oxford University Press, 1975.
[11] LEIBNIZ, G.W. ; « *Nouveaux essais sur l'entendement humain* », coll. GF, Flammarion, 1990. Voir particulièrement le livre IV intitulé *de la connaissance*
[12] HUME, D. ; « *Enquête sur l'entendement humain* », traduit par A. Leroy, coll. GF, Flammarion, 1983.
[13] KANT, E. ; « *Critique de la raison pure* », édité par F. Alquié, traduit par A. Delamarre et F. Marty à partir de la traduction de J. Barni, Coll. Folio Essais, Gallimard, Paris, 1980.
[14] HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962.

Cet ouvrage fit date ; Castañeda dit à son propos : « Ce livre contient probablement la plus importante contribution à la technique philosophique depuis l'invention par C.I. Lewis du système d'implication stricte »[15].

Le système développé par Hintikka se basait sur la logique modale[16], la logique de croyance et de connaissance pouvant en effet être considérée comme une instance de cette logique plus générale.

De par ce fait, une deuxième étape fut atteinte lorsque S. Kripke proposa une formalisation standard pour l'interprétation sémantique des systèmes de logique modale[17] ; la logique de croyance (ou logique doxastique) et de connaissance (ou logique épistémique) disposait à présent d'une théorie sémantique standard.

On peut remarquer que, bien que publié avant l'article de Kripke, l'ouvrage de Hintikka proposait pour la logique de connaissance et de croyance une théorie sémantique en terme de mondes possibles très proche de celle du philosophe américain.

A partir de là, les logiques de croyances et de connaissances se développèrent très vite et dans des chemins divers. Il nous a paru intéressant de citer quelques unes de ces directions, sans aucunement prétendre être exhaustif, mais afin de renseigner le lecteur sur les notions connexes ou apparentées à la problématique de ce travail.

Un des premiers apports, la formalisation de la notion de *connaissance commune* est dû à D. Lewis[18]. Il s'agit ici d'une formalisation de la connaissance d'un **groupe** d'agents et non d'un seul agent. On dit qu'un groupe d'agents possède la connaissance commune d'un fait φ si tout le monde connaît φ et si tout le monde sait que tout le monde connaît φ et si tout le monde sait que tout le monde sait que tout le monde connaît φ et ainsi de suite.

La première axiomatisation de la notion de connaissance commune fut apportée par D.J. Lehmann[19].

A l'autre extrémité du spectre de la connaissance d'un groupe, se trouve la notion de *connaissance distribuée* qui fut introduite par Halpern et Moses[20]. Un groupe possède une connaissance distribuée d'un fait φ si, en réunissant toutes les connaissances de chacun des membres du groupe, il peut déduire le fait φ. Par exemple, si Linda sait que Kévin possède une Porsche ou une Ferrari et que Cindy sait que Kévin ne possède pas de Ferrari, alors le groupe composé de Cindy et Linda possède la connaissance distribuée que Kévin possède une Porsche.

Une autre notion nouvelle, importante dans le cadre de ce travail, est la notion de *modalité graduée* (graded modalities) qui fut d'abord analysée par K. Fine[21] puis

---

[15] CASTAÑEDA, H.-N. ; « Review of Knowledge and Belief », *Journal of symbolic logic*, 1964, p.132.

[16] WRIGHT, G.H. von ; « *An essay in modal logic* », North Holland, Amsterdam, 1951.
POPKORN, S. ; « *First steps in modal logic* », Cambridge University Press, Cambridge, 1994.

[17] KRIPKE, S. ; « Semantic analysis of modal logic », *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 1963, pp.67-96.

[18] LEWIS, D. ; « *Convention, A Philosophical Study* », Harvard University Press, Cambridge, 1969.

[19] LEHMANN, D.J. ; « Knowledge, common knowledge and related puzzles », *Proc. 3rd ACM symp. on principles of distributed computing*, 1984, pp. 62-67.

[20] HALPERN, J.Y. ; MOSES, Y. ; « Knowledge and common knowledge in a distributed environment », *Journal of the ACM* 37(3), 1990, pp.549-587.

[21] FINE K. ; « In so many possible worlds », *Notre Dame journal of formal logic* 13(4), 1972, pp.516–520

développée et axiomatisée par J.J.Ch. Meyer et W. van der Hoek[22]. L'apport de cette notion permet de spécifier des connaissances (ou plutôt des croyances) qui sont plus fiables et d'autres qui sont moins fiables. Ceci permet à un agent d'accorder plus ou moins de crédit aux informations dont il dispose selon le type de croyance d'où l'information provient.

Un problème intéressant qui fut traité en profondeur est celui de *l'omniscience logique* qui fut dès le début remarqué par Hintikka[23]. On peut l'expliquer comme suit : dans une théorie sémantique de Kripke, les agents (dont on formalise la connaissance et la croyance) sont supposés être des raisonneurs parfaits, ils sont supposés être logiquement omniscients, c'est à dire qu'ils connaissent toutes les conséquences de leur connaissance, ou encore qu'ils savent déduire tous les théorèmes déductibles à partir de l'ensemble des axiomes de base du système.

Ce présupposé est évidement irréaliste ; pour s'en convaincre, il suffit de réfléchir un bref instant sur la façon dont un être humain raisonne. Un être humain n'est bien évidemment pas logiquement omniscient ; « *une personne peut connaître toutes les règles du jeu d'échecs sans pour autant savoir si ou comment les blancs peuvent gagner à tout coup* »[24]. Plusieurs moyens ont été envisagés pour éviter le problème de l'omniscience logique ; nous n'en citerons que quelques uns : le modèle de déduction de croyance de Konolidge[25]qui s'écarte de la sémantique de Kripke, la notion de croyance explicite de Levesque[26] où un agent ne doit pas nécessairement connaître les conséquences de ses croyances dont il est conscient, Rantala[27]et ses modèles basés sur la théorie des *mondes impossibles* de Kripke[28] ou les mondes impossibles sont des états qui peuvent se comporter de façon 'illogique', Fagin et Halpern[29] et leur notion de *conscience* (awareness) qui leur permet d'indiquer de quelles formules un agent a la disposition pour ainsi pouvoir renvoyer certaines conséquences indésirables dans l'inconnu.

Un problème d'un genre différent est l'introduction du temps et d'opérateurs modaux pour le représenter, et ce afin de pouvoir spécifier la façon dont les croyances changent au cours du temps. Kraus et Lehman[30] ont particulièrement étudié cette question.

A propos du changement des croyances, Alchourron, Gärdenfors et Makinson[31]ont proposé un système standard (l'axiomatisation AGM pour les changements de

---

[22]HOEK, W. van der ; MEYER, J.-J.CH. ; « Graded modalities in epistemic logic », *Logique et analyse* 133-134 (édition spéciale pour le symposium international sur la logique épistémique), 1991, pp.251–270.

[23] HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962.

[24] FAGIN, R. ; HALPERN, J.Y. ; MOSES, Y. ; VARDI, M.Y. ; « *Reasoning about knowledge* », MIT Press, Cambridge, 1995, p.309.

[25] KONOLIDGE, K. ; « *A deduction model of belief* », Pitman / Morgan Kaufmann, London / Los Altos, 1986.

[26] LEVESQUE, H.J. ; « A logic of implicit and explicit belief », *Proceedings of the national conference on artificial intelligence*, 1984, pp.198-202.

[27] RANTALA, V. ; « Impossible world semantics and logical omniscience », *Acta philosophica fennica* 35, 1982, pp.106-115.

[28] KRIPKE, S. ; « Semantic analysis of modal logic II : non-normal modal propositional calculi », in *Symposium on the theory of models*, North-Holland, Amsterdam, 1965.

[29] FAGIN, R. ; HALPERN, J.Y. ; « Belief, awareness and limited reasoning », *Artificial intelligence* 34, 1988, pp.39-76.

[30] KRAUS, S. ; LEHMANN, D. ; « Knowledge, belief and time », *Theoretical Computer Science* 58, 1988, pp.155-174.

[31] ALCHOURRON, C.E. ; GÄRDENFORS, P. ; MAKINSON, D. ; « On the logic of theory change : partial meet contraction and revision functions », *Journal of symbolic logic* 50, 1985, pp.510-530.

croyance) où, lorsqu'à la suite de l'acquisition d'une nouvelle information une contradiction apparaît dans les croyances d'un agent, on élimine la croyance la plus ancienne qui pose problème et priorité est donnée à la croyance la plus récemment acquise.

Le concept d'acquisition d'information nous emmène du côté des systèmes de la *logique dynamique* (ou logique de l'action) qui permettent de modéliser des actions que les agents peuvent entreprendre sur base de leurs connaissances et de leurs croyances et qui peuvent avoir une rétroaction sur ces mêmes connaissances ou croyances. Ce type de système est entre autres développé par B. Moore[32]et par B. van Linder[33].

Nous pouvons également citer les systèmes de *logique autoépistémique* qui s'occupent principalement des connaissances introspectives d'un agent, de la façon dont un agent raisonne à propos de ses propres connaissances ou de son ignorance. De tels systèmes furent étudiés par Moore[34] et Marek et Truszczynski[35].

Pour terminer cette introduction au domaine, il nous reste encore à parler des systèmes de logique floue développés pour les logiques de croyance et de connaissance. Ces systèmes se caractérisent par le fait qu'ils sont basés non pas sur une logique bivalente (une proposition est soit vraie, soit fausse) mais sur une logique multivalente ou les différentes possibilités de valeur sémantique d'une proposition sont distribuées de façon probabiliste. On trouve de tels systèmes chez L. Zadeh[36].

---

[32] MOORE, R.C. ; « A formal theory of Knowledge and action », in *Formal theories of the commonsense world*, édité par J.R. Hobbs et R.C. Moore, Ablex, Norwood, New Jersey, 1985, pp.319–358.

[33] VAN LINDER, B. ; « *Modal logics for rational agents* », Proefschrift Universiteit Utrecht, Faculteit Wiskunde en Informatica, Utrecht, 1996.

[34] MOORE, R.C. ; « Possible-world semantics for autoepistemic logic », *Proceedings of the non-monotonic reasoning workshop*, New Paltz NY, 1984, pp.344-354.

[35] MAREK, W. ; TRUSZCZYNSKI, M. ; « Autoepistemic logic », *Journal of the ACM* 38(3), 1991, pp.588-619.

[36] ZADEH, L. ; « Knowledge representation in fuzzy logic », *Tkde* 1, 1989, pp.89-100.

Chapitre 2 :
Le système de Van Linder

Le langage multi-modal formant le noyau de la formalisation du système comprend des opérateurs modaux pour représenter la connaissance ainsi que les différents niveaux de croyance d'un agent.
Remarquons que le système développé dans son intégralité est plus complexe que la partie que nous allons analyser ici, mais nous avons pris la décision de simplifier tout ce qui ne concerne pas directement le paradoxe.
Notons que Van Linder développe un système à agents multiples, et ce afin de prendre en compte les interactions possibles entre différents agents. Nous ne tiendrons pas compte de cette possibilité dans notre analyse, car elle n'a aucune incidence sur les conséquences que nous voulons étudier.
De même, nous avons décider d'éliminer du système, tout ce qui concerne la *logique dynamique* (ou logique de l'action). Nous nous focaliserons désormais sur les parties purement épistémiques et doxastiques du système.
Pour suivre la manière de représenter de Hintikka[37], fondateur du domaine, on utilise l'opérateur **K** pour représenter la connaissance d'un agent. $\mathbf{K}\varphi$ représente le fait qu'un agent *sait* que la proposition φ est valide. La caractéristique principale de la connaissance dans ce système est le fait qu'elle soit *véridique*, c'est à dire que si un agent *sait* quelque chose, ce quelque chose est vrai.
En ce qui concerne la représentation de la croyance d'un agent, Van Linder construit un système assez étoffé où il distingue différents niveaux de croyance pour un agent, auxquels sont associés différents niveaux de crédibilité.
Les différents niveaux de croyance relevés sont :

- Les *croyances observationelles* d'un agent, ce sont les plus crédibles. Ce sont les croyances qu'un agent s'est forgé sur base de ses observations.
- Les *croyances communicationnelles* d'un agent, elles sont moins crédibles que les croyances observationnelles. Ce sont les croyances qu'un agent a acquis sur base de communications avec d'autres agents.
- Les *croyances par défaut* d'un agent, ce sont les moins crédibles de toutes. Ce sont les croyances qu'un agent se forme, sans aucune information extérieure, par préjugé.

Remarquons que la connaissance peut, elle aussi, se définir en terme de crédibilité. La connaissance est la croyance *vraie*. Elle possède le plus haut niveau de crédibilité.

Pour représenter l'ordre de crédibilité de l'information dont dispose un agent, Van Linder propose de structurer cette information dans quatre ensembles situés les uns dans les autres.

---

[37] HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962.

L'ensemble le plus intérieur de cette structure contient la *connaissance* de l'agent (celle-ci est l'information la plus crédible dont il dispose). L'ensemble directement supérieur qui inclut la connaissance d'un agent est l'ensemble qui contient les croyances observationnelles de l'agent. L'ensemble directement supérieur est l'ensemble qui contient les croyances communicationnelles de l'agent. Enfin, l'ensemble le plus extérieur, celui qui comprend tous les autres, est l'ensemble qui contient les croyances par défaut de l'agent.
La figure 1 est une représentation de cette imbrication d'ensembles.

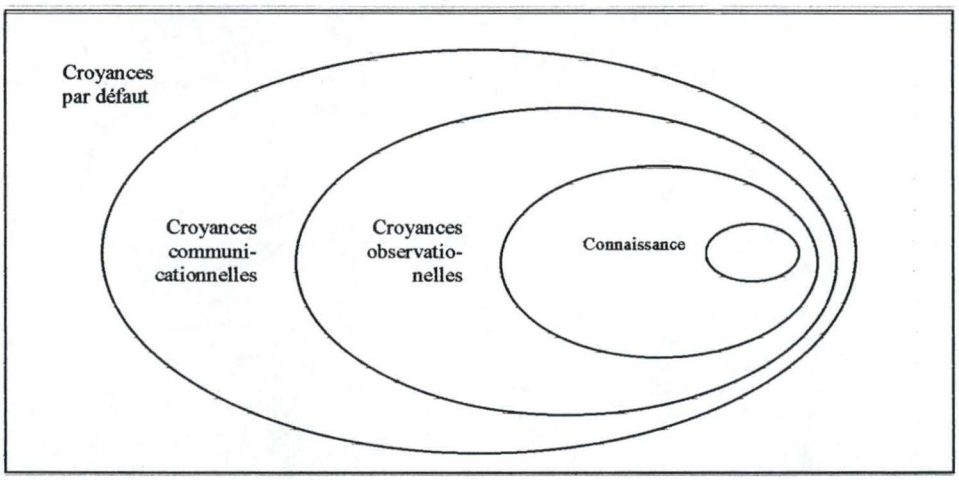

Figure 1 : Ordonnancement des formules selon le niveau de crédibilité

Nous pouvons donc, en poursuivant dans l'esprit de Hinitkka, utiliser des opérateurs $\mathbf{B}^x$ pour représenter les différentes notions de croyance que nous avons formalisées.
$\mathbf{B}^o\varphi$ indique que l'agent *croit* de manière *observationnelle* (sur base des observations qu'il a faites) que la proposition φ est valide, notons qu'une croyance observationnelle est considéré comme étant une croyance vraie (quasiment au même titre que la connaissance). $\mathbf{B}^c\varphi$ indique que l'agent *croit* de manière *communicationnelle* (sur base d'informations qui lui ont été communiquées de l'extérieur) que la proposition φ est valide. $\mathbf{B}^d\varphi$ indique que l'agent *croit par défaut* que la proposition φ est valide.
Nous pouvons donc maintenant, à l'aide de l'opérateur épistémique et des différents opérateurs doxastiques, nous appliquer à définir le langage, noyau du système.

Définition du langage

DEFINITION : Le langage $\mathbf{L}(\overline{\Pi})$ est fondé sur l'ensemble $\overline{\Pi}$ des variables propositionelles. L'alphabet contient les connecteurs $\neg$ (non) et $\wedge$ (et), l'opérateur épistémique **K** et les opérateurs doxastiques $\mathbf{B}^o$, $\mathbf{B}^c$ et $\mathbf{B}^d$.

DEFINITION : Le langage $\mathbf{L}(\overline{\Pi})$ est le plus petit ensemble qui contient $\overline{\Pi}$ tel que

- Si $\varphi \in \mathbf{L}(\overline{\Pi})$ et $\psi \in \mathbf{L}(\overline{\Pi})$ alors $\neg\varphi \in \mathbf{L}(\overline{\Pi})$ et $\varphi \wedge \psi \in \mathbf{L}(\overline{\Pi})$
- Si $\varphi \in \mathbf{L}(\overline{\Pi})$ alors $\mathbf{K}\varphi \in \mathbf{L}(\overline{\Pi})$

- Si $\varphi \in \mathbf{L}(\Pi)$ alors $\mathbf{B}^{o}\varphi \in \mathbf{L}(\Pi)$, $\mathbf{B}^{c}\varphi \in \mathbf{L}(\Pi)$, $\mathbf{B}^{d}\varphi \in \mathbf{L}(\Pi)$

Des constructions additionnelles sont introduites par des abréviations définitionelles :

$\varphi \vee \psi \quad = \quad \neg(\neg\varphi \wedge \psi)$
$\varphi \rightarrow \psi \quad = \quad \neg\varphi \vee \psi$
$\varphi \leftrightarrow \psi \quad = \quad (\varphi \rightarrow \psi) \wedge (\varphi \rightarrow \psi)$
$\top \quad = \quad p \vee \neg p$ pour $p \in \Pi$ arbitraire
$\bot \quad = \quad \neg\top$

Interprétation sémantique en modèles de Kripke

La grande majorité des interprétations proposées pour les langages modaux est basée sur une théorie des modèles dans laquelle toutes les conclusions qui sont dérivables du système logique sont valides.
Cette théorie des modèles est la sémantique des *mondes possibles*, formalisée par Saul Kripke en 1963[38].
Pour se faire une idée de la sémantique des mondes possibles, introduisons-la par un exemple[39] :

Considérons un étudiant dans un auditoire ne possédant pas de fenêtres donnant sur l'extérieur. Il se demande s'il pleut à l'extérieur. Représentons par r la proposition *il pleut à l'extérieur*. Comme il ne peut pas voir dehors, il considère deux situations, deux *mondes* possibles. L'un dans lequel r est valide, l'autre dans lequel $\neg r$ est valide. Il ne **sait** donc pas si r est valide. D'un autre côté, dans les deux situations, il est valide que 2 + 2 = 4. En supposant que ces deux situations soient les seules considérées possibles, il **sait** que 2 + 2 = 4.

On peut donc discerner une interprétation de la connaissance en termes de mondes possibles ; en général, la propriété de **connaître** une proposition φ est modélisé par le fait que cette proposition est valide dans tous les mondes possibles (c'est-à-dire les mondes qui sont considérés comme étant possibles sur base de la connaissance d'un agent).
En règle générale, ce type d'interprétation, ces modèles, consistent en : un ensemble non-vide de mondes possibles, une valuation sur les variables propositionelles indiquant la valeur de vérité des propositions atomiques dans les mondes possibles et une série de relations d'accessibilité entre ces mondes.
L'interprétation de ces relations d'accessibilité dépend du champ d'application du système logique. Dans le cas qui nous concerne, les relations dénotent l'accessibilité épistémique ou doxastique.
Donc les modèles utilisés pour interpréter les formules du langage $\mathbf{L}(\Pi)$ contiennent un ensemble S de mondes possibles, représentant des états existants ou hypothétiques, une valuation π sur les éléments de Π, indiquant quelles propositions atomiques sont

[38] KRIPKE, S. ; « Semantic analysis of modal logic », *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 1963, pp.67-96.
[39] L'exemple est tiré de : MEYER, J.-J.CH. ;VAN DER HOEK, W. ; « *Epistemic logic for AI and computer science* », Cambridge University Press, New York, 1995, pp.3-4.

vraies dans quels mondes possibles, et les relations R, B^o, B^c, B^d qui dénotent l'accessibilité épistémique et les différentes accessibilités doxastiques.
Étant donné les différents niveaux de crédibilité attribués aux croyances et les différences d'interprétation entre la croyance et la connaissance, il nous a semblé intéressant de définir d'abord le modèle d'interprétation en ne prenant en compte que les formules épistémiques, puis d'étendre la définition aux formules doxastiques.

Définition du modèle d'interprétation pour les formules épistémiques

DEFINITION : Un modèle M pour $\mathbf{L}(\Pi)$ est un n-uplet contenant au moins les éléments suivants :

- Un ensemble S non-vide de mondes possibles (ou états).
- Une valuation $\pi : \Pi \times S \rightarrow \{0,1\}$ sur les variables propositionelles.
- Une relation R d'accessibilité épistémique entre deux mondes : $R \subseteq S \times S$. Cette relation doit être une relation d'équivalence.

Nous indiquerons par Ω la classe contenant tous les modèles pour $\mathbf{L}(\Pi)$. La lettre M représente un modèle arbitraire et les lettres *s*, *s'*, *s''* sont utilisées pour représenter des éléments arbitraires de l'ensemble des mondes possibles.

La relation R indique quelles paires de mondes sont indifférenciables pour un agent **sur base de sa connaissance**. Si $(s, s') \in R$, alors si *s* est la description du monde actuel, *s'* pourrait l'être également en considérant ce que l'agent connaît.
R doit être une relation d'équivalence, donc il faut que R soit réflexive ($(s, s) \in R$) et Euclidienne (si $(s, s') \in R$ et $(s, s'') \in R$ alors $(s', s'') \in R$). Notons qu'une relation Euclidienne est automatiquement symétrique et transitive[40]. Nous reviendrons plus tard sur le caractère Euclidien et plus particulièrement symétrique de la relation.
Les formules du langage $\mathbf{L}(\Pi)$ sont interprétées sur les mondes possibles dans les modèles de Ω. Les variables propositionelles sont interprétées directement en utilisant la valuation. Une variable *p* est vraie dans un monde *s* si et seulement si $\pi(p, s)$ donne la valeur 1 (vrai). Les négations et les conjonctions sont interprétées de la même manière qu'en logique classique : une formule $\neg\varphi$ est vraie dans un monde *s* si et seulement si φ n'est pas vraie dans *s*, une formule $\varphi \wedge \psi$ est vraie dans un monde *s* si et seulement si φ est vraie dans *s* et ψ est vraie dans *s*. Les formules de connaissance $\mathbf{K}\varphi$ sont interprétées en utilisant la relation R d'accessibilité épistémique : un agent **sait** que φ dans *s* si et seulement si φ est vraie dans tous les mondes possibles que l'agent considère comme compatibles de façon épistémique (c'est-à-dire sur base de sa connaissance) avec *s*.
Nous pouvons donc maintenant clarifier les choses en donnant une définition formelle des relations entre les formules du langage et les modèles.

DEFINITION : La relation binaire $\models$ (validité) entre une formule et une paire M, *s* composée d'un modèle M et d'un monde *s* dans M est définie récursivement de façon suivante :

[40] La façon classique de définir une relation d'équivalence est de dire que c'est une relation réflexive, symétrique et transitive

$M, s \models p \quad \Leftrightarrow \pi(p, s) = 1$ pour $p \in \Pi$
$M, s \models \neg\varphi \quad \Leftrightarrow$ non$(M, s \models \varphi)$
$M, s \models \varphi \wedge \psi \quad \Leftrightarrow M, s \models \varphi$ et $M, s \models \psi$
$M, s \models \mathbf{K}\varphi \quad \Leftrightarrow \forall s' \in S((s, s') \in R \Rightarrow M, s' \models \varphi)$

Pour un modèle donné M, nous pouvons également définir l'ensemble des alternatives épistémiques d'un monde : $[s]_R = \{s' \in S \mid (s, s') \in R\}$.

Extension aux formules doxastiques

Pour l'interprétation des opérateurs doxastiques en terme de mondes possibles, Van Linder utilise une sémantique basée sur ce qu'il appelle des *clusters de croyance*[41].
Les clusters sont des ensembles de mondes situés les uns dans les autres (de façon inverse aux ensembles de formules) qui représentent chacun un ensemble de croyances d'un certain niveau de crédibilité. On parlera donc de cluster de connaissance, de cluster observationnel, de cluster communicationnel et de cluster de défaut.
Pour mieux comprendre la notion de cluster, on peut se représenter la hiérarchie qui les ordonne entre eux comme étant duale de celle qui ordonne les ensembles de formules. Le plus grand des clusters, celui qui contient les autres, est le cluster de connaissance, or l'ensemble des formules de connaissance est le plus petit. La relation duale entre les deux s'explique comme suit : moins un agent connaît de choses, plus l'ensemble des mondes qu'il ne saura pas différencier les uns des autres sur la base de sa seule connaissance sera grand. Et c'est exactement ce qu'est le cluster de connaissance : l'ensemble des mondes que l'agent ne sait pas différencier sur base de sa connaissance.
Plus généralement, on peut dire qu'une formule est crue avec un certain degré de crédibilité si et seulement si elle est valide dans tous les mondes possibles du cluster associé.
La figure 2 permet de se donner une première idée de l'imbrication des clusters. On peut ainsi la mettre en rapport avec la figure 1.

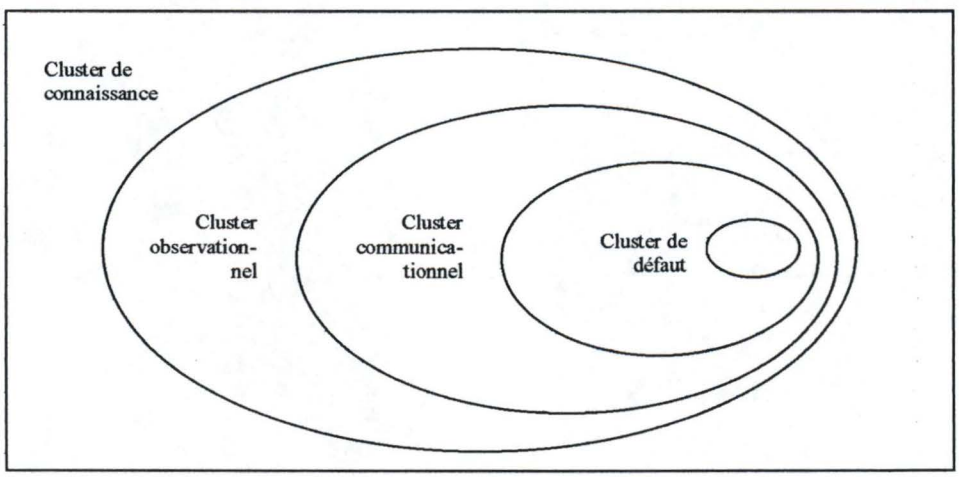

Figure 2 : Inclusion des clusters de croyances

[41] Le terme utilisé est « belief clusters »

Nous avons évoqué l'imbrication des clusters dont nous avons fait une ébauche à la figure 2 où le cluster de défaut est sous-ensemble du cluster communicationnel, qui est sous-ensemble du cluster observationnel, qui est lui même sous-ensemble du cluster de connaissance.

Le système de Van Linder est un peu plus complexe que cela. Il dit en effet qu'un cluster de connaissance peut contenir plusieurs clusters observationnels. On peut expliquer cela par le fait qu'on ne sait pas, dans l'ensemble des mondes, quel est le monde **actuel** (réel), et donc, dans des mondes différents, des observations différentes peuvent être faites, ce qui implique la nécessité d'avoir plusieurs clusters observationnels. Ainsi il se pourrait, par exemple, que dans un monde possible un agent fasse l'observation que la porte qui se trouve devant lui est fermée, alors que dans un autre monde possible, l'agent fasse l'observation que la porte est ouverte. Etant donné que les deux mondes possibles font partie du même cluster de connaissance, il convient d'approfondir un peu la différence entre les formules qu'un agent sait (formules de connaissance) des formules qu'un agent croit observationnellement. Nous avons dit plus haut que les formules de croyance observationnelle pouvaient être considérées au même titre que les formules de connaissance. Nous pouvons cependant, à la suite de Van Linder[42], penser la différence entre elles deux en se référant aux deux types de connaissance définis par Kant : « *C'est donc une question qui a encore besoin d'une recherche plus poussée ... que celle de savoir si il y a une telle connaissance indépendante de l'expérience et même de toutes les impressions des sens. On nomme* ***a priori*** *de telles connaissances, et on les distingue des connaissances empiriques, qui ont leur source* ***a posteriori****, c'est-à-dire dans l'expérience. ... Nous entendrons donc en ce qui suit par connaissances* ***a priori****, non celles qui ont lieu indépendamment de telle ou telle expérience, mais celles qui sont absolument indépendantes de toute expérience. Leur sont opposées les connaissances empiriques, ou celles qui ne sont possibles* ***qu'a posteriori****, c'est à dire par expérience* »[43].

Nous pouvons donc maintenant considérer les formules de connaissance (qui sont les plus certaines) comme de la connaissance de type *a priori.*

A l'opposé, les formules de croyance observationnelle peuvent être considérées comme des connaissances a posteriori au sens kantien du terme.

Ceci précisé, nous pouvons continuer l'analyse de l'interprétation sémantique du système.

A la différence du cluster de connaissance, chaque cluster observationnel ne contient qu'un seul cluster communicationnel qui ne contient lui-même qu'un seul cluster de défaut. L'unicité du cluster de défaut peut s'expliquer comme suit : les croyances par défaut sont des croyances qu'un agent acquiert de façon arbitraire. Ce sont des « préjugés » qu'il assume sans aucune référence objective à la réalité. On peut donc s'imaginer qu'un agent « complète » ses croyances plus « objectives » en adoptant des croyances par défaut. On peut alors considérer qu'un agent, pour toute situation donnée, complètera toujours ses croyances de la même manière, en adoptant les mêmes croyances par défaut.

Nous pouvons justifier l'unicité du cluster communicationnel en remarquant que toute communication doit nécessairement passer par une modification de la réalité physique

---

[42] VAN LINDER, B. ; « *Modal logics for rational agents* », Proefschrift Universiteit Utrecht, Faculteit Wiskunde en Informatica, Utrecht, 1996, p.109.

[43] KANT, E. ; « *Critique de la raison pure* », édité par F. Alquié, traduit par A. Delamarre et F. Marty à partir de la traduction de J. Barni, Coll. Folio Essais, Gallimard, Paris, 1980, p.64.

observable par l'agent (présence d'ondes sonores pour la parole, d'ondes lumineuses pour une communication écrite...).
Par conséquent, la communication est entièrement déterminée par l'observation et supposer une communication différente (un autre cluster communicationnel) oblige à supposer des observations différentes (un autre cluster observationnel). L'unicité du cluster communicationnel dans un cluster observationnel est donc bien justifiée.

Nous pouvons dès lors compléter la définition du modèle d'interprétation en l'étendant aux formules doxastiques.

DEFINITION : Un modèle M pour **L**(Π) est un n-uplet contenant au moins les éléments définis plus haut et les éléments suivants :

- Une relation B^o d'accessibilité doxastique observationnelle entre deux mondes : $B^o \subseteq S \times S$. Cette relation doit être une relation d'équivalence.
- Une fonction B^c : $S \rightarrow \wp(S)$[44] qui donne les alternatives doxastiques communicationnelles de l'agent dans un monde.
- Une fonction B^d : $S \rightarrow \wp(S)$ qui donne les alternatives doxastiques par défaut de l'agent dans un monde.

Les fonctions B^c et B^d sont telles que $\forall\, s, s' \in S$:

- $B^d(s) \neq 0$
- $B^d(s) \subseteq B^c(s) \subseteq [s]_{B^o}$[45] $\subseteq [s]_R$
- Si $s' \in [s]_{B^o}$ alors $B^c(s') = B^c(s)$ et $B^d(s') = B^d(s)$

DEFINITION : La relation binaire $\models$ entre une formule du langage et une paire M, s composée d'un modèle M et d'un monde s dans M est définie récursivement de façon suivante pour les formules doxastiques:

$$M, s \models \mathbf{B}^o\varphi \quad \Leftrightarrow \forall s' \in S\,((s, s') \in B^o \quad \Rightarrow M, s' \models \varphi)$$
$$M, s \models \mathbf{B}^c\varphi \quad \Leftrightarrow \forall s' \in S\,(s' \in B^c(s) \quad \Rightarrow M, s' \models \varphi)$$
$$M, s \models \mathbf{B}^d\varphi \quad \Leftrightarrow \forall s' \in S\,(s' \in B^d(s) \quad \Rightarrow M, s' \models \varphi)$$

Axiomes du système

Nous disposons maintenant d'une définition formelle du langage ainsi que d'une théorie sémantique. Nous pouvons donc passer en revue les axiomes de base du système :

---

[44] Nous utiliserons désormais le symbole « $\wp$ » pour indiquer l'ensemble des parties d'un ensemble donné.
$\wp(S)$ désigne donc l'ensemble des parties de l'ensemble S

[45] $[s]_{B^o}$ se définit de manière analogue à $[s]_R$

$\forall \varphi, \psi \in \mathbf{L}(\Pi)$, le système possède les axiomes suivants :

1. $\models \mathbf{K}(\varphi \rightarrow \psi) \rightarrow (\mathbf{K}\varphi \rightarrow \mathbf{K}\psi)$ (K)
2. $\models \varphi \Rightarrow \models \mathbf{K}\varphi$ (N)
3. $\models \mathbf{K}\varphi \rightarrow \varphi$ (T)
4. $\models \mathbf{K}\varphi \rightarrow \mathbf{K}\mathbf{K}\varphi$ (4)
5. $\models \neg\mathbf{K}\varphi \rightarrow \mathbf{K}\neg\mathbf{K}\varphi$ (5)

Posons $\mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$. $\forall \varphi, \psi \in \mathbf{L}(\Pi)$, le système possède également les axiomes suivants :

6. $\models \mathbf{X}(\varphi \rightarrow \psi) \rightarrow (\mathbf{X}\varphi \rightarrow \mathbf{X}\psi)$ (K)
7. $\models \varphi \Rightarrow \models \mathbf{X}\varphi$ (N)
8. $\models \mathbf{B}^o \varphi \rightarrow \varphi$ (T)
9. $\models \mathbf{X}\varphi \rightarrow \mathbf{X}\mathbf{X}\varphi$ (4)
10. $\models \neg\mathbf{X}\varphi \rightarrow \mathbf{X}\neg\mathbf{X}\varphi$ (5)
11. $\models \neg(\mathbf{X}\varphi \wedge \mathbf{X}\neg\varphi)$ (D)

Il y a un dernier axiome qui est nécessaire pour formaliser la hiérarchie des clusters de croyance :

12. On définit la relation d'ordonnancement $>$ sur les opérateurs par : $\mathbf{K} > \mathbf{B}^o > \mathbf{B}^c > \mathbf{B}^d$; définissons $\geq$ comme étant la fermeture réflexive et transitive de $>$. Alors, $\forall \mathbf{X}, \mathbf{Y} \in \{ \mathbf{K}, \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$, $\forall \varphi \in \mathbf{L}(\Pi)$, si $\mathbf{X} \geq \mathbf{Y}$ alors $\models \mathbf{X}\varphi \rightarrow \mathbf{Y}\varphi$

Les axiomes 1 et 6 (axiomes-K) et les axiomes 2 et 7 (règle de nécessitation) formalisent le fait que les opérateurs épistémiques et doxastiques sont des opérateurs modaux normaux. Les axiomes 3 et 8 (axiomes de véridicité) assurent que la connaissance et les croyances observationelles soient des croyances vraies.
Les axiomes 4 et 9 indiquent que les opérateurs satisfont à *l'introspection positive*, les axiomes 5 et 10 indiquent que les opérateur satisfont à *l'introspection négative*, l'axiome 11 indique que les croyances d'un agent sont cohérentes.
Finalement, remarquons que l'implication $\models \mathbf{B}^c\varphi \rightarrow \mathbf{B}^d\varphi$ qui découle de l'axiome 12 peut quelque fois sembler contraire à l'intuition, en effet, il est très possible que dans ses croyances par défaut, un agent (particulièrement un agent humain) adopte des préjugés forts qui ne puissent pas être invalidés par une communication. Mais c'est loin d'être toujours le cas et dans le cadre de ce travail, nous pouvons accepter complètement l'axiome 12.
Les raisons pour lesquelles on utilise ces noms pour décrire ces axiomes sont purement historiques[46].

A partir de ces axiomes, Van Linder déduit quelques théorèmes supplémentaires que nous allons exposer. Nous nous intéresserons ensuite plus particulièrement à un de ces théorèmes.

[46] Ils sont tirés de la classification de Chellas in : CHELLAS, B.F. ; « *Modal logic. An introduction* », Cambridge University Press, Cambridge, 1980

Quelques théorèmes supplémentaires

Avant de pouvoir exposer les théorèmes supplémentaires du système, il est indispensable de définir auparavant la notion de suites doxastiques[47].

DEFINITION : une formule χ est une suite doxastique s'il existe un $\varphi \in \mathbf{L}(\Pi)$ et des opérateurs $\mathbf{X}_1, ..., \mathbf{X}_m \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d, \neg\mathbf{B}^o, \neg\mathbf{B}^c, \neg\mathbf{B}^d \}$ avec $m > 0$, tels que $\chi = \mathbf{X}_1...\mathbf{X}_m\varphi$.

Nous pouvons donc maintenant exposer les nouveaux théorèmes :

$\forall \chi$[48] $\in \mathbf{L}(\Pi), \forall \varphi \in \mathbf{L}(\Pi), \forall \mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$:

1. $\models \mathbf{X}\chi \leftrightarrow \chi$
2. $\models \boldsymbol{XK\varphi \leftrightarrow K\varphi}$ *! !!*
3. $\models \mathbf{X}\neg\mathbf{K}\varphi \leftrightarrow \neg\mathbf{K}\varphi$
4. $\mathbf{X}\varphi \rightarrow \mathbf{KX}\varphi$ n'est pas valide pour tout $\varphi \in \mathbf{L}(\Pi)$
5. $\neg\mathbf{X}\varphi \rightarrow \mathbf{K}\neg\mathbf{X}\varphi$ n'est pas valide pour tout $\varphi \in \mathbf{L}(\Pi)$

Le théorème 2 est celui qui nous a posé problème ; c'est celui dont l'analyse est au cœur de ce travail. Il semble bien, en effet, que son énoncé soit intuitivement **paradoxal**.
Nous allons l'examiner de plus près.

[47] Cette notion, définie dans : VAN DER HOEK, W. ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, pp.173-195, est reprise par Van Linder qui l'étend pour prendre en compte les éléments de son système.

[48] χ est une suite doxastique.

Chapitre 3 : Analyse du paradoxe

Dans cette section, nous allons tenter de comprendre ce qui a pu entraîner le paradoxe dans le système, quels sont les axiomes qui pourraient éventuellement causer son apparition . Nous essayerons également de prouver sa validité formelle par rapport au système. Mais tout d'abord, nous allons énoncer le paradoxe en langage naturel et montrer en quoi il nous a semblé contre-intuitif.

Énoncé du paradoxe

Reprenons le théorème en cause et essayons de le formuler en langage naturel :

$\forall \varphi \in \mathbf{L}(\Pi), \forall \mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$:

$\models \mathbf{XK}\varphi \leftrightarrow \mathbf{K}\varphi$

Si on pose $X = \mathbf{B}^d$, et qu'on considère l'implication de gauche à droite ($\rightarrow$), le théorème s'énonce comme suit : Si un agent croit par défaut qu'il connaît φ, alors il connaît φ. Il semble en effet parfaitement contre-intuitif que lorsqu'un agent **croit** connaître quelque chose, cela implique qu'il connaisse ce quelque chose. Particulièrement lorsqu'il s'agit d'une croyance par défaut, la moins fiable de toutes les croyances.
A ce propos, il est utile de remarquer que le théorème en question ne pose pas de problème sous la forme où l'on pose $X = \mathbf{B}^o$.
L'implication de gauche à droite s'énonce en effet alors sous la forme suivante :

$\models \mathbf{B}^o \mathbf{K}\varphi \rightarrow \mathbf{K}\varphi$

Ce qui se justifie étant donné la véracité des connaissances observationnelles.
Le paradoxe ne survient donc que lorsqu'on pose dans le théorème $X = \mathbf{B}^c$ ou $X = \mathbf{B}^d$.
Nous allons ainsi focaliser notre attention sur ces cas spécifiques.
Avant d'analyser plus en détail les causes possibles de l'apparition de ce paradoxe, il nous a paru pertinent de montrer que, d'un point de vue purement formel, ce théorème est juste et démontrable dans le système de Van Linder.

Preuve du paradoxe

Nous nous contenterons de la preuve de l'implication de gauche à droite ($\rightarrow$). C'est en effet dans ce sens que le théorème pose problème.

Preuve syntaxique :

$\forall \varphi \in \mathbf{L}(\Pi), \forall \mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$:

$\mathbf{XK}\varphi \rightarrow \neg\mathbf{X}\neg\mathbf{K}\varphi$	(par l'axiome 11 : D)
$\neg\mathbf{X}\neg\mathbf{K}\varphi \rightarrow \neg\mathbf{K}\neg\mathbf{K}\varphi$	(par la contraposée de l'axiome 12)
$\neg\mathbf{K}\neg\mathbf{K}\varphi \rightarrow \mathbf{K}\varphi$	(par la contraposée de l'axiome 5 : introspection nég.)

on a donc $\mathbf{XK}\varphi \rightarrow \mathbf{K}\varphi$.
C.Q.F.D.

Causes du paradoxe

Il semble bien que l'on puisse assez facilement remonter à la source du paradoxe. Ce qui permet d'affirmer le fameux deuxième théorème (et de le prouver), c'est l'emploi des axiomes d'introspection négative.
Il est entendu dans la littérature spécialisée que l'introduction dans un système de l'axiome d'introspection négative - qui veut que si un agent ne connaît pas φ, il sache qu'il ne connaisse pas φ - ne soit pas directement justifié intuitivement ou philosophiquement : « *...d'un autre côté, l'axiome d'introspection négative est très controversé d'un point de vue philosophique. Si un agent ignore la valeur de vérité d'une assertion, il est très improbable, en particulier pour les agents humains, qu'il connaisse son ignorance.* »[49]. L'axiome d'introspection négative est introduit en fonction des améliorations techniques qu'il apporte au système : « *...Cependant, le système **S5**[50] possède des propriétés techniques plus utiles que les autres systèmes ... Pour ces raisons, **S5** est de loin le système logique pour la connaissance le plus populaire parmi les informaticiens et les chercheurs en intelligence artificielle...* »[51], « *... pour beaucoup de nos applications, les axiomes de **S5** semblent les plus appropriés, bien que les philosophes aient protesté de façon virulente contre eux (et plus particulièrement contre l'axiome d'introspection négative)...* »[52].
Du point de vue de la relation d'accessibilité épistémique, ce dont découle l'axiome d'introspection négative, c'est la transitivité et la **symétrie** de la relation.
On peut d'ailleurs faire la preuve sémantique de cet axiome[53] :

Van Linder définit sa relation R comme une relation d'équivalence ; il faut donc bien qu'elle soit réflexive, symétrique et transitive

Il faut prouver que : $\models \neg\mathbf{K}\varphi \rightarrow \mathbf{K}\neg\mathbf{K}\varphi$.

Supposons M, $s \models \neg\mathbf{K}\varphi$ avec $s \in$ S.

---

[49] MEYER, J.-J.CH. ;VAN DER HOEK, W. ; « *Epistemic logic for AI and computer science* », Cambridge University Press, New York, 1995, p.23.
[50] C'est le système qui possède l'axiome d'introspection négative
[51] MEYER, J.-J.CH. ;VAN DER HOEK, W. ; « *Epistemic logic for AI and computer science* », Cambridge University Press, New York, 1995, p.23.
[52] FAGIN, R. ; HALPERN, J.Y. ; MOSES, Y. ; VARDI, M.Y. ; « *Reasoning about knowledge* », MIT Press, Cambridge, 1995, p.56.
[53] cette preuve est tirée de : FAGIN, R. ; HALPERN, J.Y. ; MOSES, Y. ; VARDI, M.Y. ; « *Reasoning about knowledge* », MIT Press, Cambridge, 1995, p.33.

Alors, pour s' tel que $(s, s') \in R$, on a M, $s' \models \neg \varphi$.
Pour s'' tel que $(s, s'') \in R$, comme la relation est symétrique, on a $(s'', s) \in R$.
Comme la relation est transitive on a également $(s'', s') \in R$.
D'où il suit (par la définition de l'opérateur **K**) que M, $s'' \models \neg \mathbf{K}\varphi$.
Comme cela est vrai pour tout s'' tel que $(s, s'') \in R$,
on obtient M, $s \models \neg \mathbf{K}\varphi \rightarrow \mathbf{K}\neg \mathbf{K}\varphi$.
C.Q.F.D.

Il est donc raisonnable de penser, comme le fait la littérature spécialisée que : « ...*la validité de l'axiome d'introspection négative découle du fait que la relation est symétrique*... »[54].
Hintikka lui-même considère que la relation d'accessibilité épistémique[55] est bien réflexive et transitive : « ...*la relation est réflexive, ... ,de la même façon, la relation apparaît comme transitive*... »[56], par contre il ne pense pas qu'elle soit symétrique : « ... *on peut voir que la relation n'est pas symétrique*... »[57]
Il explique ce fait comme suit[58] : « ... *rappelons nous qu'un modèle μ_2 est une alternative épistémique d'un modèle μ_1 si et seulement si, intuitivement parlant, il n'y a rien dans le monde décrit par μ_2 qui soit incompatible avec ce qu'un agent connaît dans le monde décrit par μ_1. Maintenant, il n'est évidemment pas exclu par ce que je connais maintenant que je devrais connaître plus que je ne connais. Mais cette connaissance supplémentaire pourrait très bien être incompatible avec ce qui est encore possible maintenant, pour autant que je le sache.* »[59]. Hintikka refuse donc la symétrie de la relation à cause du fait que, si un monde possible est une alternative épistémique du monde actuel, il se pourrait très bien que dans ce monde possible un agent ait une connaissance supplémentaire qui rende ce monde incompatible avec le monde actuel. La relation ne serait donc pas symétrique.
En fait, la symétrie de la relation dépend de la manière dont on définit le sens de celle-ci.
Si on considère, comme le fait Van Linder[60], que la relation d'accessibilité épistémique indique les paires de mondes indifférenciables pour un agent sur base de sa connaissance, alors bien évidemment, la relation est symétrique et on ne voit pas très bien comment on pourrait connaître plus dans un monde que dans l'autre puisqu'ils sont sensés être pareils du point de vue de la connaissance de l'agent.
Par contre, si on considère qu'un monde $\mu 1$ est en relation avec un monde $\mu 2$ si et seulement si il n'y a rien d'incompatible dans ce monde $\mu 2$ avec ce que l'agent sait dans $\mu 1$, alors la critique de Hintikka prend tout son sens. En effet, cela n'implique en

[54] *ibid.*
[55] Il appelle cette relation : « alternativeness relation ». Notons que, bien que la formalisation de la sémantique des mondes possibles par S. Kripke ne date que de 1963, Hintikka propose déjà en 1962 une sémantique pour son système qui est très proche de la façon de faire de Kripke.
[56] HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962,p.45.
[57] *Ibid.*
[58] N'oublions pas que Hintikka est philosophe et non chercheur en intelligence artificielle. C'est pour ça qu'il n'introduit dans son système que ce qui lui semble pouvoir être justifié de façon intuitive. Il n'est pas intéressé, comme peuvent l'être des informaticiens, par les belles caractéristiques techniques d'un système.
[59] HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962, p.45.
[60] VAN LINDER, B. ; « *Modal logics for rational agents* », Proefschrift Universiteit Utrecht, Faculteit Wiskunde en Informatica, Utrecht, 1996, p.15.

rien le fait que les deux mondes dussent être pareils au niveau de la connaissance de l'agent. Il se pourrait donc bien qu'en μ2 l'agent ait acquis une connaissance supplémentaire qui empêcherait ainsi la symétrie de la relation.
Nous allons essayer de développer notre système dans ce sens et supprimer la symétrie.
Nous pouvons donc tenter de résoudre le paradoxe. Pour ce faire, nous allons suivre la thèse selon laquelle le paradoxe est introduit par la symétrie de la relation d'accessibilité épistémique dont découle l'axiome d'introspection négative.
Nous allons donc redévelopper le système, sans y mettre ces éléments qui semblent être la source du paradoxe et qui, de plus, ne semblent pas justifiés philosophiquement.
Nous supprimerons donc de notre nouveau système la symétrie de la relation et les axiomes d'introspection négative du point de vue épistémique et du point de vue doxastique.
Le paradoxe ne découle que de l'introduction de l'axiome d'introspection négative au point de vue épistémique, mais, pour garder la cohérence de notre nouveau système, nous avons décidé d'en éliminer également l'axiome d'introspection négative au point de vue doxastique.

Chapitre 4 :
Le nouveau système

Nous gardons le langage utilisé par Van Linder afin de faciliter la comparaison.

Définition du langage

DEFINITION : Le langage $\mathbf{L}(\Pi)$ est fondé sur l'ensemble Π des variables propositionelles. L'alphabet contient les connecteurs $\neg$ et $\wedge$, l'opérateur épistémique $\mathbf{K}$ et les opérateurs doxastiques $\mathbf{B}^o$, $\mathbf{B}^c$ et $\mathbf{B}^d$.

DEFINITION : Le langage $\mathbf{L}(\Pi)$est le plus petit ensemble qui contient Π tel que

- Si $\varphi \in \mathbf{L}(\Pi)$ et $\psi \in \mathbf{L}(\Pi)$ alors $\neg\varphi \in \mathbf{L}(\Pi)$ et $\varphi \wedge \psi \in \mathbf{L}(\Pi)$
- Si $\varphi \in \mathbf{L}(\Pi)$ alors $\mathbf{K}\varphi \in \mathbf{L}(\Pi)$
- Si $\varphi \in \mathbf{L}(\Pi)$ alors $\mathbf{B}^o\varphi \in \mathbf{L}(\Pi)$, $\mathbf{B}^c\varphi \in \mathbf{L}(\Pi)$, $\mathbf{B}^d\varphi \in \mathbf{L}(\Pi)$

Des constructions additionnelles sont introduites par des abréviations définitionnelles :

$\varphi \vee \psi \quad = \quad \neg(\neg\varphi \wedge \psi)$
$\varphi \rightarrow \psi \quad = \quad \neg\varphi \vee \psi$
$\varphi \leftrightarrow \psi \quad = \quad (\varphi \rightarrow \psi) \wedge (\varphi \rightarrow \psi)$
$\top \quad = \quad p \vee \neg p$ pour $p \in \Pi$ arbitraire
$\bot \quad = \quad \neg\top$

Nous pouvons maintenant donner l'interprétation sémantique de notre nouveau système. Celle-ci est différente de celle de Van Linder.

Interprétation sémantique

Remarque préliminaire : Nous identifierons pour la suite les relations qui appartiennent à $\wp(S \times S)$ avec les fonctions $S \rightarrow \wp(S)$ et nous noterons donc indifféremment xRy ou $y \in R(x)$ pour indiquer que le monde y est une alternative épistémique du monde x.

DEFINITION : Un modèle M pour $\mathbf{L}(\Pi)$ est un n-uplet contenant au moins les éléments suivants :

- Un ensemble S non-vide de mondes possibles (ou états).
- Une valuation $\pi : \Pi \times S \rightarrow \{0,1\}$ sur les variables propositionelles.

- Une relation R d'accessibilité épistémique entre deux mondes : $R : S \rightarrow \wp(S)$. Cette relation doit être **réflexive et transitive**.
- Une relation B^o d'accessibilité doxastique observationnelle entre deux mondes : $B^o : S \rightarrow \wp(S)$. Cette relation doit être **réflexive et transitive**.
- Une fonction $B^c : S \rightarrow \wp(S)$ qui donne les alternatives doxastiques communicationnelles de l'agent dans un monde.
- Une fonction $B^d : S \rightarrow \wp(S)$ qui donne les alternatives doxastiques par défaut de l'agent dans un monde.

Les fonctions B^c et B^d sont telles que $\forall\, s, s' \in S$:

1. $B^d(s) \neq 0$
2. $B^d(s) \subseteq B^c(s) \subseteq B^o(s) \subseteq R(s)$
3. Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$[61]

DEFINITION : La relation binaire $\models$ (validité) entre une formule du langage et une paire M, s composée d'un modèle M et d'un monde s dans M est définie récursivement de façon suivante :

$$
\begin{array}{lll}
M, s \models p & \Leftrightarrow \pi(p, s) = 1 \text{ pour } p \in \Pi & \\
M, s \models \neg\varphi & \Leftrightarrow \text{non}(M, s \models \varphi) & \\
M, s \models \varphi \wedge \psi & \Leftrightarrow M, s \models \varphi \text{ et } M, s \models \psi & \\
M, s \models \mathbf{K}\varphi & \Leftrightarrow \forall\, s' \in S((s, s') \in R & \Rightarrow M, s' \models \varphi) \\
M, s \models \mathbf{B}^o\varphi & \Leftrightarrow \forall\, s' \in S\ ((s, s') \in B^o & \Rightarrow M, s' \models \varphi) \\
M, s \models \mathbf{B}^c\varphi & \Leftrightarrow \forall\, s' \in S\ (s' \in B^c(s) & \Rightarrow M, s' \models \varphi) \\
M, s \models \mathbf{B}^d\varphi & \Leftrightarrow \forall\, s' \in S\ (s' \in B^d(s) & \Rightarrow M, s' \models \varphi)
\end{array}
$$

Remarquons que les relations d'accessibilité épistémique et d'accessibilité doxastique observationelle ont été redéfinies. Elles ne sont plus des relations d'équivalence, nous les avons redéfinies comme étant juste réflexives et transitives.
Notons aussi que les relations B^c et B^d sont transitives par les conditions 2 et 3, en effet :

$\forall$ s, s', s'' $\in$ S :

$sB^cs' \Rightarrow sRs'$	par la condition 2
$\Rightarrow B^c(s') \subseteq B^c(s)$	par la condition 3
$\Rightarrow (s'B^cs'' \Rightarrow sB^cs'')$	ce qui est la définition de la transitivité

C.Q.F.D.

Remarquons également que la dernière condition sur les relations d'accessibilité doxastique communicationnelle et par défaut (la condition numéro 3) n'apparaît pas chez Van Linder[62].

Il nous reste donc à exposer les axiomes de base de notre système.

---

[61] L'apparition de cette condition dans notre nouveau système est expliquée à la page suivante
[62] *ibid.*

Axiomes du système

$\forall \varphi, \psi \in \mathbf{L}(\Pi)$, le système possède les axiomes suivants :

1. $\models \mathbf{K}(\varphi \rightarrow \psi) \rightarrow (\mathbf{K}\varphi \rightarrow \mathbf{K}\psi)$ (K)
2. $\models \varphi \Rightarrow \models \mathbf{K}\varphi$ (N)
3. $\models \mathbf{K}\varphi \rightarrow \varphi$ (T)
4. $\models \mathbf{K}\varphi \rightarrow \mathbf{KK}\varphi$ (4)

Posons $\mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$. $\forall \varphi, \psi \in \mathbf{L}(\Pi)$, le système possède également les axiomes suivants :

5. $\models \mathbf{X}(\varphi \rightarrow \psi) \rightarrow (\mathbf{X}\varphi \rightarrow \mathbf{X}\psi)$ (K)
6. $\models \varphi \Rightarrow \models \mathbf{X}\varphi$ (N)
7. $\models \mathbf{B}^o \varphi \rightarrow \varphi$ (T)
8. $\models \mathbf{X}\varphi \rightarrow \mathbf{XX}\varphi$ (4)
9. $\models \neg(\mathbf{X}\varphi \wedge \mathbf{X}\neg\varphi)$ (D)
10. On définit la relation d'ordonnancement $>$ sur les opérateurs par : $\mathbf{K} > \mathbf{B}^o > \mathbf{B}^c > \mathbf{B}^d$; définissons $\geq$ comme étant la fermeture réflexive et transitive de $>$. Alors, $\forall \mathbf{X}, \mathbf{Y} \in \{ \mathbf{K}, \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$, $\forall \varphi \in \mathbf{L}(\Pi)$, si $\mathbf{X} \geq \mathbf{Y}$ alors $\models \mathbf{X}\varphi \rightarrow \mathbf{Y}\varphi$

En plus de ces dix axiomes de base pour notre système, il nous a semblé essentiel d'ajouter un axiome assez standard dans le domaine, et qui permet de relier les modalités de connaissance et de croyance entre elles.
Cet axiome s'énonce comme suit :

11. $\models \mathbf{X}\varphi \rightarrow \mathbf{KX}\varphi$ (A15)

Nous l'appellerons l'axiome A15 pour rester cohérent avec la notation employée dans la littérature[63].
Notons que l'axiome 8 se déduit assez facilement des axiomes 10 et 11 de la façon suivante :

- $\mathbf{X}\varphi \rightarrow \mathbf{KX}\varphi$ par l'axiome 11 (A15)
- $\mathbf{KX}\varphi \rightarrow \mathbf{XX}\varphi$ par l'axiome 10

Nous n'avons inclus l'axiome 8 dans nos axiomes de base que pour faciliter son emploi dans la suite du texte.
Remarquons également que l'adoption de l'axiome A15 dans notre système entraîne automatiquement l'apparition de la condition sémantique n° 3 (Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$) dans notre modèle d'interprétation. En effet, cette condition sémantique *correspond* à l'axiome A15 lorsque $\mathbf{X} \in \{ \mathbf{B}^c, \mathbf{B}^d \}$. Nous allons

[63] KRAUS, S. ; LEHMANN, D. ; « Knowledge, belief and time », *Theoretical Computer Science*, 58, 1988, pp.155-174.
voir aussi VAN DER HOEK, W. ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, pp.173-195.

en faire la preuve par une technique de logique modale[64] que l'on nomme : preuve de correspondance.

Preuve de correspondance entre l'axiome A15 et la condition sémantique n°3 de notre système

Remarques préliminaires :

Pour cette preuve, nous devons introduire la notion de *structure*. Une structure peut être considéré comme un modèle sans valuation, ce qui permet de parler de structure pour toutes valuations.
Une formule sera dite valide dans une structure F si et seulement si elle est valide dans chaque interprétation obtenue par ajout d'une valuation quelconque.
D'un point de vue technique, il est bon de préciser que, lorsqu'à partir d'hypothèses on veut prouver une thèse, et que cette thèse est une implication, il suffit de rajouter la partie droite de l'implication dans les hypothèses et de prouver la partie gauche de l'implication. Nous emploierons cette technique dans la preuve.
Nous ne prouverons la correspondance que pour la condition n°3 sous la forme suivante : Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$. La preuve qui montre qu'il y a également $B^d(s') \subseteq B^d(s)$ est exactement similaire.

Preuve :

Pour toute la suite de la preuve nous utiliserons une structure F comprenant l'ensemble S des mondes et les relations R, B^o, B^c, B^d.
Il faut prouver que A15 est valide dans F $\Leftrightarrow$ F satisfait la condition n° 3. Pour ce faire, nous allons d'abord exprimer A15 et la condition n° 3 en terme de mondes possibles et d'accessibilité.
La condition n° 3 (Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$) peut s'exprimer ainsi : Si $\forall s \forall s'$ (s, s' $\in$ S) on a sRs' (on peut atteindre s' par une relation R à partir du point s), alors $\forall s_2$ ($s_2 \in$ S) $s'B^c s_2 \Rightarrow sB^c s_2$ (si on peut atteindre s_2 par une relation B^c à partir de s', alors on peut atteindre s_2 par une relation B^c à partir de s).
L'axiome A15 s'exprime de façon suivante en terme de mondes possibles et de relations d'accessibilité : ($\forall V \forall w$ (F, V), w $\models$ A15) avec V :une valuation, w $\in$ S.
Ce qui donne : ($\forall V$ si (F, V), w $\models \mathbf{B}^c\varphi$ alors (F, V), w $\models \mathbf{KB}^c\varphi$) avec $\varphi \in \mathbf{L}(\Pi)$.
C'est-à-dire : ($\forall V$ si $\forall w_1$, $wB^c w_1 \Rightarrow$ (F, V), $w_1 \models \varphi$ alors $\forall w_2$, $wRw_2 \Rightarrow$ ($\forall w_3$, $w_2B^c w_3 \Rightarrow$ (F, V), $w_3 \models \varphi$)) avec $\varphi \in \mathbf{L}(\Pi)$ et $w_1, w_2 \in$ S.

Il faut donc prouver A15 est valide dans F $\Leftrightarrow$ F satisfait la condition n° 3 :

- Commençons par prouver l'implication à gauche ($\Leftarrow$) :

  En utilisant la technique de passage en hypothèse des membres de gauche d'une implication qu'on doit prouver, on a donc les hypothèses suivantes :

[64] POPKORN, S. ; « *First steps in modal logic* », Cambridge University Press, Cambridge, 1994.

a) Condition n° 3 (Si $\forall s \forall s'$, sRs alors $\forall s_2$, $s'B^c s_2 \Rightarrow sB^c s_2$)
b) $\forall w_1$, $wB^c w_1 \Rightarrow (F, V)$, $w_1 \models \varphi$
c) wRw_2
d) $w_2 B^c w_3$

et il faut prouver :

$(F, V) \models_{w3} \varphi$

Supposons $s = w$; $s' = w_2$; $s_2 = w_3$.
Par c) on obtient wRw_2 et par d) on obtient $w_2 B^c w_3$
On a donc sRs' et $s'B^c s_2$, d'où par a), on obtient $sB^c s_2$, c'est à dire $wB^c w_3$
On obtient ainsi $wB^c w_3$ et donc par b) on obtient $(F, V), w_3 \models \varphi$.
On a donc $(F, V), w_3 \models \varphi$.
C.Q.F.D.

- Prouvons maintenant l'implication à droite ($\Rightarrow$) :

En utilisant la technique de passage en hypothèse des membres de gauche d'une implication qu'on doit prouver, on a donc les hypothèses suivantes :

a) ($\forall V$ si $\forall w_1$, $wB^c w_1 \Rightarrow (F,V)$, $w_1 \models \varphi$ alors $\forall w_2$, $wRw_2 \Rightarrow (\forall w_3$, $w_2 B^c w_3 \Rightarrow (F,V)$, $w_3 \models \varphi$) : c'est l'axiome A15
b) sRs'
c) $s'B^c s_2$

et il faut prouver :

$sB^c s_2$

Supposons $w = s$; $w_2 = s'$; $w_3 = s_2$
Supposons aussi une valuation V telle qu'on a φ valide en tout monde $x \in S$ si on a $wB^c x$ (c.à.d. x accessible par B^c à partir de w)
La première partie de l'hypothèse a) devient donc : $\forall w_1$, $wB^c w_1 \Rightarrow wB^c w_1$ par la valuation que nous avons choisie. Cette première partie est trivialement valide, on peut donc la laisser tomber.
L'hypothèse a) devient donc : Si $\forall w_2$, wRw_2 alors ($\forall w_3$, $w_2 B^c w_3 \Rightarrow (F,V)$, $w_3 \models \varphi$).
Et on a donc pour la première partie de cette hypothèse : $\forall w_2$, wRw_2, c'est à dire sRs', or sRs' est valide par b) ; on peut donc de nouveau laisser tomber la première partie de la nouvelle hypothèse a) qui devient ainsi : Si $\forall w_3$, $w_2 B^c w_3$ alors (F,V), $w_3 \models \varphi$.
Et on a donc pour la première partie de cette hypothèse : $\forall w_3$, $w_2 B^c w_3$, c'est-à-dire $s'B^c s_2$, or $s'B^c s_2$ est valide par c) ; on laisse donc de nouveau tomber la première partie de l'hypothèse a) qui devient ainsi (F,V), $w_3 \models \varphi$
On a donc $(F,V) \models_{w3} \varphi$, ce qui revient au même que $(F,V) \models_{s2} \varphi$.
Par la valuation que nous avons choisi, on obtient alors bien : $sB^c s_2$
C.Q.F.D.

Preuve de la disparition du paradoxe

En reconstruisant le système, nous avons supprimé les axiomes d'introspection négative du point de vue épistémique comme du point de vue doxastique.
Il nous reste à montrer que le paradoxe n'apparaît plus dans ce système, au vu de ce que nous avons enlevé.

Il suffit, pour ce faire, d'utiliser la sémantique pour donner un modèle qui vérifie les conditions de notre nouveau système et où le théorème $\mathbf{XK}\varphi \rightarrow \mathbf{K}\varphi$ soit faux si on pose $\mathbf{X} \in \{ \mathbf{B}^c, \mathbf{B}^d \}$:

Posons l'ensemble des états : $S = \{w0, w1, w2, w3\}$;
Posons la valuation π : $\varphi \rightarrow 1$ en $\{w0\}$, $\{w2\}$,$\{w3\}$
$\varphi \rightarrow 0$ en $\{w1\}$;
Posons la relation R : $\{(w0, w1)$; $(w0, w2)$; $(w0, w3)$; $(w1, w2)$; $(w1, w3)$;
$(w2, w3)$; $(y, y)\}$
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$
Posons la relation B^c: $\{(w0, w2)$; $(w0, w3)$; $(w1, w2)$; $(w1, w3)$; $(w2, w3)$;
$(w3, w3)\}$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w2, w3\}$
  $B^c(w1) = \{w2, w3\}$
  $B^c(w2) = \{w3\}$
  $B^c(w3) = \{w3\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  $w3 \in R(w0)$ et $B^c(w3) \subseteq B^c(w0)$
  $w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
  $w3 \in R(w1)$ et $B^c(w3) \subseteq B^c(w1)$
  $w3 \in R(w2)$ et $B^c(w3) \subseteq B^c(w2)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\mathbf{XK}\varphi$ vrai et $\mathbf{K}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

Le paradoxe est donc bien supprimé dans notre nouveau système. Mais, pour compléter celui-ci et s'assurer de son efficacité, nous avons considéré nécessaire de vérifier si certaines conséquences du système basé sur **S5**, restaient valides.
A la suite de Van der Hoek[65], nous avons dégagé une série de théorèmes à vérifier.
Nous en avons donc fait les preuves.

[65] VAN DER HOEK, W. ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, p.176

Preuves de certaines conséquences dans le nouveau système

Liste des théorèmes à vérifier

Les théorèmes à vérifier que Van der Hoek dégage[66] et qui sont pertinents pour notre système sont les suivants :

Posons $\mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$, $\forall \varphi \in \mathbf{L}(\Pi)$:

1) $\mathbf{K}\neg\varphi \rightarrow \neg\mathbf{X}\varphi$

Ce théorème montre la cohérence des croyances et de la connaissance, nous allons démontrer sa validité

2) $\mathbf{X}\varphi \leftrightarrow \mathbf{KX}\varphi$

L'implication à droite de ce théorème combine les modalités de croyance et de connaissance, l'implication à gauche nous rappelle la véracité de la connaissance. Nous allons démontrer la validité de ce théorème.

3) $\neg\mathbf{X}\varphi \leftrightarrow \mathbf{K}\neg\mathbf{X}\varphi$

L'implication à droite de ce théorème est une introspection négative de la connaissance sur les croyances, nous allons démontrer qu'elle n'est pas valide.

4) $\mathbf{K}\varphi \leftrightarrow \mathbf{XK}\varphi$

L'implication à droite de ce théorème est une introspection positive des croyances sur la connaissance, nous allons démontrer qu'elle est valide.
L'implication à gauche est notre paradoxe, il n'est évidemment pas valide.

5) $\neg\mathbf{K}\varphi \leftrightarrow \mathbf{X}\neg\mathbf{K}\varphi$

L'implication à droite est une introspection négative de la croyance sur la connaissance, nous allons démontrer qu'elle n'est pas valide.
Nous démontrerons aussi que l'implication à gauche est valide.

6) $\mathbf{X}\varphi \leftrightarrow \mathbf{XX}\varphi$

L'implication à droite est l'introspection positive pour les croyances, elle est évidemment valide.
L'implication à gauche se décompose en une multitude de théorèmes selon les modalités de croyance qui sont combinées. Le sens de ces théorèmes n'est pas toujours évident et nous démontrerons leur validité ou leur non-validité en les examinant un à un.

---

[66] VAN DER HOEK, W. ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, p.176

7) $\neg \mathbf{X}\varphi \leftrightarrow \mathbf{X}\neg\mathbf{X}\varphi$

Ce théorème se décompose également en une série d'autres théorèmes dont il faudra démontrer si ils sont valides ou pas en les examinant un à un.

Preuves de la validité ou de la non-validité des théorèmes

Remarque préliminaire : toutes les preuves de validité d'un théorème que nous donnons ici, sont des preuves de type syntaxique. A l'opposé, nous utilisons la sémantique pour démontrer la non-validité d'un théorème. Il suffit en effet dans ce cas, de donner un seul contre exemple qui vérifie toutes les conditions sémantiques de notre système et dans lequel le théorème est faux.
Les preuves des théorèmes sont présentées dans l'ordre de la sous-section précédente.

1. Posons $\mathbf{X} \in \{ \mathbf{B}^{\circ}, \mathbf{B}^{c}, \mathbf{B}^{d} \}, \forall \varphi \in \mathbf{L}(\Pi)$, prouvons que $\mathbf{K}\neg\varphi \rightarrow \neg\mathbf{B}\varphi$:

   $\mathbf{K}\neg\varphi \rightarrow \mathbf{X}\neg\varphi$ par l'axiome 10
   $\mathbf{X}\neg\varphi \rightarrow \neg\mathbf{X}\varphi$ par l'axiome 9 (D)

   On a donc bien : $\mathbf{K}\neg\varphi \rightarrow \neg\mathbf{X}\varphi$.
   C.Q.F.D.

2. Posons $\mathbf{X} \in \{ \mathbf{B}^{\circ}, \mathbf{B}^{c}, \mathbf{B}^{d} \}, \forall \varphi \in \mathbf{L}(\Pi)$

   - Prouvons que $\mathbf{X}\varphi \rightarrow \mathbf{K}\mathbf{X}\varphi$:

     $\mathbf{X}\varphi \rightarrow \mathbf{K}\mathbf{X}\varphi$ par l'axiome 11 (A15)

   - Prouvons que $\mathbf{K}\mathbf{X}\varphi \rightarrow \mathbf{X}\varphi$:

     $\mathbf{K}\mathbf{X}\varphi \rightarrow \mathbf{X}\varphi$ par l'axiome 3 (T)

   On a donc bien : $\mathbf{X}\varphi \leftrightarrow \mathbf{K}\mathbf{X}\varphi$.
   C.Q.F.D.

3. Posons $\mathbf{X} \in \{ \mathbf{B}^{\circ}, \mathbf{B}^{c}, \mathbf{B}^{d} \}, \forall \varphi \in \mathbf{L}(\Pi)$

   - Prouvons que $\neg\mathbf{X}\varphi \rightarrow \mathbf{K}\neg\mathbf{X}\varphi$ n'est pas valide pour tout φ :

     Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

     Posons l'ensemble des états : S = {w0, w1, w2 } ;
     Posons la valuation π : $\varphi \rightarrow 1$ en {w0}, {w2}
     $\varphi \rightarrow 0$ en {w1} ;
     Posons la relation R : {(w0, w1) ; (w0, w2) ; (y, y)}
     avec $y \in S$ (pour exprimer la réflexivité de R) ;
     Posons la relation B° : $B^{\circ} = R$;
     Posons la relation B^{c} : {(w0, w1) ; (w0, w2) ; (w1, w1) ; (w2, w2)} ;

Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1, w2\}$
  $B^c(w1) = \{w1\}$
  $B^c(w2) = \{w2\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg\mathbf{X}\varphi$ vrai et $\mathbf{K}\neg\mathbf{X}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ ;

- Prouvons que $\mathbf{K}\neg\mathbf{X}\varphi \rightarrow \neg\mathbf{X}\varphi$:

  $\mathbf{K}\neg\mathbf{X}\varphi \rightarrow \neg\mathbf{X}\varphi$ par l'axiome 3 (T)

Le théorème $\neg\mathbf{X}\varphi \leftrightarrow \mathbf{K}\neg\mathbf{X}\varphi$ n'est donc pas valable pour tout φ.
C.Q.F.D.

4. Posons $\mathbf{X} \in \{ \mathbf{B}^\circ, \mathbf{B}^c, \mathbf{B}^d \}$, $\forall \varphi \in \mathbf{L}(\Pi)$

- Prouvons que $\mathbf{K}\varphi \rightarrow \mathbf{XK}\varphi$:

  $\mathbf{K}\varphi \rightarrow \mathbf{KK}\varphi$ par l'axiome 4 (introspection positive)
  $\mathbf{KK}\varphi \rightarrow \mathbf{XK}\varphi$ par l'axiome 10

- Prouvons que $\mathbf{XK}\varphi \rightarrow \mathbf{K}\varphi$ n'est pas valable pour tout φ :

  Cette preuve est triviale, vu que le théorème est l'exacte expression du paradoxe que nous voulions éliminer du système. Pour s'en convaincre, il suffit de se reporter à la preuve qui a déjà été faite[67]
  Le théorème ne peut donc être valide.

Le théorème $\mathbf{K}\varphi \leftrightarrow \mathbf{XK}\varphi$ n'est donc pas valable pour tout φ.
C.Q.F.D.

[67] cfr. p.26

5. Posons $\mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}, \forall \varphi \in \mathbf{L}(\Pi)$

   - Posons $\mathbf{Y} \in \{\mathbf{B}^c, \mathbf{B}^d\}, \forall \varphi \in \mathbf{L}(\Pi)$

     1. Prouvons que $\neg\mathbf{K}\varphi \rightarrow \mathbf{Y}\neg\mathbf{K}\varphi$ n'est pas valable pour tout φ :

        Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

        Posons l'ensemble des états : S = { w0, w1, w2 } ;
        Posons la valuation π : $\varphi \rightarrow 1$ en {w0}, {w2}
        $\varphi \rightarrow 0$ en {w1} ;
        Posons la relation R : {(w0, w1) ; (w0, w2) ; (y, y)}
        avec y ∈ S (pour exprimer la réflexivité de R) ;
        Posons la relation B^o : $B^o = R$;
        Posons la relation B^c : {(w0, w1) ; (w0, w2) ; (w1, w1) ; (w2, w2)} ;
        Posons la relation $B^d = B^c$;

        On vérifie ainsi donc bien que :

        - $B^d(s) \neq 0$;

          En effet : $B^d = B^c$ et $B^c(w0) = \{w1, w2\}$
          $B^c(w1) = \{w1\}$
          $B^c(w2) = \{w2\}$

        - $B^d(s) \subseteq B^c(s) \subseteq B^o(s) \subseteq R(s)$;

          En effet : $B^d = B^c$, $B^o = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

        - Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

          En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
          $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
          et : $B^d = B^c$

        Et on peut donc bien avoir :
        $\neg\mathbf{K}\varphi$ vrai et $\mathbf{X}\neg\mathbf{K}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

     2. Prouvons que $\neg\mathbf{K}\varphi \rightarrow \mathbf{B}^o\neg\mathbf{K}\varphi$ n'est pas valable pour tout φ :

        Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

        Posons l'ensemble des états : S = { w0, w1, w2 } ;
        Posons la valuation π : $\varphi \rightarrow 1$ en {w0}, {w1}
        $\varphi \rightarrow 0$ en {w2} ;

Posons la relation R : {(w0, w1) ; (w0, w2) ; (y, y)}
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons: $B^d = B^c = B^o = R$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = R$ et $R(w0) = \{w1, w2\}$
  $R(w1) = \{w1\}$
  $R(w2) = \{w2\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^o(s) \subseteq R(s)$;

  En effet : $B^d = B^c = B^o = R$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg\mathbf{K}\varphi$ vrai et $\mathbf{B}^o\neg\mathbf{K}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

- Prouvons que $\mathbf{X}\neg\mathbf{K}\varphi \rightarrow \neg\mathbf{K}\varphi$:

$\mathbf{X}\neg\mathbf{K}\varphi \rightarrow \neg\mathbf{X}\mathbf{K}\varphi$	par l'axiome 9 (D)
$\neg\mathbf{X}\mathbf{K}\varphi \rightarrow \neg\mathbf{K}\mathbf{K}\varphi$	par la contraposée de l'axiome 10
$\neg\mathbf{K}\mathbf{K}\varphi \rightarrow \neg\mathbf{K}\varphi$	par la contraposée de l'axiome 4 (introspection positive)

Le théorème $\neg\mathbf{K}\varphi \leftrightarrow \mathbf{X}\neg\mathbf{K}\varphi$ n'est donc pas valable pour tout φ.
C.Q.F.D

6. Posons $\mathbf{X} \in \{ \mathbf{B}^o, \mathbf{B}^c, \mathbf{B}^d \}$, $\forall \varphi \in \mathbf{L}(\Pi)$

- Prouvons que $\mathbf{X}\varphi \rightarrow \mathbf{X}\mathbf{X}\varphi$:

$\mathbf{X}\varphi \rightarrow \mathbf{X}\mathbf{X}\varphi$	par l'axiome 8 (introspection positive)

- Il convient de se montrer prudent pour la suite des preuves du théorème 6. En effet, ce théorème permet de combiner plusieurs modalités de croyance de degrés divers de fiabilité. Il est donc nécessaire d'examiner spécifiquement chacune des combinaisons possibles. Néanmoins, en posant partout dans nos contre exemples : $B^d = B^c$, nous pouvons réduire le nombre de possibilité en considérant que ce qui marche pour une relation B^c marche également pour une relation B^d.

La deuxième partie de ce théorème s'énonce sous sa forme la plus générale de façon suivante : $\mathbf{XX}\varphi \rightarrow \mathbf{X}\varphi$
Nous pouvons poser pour la suite l'opérateur **Y** tel que : $\mathbf{Y} \in \{ \mathbf{B}^c, \mathbf{B}^d \}$
$\forall \varphi \in \mathbf{L}(\Pi)$:

1. Prouvons que $\mathbf{YY}\varphi \rightarrow \mathbf{Y}\varphi$ n'est pas valable pour tout φ :

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : $S = \{ w0, w1, w2 \}$;
   Posons la valuation π : $\varphi \rightarrow 1$ en $\{w0\}, \{w2\}$
   $\varphi \rightarrow 0$ en $\{w1\}$;
   Posons la relation R : $\{(w0, w1) ; (w0, w2) ; (w1, w2) ; (y, y)\}$
   avec $y \in S$ (pour exprimer la réflexivité de R) ;
   Posons la relation B° : $B^\circ = R$;
   Posons la relation B^c: $\{(w0, w1) ; (w0, w2) ; (w1, w2) ; (w2, w2)\}$;
   Posons la relation $B^d = B^c$;

   On vérifie ainsi donc bien que :

   - $B^d(s) \neq 0$;

     En effet : $B^d = B^c$ et $B^c(w0) = \{w1, w2\}$
     $B^c(w1) = \{w2\}$
     $B^c(w2) = \{w2\}$

   - $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

     En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

   - Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

     En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
     $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
     $w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
     et : $B^d = B^c$

   Et on peut donc bien avoir :
   $\mathbf{YY}\varphi$ vrai et $\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

2. Prouvons que $\mathbf{YY}\varphi \rightarrow \mathbf{B}^\circ\varphi$ n'est pas valable pour tout φ :

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : $S = \{ w0, w1, w2 \}$;

Posons la valuation π : $\varphi \rightarrow 1$ en {w1}, {w2}
$\varphi \rightarrow 0$ en {w0} ;
Posons la relation R : {(w0, w1) ; (w0, w2) ; (w1, w2) ; (y, y)}
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : {(w0, w1) ; (w0, w2) ; (w1, w2) ; (w2, w2)} ;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1, w2\}$
  $B^c(w1) = \{w2\}$
  $B^c(w2) = \{w2\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  $w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\mathbf{YY}\varphi$ vrai et $\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

3. Prouvons que $\mathbf{B}^\circ\mathbf{Y}\varphi \rightarrow \mathbf{B}^\circ\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : S = { w0, w1 } ;
Posons la valuation π : $\varphi \rightarrow 1$ en {w1}
$\varphi \rightarrow 0$ en {w0} ;
Posons la relation R : {(w0, w1) ; (y, y)}
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : {(w0, w1) ; (w1, w1)} ;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
$B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
YYφ vrai et $\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

4. Prouvons que $\mathbf{YB}^\circ\varphi \rightarrow \mathbf{B}^\circ\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : $S = \{ w0, w1 \}$;
Posons la valuation π : $\varphi \rightarrow 1$ en $\{w1\}$
$\varphi \rightarrow 0$ en $\{w0\}$;
Posons la relation R : $\{(w0, w1) ; (y, y)\}$
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : $\{(w0, w1) ; (w1, w1)\}$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
YYφ vrai et $\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

5. Prouvons que $\mathbf{B}^{\circ}\mathbf{Y}\varphi \rightarrow \mathbf{Y}\varphi$:

$\mathbf{B}^{\circ}\mathbf{Y}\varphi \rightarrow \mathbf{Y}\varphi$	par l'axiome 7 (T)

6. Prouvons que $\mathbf{B}^{\circ}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{Y}\varphi$:

$\mathbf{B}^{\circ}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{B}^{\circ}\varphi$	par l'axiome 7 (T)
$\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{Y}\varphi$	par l'axiome 10

7. Prouvons que $\mathbf{B}^{\circ}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{B}^{\circ}\varphi$:

$\mathbf{B}^{\circ}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{B}^{\circ}\varphi$	par l'axiome 7 (T)

8. Prouvons que $\mathbf{Y}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{Y}\varphi$:

$\mathbf{B}^{\circ}\varphi \rightarrow \varphi$	par l'axiome 7(T)
$\mathbf{Y}(\mathbf{B}^{\circ}\varphi \rightarrow \varphi)$	par l'axiome 6(N)
$\mathbf{Y}\mathbf{B}^{\circ}\varphi \rightarrow \mathbf{Y}\varphi$	par l'axiome 5(K) et le Modus Ponens

Le théorème $\mathbf{X}\varphi \leftrightarrow \mathbf{X}\mathbf{X}\varphi$ n'est donc pas valable pour tout φ
C.Q.F.D.

7. Posons $\mathbf{X} \in \{ \mathbf{B}^{\circ}, \mathbf{B}^{c}, \mathbf{B}^{d} \}, \forall \varphi \in \mathbf{L}(\Pi)$

- Le théorème 7 combine également plusieurs modalités de croyance. Nous allons d'abord traiter l'implication vers la droite en décomposant les combinaisons possibles comme nous l'avons fait pour la deuxième partie du théorème 6.

L'implication à droite du théorème s'exprime sous sa forme la plus générale de façon suivante :
$\neg\mathbf{X}\varphi \rightarrow \mathbf{X}\neg\mathbf{X}\varphi$
Nous pouvons poser pour la suite l'opérateur $\mathbf{Y}$ tel que : $\mathbf{Y} \in \{ \mathbf{B}^{c}, \mathbf{B}^{d} \}$
$\forall \varphi \in \mathbf{L}(\Pi)$:

1. Prouvons que $\neg\mathbf{Y}\varphi \rightarrow \mathbf{Y}\neg\mathbf{Y}\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : S = { w0, w1, w2 } ;
Posons la valuation π : $\varphi \rightarrow 1$ en {w0}, {w2}
$\varphi \rightarrow 0$ en {w1} ;
Posons la relation R : {(w0, w1) ; (w0, w2) ; (w1, w2) ; (y, y)}
avec y ∈ S (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^{\circ} = R$;
Posons la relation B^{c} : {(w0, w1) ; (w0, w2) ; (w1, w2) ; (w2, w2)} ;
Posons la relation $B^{d} = B^{c}$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1, w2\}$
  $B^c(w1) = \{w2\}$
  $B^c(w2) = \{w2\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  $w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{Y}\varphi$ vrai et $\mathbf{Y}\neg\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

2. Prouvons que $\neg\mathbf{B}^\circ\varphi \rightarrow \mathbf{Y}\neg\mathbf{Y}\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : $S = \{ w0, w1 \}$;
Posons la valuation π : $\varphi \rightarrow 1$ en $\{w1\}$
$\varphi \rightarrow 0$ en $\{w0\}$;
Posons la relation R : $\{(w0, w1) ; (y, y)\}$
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : $\{(w0, w1) ; (w1, w1)\}$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{B}^\circ\varphi$ vrai et $\mathbf{Y}\neg\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

3. Prouvons que $\neg \mathbf{B}^\circ\varphi \rightarrow \mathbf{B}^\circ\neg\mathbf{Y}\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : $S = \{ w0, w1 \}$;
Posons la valuation π : $\varphi \rightarrow 1$ en $\{w1\}$
$\varphi \rightarrow 0$ en $\{w0\}$;
Posons la relation R : $\{(w0, w1) ; (y, y)\}$
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : $\{(w0, w1) ; (w1, w1)\}$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{B}^\circ\varphi$ vrai et $\mathbf{B}^\circ\neg\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

4. Prouvons que $\neg \mathbf{B}^\circ\varphi \rightarrow \mathbf{Y}\neg\mathbf{B}^\circ\varphi$ n'est pas valable pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : $S = \{ w0, w1 \}$;

Posons la valuation π : $\varphi \rightarrow 1$ en {w1}
$\varphi \rightarrow 0$ en {w0} ;
Posons la relation R : {(w0, w1) ; (y, y)}
avec y $\in$ S (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : {(w0, w1) ; (w1, w1)} ;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : w1 $\in$ R(w0) et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{B}^\circ\varphi$ vrai et $\mathbf{Y}\neg \mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

5. Prouvons que $\neg \mathbf{B}^\circ\varphi \rightarrow \mathbf{B}^\circ\neg \mathbf{B}^\circ\varphi$ n'est pas valide pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : S = { w0, w1 } ;
Posons la valuation π : $\varphi \rightarrow 1$ en {w0}
$\varphi \rightarrow 0$ en {w1} ;
Posons la relation R : {(w0, w1) ; (y, y)}
avec y $\in$ S (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : $B^c = B^\circ$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w0, w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{B}^\circ\varphi$ vrai et $\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

6. Prouvons que $\neg\mathbf{Y}\varphi \rightarrow \mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi$ n'est pas valide pour tout φ :

Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

Posons l'ensemble des états : $S = \{ w0, w1 \}$;
Posons la valuation π : $\varphi \rightarrow 1$ en $\{w0\}$
$\varphi \rightarrow 0$ en $\{w1\}$;
Posons la relation R : $\{(w0, w1) ; (y, y)\}$
avec $y \in S$ (pour exprimer la réflexivité de R) ;
Posons la relation B° : $B^\circ = R$;
Posons la relation B^c : $B^c = B^\circ$;
Posons la relation $B^d = B^c$;

On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w0, w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg\mathbf{Y}\varphi$ vrai et $\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

7. Prouvons que $\neg \mathbf{Y}\varphi \rightarrow \mathbf{B}^{o}\neg \mathbf{Y}\varphi$ n'est pas valide pour tout φ :

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : S = { w0, w1 } ;
   Posons la valuation π : $\varphi \rightarrow 1$ en {w0}
   $\varphi \rightarrow 0$ en {w1} ;
   Posons la relation R : {(w0, w1) ; (y, y)}
   avec y $\in$ S (pour exprimer la réflexivité de R) ;
   Posons la relation B^{o} : $B^{o} = R$;
   Posons la relation B^{c} : $B^{c} = B^{o}$;
   Posons la relation $B^{d} = B^{c}$;

   On vérifie ainsi donc bien que :

   - $B^{d}(s) \neq 0$;

     En effet : $B^{d} = B^{c}$ et $B^{c}(w0) = \{w0, w1\}$
     $B^{c}(w1) = \{w1\}$

   - $B^{d}(s) \subseteq B^{c}(s) \subseteq B^{o}(s) \subseteq R(s)$;

     En effet : $B^{d} = B^{c}$, $B^{o} = R$ et $B^{c}(s) \subseteq R(s)$ pour tout $s \in S$

   - Si $s' \in R(s)$ alors $B^{c}(s') \subseteq B^{c}(s)$ et $B^{d}(s') \subseteq B^{d}(s)$

     En effet : $w1 \in R(w0)$ et $B^{c}(w1) \subseteq B^{c}(w0)$
     et : $B^{d} = B^{c}$

   Et on peut donc bien avoir :
   $\neg \mathbf{Y}\varphi$ vrai et $\mathbf{B}^{o}\neg \mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

8. Prouvons que $\neg \mathbf{Y}\varphi \rightarrow \mathbf{Y}\neg \mathbf{B}^{o}\varphi$ n'est pas valide pour tout φ :

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : S = { w0, w1 } ;
   Posons la valuation π : $\varphi \rightarrow 1$ en {w0}
   $\varphi \rightarrow 0$ en {w1} ;
   Posons la relation R : {(w0, w1) ; (y, y)}
   avec y $\in$ S (pour exprimer la réflexivité de R) ;
   Posons la relation B^{o} : $B^{o} = R$;
   Posons la relation B^{c} : $B^{c} = B^{o}$;
   Posons la relation $B^{d} = B^{c}$;

   On vérifie ainsi donc bien que :

- $B^d(s) \neq 0$;

  En effet : $B^d = B^c$ et $B^c(w0) = \{w0, w1\}$
  $B^c(w1) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\neg \mathbf{Y}\varphi$ vrai et $\mathbf{Y}\neg\mathbf{B}^\circ\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

- L'implication à gauche du théorème 7 combine évidement elle aussi plusieurs modalités de croyance. Nous allons donc la traiter de la même façon que l'implication à droite, c'est-à-dire en décomposant les différentes combinaisons possibles

L'implication à gauche du théorème s'exprime sous sa forme la plus générale de façon suivante :
$\mathbf{X}\neg\mathbf{X}\varphi \rightarrow \neg\mathbf{X}\varphi$
Nous pouvons poser pour la suite l'opérateur **Y** tel que : $\mathbf{Y} \in \{ \mathbf{B}^c, \mathbf{B}^d \}$
$\forall \varphi \in \mathbf{L}(\Pi)$:

1. Prouvons que $\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{Y}\varphi$ n'est pas valide pour tout φ

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : $S = \{ w0, w1, w2 \}$;
   Posons la valuation π : $\varphi \rightarrow 1$ en $\{w0\}$, $\{w1\}$
   $\varphi \rightarrow 0$ en $\{w2\}$;
   Posons la relation R : $\{(w0, w1) ; (w0, w2) ; (w1, w2) ; (w2, w1) ; (y, y)\}$
   avec $y \in S$ (pour exprimer la réflexivité de R) ;
   Posons la relation B° : $B^\circ = R$;
   Posons la relation B^c : $\{(w0, w1) ; (w1, w1) ; (w2, w1)\}$;
   Posons la relation $B^d = B^c$;

   On vérifie ainsi donc bien que :

   - $B^d(s) \neq 0$;

     En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$

$B^c(w1) = \{w1\}$
$B^c(w2) = \{w1\}$

- $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

  En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

- Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

  En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
  $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$
  $w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
  $w1 \in R(w2)$ et $B^c(w1) \subseteq B^c(w2)$
  et : $B^d = B^c$

Et on peut donc bien avoir :
$\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi$ vrai et $\neg\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

2. Prouvons que $\mathbf{Y}\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{Y}\varphi$ n'est pas valide pour tout φ

   Il faut donc définir un contre exemple qui vérifie toutes les conditions du système :

   Posons l'ensemble des états : $S = \{ w0, w1, w2 \}$;
   Posons la valuation π : $\varphi \rightarrow 1$ en $\{w0\}, \{w1\}$
   $\varphi \rightarrow 0$ en $\{w2\}$;
   Posons la relation R : $\{(w0, w1) ; (w0, w2) ; (w1, w2) ; (w2, w1) ; (y, y)\}$
   avec $y \in S$ (pour exprimer la réflexivité de R) ;
   Posons la relation B° : $B^\circ = R$;
   Posons la relation B^c : $\{(w0, w1) ; (w1, w1) ; (w2, w1)\}$;
   Posons la relation $B^d = B^c$;

   On vérifie ainsi donc bien que :

   - $B^d(s) \neq 0$;

     En effet : $B^d = B^c$ et $B^c(w0) = \{w1\}$
     $B^c(w1) = \{w1\}$
     $B^c(w2) = \{w1\}$

   - $B^d(s) \subseteq B^c(s) \subseteq B^\circ(s) \subseteq R(s)$;

     En effet : $B^d = B^c$, $B^\circ = R$ et $B^c(s) \subseteq R(s)$ pour tout $s \in S$

   - Si $s' \in R(s)$ alors $B^c(s') \subseteq B^c(s)$ et $B^d(s') \subseteq B^d(s)$

     En effet : $w1 \in R(w0)$ et $B^c(w1) \subseteq B^c(w0)$
     $w2 \in R(w0)$ et $B^c(w2) \subseteq B^c(w0)$

$w2 \in R(w1)$ et $B^c(w2) \subseteq B^c(w1)$
$w1 \in R(w2)$ et $B^c(w1) \subseteq B^c(w2)$
et : $B^d = B^c$

Et on peut donc bien avoir :
$\mathbf{Y}\neg\mathbf{B}^\circ\varphi$ vrai et $\neg\mathbf{Y}\varphi$ faux en w0, ce qui montre que le théorème n'est pas valide pour tout φ

3. Prouvons que $\mathbf{Y}\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$:

$\mathbf{Y}\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\mathbf{Y}\varphi$	par l'axiome 9
$\neg\mathbf{Y}\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$	par la contraposée de l'axiome 8

4. Prouvons que $\mathbf{B}^\circ\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$:

$\mathbf{B}^\circ\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$	par l'axiome 7 (T)

5. Prouvons que $\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$:

$\mathbf{B}^\circ\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$	par l'axiome 7(T)

6. Prouvons que $\mathbf{B}^\circ\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$:

$\mathbf{B}^\circ\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$	par l'axiome 7 (T)
$\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$	par la contraposée de l'axiome 10

7. Prouvons que $\mathbf{Y}\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$

$\mathbf{Y}\neg\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{Y}\mathbf{B}^\circ\varphi$	par l'axiome 9
$\neg\mathbf{Y}\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{B}^\circ\mathbf{B}^\circ\varphi$	par la contraposée de l'axiome 10
$\neg\mathbf{B}^\circ\mathbf{B}^\circ\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$	par la contraposée de l'axiome 8

8. Prouvons que $\mathbf{Y}\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$

$\mathbf{Y}\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\mathbf{Y}\varphi$	par l'axiome 9
$\neg\mathbf{Y}\mathbf{Y}\varphi \rightarrow \neg\mathbf{Y}\varphi$	par la contraposée de l'axiome 8
$\neg\mathbf{Y}\varphi \rightarrow \neg\mathbf{B}^\circ\varphi$	par la contraposée de l'axiome 10

Nous en avons ainsi fini avec les preuves des diverses conséquences du système. Nous pouvons observer que notre système diffère quelque peu du système classique développé par van der Hoek[68] et qui est basé sur S5.
Ceci est évidemment causé par la disparition de l'axiome d'introspection négative.

[68] VAN DER HOEK, W. ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, pp.173-195.

Conclusion

Le nouveau système que nous avons développé dans ce travail est relativement peu complexe du point de vue des conditions sémantiques et du nombre d'axiomes de base, comparé aux systèmes standards que la littérature spécialisée propose.
Il possède sans doute des propriétés techniques moins intéressantes que le système de Van Linder, mais son apport essentiel nous a semblé être le fait qu'il soit plus conforme au langage naturel et à l'intuition. Le paradoxe que nous avons découvert chez Van Linder n'apparaît plus ; du point de vue philosophique, cela nous a paru digne d'analyse.
En ce qui concerne l'inscription de ce travail dans une perspective informatique plus globale, vu que le système que nous avons construit possède moins de propriétés techniques intéressantes que d'autres systèmes standards, des exemples possibles de son application directe dans un cas particulier précis ne nous sont pas clairement apparus, bien qu'il soit possible d'imaginer qu'un jour apparaîtra le besoin de modéliser des agents virtuels qui ne devront pas se comporter selon l'axiome paradoxal que nous avons supprimé.
L'apport de cette analyse nous semble donc bien se situer essentiellement à un niveau plus philosophique. Nous avons identifié un paradoxe dans un système existant, nous l'avons analysé et nous avons développé un système qui semble plus conforme à l'intuition.
Pour conclure, nous nous permettons de nous impliquer un peu pour dire que l'intérêt de cette étude a surtout été beaucoup important à un niveau personnel. Nous avons appris la base de la logique modale et toute une série de techniques de manipulation et de preuve dans ce domaine. Nous avons pu explorer relativement en profondeur ce qui se fait dans cette logique modale particulière qu'est la logique de croyance et de connaissance, nous avons ainsi beaucoup appris. Le tout fut donc fort enrichissant.

BIBLIOGRAPHIE

Ouvrages :

- ARISTOTE ; « *La métaphysique* », traduit par J. Tricot, Vrin, Paris, 1986.
- SAINT BONAVENTURE ; « *Itinéraire de l'esprit vers Dieu* », Vrin, Paris, 1924.
- CHELLAS, B.F. ; « *Modal logic. An introduction* », Cambridge University Press, Cambridge, 1980.
- DESCARTES, R. ; « *Oeuvres et lettres* », Bibliothèque de la Pléiade, Gallimard, 1953.
- FAGIN, R. ; HALPERN, J.Y. ; MOSES, Y. ; VARDI, M.Y. ; « *Reasoning about knowledge* », MIT Press, Cambridge, 1995.
- GILLET, E. ; « *Contributions à la logique de connaissance – Omniscience logique et connaissance faible* », thèse présentée pour l'obtention du grade de Docteur en Philosophie et Lettres, Université de Liège, 1992.
- GOCHET, P. ; GRIBOMONT, P. ; « *Logique – Méthodes pour l'informatique fondamentale* », Hermès, Paris, 1990.
- HINTIKKA, J. ; « *Knowledge and belief – An introduction to the logic of the two notions* », Cornell University Press, Ithaca, NY, 1962.
- HOEK, W. van der ; MEYER, J.-J.CH. ; « *Epistemic logic for AI and computer science* », Cambridge University Press, New York, 1995.
- HUME, D. ; « *Enquête sur l'entendement humain* », traduit par A. Leroy, coll. GF, Flammarion, 1983.
- KANT, E. ; « *Critique de la raison pure* », édité par F. Alquié, traduit par A. Delamarre et F. Marty à partir de la traduction de J. Barni, Coll. Folio Essais, Gallimard, Paris, 1980.
- KONOLIDGE, K. ; « *A deduction model of belief* », Pitman / Morgan Kaufmann, London / Los Altos, 1986.
- LEIBNIZ, G.W. ; « *Nouveaux essais sur l'entendement humain* », coll. GF, Flammarion, 1990.

- LEWIS, D. ; « *Convention, A Philosophical Study* », Harvard University Press, Cambridge, 1969.
- LINDER, B. van ; « *Modal logics for rational agents* », Proefschrift Universiteit Utrecht, Faculteit Wiskunde en Informatica, Utrecht, 1996.
- LOCKE, J. ; « *An essay concerning human understanding* », édité par P. Nidditch, Oxford University Press, 1975.
- PARMENIDE ; « *De la nature* », traduit par J.P.Dumont, in *les présocratiques*, Bibliothèque de la Pléiade, Gallimard, 1988.
- PLOTIN ; « *Ennéades* », traduit par E. Bréhier, Les Belles Lettres, 1989.
- POPKORN, S. ; « *First steps in modal logic* », Cambridge University Press, Cambridge, 1994.
- SAINT THOMAS D'AQUIN ; « *Somme contre les gentils* », traduit par R. Bernier, M. Corvez, L. J. Moreau, 4 vol., Lethielleux, 1951-1961.
- SPINOZA, B. ; « *L'Ethique* », traduit par R. Misrahi, Presses Universitaires de France, 1990.
- WRIGHT, G.H. von ; « *An essay in modal logic* », North Holland, Amsterdam, 1951.

Articles :

- ALCHOURRON, C.E. ; GÄRDENFORS, P. ; MAKINSON, D. ; « On the logic of theory change : partial meet contraction and revision functions », *Journal of symbolic logic* 50, 1985, pp.510-530.
- CASTAÑEDA, H.-N. ; « Review of Knowledge and Belief », *Journal of symbolic logic*, 1964, pp.132-134.
- FAGIN, R. ; HALPERN, J.Y. ; « Belief, awareness and limited reasoning », *Artificial intelligence* 34, 1988, pp.39-76.
- FINE K. ; « In so many possible worlds », *Notre Dame journal of formal logic* 13(4), 1972, pp.516–520
- HALPERN, J.Y. ; MOSES, Y. ; « Knowledge and common knowledge in a distributed environment », *Journal of the ACM* 37(3), 1990, pp.549-587.
- HOEK, W. van der ; « Systems for knowledge and belief », *Journal of logic and computation*, 3(2), 1993, pp.173-195.

- HOEK, W. van der ; MEYER, J.-J.CH. ; « Graded modalities in epistemic logic », *Logique et analyse* 133-134 (édition spéciale pour le symposium international sur la logique épistémique), 1991, pp.251–270.

- KRAUS, S. ; LEHMANN, D. ; « Knowledge, belief and time », *Theoretical Computer Science* 58, 1988, pp.155-174.

- KRIPKE, S. ; « Semantic analysis of modal logic », *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 1963, pp.67-96.

- KRIPKE, S. ; « Semantic analysis of modal logic II : non-normal modal propositional calculi », in *Symposium on the theory of models*, North-Holland, Amsterdam, 1965.

- LEHMANN, D.J. ; « Knowledge, common knowledge and related puzzles », *Proc. 3rd ACM symp. on principles of distributed computing*, 1984, pp. 62-67.

- LEVESQUE, H.J. ; « A logic of implicit and explicit belief », *Proceedings of the national conference on artificial intelligence*, 1984, pp.198-202.

- MAREK, W. ; TRUSZCZYNSKI, M. ; « Autoepistemic logic », *Journal of the ACM* 38(3), 1991, pp.588-619.

- MOORE, R.C. ; « Possible-world semantics for autoepistemic logic », *Proceedings of the non-monotonic reasoning workshop*, New Paltz NY, 1984, pp.344-354.

- MOORE, R.C. ; « A formal theory of Knowledge and action », in *Formal theories of the commonsense world*, édité par J.R. Hobbs et R.C. Moore, Ablex, Norwood, New Jersey, 1985, pp.319–358.

- RANTALA, V. ; « Impossible world semantics and logical omniscience », *Acta philosophica fennica* 35, 1982, pp.106-115.

- ZADEH, L. ; « Knowledge representation in fuzzy logic », *Tkde* 1, 1989, pp.89–100.

Table des matières

INTRODUCTION ---- 2

CHAPITRE 1 : PANORAMA HISTORIQUE ET THÉMATIQUE DU DOMAINE ---- 4

CHAPITRE 2 : LE SYSTÈME DE VAN LINDER ---- 8

DÉFINITION DU LANGAGE ---- 9
INTERPRÉTATION SÉMANTIQUE EN MODÈLES DE KRIPKE ---- 10
Définition du modèle d'interprétation pour les formules épistémiques ---- *11*
Extension aux formules doxastiques ---- *12*
AXIOMES DU SYSTÈME ---- 14
QUELQUES THÉORÈMES SUPPLÉMENTAIRES ---- 16

CHAPITRE 3 : ANALYSE DU PARADOXE ---- 17

ÉNONCÉ DU PARADOXE ---- 17
PREUVE DU PARADOXE ---- 17
CAUSES DU PARADOXE ---- 18

CHAPITRE 4 : LE NOUVEAU SYSTÈME ---- 21

DÉFINITION DU LANGAGE ---- 21
INTERPRÉTATION SÉMANTIQUE ---- 21
AXIOMES DU SYSTÈME ---- 23
PREUVE DE CORRESPONDANCE ENTRE L'AXIOME A15 ET LA CONDITION SÉMANTIQUE N°3 DE NOTRE SYSTÈME ---- 24
Remarques préliminaires : ---- *24*
Preuve : ---- *24*
PREUVE DE LA DISPARITION DU PARADOXE ---- 26
PREUVES DE CERTAINES CONSÉQUENCES DANS LE NOUVEAU SYSTÈME ---- 28
Liste des théorèmes à vérifier ---- *28*
Preuves de la validité ou de la non-validité des théorèmes ---- *29*

CONCLUSION ---- 45

BIBLIOGRAPHIE ---- 46