Bulletin d'analyse phénoménologique VI 8 (Actes 3), p. 64-82
ISSN 1782-2041 http://popups.ulg.ac.be/bap.htm

La logique des verbes intentionnels

Par PAUL GOCHET
Université de Liège

1. Introduction

Dans l'encyclopédie *Wikipedia* accessible sur Internet, les thèses de Brentano sont résumées dans les trois phrases suivantes :

> L'intentionnalité est une caractéristique de la conscience.
>
> L'intentionnalité peut être caractérisée par quelques formules : « contenir quelque chose (pas forcément réel) à titre d'objet », « être à propos de quelque chose », « avoir un objet immanent ».
>
> Selon Franz Brentano, l'intentionnalité est le critère permettant de distinguer les « faits » psychiques des « faits » physiques : tout fait psychique est intentionnel, c'est-à-dire qu'il contient quelque chose à titre d'objet bien que ce soit toujours d'une manière différente (croyance, jugement, perception, conscience, désir, haine, etc.).

Comme le souligne Denis Seron, l'intentionnalité pour Brentano est une relation, mais ce n'est pas une relation comme les autres dans la mesure où elle n'exige pas l'existence de tous ses termes [Seron 2010]. C'est pour capter cette particularité que Brentano emprunte à saint Thomas le concept d'*inexistence intentionnelle* [McCormick 1981]. Par exemple, on peut penser à un objet non réel comme la planète inexistante Vulcain dont certains astronomes du XIX^e^ siècle avaient postulé l'existence pour expliquer les perturbations de l'orbite de Mercure.

Il est clair qu'une relation intentionnelle comme « penser à » aura une logique à part qui interdira d'inférer de « le sujet A pense à Vulcain », la conclusion « il existe un objet (réel) auquel le sujet A pense ». Nous ne traiterons pas ici des logiques inventées pour rendre compte des raisonne-

ments sur les êtres inexistants [Gochet 2010].

Brentano subsume sous le concept de relation intentionnelle des actes mentaux qui s'expriment en langue naturelle par des verbes qui admettent des constructions grammaticales différentes. L'acte mental de *penser à* s'exprime au moyen d'un verbe intransitif qui unit un terme singulier à un autre. L'acte mental de *désirer* s'exprime soit au moyen d'un verbe transitif qui unit un terme singulier à un autre, soit au moyen d'un verbe qui unit un terme singulier à une phrase complète enchâssée dans la première et introduite par la conjonction « que ». Comparez « A désire le gâteau [qui est en face de lui] » à « A désire *que* le gâteau [qui est en face de lui] lui soit remis ».

Pour capter les inférences autorisées par ces deux constructions, on aura besoin de faire appel à des chapitres différents de la logique. Dans la première construction, « désire » est traité comme un prédicat binaire de la logique du premier ordre. Dans la deuxième construction, « désire » est traité comme un opérateur de la logique modale propositionnelle. Russell appelle les actes ou états mentaux exprimés par les verbes intentionnels « croire », « désirer », « douter », etc., des *attitudes propositionnelles* [Russell 1940, 1969, p. 29]. Cette thèse a été reprise par les philosophes de l'école analytique Nathan Salmon et Scott Soames dans l'introduction à leur anthologie [Salmon & Soames 1988, p. 1].

Il existe même une troisième construction grammaticale faisant intervenir des verbes intentionnels. Elle a étudiée par Jon Barwise et John Perry pour ses particularités logiques. Il s'agit de la construction *verbe de perception + infinitif*. L'énoncé « Whitehead a vu Russell *faire un clin d'œil* » obéit à la logique extensionnelle (les expressions coextensives y sont interchangeables *salva veritate*), mais ce n'est pas le cas de l'énoncé « Whitehead a vu *que* Russell faisait un clin d'œil ». Ici il faut faire appel à la logique intensionnelle qui interdit l'échange d'expressions coextensives [Barwise et Perry 1983, Gochet 1983].

Dans cet article, nous nous limiterons à l'étude de la logique des constructions grammaticales de la deuxième espèce : *la logique des attitudes propositionnelles*. Avant d'entamer cette étude, il convient d'examiner des objections récentes qui mettent en cause l'idée même d'attitude propositionnelle conçue comme une relation binaire reliant un agent à une proposition.

En 2003, Frederike Moltmann a publié un article intitulé « Propositional attitudes without propositions » dans lequel elle présente de nombreuses données linguistiques qui militent en faveur de la thèse soutenue autrefois par Russell (dans la deuxième décennie du vingtième siècle), thèse selon laquelle une subordonnée introduite par « que » n'exprime pas en

premier lieu une relation binaire entre un agent et une *proposition*, mais spécifie plutôt une relation entre un agent et des *constituants* de la proposition.

Dans une contribution aux *Proceedings* de l'*Aristotelian Society* de 2009, Trenton Merricks avance des arguments philosophiques nouveaux étayant la thèse que si la *croyance* est bel et bien une attitude propositionnelle, la *crainte* ou le *désir*, en revanche, ne sont jamais des attitudes propositionnelles. À l'appui de la thèse de Merricks, on peut invoquer aussi la notion de *direction d'ajustement* introduite initialement par Searle pour expliquer la différence entre les conditions de satisfaction d'*actes de langage* comme l'assertion et l'ordre, mais étendue ensuite par le même auteur aux *états intentionnels* [Searle 1983]. La croyance, comme l'assertion, doit se conformer à la réalité. Au contraire le désir comme l'ordre est satisfait si c'est la réalité qui se conforme à eux (inversion de la direction d'ajustement précédente).

Dans la Section 2, nous exposerons et commenterons la sémantique formelle générale de l'intentionnalité proposée par Graham Priest en 2005 dans *Towards Non-Being. The Logic and Metaphysics of Intentionality*.

2. La sémantique formelle de l'intentionnalité

Priest commence par construire une langue formelle pour la logique de l'intentionnalité qui comporte, outre les connecteurs usuels de la logique propositionnelle standard, un nouvel opérateur, *l'opérateur intentionnel* Ψ qui se combine, à sa gauche, avec un terme singulier *t* et, à sa droite, avec une formule A. On représente formellement les expressions d'attitude propositionnelle telles que « croit que », « sait que », « désire que », craint que », etc., à l'aide de cet opérateur intentionnel unique.

Priest signale une particularité logique intéressante de l'opérateur d'intentionnalité: *il n'est pas fermé sous la relation d'implication logique*.

Un opérateur Θ est dit fermé sous l'*implication logique* s'il vérifie le schéma suivant :

Si $\vdash A \supset B$
alors $\vdash \Theta A \supset \Theta B$

L'*opérateur modal* « nécessairement » ($\Box$) satisfait cette condition. Voici un exemple :

$\vdash (p \supset r) \supset ((p \wedge q) \supset r)$

$\vdash \Box (p \supset r) \supset \Box ((p \wedge q) \supset r)$

L'opérateur intentionnel Ψ, au contraire, ne la satisfait pas. On peut avoir :

$\vdash A \supset B$
et non $\vdash t\Psi\ A \supset t\Psi\ B$

Priest donne l'exemple suivant :

« Je mange mon gâteau » implique « je n'ai plus mon gâteau », mais « je désire manger mon gâteau » n'implique pas « je désire ne plus avoir mon gâteau ». Commentant l'exemple, il écrit : « Je désire à la fois avoir mon gâteau et le manger. Irrationnel ? Peut-être, mais les gens sont comme ça » [Priest, *ibid.* 21]. On a donc bien ici un cas où ce schéma de fermeture sous l'implication logique n'est pas vérifié.

Remarque : dans l'exemple de Priest, l'implication qui lie « je mange mon gâteau » à « je n'ai plus mon gâteau » n'est pas une implication logique au sens strict du terme (un conditionnel valide en vertu du sens des connecteurs qu'il contient), mais c'est toute de même une implication analytique (un conditionnel valide en vertu du *sens* des mots « manger » et « avoir »).

Priest a entrepris de construire une sémantique formelle qui permette de donner un *sens*, c'est-à-dire des *conditions de vérité*, aux formules contenant l'opérateur Ψ. Une sémantique formelle comporte deux parties : (1) la construction d'un modèle, (2) la formulation d'une définition récursive de la vérité pour le langage formel proposé.

Le *modèle de Priest* pour la logique de l'intentionnalité s'inspire au départ du *modèle de Kripke* [Kripke 1963] pour la logique modale. Un modèle de Kripke est constitué d'un ensemble de *mondes possibles P*, d'une relation d'accessibilité *R* entre ces mondes et d'une fonction d'interprétation ℑ qui assigne à chaque proposition atomique la classe des mondes dans lesquels cette proposition est vraie. (*P*, *R*, ℑ). La relation d'accessibilité diffère selon l'opérateur modal qu'il s'agit d'interpréter (opérateur temporel, épistémique, déontique, etc.) [Blackburn *et al.*2001]. Le monde réel désigné par @ figure parmi les mondes possibles.

Priest enrichit le modèle de Kripke en introduisant un *ensemble de mondes impossibles I*, des mondes inventés par le logicien finlandais Veikko Rantala [Rantala 1982a, 1982b] pour traiter un problème bien connu en logique épistémique, le *problème de l'omniscience logique*. À ces mondes possibles et impossibles, Priest ajoute une troisième catégorie de mondes, les mondes ouvert *O* (ouverts sur la relation d'implication logique). Ces mondes,

écrit Priest, « réalisent comment les choses sont censées être pour constituer le contenu d'états intentionnels quelconques » [Priest 2005, 21].

Les mondes impossibles sont des mondes dans lesquels les lois logiques peuvent être différentes. Différentes en quoi ? Priest répond que lorsqu'on évalue une formule — par exemple lorsqu'on assigne une valeur de vérité à un conditionnel ($p \supset q$) relativement à un monde impossible —, on peut aller jusqu'à se donner la liberté de le traiter comme un atome *r*. Les mondes ouverts sont plus anarchiques encore : « Les conditionnels pouvaient se comporter arbitrairement relativement aux mondes impossibles, toutes les formules peuvent se comporter arbitrairement quand on les évalue relativement aux mondes ouverts » [Priest, *ibid*., 22].

Examinons à présent la *définition récursive de la vérité* en nous bornant à la clause servant à interpréter l'opérateur modal chez Kripke et l'opérateur intentionnel chez Priest. Pour interpréter l'opérateur modal de nécessité, on utilise le clause suivante qui est standard :

> $\left|\left|\Box A\right|\right|$ est vrai dans *w* ssi pour tous les *w'* $\in$ *P* tels que *wRw'*, $\left|\left|A\right|\right|$ est vrai dans *w'*

Ce qui se lit : « nécessairement A » est vrai dans le monde *w* si et seulement si A est vrai dans tous les mondes *w'* accessibles à partir de *w*.

Hintikka a identifié le pouvoir explicatif de clauses de ce genre : « Pour dire l'essentiel brièvement et crûment, en marchant d'un monde à ses alternatives, nous pouvons réduire les conditions de vérité des énoncés modaux aux conditions de vérité d'énoncés non modaux » [Hintikka 1973, 193].

Pour interpréter l'opérateur intentionnel Ψ, Priest procède de la même manière, mais en modifiant les mondes possibles et la relation d'accessibilité. Ce qui est nouveau c'est que la relation d'accessibilité a désormais accès aux *mondes ouverts* au sens défini plus haut. On désigne cette nouvelle relation d'accessibilité par R^{Ψ}. La clause qui donne les conditions de vérité de $t\ \Psi$ A s'énonce ainsi :

> $\left|\left|t\ \Psi A\right|\right|$ est vrai dans *w* ssi pour tout *w'* $\in$ *W* tel que *w* R^{Ψ} *w'*, $\left|\left|A\right|\right|$ est vrai dans *w'*

Si le verbe intentionnel « désire » est substitué à Ψ, nous obtenons l'instance suivante de la clause :

L'énoncé « l'agent *t désire que* A » est vrai dans *w* si et seulement si l'énoncé « A » est vrai dans tous les mondes *w*' compatibles avec les désirs de *t*.

3. Un exemple abstrait de non-fermeture sous l'implication logique

Priest a montré *sur un exemple abstrait* que sa théorie des modèles pour les opérateurs intentionnels permettait de rendre compte de la non-fermeture en forgeant un contre-modèle approprié. Cette dernière tâche est assez délicate et permet de comprendre rétrospectivement pourquoi Priest se donne licence d'interpréter des formules moléculaires comme si elles étaient des atomes et pourquoi il s'autorise à utiliser une sémantique qui n'obéit pas au *principe de compositionnalité* [Gochet 2008] selon lequel le sens (ici l'extension) du tout est fonction du sens des parties et de la structure grammaticale.

Dans l'exemple traité dans son livre, Priest cherche à construire un modèle dans lequel

$(Pa \wedge Qa) \supset Pa$ est valide mais $t\ \Psi(Pa \wedge Qa) \supset t\ \Psi\ Pa$ ne l'est pas.

La formule $(Pa \wedge Qa) \supset Pa$ est vraie par hypothèse. Il faut, à présent, trouver une interprétation dans laquelle la formule $t\ \Psi\ (Pa \wedge Qa)$ soit vraie et la formule $t\ \Psi\ Pa$ soit fausse.

Le contre-modèle de Priest contient seulement deux mondes : le monde réel @ qui est fermé sous l'implication logique et le monde ouvert *w*. La relation d'accessibilité conduit de @ à *w* et nulle part ailleurs. Le contre-modèle est représenté dans la figure ci-dessous :

$@ \bullet \rightarrow \bullet\ w$

En vertu de la clause qui définit Ψ, montrer que $t\ \Psi\ (Pa \wedge Qa)$ est vrai en @, *se réduit* à montrer que $(Pa \wedge Qa)$ est vrai en *w*. Rappelons-nous le commentaire de Hintikka. De manière analogue, montrer que $t\ \Psi\ Pa$ est faux en @ *se réduit* à montrer que Pa est faux en *w*.

Sachant que *w* est un monde ouvert, la formule moléculaire $(Pa \wedge Qa)$ peut être traitée comme *atomique.* Profitons de cette latitude et traitons cette formule comme si elle avait la forme Rab et donnons-lui une interprétation qui la rende vraie. Dans ce but, Priest assigne à $Px_1 \wedge Gx_2$ (c'est-à-dire à Rx_1x_2) l'extension D×D et assigne aux termes singuliers « a » et « b » des

individus qui sont membres du domaine D. Sous cette interprétation $t\ \Psi$ (Pa $\wedge$ Qa) est vrai.

Il faut, pour terminer, donner une interprétation à Pa qui rende cette formule fausse dans les mondes ouverts. À cette fin, Priest assigne l'extension $\varnothing$ à Px_1 et assigne un individu membre du domaine D au terme singulier « a ». Manifestement, dans cette interprétation *Pa* est fausse en *w* et $t\ \Psi$Pa est fausse en @.

On serait tenté d'objecter que dès que la formule (Pa $\wedge$ Qb) est traitée comme atomique, elle ne peut plus être considérée comme étant l'antécédent de (Pa $\wedge$ Qa) $\supset$ Pa. Cette objection n'est cependant pas valable. Priest ne change pas la *forme syntaxique* des formules qu'il considère comme atomiques, mais leur assigne *une sémantique non compositionnelle* grâce à laquelle il lui est possible de construire le contre-modèle dont il a besoin. [Je remercie Shahid Rahman pour son aide substantielle dans l'élucidation de ce point difficile.]

4. La portée et les limites de la logique de l'intentionnalité de Priest

Priest s'est fixé pour objectif de caractériser *l'intentionnalité en général*. Il travaille avec un seul opérateur intentionnel (Ψ) et se donne une seule relation d'accessibilité *sans structure particulière*. On ne doit donc pas reprocher à la sémantique de Priest d'ignorer les *différences* entre les diverses attitudes propositionnelles (croyance, crainte, désir). Ce n'était pas son propos de les représenter. D'autres s'en chargeront dans le cadre de l'architecture BDI (*Belief, Desire, Intention*). Les logiciens qui ont élaboré cette architecture se sont également intéressés à la non-fermeture sous d'autres implications que l'implication logique (implication causale, implication causale à laquelle croit l'agent). Nous examinerons leur apport dans la section 7.

Priest fait néanmoins une certaine place à des attitudes propositionnelles spécifiques telles que « être rationnellement engagé à *p* » qui, contrairement à l'intentionnalité en général, sont fermées sous l'implication logique.

Dans la première édition de *Introduction to Non-Classical Logic*, ouvrage paru en 2001, Priest a présenté une méthode de preuve par tableaux (N4) pour la sémantique qui sera proposée quatre ans plus tard dans *Towards Non-Being*. En raison des caractéristiques de la sémantique que avons décrites au début de cette section, la logique N4 est — Priest le reconnaît de bonne grâce — « presque triviale ».

On peut sans doute contester l'affirmation de Priest selon laquelle les agents humains peuvent héberger dans leur esprit des *désirs contradictoires* tels que désirer à la fois avoir son gâteau et le manger. Sommes-nous donc si irrationnels ? On verra plus loin que Caroline Semmling et Heinrich Wansing ont formalisé la notion plus faible de *désirs conflictuels*.

L'exemple abstrait de non-fermeture sous l'implication logique de Priest est très éclairant, mais c'est un exemple *inventé pour illustrer* la sémantique formelle proposée et non un exemple *observé indépendamment* de cette sémantique et susceptible de la justifier. Or il se trouve qu'il existe bel et bien un exemple concret et naturel de non-fermeture sous l'implication logique. Celui-ci a été étudié par les logiciens qui ont développé la logique de l'opérateur *Stit* (*see to it that*). Nous l'examinerons dans la prochaine section. Mais — fait troublant — la sémantique formelle de Priest construite pour rendre compte de la non-fermeture de Ψ sous l'implication logique ne rend pas compte de la non-fermeture de *Stit* sous cette même implication attestée par l'exemple en question.

5. Un exemple concret de non-fermeture sous l'implication logique

Dans *Facing the future, Agents and Choices in Our Indeterminist World*, Nuel Belnap, Michael Perloff et Min Xu offrent l'exemple suivant de non-fermeture sous l'implication logique: « S'il y a au moins un homme blessé qui est pansé, alors il y a au moins un homme blessé » n'implique pas « Si *vous veillez à ce qu*'il y ait au moins un homme blessé qui est pansé, alors, *vous veillez à* ce qu'il y ait au moins un homme blessé ». Formellement la non-validité de l'inférence s'écrit ainsi :

$$\models (A \wedge B) \supset A \text{ mais } \not\models \textit{Stit}\ (A \wedge B) \supset \textit{Stit}\ A$$

Les auteurs précités soulignent qu'il n'y a rien de paradoxal dans cette absence de validité du deuxième conditionnel ci-dessus et qu'on n'est nullement en présence d'une logique bizarre ou d'une subtilité grammaticale, mais qu'on est en présence, au contraire, de quelque chose qui est à la fois « typique » et « profondément inscrit dans l'idée d'action basée sur un choix » [Belnap *et al*. 2001, 40].

La plupart des philosophes qui traitent d'ontologie se préoccupent de *ce qui est* [Quine 1960]. Belnap et ses co-auteurs se préoccupent de *ce que les gens font* [Antoniol 1998]. Ils ont élaboré à la fois une théorie des

modèles et une théorie de la preuve pour la logique *stit* qui se révélera ultérieurement étroitement liée à la logique de l'intentionnalité.

Ici aussi, une théorie des modèles doit être construite pour expliquer le manque de fermeture de l'opérateur *stit* sous l'implication logique. Avant d'esquisser la théorie des modèles pour la logique *stit*, il convient de rappeler comment on exprime l'opérateur *stit*, ou plutôt l'opérateur *astit* (*achievement stit operator*). Pour Belnap et ses co-auteurs, [α *astit* : A] signifie « le fait (momentané) que A est garanti par un choix antérieur de l'agent α » [Belnap *ibid*. 33].

Pour représenter en sémantique formelle l'*achievement stit*, on construit une *structure*, c'est-à-dire un n-uple, <*A*, *T*, *I*, ≤, *C*> dont le premier élément est un ensemble d'*agents* conçus comme faisant des choix dans le temps. Le second élément est un *arbre*, dont les branches représentent les différents cours des événements (appelés histoires) qui peuvent se réaliser ou non, en fonction des choix faits par les agents. Le troisième élément est un ensemble d'*instants*. Le quatrième est la *relation* « antérieur ou simultané à » défini sur *I*. Un instant est un ensemble de moments contemporains traversant différentes histoires. Le cinquième est une fonction de *choix*. Elle associe à chaque agent α et à chaque moment *m* une partition des histoires $H_{(m)}$ à travers *m*. L'agent doit choisir entre différentes classes d'équivalences d'histoires. Belnap et ses co-auteurs soutiennent que les structures exhibées par la sémantique de l'opérateur *stit* « ne sont pas de simples curiosités mathématiques, mais qu'elles décrivent — à une idéalisation près — le monde dans lequel les agents agissent » [Belnap *et al*., *ibid*., 36].

Pour transformer une *structure* (*frame*) en *modèle*, nous avons besoin d'une fonction d'interprétation ℑ qui associe chaque formule atomique à un ensemble de paires *m*/*h* (paires dont le premier terme est un moment et dont le deuxième est le cours de l'histoire auquel le moment appartient).

Pour obtenir une sémantique, il reste à formuler les clauses d'une définition récursive de la vérité. Ce qui est neuf ici, c'est que la formule *stit A* est évaluée non pas *relativement à* un moment *m*, mais *relativement à* une paire *m*/*h*.

La clause qui interprète *stit A* doit satisfaire deux exigences : (1) la vérité de *A* maintenant résulte d'un choix antérieur de l'agent (exigence positive), (2) la vérité de *A* n'était pas déjà fixée au moment précédent (exigence négative).

Le modèle ébauché plus haut et les deux exigences qui viennent d'être mentionnées *expliquent pourquoi* l'inférence citée au début de cette section est non valide. L'exigence 2 a été transgressée. En effet, on ne peut faire

advenir au temps $t + 1$ l'état de chose décrit par « il y a un homme blessé » si l'homme était déjà blessé au temps t.

On observera que l'explication du manque de fermeture sous l'implication logique requiert une structure temporelle particulière. Cela jette un doute sur la *généralité* de la structure construite par Priest pour expliquer le manque de fermeture sous l'implication logique manifestée par les constructions intentionnelles formalisées par l'opérateur Ψ.

6. Implication logique, implication causale et implication causale crue *(believed causal implication)*

Rappelons d'abord ce qui distingue l'implication causale de l'implication logique. L'implication logique est un conditionnel vérifonctionnellement valide [Quine 1982], c'est-à-dire une formule qui est valide en vertu des connecteurs vérifonctionnels qu'elle contient [Nous ne dirons rien dans cet article de la validité quantificationnelle]. La formule ci-dessous illustre la validité véritconditionnelle :

$$(p \supset r) \supset ((p \wedge q) \supset r)$$

Comment alors certaines de ses instances sont-elles non valides ? [Nous remercions Joost Joosten et Frank Veltman de nous avoir indiqué le problème.] Considérons l'inférence suivante :

> Si j'absorbe du poison [en dose suffisante] je meurs, donc si j'absorbe du poison et que j'absorbe du contrepoison, je meurs.

Ce conditionnel n'est pas valide. Or il a l'air d'être une instance de la formule citée plus haut, mais ce n'est pas le cas. Les conditionnels à gauche et à droite de *donc* sont des conditionnels non vérifonctionnels. Le conditionnel causal est *non monotone*. On ne peut enrichir la prémisse sans risquer de rendre fausse la conclusion. Dès lors, on ne peut représenter le conditionnel causal par le connecteur vérifonctionnel $\supset$ sauf si on place ce dernier dans la portée de l'opérateur « nécessairement » ou « inévitablement » : « il est inévitable que $\varphi \supset \psi$ », mais alors il est clair qu'il ne s'agit plus du connecteur vérifonctionnel de la logique propositionnelle classique.

On peut obtenir une construction encore plus complexe en mettant le conditionnel causal dans la portée de l'opérateur épistémique BEL. On for-

malise alors l'implication causale « crue » (*believed causal implication*) grâce à la formule : BEL (inévitable $\varphi \supset \psi$).

Cette mise au point était nécessaire au moment d'examiner les cas de non-fermeture de l'*intention* et du *but* sous la *believed causal implication* qui ont été fort discutés dans la communauté des chercheurs en intelligence artificielle.

7. La logique de la croyance, du désir et de l'intention dans l'architecture BDI

Dans un article fameux publié dans *Artificial Intelligence*, Philippe Cohen and Hector Levesque ont construit un exemple qui montre que la locution verbale « a l'intention de » n'est pas fermée sur la *believed causal implication*. L'exemple fait aujourd'hui partie du folklore du sujet et nous le citons *in extenso* :

> […] un agent avait l'intention d'avoir sa dent plombée. Ignorant tout des anesthésiques (on pourrait supposer que ceci se passe au moment où les anesthésiques viennent tout juste d'être introduits en dentisterie), l'agent croit qu'il est toujours le cas que si on lui plombe une dent, il aura mal. On pourrait à la rigueur dire que l'agent *choisit* d'avoir mal. Néanmoins, on n'est pas prêt à dire qu'il a l'intention d'avoir mal » [Cohen & Levesque 1990, 251].

L'agent a l'intention de se faire plomber une dent. Il croit que se faire plomber une dent implique causalement avoir mal. Mais il n'a pas l'intention d'avoir mal. L'intention n'est donc pas fermée sur la *believed causal implication*. Il incombe au logicien de proposer une sémantique formelle qui rende compte de cette non-fermeture.

Un an après Cohen et Levesque, deux logiciens appartenant à la communauté des chercheurs en Intelligence artificielle, Anand Rao et Michael Georgeff, ont élaboré un formalisme, *l'architecture BDI* , qui résout ce problème.

Le formalisme de Rao et Georgeff est une logique modale du premier ordre qui contient la logique temporelle CTL* (*Computer Tree Logic*). Au lieu de traiter chaque monde possible comme une séquence linéaire d'événements, ils considèrent chaque monde possible comme un arbre (temps ramifié).

Les opérateurs modaux « BEL », « GOAL » et « INTEND » sont interprétés en termes de relations d'accessibilité : R_B pour « BEL », R_G pour « GOAL », R_I pour « INTEND ». Le traitement de la croyance est standard :

une formule est dite *crue au moment t* par un agent si et seulement si elle est vraie dans tous les mondes accessibles par la relation R_B [Gochet & Gribomont 2006]. Rao et Georgeff suivent le même canevas pour l'interprétation de l'opérateur « GOAL » : « la relation d'accessibilité pour les buts spécifie les situations dans lesquelles l'agent *désire* être […]. Un agent a un but φ au moment *t* si et seulement si φ est vrai dans tous les mondes accessibles à l'agent par la relation R_G au temps *t* » [Rao & Georgeff 1991, 477].

Les relations d'accessibilité R_G et R_I n'ont d'autre propriété que la sérialité. Elles ont donc une structure assez pauvre qui ne permettrait pas, en partant de la relation d'accessibilité, de dire quel opérateur elles interprètent. L'intérêt de la sémantique formelle de Rao et Georgeff est ailleurs. Il est dans les axiomes qui définissent les relations unissant les opérateurs intentionnels entre eux (la croyance au but, le but à l'intention). Un de ces axiomes énonce que les agents croient que leurs buts peuvent être atteints dans le futur.

L'objectif visé par la sémantique de Rao et Georgeff n'est pas de donner une analyse conceptuelle satisfaisante de la croyance, du désir et de l'intention. Il est de servir à la construction d'une modèle dans lequel l'opérateur GOAL et l'opérateur INTEND ne sont pas fermés sous l'implication causale crue. Or cet objectif est atteint. Il est aisé de construire un modèle dans lequel, par exemple, la conjonction ci-dessous est satisfaisable si l'on substitue « se faire plomber une dent » à φ et « avoir mal » à ψ :

$$\text{GOAL } \varphi \wedge \text{BEL } (\textit{inévitable } \varphi \supset \psi) \wedge \neg \text{GOAL } \psi$$

8. Les combinaisons de l'architecture BDI et de la logique STIT

Semmling et Wansing soutiennent que les désirs peuvent entrer en conflit sans pourtant être incohérents. Je peux tout à la fois *désirer* faire don d'un de mes reins pour sauver mon frère, et *désirer ne pas faire ce don* qui affectera mon intégrité physique. Cependant, en tant qu'agent rationnel, il n'est pas vrai que je désire tout à la fois faire don et ne pas faire don d'un rein, ce qui reviendrait à entretenir dans mon esprit des désirs incohérents. En termes formels, (DES $p \wedge$ DES$\neg p$) n'implique pas DES ($p \wedge \neg p$).

Pour représenter la différence entre désirs conflictuels et désirs incohérents, on a besoin d'une logique dans laquelle la loi de logique modale classique ($\Box p \wedge \Box q) \supset \Box (p \wedge q)$ cesse d'être valide. Une telle logique existe, mais elle exige de nous que nous abandonnions la *sémantique relationnelle* de Kripke au profit de la *sémantique des voisinages* de Montague et de Scott

lorsque nous construisons une sémantique formelle pour l'opérateur « désirer » [Chellas 1980, 210 ; Straetmans 1991]. C'est ce que font Semmling et Wansing.

9. Une combinaison de la logique Stit avec la logique épistémique

Jusqu'ici il a été question du manque de fermeture des constructions intentionnelles sous *l'implication logique* puis sous la *believed causal implication*. Maintenant nous allons voir, dans les travaux de Jan Broersen, un exemple de non-fermeture de constructions intentionnelles sous *l'implication causale*.

Broersen construit un langage formel qui contient les constructions suivantes en plus des connecteurs vérifonctionnels de la logique propositionnelle classique :

> L'opérateur □ φ exprime *la nécessité historique*.
>
> L'opérateur *X* φ exprime que, *dans l'état prochain*, φ.
>
> L'opérateur [*A xstit*] φ signifie que les agents du groupe *A* conjointement *veillent à ce que*, dans l'état prochain, φ.
>
> L'opérateur K_a φ est l'opérateur standard de *connaissance.*
>
> L'opérateur [*a xint*] φ signifie que l'agent *a* exécute l'action qui sera réalisée dans l'état prochain. On peut le lire ainsi : « l'agent *a* exécute *intentionnellement* l'action qui produit l'état de chose φ ».

Une structure est fournie dans laquelle le *domaine* est un ensemble d'états dynamiques, c'est-à-dire de n-uples <*s,h*> où *s* est un membre d'un ensemble d'états de chose (*S*), *h* est un membre d'un ensemble d'histoires (*H*) et *s* est un membre de l'histoire *h*. Des *relations d'accessibilité* sont utilisées pour interpréter les opérateurs modaux et un système axiomatique (correct et complet) est présenté. Nous n'entrerons pas dans ces détails techniques. Nous nous bornerons à préciser la *portée philosophique* de la nouvelle sémantique formelle proposée par Broersen.

Les intentions (dans le sens courant du terme) ne sont plus traitées comme des *états mentaux.* Elles sont traitées comme des *modes d'action* et cela contribuera à expliquer pourquoi les intentions ne sont pas fermées sur les *effets collatéraux* des actions. (L'aviateur en mission a l'intention de bombarder un arsenal. Bombarder l'arsenal causera la destruction de l'hôpi-

tal qui en est proche (effet collatéral), mais l'aviateur n'a pas pour autant l'intention de détruire l'hôpital.)

Ensuite Broersen construit un nouvel opérateur en combinant l'opérateur *K* avec l'opérateur *stit* : l'opérateur composé qui en résulte, à savoir $K_a[a\ xstit]\varphi$, signifie que l'agent *a* accomplit *sciemment* l'action qui produit l'état de chose φ.

Faire intentionnellement φ implique logiquement *faire sciemment* φ, mais non réciproquement. Comme l'auteur le fait remarquer, un agent qui tue quelqu'un pour défendre sa propre vie tue sciemment, mais ne tue pas intentionnellement.

En exploitant la différence entre *faire sciemment* et *faire intentionnellement*, on peut apporter une solution au problème classique de la visite chez le dentiste. Broersen résume élégamment ce point dans le passage suivant :

> […] (1) l'agent visite intentionnellement le dentiste, (2) il visite sciemment le dentiste qui lui fera mal, et (3) il ne connaît pas une manière de visiter le dentiste qui lui permettrait de ne pas avoir mal » .De ces prémisses nous pouvons déduire que l'agent fait sciemment quelque chose qui entraîner une souffrance, mais nous ne pouvons pas déduire que l'agent a l'intention d'avoir mal [Broersen, à paraître].

On notera qu'une analyse satisfaisante des constructions intentionnelles en jeu dans la visite chez le dentiste exige que l'on tienne compte non seulement des *implications logiques* mais aussi *des implications causales* (*crues ou non*).

En combinant à leur tour la logique *Stit* avec *la logique épistémique*, Emiliano Lorini et François Schwartzentruber ont pu élaborer une logique qui permet d'étudier les raisonnements qui reposent sur deux attitudes propositionnelles souvent ignorées des philosophes : le « regret » et l'acte mental de se réjouir de quelque chose qui a eu lieu [Lorini et Schartzentruber, à paraître].

10. Recherches futures : un problème ouvert

Depuis près d'un demi-siècle, on étudie intensément la logique de « savoir que ». L'étude logique de « savoir comment » a commencé plus tard. Les analyses logiques de *savoir comment* proposées par les logiciens contiennent généralement, parmi leurs composants, le concept de *savoir que*. Or une

analyse philosophique récente de la distinction entre ces deux savoirs a fait surgir un problème inattendu auquel nous proposerons une solution.

Robert Moore, à qui l'on doit la première formalisation du *savoir comment*, observe qu'un agent peut savoir comment produire un état de chose *p* en exécutant une action complexe *A* sans savoir dès le début ce qu'il devra faire à chaque étape. Tout ce qu'il doit savoir, c'est ce qu'il doit faire à la première étape et *savoir qu'il* saura à chaque étape ultérieure ce qu'il devra faire [Moore 1985, 1995]. On voit dès à présent que le *savoir que* est imbriqué dans le *savoir comment*.

Cette imbrication est évidente aussi dans les analyses ultérieures consacrées par Munindar Singh dans sa thèse de doctorat sur le *savoir comment* publiée en 1994. La forme la plus achevée de sa définition du savoir comment, on la trouve dans un article qu'il a publié dans un livre collectif en 1999. Dans cet article, Singh présente le lemme d'équivalence suivant comme une caractérisation du *savoir comment.* Nous le formulons d'abord en langage naturel.

L'agent *x sait comment* produire l'état de chose (ou l'événement) *p* si et seulement si ou bien l'agent *x* sait que *p* est déjà le cas ou bien il y a une action possible *a* telle que, dans au moins un scénario, *x* sait que l'action *a* est exécutée et que pour tous les scenarii, si *a* est exécutée, *x* sait comment produire l'état de chose *p.*

Cette caractérisation du *savoir comment* est condensée dans la formule ci-dessous [Singh 1999], où E et A sont des quantificateurs de scenarii et « [ ] » se lit « si… alors- - - » :

$$K_t p \vee (\bigvee a : K_t (E \langle a \rangle \mathit{true} \wedge A [a] K_h p) \Leftrightarrow K_h p$$ [Singh 1999].

On peut savoir comment écrire un roman *(accomplishment*) ou savoir comment piloter un avion (*action*). Concentrons-nous sur un cas intermédiaire : savoir comment exécuter une certaine danse sur scène. Ce cas appartient à l'espèce de savoir comment que Frank Lihoreau appelle *savoir comment impliquant une capacité* [*ability-entailing knowledge-how*].

Les capacités et le savoir-faire impliquant des capacités sont *extensionnels* [Lihoreau 2008]. L'exemple suivant, dû à Carr et discuté par Lihoreau, le montre clairement. Imaginons qu'un danseur fameux exécute devant le public un article de son répertoire auquel il a donné le titre « Improvisation No 15 ». Imaginons que la suite des pas de danse qu'il exécute se révèle être, à son insu, une réplique parfaite des mouvements de la version dansée de l' « Élégie » de Gray. Nous serions prêts à dire, au sens de savoir comment

qui implique une capacité, que le danseur *sait comment* exécuter la version dansée de l' « Élégie » de Gray, mais nous ne serions pas prêts à concéder que le danseur *sait qu'*il exécute une version dansée de l' « Élégie » de Gray.

Considérons l'énoncé suivant : (1) « l'agent x sait comment on exécute l'improvisation No 15 ». La phrase subordonnée (2) « on exécute l'improvisation No 15 » n'est pas assertée, mais cela ne l'empêche pas d'avoir une valeur de vérité (comme c'est le cas aussi pour l'antécédent et le conséquent des conditionnels). Elle a la même valeur de vérité que la phrase (3) « on exécute la version dansée de l' "Élégie" de Gray ». Nous pouvons donc substituer *salva veritate* (3) à (2) dans (1). Le contexte formé par « l'agent x sait comment » est extensionnel. Au contraire, le contexte formé par « l'agent x sait que » est intensionnel. On ne peut substituer « il exécute une version dansée de l' "Élégie" de Gray » à « il exécute l'improvisation No 15 » bien que les deux énoncés aient la même valeur de vérité.

Certains seraient tentés de dire que l'analyse du *savoir comment* proposée par Lihoreau révèle un défaut dans la définition du *savoir comment* de Singh. Celui-ci utiliserait de manière équivoque la variable « p », sachant que cette variable apparaît à la fois dans la portée de l'opérateur intensionnel « Kt » et dans la portée de l'opérateur extensionnel « Kh ». Plus précisément on pourrait objecter que, dans le premier contexte, la variable prend comme valeurs des intensions et, dans le deuxième, des extensions.

L'objection n'est pas valable. Comme la lettre « p » n'est pas liée par un quantificateur dans le lemme de Singh, on peut la traiter comme un marque-place (ce que Quine appelle une « lettre schématique »). Or, contrairement aux variables, les marque-place n'ont pas de domaine de valeurs, ils admettent seulement des substituts (à savoir les constantes propositionnelles, c'est-à-dire les phrases déclaratives). Ceci laisse intacts les acquis de Lihoreau : les conditions d'échange de constantes propositionnelles dans la portée de « Kt » sont plus restrictives que les conditions d'échange dans la portée de « Kh », ce qu'on exprime parfois en disant que le premier opérateur crée un contexte opaque et le deuxième un contexte transparent.

Bibliographie

Antoniol L., *Things People Do*, PhD, Stirling, Stirling University, 1998.
Barwise J. et Perry J., *Situations and Attitudes*, Cambridge Mass., MIT Press, 1983.
Belnap N., Perloff M., Xu M., *Facing the Future*, Oxford, O.U.P., 2001.

Blackburn P., Rijke M. de, Venema Y., *Modal Logic*, Cambridge, C.U.P., 2001.
Broersen J., « A complete STIT Logic for Knowledge and Action, and Some of Its Applications », dans M. Baldoni *et al.* (éds.), *Lecture Notes in Artificial Intelligence*, 5397 (2009), p. 47-59.
Broersen J., « An *xstit*-Logic Analysis of Intentional Action », à paraître dans *The Journal of Philosophical Logic*.
Chellas B., *Modal Logic, an Introduction*, Cambridge, C.U.P., 1980.
Cohen Ph. et Levesque H., « Intention Is Choice With Commitment », *Artificial Intelligence* 42, (1990), p. 213-261.
Lorini E. et Schwarzentruber F., « A Logic for Reasoning about Counterfactual Emotions », à paraître.
Moltmann F., « Propositional Attitudes Without Propositions », *Synthese*, 135 (2003), p. 77-118.
Gochet P., « La sémantique des situations », *Épistémologie, Langage, Histoire*, tome 5, fasc. 2 (1983), p. 195-211.
Gochet P., « La formalisation du savoir-faire », avec des commentaires de M. Cozic, P. Egré, G. Sandu, dans Th. Martin et Ph. Mongin (éds.), *Logique épistémique et philosophie des mathématiques*, Paris, Vuibert (coll. « Philosophie des sciences »), 2007, p. 3-24.
Gochet P., « L'impact de la philosophie et de la logique sur la linguistique », 75 (2008), p. 6-14.
Gochet P., « La théorie de l'objet de Meinong à la lumière de la logique actuelle », Bour & *alii*, *Construction*, London, College Publications, 2010, p. 359-368.
Gochet P. et Gribomont P., « Epistemic Logic », dans Dov Gabbay et John Woods (éds.), *Handbook of the History of Logic*, vol. 7, Amsterdam, Elsevier, 2006, p. 99-195.
Hintikka J., « Grammar and Logic : Some Borderline Problems », dans J. Hintikka, J.M.E. Moravcsik et P. Suppes (éds.), *Approaches to Natural Language*, Dordrecht, D. Reidel, 1973, p. 197-214.
Kripke S., « Semantical Considerations on Modal Logic », dans L. Linsky (éd.), *Reference and Modality*, Oxford, O.U.P., 1963, 1971, p. 62-72.
Lihoreau F., « Knowledge-How and Ability », *Grazer Philosophische Studien*, 77 (2008), p. 263-305.
McCormick P., « Sur le développement du concept d'intentionnalité chez Brentano et Husserl », *Philosophiques*, 8 (1981), p. 227-237.
Merricks T., « Propositional Attitudes ? », *Aristotelian Society Proceedings*, vol. CIX (2009), p. 207-232.
Moore R., « A formal Theory of Knowledge and Action », 1985, dans R. Moore, *Logic and Representation*, Stanford, CSLI Lecture Notes, 1995, p. 27-70.
Priest G., *An Introduction to Non-Classical Logic*, Cambridge, C.U.P., 2001.
Priest G., *Towards Non-Being, The logic and metaphysics of Intentionality*, Oxford, O.U.P. 2005.
Priest G., *An Introduction to Non-Classical Logic*, second extended edition, Cambridge, C.U.P., 2008.

Quine W.V.O., *Methods of Logic*, 1950, 4e éd., 1980, Cambridge Mass., Harvard University Press, 1980 ; tr. fr. de la 3e éd. par Maurice Clavelin, Paris, Armand Colin, 1973.
Quine W.V.O., *Word and Object*, Cambridge Mass., MIT Press, 1960 ; tr. fr. de J. Dopp et P. Gochet, 2e éd., Paris, Flammarion, 1999.
Rao A.S. et Georgeff M.P., « Modeling Rational Agents Within a B.D.I. Architecture », dans J. Allen, R. Fikes, E. Sandewall (éds.), *Principles of Knowledge, Representation and Reasoning*, San Mateo, Morgan Kaufman, 1991, p. 473-484.
Rantala V., « Impossible Worlds Semantics and Logical Omniscience », *Acta Philosophica Fennica*, 35 (1982), p. 106-115.
Rantala V., « Non-Normal Worlds and Propositional Attitudes », *Studia Logica*, XLI (1982), p. 41-65.
Russell B., *Inquiry into Meaning and Truth*, 1940, tr. fr. de Ph. Devaux, Paris, Flammarion, 1969.
Searle J., *Intentionality, An Essay in the Philosophy of Mind*, Cambridge, C.U.P., 1983.
Salmon N. et Soames S., (éds.), *Propositions and Attitudes*, Oxford, O.U.P., 1988.
Semmling C. et Wansing H., « From BDI AND stit TO bdi-stit LOGIC », *Logic and Logical Philosophy*, 17 (1998), p. 189-211.
Semmling C. et Wansing H., « A sound and complete system of bdi-stit logic », dans M. Perlis (éd.), *Logical Yearbook* 2008, Londres, College Publications, 2009, p. 193-210.
Seron D., « L'ontologie de l'acte intentionnel : Un point de rencontre entre réalisme métaphysique et philosophie de l'esprit » (journée d'étude « Métaphysique analytique : thèmes et enjeux »), ULB, 29/1/2010.
Singh M., *Multiagent Systems. A Theoretical Framework for Intentions, Know-How, and Communications*, Lecture Notes in Computer Science 799, Berlin, Springer 1994.
Singh M., « Know-How », dans M. Wooldridge et A. Rao (éds.) *Foundations of Rational Agency*, Dordrecht, Kluwer Academic Publishers, 1999, p. 105-132.
Straetmans M., *Logiques modales non normales et logiques modales non monotones*, M.A. thèse, Liège, Département de mathématique de l'Université de Liège (1991).

Remerciements

Je remercie mes collègues A. Dewalque et D. Seron, organisateurs du Séminaire sur l'intentionnalité, de leur aimable invitation et Bruno Leclercq, président de séance, dont les questions lors de la discussion m'ont suggéré des remaniements. Je remercie aussi de leur aide les professeurs Pascal Gri-

bomont, Angel Nepomuceno, Shahid Rahman, Jacques Riche et les directeurs de recherche Hans van Ditmarsch, Andreas Herzig ainsi que le Dr Joost Joosten.