1.2 Théologie et modalité

1.2.1 Théologie

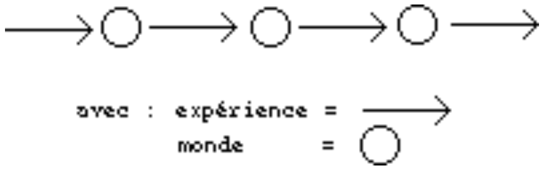
J'entends par *théologie*[1] : l'étude ou les discours sur les *états*, les *mondes* ou les *situations possibles* (je dirai simplement *monde*) en relation avec les possibilités de conscience, de vie, et de survie de soi (de son âme) et des autres (de leur âmes). L'*âme* est conçue de façon indexicale, en gros, *votre* âme, c'est ce à quoi *vous* tenez, c'est ce que vous estimez nécessaire de préserver après une expérience pour pouvoir penser a priori que vous allez survivre à cette expérience.

Lorque le soi survit (lorqu'*on* survit), le soi *(on)* survit dans un monde, à une expérience personnelle. Schématiquement :

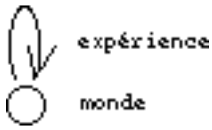
---

[1] Le terme *théologie* serait apparu la première fois chez Platon comme science (discours) sur les Dieux. Aristote l'a considérée comme une science, et St Thomas a élaboré cet aspect (Benoît 1989). L'adoption du dualisme cartésien permet une séparation de la théologie et de la science. Il me semble que cette séparation appauvrit autant la science que la théologie.

Je suis d'accord avec Gardner lorqu'il définit la *foi* (religieuse) comme étant surtout une foi en Dieu ***et*** l'immortalité. J'accepte l'analyse de Gardner (et l'analyse d' Unamuno auquelle Gardner se réfère) qui rend la croyance en l'existence de Dieu pratiquement inséparable de la croyance en l'immortalité (Gardner 1983, page 209, 211). Gardner cite Kant à ce sujet. Je suppose qu'il se réfère à la critique de la raison pratique où une dérivation de la croyance en Dieu est *pratiquement* tirée de la croyance en l'immortalité. Dieu serait ce qui, ou celui qui, pourvoit à l'immortalité de l'âme, ou à la possibilité de cette immortalité. Cette conception minimale se retrouve sous des formes très variées dans pratiquement toutes les religions aussi bien orientales qu'occidentales, révélées que naturelles. Exemple : l' immortalité est une sorte de but pratique dans le taoïsme chinois, quelque chose auquelle on essaie d'échapper dans l'hindouisme (en coupant le cycle des réincarnations successives), une récompense ou une punition infinie dans de nombreuses religions chrétiennes. Un non-croyant en Dieu (qu'il soit agnostique ou athée), qui parvient cependant à concevoir une forme d'immortalité possible, peut substituer le terme "théologie" par le terme "métaphysique", ce terme désignant alors la tentative d'appréhension des lieux possibles où l'on émerge, d'une façon ou d'une autre, lorsque l'on survit à une expérience. La nature ontologique de tels lieux est laissée pour la troisième partie.

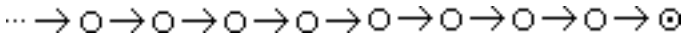


Le monde dans lequel on survit ne doit pas nécessairement être considéré comme distinct du monde où l'expérience est tentée.



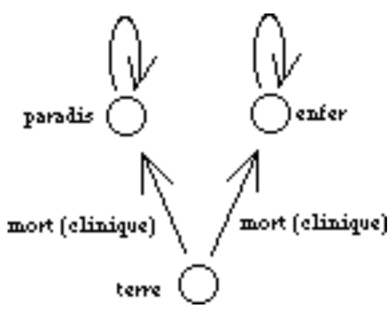
Prenons par exemple l'expérience du sommeil. Lorqu'on s'éveille le matin, il semble qu'on a survécu au sommeil, il semble qu'on accède à un monde possible, le *monde* quotidien, paramétré ou non par la date du jour (ou encore on accède à l'*état* de veille de ce jour). Autre exemple, l'activité de la veille. Lorqu'on se couche le soir, il semble qu'on a survécu aux activités du jour, en entrant dans son lit, on accède à un monde (état) possible (son lit, l'état endormi, ou le monde des rêves selon la conception qu'on se fait du sommeil). Etc.

Les théologiens s'intéressent à la question de la survie après la *mort*. Le mot "mort", dans ce contexte, est ambigu. Il peut désigner d'une part la mort *clinique*. C'est celle dont juge le médecin ou celui qui signe le certificat de decès. Il peut désigner d'autre part ce qu'on pourrait appeler la mort absolue qui est ce qui arrive lorqu'on aboutit à un dernier monde, ou un dernier état, c-à-d un monde (état) dans lequel (duquel) plus aucune expérience n'est permise. Schèmatiquement :



Seul le croyant athée peut décider d'identifier systématiquement les deux désignations. L'expression "survivre après la mort" est contradictoire seulement si la mort désigne la mort absolue.

Exemple : le diagramme suivant peut représenter une forme de croyance religieuse :



1.2.2 La logique modale

Aristote (384-322 ac) est connu pour sa contribution fondamentale dans l'histoire de la logique (voir Kneale et Kneale 1962). Aristote a considéré quelques relations d'inférence, sous forme de syllogismes, entre des propositions vraies (et fausses), mais est à l'origine des premières modalités, comme *possible, nécessaire, contingent...* A l'exception des scholastiques au moyen âge, et de l'utilisation plus ou moins formelle ou intuitive des modalités chez les théologiens, l'extension modale de la logique classique s'est développée principalement au 20ème siècle sous l'impulsion des travaux de Lewis sur la déduction, mais pas exclusivement, voir par exemple Feys 1937.

La notion de *monde possible* était apparue explicitement chez Leibnitz. C'est finalement Kripke (1963a, 1963b) qui propose une interprétation des logiques modales en terme de mondes possibles *accessibles*[2].

1°) Syntaxe

Les logiques modales que considère Kripke sont des extensions de la logique classique. On dispose d'un **alphabet** :

&, v, ¬, ->, p, q, r, ..., ☐, ◊,), (,⊥,T.

p, q, r, ... sont les variables propositionnelles, "☐p" est destiné à être interprété par une modalité, comme nécessaire p, partout p, toujours p, obligatoire p, croire p (cf l'introduction), savoir p, etc.

On dispose d'un **ensemble de formules**. C'est l'ensemble des formules générées par un nombre fini d'applications exclusives des règles suivantes :

1) ⊥, ,T, p, q, r, ... sont des formules (appelée formules atomiques)

2) si X est une formule, (¬ X) (☐ X) (◊ X) sont des formules (dites composées)

3) si X et Y sont des formules, (X & Y), (X -> Y), (X v Y) sont des formules (dites composées)

Exemples : (¬p) est une formule, "((☐p)->p)", "(¬((☐p)->p))" sont des formules, "(¬ & p☐ " n'est pas une formule.

---

[2] Kripke, mais aussi Scott et Montague et bien d'autre, prolongent ainsi les travaux de Lewis et assoient la logique modale dans la logique mathématique. Cela permet de faire de la logique modale, dégagé de toute considérations métaphysiques et en particulier de tout présupposés ontologiques. Cela ne signifie pas qu'il *faille* oublier la motivation métaphysique ou les préoccupations ontologiques. Ce serait aussi ridicule que de croire qu'on puisse oublier les étoiles sous prétexte qu'on a compris le fonctionnement des téléscopes, ou oublier la chose sous prétexte qu'on a saisi le mot.

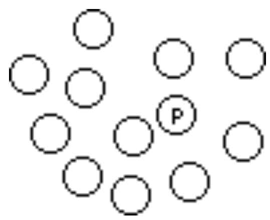
Abréviation : Nous supprimerons les parenthèses formelles pour alléger l'écriture. Par exemple l'expression "$\neg(\Box p \to p)$" est considérée comme une abréviation de la formule $(\neg((\Box p) \to p))$.

Remarque *Partout p* est intuitivement équivalent à *nulle part* $\neg p$, c-à-d $\neg$ *quelque part* $\neg p$, de même avec les couples (nécessaire, possible), (toujours, quelque fois) etc. On s'intéressera aux logiques pour lesquelles $\Box = \neg\Diamond\neg$, et réciproquement $\Diamond = \neg\Box\neg$. Remarquons qu'il en est de même pour les quantificateurs de la logique classique de prédicats $\forall$ et $\exists$ ($\forall = \neg\exists\neg$, $\exists = \neg\forall\neg$).

<u>*2°) La sémantique de Kripke*</u>

L'idée intuitive de $\Box$p, est que p est vrai dans tous les mondes possibles (ou dans tous les états possibles ou encore toutes les situations possibles, je rappelle que j'utilise monde et état comme des termes informels ou primitifs). Schématiquement, les ronds représentant les mondes :

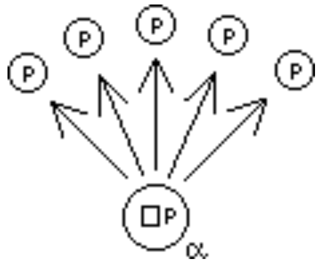
L'idée intuitive de $\Diamond$p est qu'il existe au moins un monde dans lequel p est vrai. Schématiquement :



Kripke relativise cette idée à chaque monde. Je désigne les mondes par des lettres grecques α, β, γ... ξ désignera une (méta)-variable parcourant les mondes.

L'idée de Kripke est d'introduire une relation d'accessibilité entre les mondes et d'interpréter $\Box$p dans un monde α par le fait que p est vrai dans tous les mondes *accessibles* à partir de α.

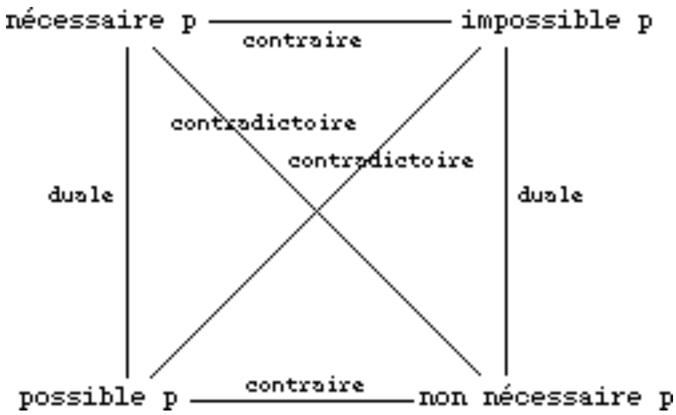
Dans les schémas qui vont suivre, les ronds représentent les mondes. Les formules écrites à l'intérieur des mondes sont vraies dans ces mondes. La relation d'accessibilité représentée par une flèche :



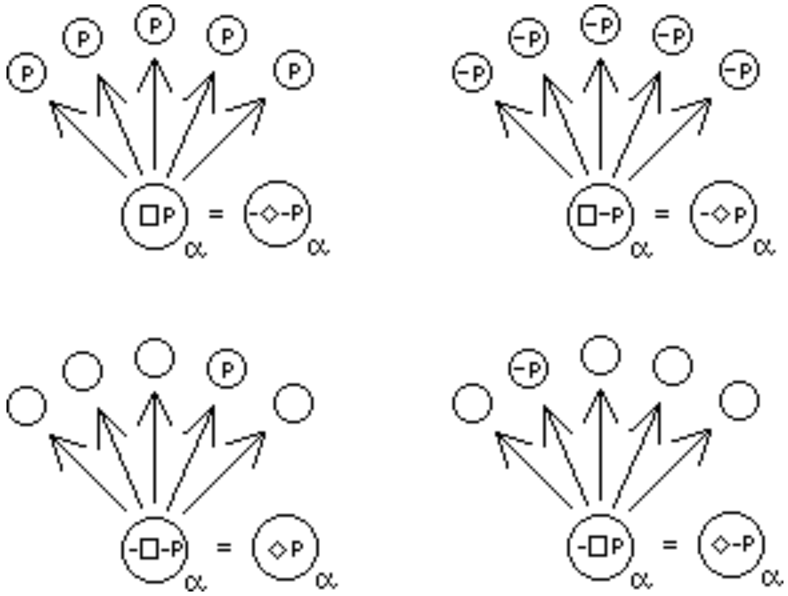
Remarquons que $\Box p$ est équivalent à $\neg\Diamond\neg p$, en particulier la proposition $\Box p$ est toujours vraie (et $\Diamond p$ est toujours fausse) dans un monde duquel ne part aucune flèche :

Un tel monde est appelé *un dernier monde.*

Illustrons par exemple, le *carré Aristotélicien* grâce auquel Aristote distinguait déjà le *contraire* de la *négation* avec les modalités (ontiques) *nécessaire* et *possible* :



et résumons l'interprétation de Kripke sur ce diagramme :



Définition

Un référentiel (W,R) est un ensemble W (dont les éléments sont appelés mondes, ou états, ou situations) muni d'une relation binaire R (appelée relation d'accessibilité).

Un modèle (W,R,V) est obtenu lorsqu'est assignée (arbitrairement) *dans chaque monde* une valeur, *vrai* ou *faux*, pour les variables propositionnelles *p*, *q*, *r*, ... Si L désigne le sous-ensemble {p, q, r, ...} de l'alphabet, l'assignation est capturée par une fonction V de $L \times W$ dans {vrai, faux}. Chaque monde est supposé obéir à la logique classique, si bien que V définit une valuation booléenne pour chaque monde.

Cela signifie que si la proposition *p* est vraie dans un monde α, et si *q* est vraie dans α, alors *p&q* est vraie dans α, etc. Je rappelle que *p->q* est classiquement vraie si *p* est faux ou si *q* est vraie (ou encore si ($p \ \& \neg q$) est fausse).

$\bot$ est une constante propositionnelle désignant, dans chaque monde le faux, et $\top$ est une constante propositionnelle désignant dans chaque monde le vrai.

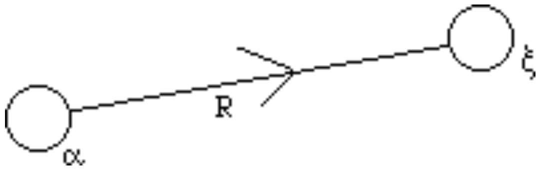
Résumons l'idée de Kripke on a :

$\Box$ A est vrai dans α ssi pour tout monde ξ tel que α R ξ, A est vrai dans ξ.

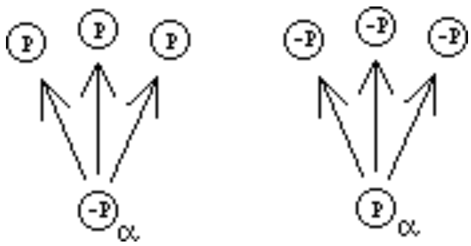
De même :

$\Diamond$A est vrai dans α ssi il existe un monde ξ tel que α R ξ et A est vrai dans ξ.

α R ξ est lu *α accède à ξ*, ou encore *ξ est accessible à partir de α*. Schématiquement :



Remarque : En logique classique non modale, la valeur de vérité d'une formule est univoquement déterminée par la valeur des sous-formules, et donc par la valeur des variables propositionnelles. Ce n'est plus le cas en logique modale. La valeur de vérité de $\Box$p, dans un monde α, ne dépend pas, a priori, de la valeur de vérité de p dans α, comme on le voit dans le diagramme :



Il n'y a donc pas moyen d'utiliser une table de vérité pour évaluer une formule modale à partir des valeurs de ses variables propositionnelles.

Définition fondamentale

Nous savons qu'en logique classique une *tautologie* est une formule qui est vraie quelle que soit la valuation de ses variables propositionnelles. Ainsi p->p, pv¬p, p->⊤, ⊥->p, sont des tautologies classiques.

On dira qu'un référentiel (W,R) respecte une formule A si, quelque soit la valuation V que l'on pourrait choisir, A est vrai dans tous les mondes de W.

Dit autrement : un référentiel (W,R) respecte une formule A si A est vraie dans tous les mondes dans tous les modèles que l'on peut construire sur le référentiel.

Conséquences :

i) Tous les référentiels respectent les tautologies classiques non modales puisque la logique classique est valable dans tous les mondes.

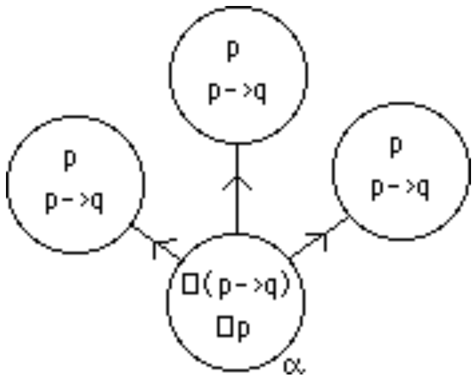
ii) Pour la même raison tous les référentiels respectent les tautologies classiques dans lesquelles on a substitué les variables propositionnelles par des formules quelconques, comme $\Box$p->$\Box$p, ◊p v ¬◊p, etc.

iii) (□ (p->q) & □ p) -> □ q est vrai dans tous les mondes de tous les modèles.

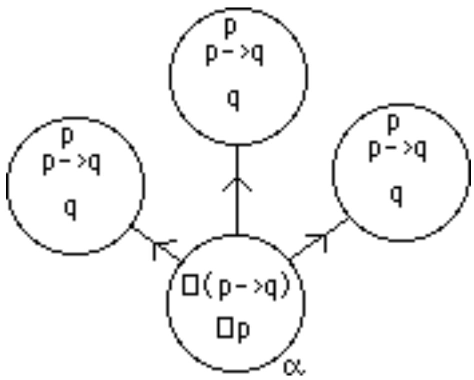
justification

Supposons que (□ (p->q) & □ p) soit vraie dans un monde α. Alors □ (p->q) et □ p sont chacune vraie dans α (par la sémantique classique *(tarskienne)* du &) :

par Kripke : □ p est vrai dans α, si p est vrai dans tous les mondes accessibles à partir de α (et donc *a fortiori* s'il n'en existe pas), de même □ (p->q) est vrai dans α si (p->q) est vrai dans tous les mondes accessibles à partir de α :



Mais chaque monde respecte la logique classique, et donc q est vrai dans tous les mondes auxquels accède α :



Mais si q est vrai dans tous les mondes accessibles à partir de α, alors, par Kripke, □ q est vrai dans α.

Donc $\Box$ q ne peut pas être faux en même temps que (($\Box$ (p->q) & $\Box$ p) est vrai) dans α, monde quelconque. ($\Box$ (p->q) & $\Box$ p)-> $\Box$ q est vrai dans tous les mondes, quelles que soient les valeurs de p et q.

Conclusion La formule ($\Box$ (p->q) & $\Box$ p)-> $\Box$ q, ou plutôt la formule (tautologiquement) équivalente $\Box$ (p->q) -> ($\Box$ p-> $\Box$ q) est respectée par tous les référentiels. On la désigne par **K** (pour Kripke).

K $\Box$ (p->q) -> ($\Box$ p-> $\Box$ q)

Toutes les formules ne sont pas respectées par tous les référentiels. Beaucoup de formules sont cependant respectées par une classe de référentiels caractérisée par la relation binaire. C'est là que réside l'intérêt de la sémantique de Kripke : associer un type de référentiel (c-à-d une relation binaire) à une formule modale. Cela permet **deux formes d'utilisation** de la logique modale :

a) **Géométrie d'abord -> Linguistique après (GL)** En commençant par l'observation, ou l'imagination de la structure des mondes et de la relation d'accessibilité (c'est la géométrie (au sens large)), et ensuite on cherche une formule modale qui caractérise cette structure (on passe à la linguistique (au sens large)).

b) **Linguistique d'abord -> Géométrie après (LG)** En utilisant l'intuition concernant une modalité pour inférer ou simplement suggérer une formule modale, et ensuite en étudiant l'éventuelle relation binaire qui caractérise la formule. Ici on peut s'inspirer de la façon dont une modalité est utilisée dans la langue naturelle.

Parce que nous allons utiliser la logique modale des deux façons dans le ***même*** domaine, je vais préalablement illustrer ces deux façons sur des exemples élémentaires.

Exemples

Enoncés :

1a) (GL) Bruxelles est le référentiel, les mondes possibles sont les endroits de Bruxelles, et le monde (endroit) β est accessible à partir de α ssi il est possible de se déplacer de α en β en un nombre fini, mais quelconque, de pas.

1b) (GL) idem que 1, mais β est accessible à partir de α ssi il est possible de se déplacer de α en β en un nombre fini, mais plus petit que 100, de pas.

2) (LG) 2a) si je sais p, p est-il vrai ?

2b) si je sais p, est-il vrai que je sais que je sais p ?

2a') si je peux savoir p, p est-il vrai ?

2b') si je peux savoir p, est-il vrai que je peux savoir que je peux savoir p ?

Résolution

1) Il est facile de se convaincre, géométrico-mentalement, que tout endroit de Bruxelles est accessible à lui-même en un nombre fini de pas (en l'occurrence 0). On a donc, partout à Bruxelles, $\xi R \xi$. On dit que la relation est réflexive. Ceci vaut pour 1a et 1b. Il reste à trouver une formule modale respectée par les référentiels réflexifs. Il est encore facile de se convaincre que si on peut aller de α en β en un nombre fini de pas, et de β en χ en un nombre fini de pas, alors on peut aller de α en χ en un nombre fini de pas. Donc pour 1a) la relation est transitive, c'est-à-dire qu'elle vérifie ($\xi R \zeta$ et $\zeta R \tau$ => $\xi R \tau$). 1b) illustre une relation d'accessibilité ***non*** transitive. En général lorsque l'on limite les ressources utilisées pour l'accessibilité, on perd la transitivité. Il reste à trouver une formule modale caractérisant les référentiels munis d'une relation d'accessibilité transitive (par abus de langage je dirai simplement référentiel transitif)

2) représentons "savoir" par $\boxed{S}$ et "peut savoir" par $\boxed{PS}$

les formules suggérées sont $\boxed{S}$ p->p, $\boxed{S}$ p-> $\boxed{S}$ $\boxed{S}$ p, $\boxed{PS}$ p->p, $\boxed{PS}$ p-> $\boxed{PS}$ $\boxed{PS}$ p. La différence entre ***savoir*** et ***pouvoir savoir***

Si on admet que *savoir p*, c'est *savoir p* dans *l'unique ici et maintenant* avec p à l'esprit, alors $\boxed{S}$ p-> $\boxed{S}$ $\boxed{S}$ p est peu vraisemblable. Par contre *peut savoir* ne limite pas le temps et l'espace (et l'énergie), bref les ressources, pour accéder au savoir, dans ce cas il est raisonnable de supposer que si l'on peut savoir p, alors on peut aussi savoir qu'on peut savoir p (c'est la différence entre *su* et *sachable*). Il reste à trouver une classe de référentiels caractérisant $\Box$ p->p, et $\Box$ p-> $\Box$ $\Box$ p. Pour cette dernière l'invocation des ressources[3] nous laisse deviner la proposition :

Proposition 1) tous les référentiels transitifs respectent $\Box$ p-> $\Box$ $\Box$ p,

2) Seuls les référentiels transitifs respectent $\Box$ p-> $\Box$ $\Box$ p.

Les référentiels transitifs caractérisent ainsi la formule modale $\Box$ p-> $\Box$ $\Box$ p. Dit de façon plus succincte :

(W,R) respecte $\Box$ p-> $\Box$ $\Box$ p <=> R est transitive

Preuve

---

[3] La transitivité sera ultérieurement liée à l'universalité des machines que l'on va considérer, ainsi que le fait qu'on ne bornera pas les ressources de ces machines (cf. le jeu infini de Carse 1986). L'axiome 4 n'est pas nécessairement relié au suffixe "able". On verra que la notion d'observ*able* (extraite de MDI) ne satisfait pas 4.

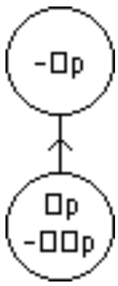
1) R est transitive => (W,R) respecte □ p->□ □ p.

je démontre par l'absurde.

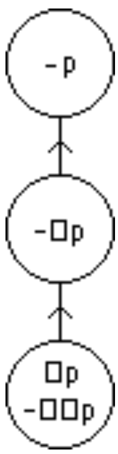
Supposons que i) R est transitive et ii) (W,R) ne respecte pas □ p->□ □ p. Dire que (W,R) ne respecte pas □ p->□ □ p, cela signifie (en vertu de la définition fondamentale) qu'il existe un modèle (W,R,V) comprenant un monde où □ p->□ □ p est fausse. Mais cette implication est fausse seulement si □ p est vrai et □ □ p est fausse dans ce monde. La suite est un dessin animé :

Voilà le monde (de W) coupable du non respect de □ p->□ □ p par (W,R) :

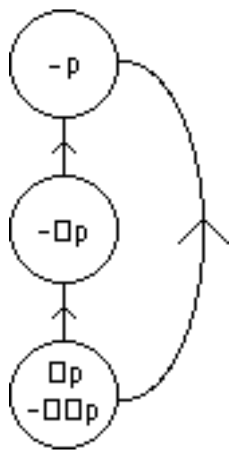
Mais ¬ □ □ p est équivalent à ◊ ¬ □ p, donc, (revoir au besoin le carré d'Aristote Kripke. Dans les dessins la négation "¬" est représentée par un tiret) :



et pour la même raison :



mais (i) on a supposé la relation transitive :

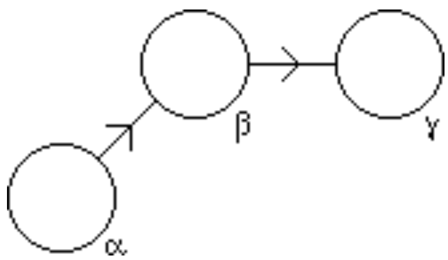


Contradiction, puisqu'on a un monde où □p est vrai et il accède à un monde où p est faux.

2) (W,R) respecte □p->□□p => R est transitive

Supposons que i) (W,R) respecte □p->□□p et ii) R n'est pas transitive. Rappelons nous que le respect d'une formule nécessite qu'elle soit partout vraie quelle que soit la valuation présentée (cf la définition fondamentale). Nous allons montrer que si R est non transitive, on peut construire une valuation témoignant que (W,R) ne respecte pas □p->□□p.

Dire que R n'est pas transitive signifie que le sous-référentiel suivant existe quelque part, plongé dans (W,R) :



Et il n'y a pas de flèche entre α et γ. Considérons la valuation suivante : p est vrai dans tous les mondes auxquels α accède, et faux partout ailleurs. (et quelconque sur les variables propositionnelles différentes de p). Par Kripke, □p est vrai dans α, et □□p aussi puisque le référentiel est supposé respecter □p->□□p, mais alors, par Kripke encore, □p est vrai dans β, et p est vrai dans γ, ce qui est en contradiction avec la valuation choisie.

De même il est facile de prouver :

proposition (W,R) respecte □p->p <=> R est réflexive

Preuve (voir plus loin).

Remarque :

Difficulté avec GL : attribuer une signification au carré cohérente avec la signification intuitive de la relation d'accessibilité utilisée au départ.

Difficulté avec LG : attribuer une signification à la relation d'accessibilité cohérente avec la signification intuitive du carré utilisée au départ.

3°) Théorie formelle

Une théorie formelle nécessite un langage formel et est donc de nature syntactique et linguistique.

Jusqu'à présent, nous avons un langage permettant de construire des formules modales et nous avons une structure géométrique permettant la caractérisation de certaines d'entre elles. Nous savons construire des *modèles* validant □ p->p, □ p->□ □ p, □ p->p & □ p->□ □ p, ainsi que des contre-exemples justifiant l'indépendance de formules. Par exemple, on peut trouver une valuation sur un référentiel transitif et non réflexif validant □ p->□ □ p et possédant un monde où □ p->p est fausse. Je définis à présent ce qu'est une théorie modale et ce qu'est une démonstration. Grâce au résultat de complétude et de correctitude (soundness) les indépendances sémantiques (faciles à visualiser) démontrent automatiquement les indépendances syntactiques.

Définition : une théorie formelle est présentée par la donnée d'un ensemble de formules accompagnée de règles[4] dite *règles d'inférence*, permettant la déduction de nouvelles formules. Une théorie est un ensemble de formules fermé pour l'application des règles d'inférence.

Définition : la *présentation* d'une théorie est l'ensemble des axiomes choisis et des règles d'inférence.

Une théorie peut avoir de nombreuses présentations.

Comme la logique modale est une extension de la logique classique, je présente d'abord une formalisation de celle-ci. Je choisis celle de Kleene (1952). A, B, C sont des métavariables représentant des formules quelconques. Il s'agit donc de schémas d'axiomes. J'utilise des schémas d'axiomes pour ne pas avoir de règles de substitution difficiles et longues à énoncer :

SCHEMA D'AXIOMES

1) A->(B->A) (principe de l'a posteriori)

2) (A->B)->((A->(B->C))->(A->C))

[4] Je donne ce qu'on appelle une présentation hilbertienne d'une logique formelle, c'est-à-dire beaucoup d'axiomes et peu de règles. La raison est que nous serons plus concernés par la notion de prouvabilité que par la notion de preuve. L'analyse de celle-ci est mieux capturée par des présentations à la Gentzen des systèmes formelles (ou leurs correspondants algébriques ou catégoriels), c-à-d (en gros) beaucoup de règles et peu d'axiomes. La majorité des résultats ne dépendront pas du type de présentation des théories.

3) $A \to (B \to A \& B)$
4) $(A \to C) \to ((B \to C) \to (A \vee B \to C))$
5) $A \to A \vee B$
6) $B \to A \vee B$
7) $A \& B \to A$
8) $A \& B \to B$
9) $(A \to B) \to ((A \to \neg B) \to \neg A)$
10) $A \vee \neg A$ (principe du tiers-exclu)

REGLES D'INFERENCE :

MP) $A, A \to B \Rightarrow B$

J'ajoute dès à présent les règles concernant les quantificateurs car nous en aurons besoin ultérieurement.

QE1) $A(t) \to \exists x A(x)$
QE2) $A(x) \to B \Rightarrow \exists x A(x) \to B$
QU1) $\forall x A(x) \to A(t)$
QU2) $B \to A(x) \Rightarrow B \to \forall x A(x)$

PA, l'arithmétique de Peano

Et tant que j'y suis voilà PA, l'Arithmétique de Peano, qui est devenue l'Escherichia Coli de l'auto-référence. Les machines auxquels on s'intéressera dans la seconde partie seront supposées à même de démontrer (au moins) les théorèmes de PA. On suppose présent dans le langage de PA (ou d'une machine) la constante individuelle 0, le symbole fonctionnel **un**-aire SUC, deux symboles **deux**-aires + et x. Les axiomes de PA, auxquels on ajoutera les axiomes standarts de l'égalité "=", sont :

1) $\neg(0 = SUC(x))$
2) $(SUC(x) = SUC(y)) \to x = y$

3) $x+0 = x$; définition récursive de l'addition
4) $x+SUC(y) = SUC(x+y)$

5) $x \times 0 = 0$; définition récursive de la multiplication
6) $x \times SUC(y) = (x \times y) + x$

accompagné du fondamental schéma d'axiomes d'induction, pour les formules du langage avec une variable libre x.

$$A(0) \& \forall x\{A(x) \to A(SUC(x))\} \quad \to \quad \forall x A(x)$$

Théories formelles modales

Pratiquement toutes les théories modales qui vont nous intéresser, comprennent, outre les axiomes propositionnels de Kleene et la règle du modus ponens MP, l'axiome K, et la règle de nécessitation : p => $\Box$ p.

Un théorème est soit un axiome, soit une formule obtenue par un nombre fini d'applications de la règle du modus ponens, ou de la règle de nécessitation, à partir d'axiomes ou de théorèmes déjà démontrés.

Théories faibles

On aura l'occasion de s'intéresser à quelques logiques plus faibles que la logique classique. Il s'agit essentiellement de la *logique intuitioniste* et de la *logique quantique*. La logique intuitioniste (formalisée par Heyting, voir Troelstra & Van Dalen 1988) peut être obtenue en remplaçant le principe du tiers-exclu par un principe plus faible : $\neg$p->(p->q). J'en parle un peu plus loin. La logique quantique peut de même être obtenue en affaiblissant le principe de l'a posteriori (voir annexe 5). Je n'utiliserai jamais ces logiques de façon directe car j'essaierai d'en rendre compte par des considérations philosophiques (tirée de MDI) en logique modale classique. Gödel (1933) et Goldblatt (1974) ont respectivement montré comment traduire la logique intuitioniste avec S4 et la logique quantique avec B (KT + p -> $\Box\Diamond$p). On reviendra sur ces considérations au moment opportun.

Théories formelles modales et sémantique de Kripke :

Proposition (Correction, soundness) Si un référentiel respecte un (ou plusieur axiomes) alors il respecte les théorèmes que l'on obtient avec ce (ou ces) axiomes par application des règles d'inférence.

Preuve

a) Le modus ponens est vérifié dans chaque monde. C'est le cas par définition de la sémantique de Kripke. En effet, chaque monde vérifie la logique classique. Donc si p est vraie dans tous les mondes, et si p->q est vraie dans tous les mondes, alors q est vraie dans tous les mondes.

b) Les référentiels respectent les formules obtenues par nécessitation sur des formules déjà respectées. En effet si p est vraie dans tous les mondes d'un référentiel W, alors en particulier, pour chaque monde x, p est vraie dans tous les mondes accessibles à partir de x, donc $\Box$ p est vraie dans tous les mondes ξ du référentiel, et donc $\Box$ p est respecté aussi.

En particulier si une proposition p est respectée par une classe de référentiel, $\Box$ p est respectée par cette classe de référentiel. Cette fermeture pour le "$\Box$" est une caractéristique importante de la sémantique de Kripke. On rencontrera des situations où cette caractéristique modale n'est pas vérifiée, et où, donc, d'autres sémantiques, comme celle de Scott et Montague (voir Chellas 1980) sont préférables.

Proposition (complétude) Le respect dans un référentiel est fermé pour la nécessitation et le modus ponens, et cela entraîne que tous les théorèmes de K (KT, K4, etc.) sont respectés dans un référentiel quelconque (réflexif. transitif, etc.) L'inverse est aussi vrai : toutes les formules respectées par les référentiels quelconques (réflexifs, transitifs) sont démontrables dans K (KT, K4, etc.). La technique classique consiste a construire un *modèle canonique* qui satisfait *tous et seulement* les <u>théorèmes</u> de K (KT, K4, etc.) et à montrer que ce modèle est quelconque (réflexif, transitif). Dans ce cas, si une formule est respectée dans tous les référentiels appropriés, elle est satisfaite alors dans le modèle canonique et est donc un théorème de la théorie correspondante.

Les logiciens depuis longtemps travaillent avec des modèles dont les mondes ou les univers sont des ensembles de formules, comme les modèles de Herbrand où les structures libres des algébristes. Un monde est défini par un ensemble consistant maximal de formules. *Consistant* est défini de façon purement syntaxique. Pour K (KT, K4, etc.), un ensemble de formules est *consistant* s'il n'existe pas de formule F_1, ..., F_n tel que K (KT, K4, etc.) démontre $\neg(F_1\&...\&F_n)$. Si E est un ensemble consistant de formules, on peut montrer que :

1) $E \cup \{F\}$ est consistant ou $E \cup \{\neg F\}$ est consistant,

2) E est contenu dans un ensemble consistant maximal, c'est à dire un ensemble consistant qui contient toute formule F ou sa négation $\neg F$ (voir Boolos 1979).

Pour définir le modèle canonique, on doit définir d'une part le référentiel, c'est-à-dire l'ensemble W des mondes et la relation R d'accessibilité, et d'autre part la valuation V.

On prend pour W, l'ensemble des mondes du modèle canonique, l'ensemble des ensembles maximaux consistants. La relation d'accessibilité R du modèle canonique est définie par

$$\alpha R \beta <=> \{F \mid \Box F \in \text{à } \alpha\} \subseteq \beta$$

On peut alors montrer, c'est peut-être le point délicat, qu'on a bien pour chaque monde $\alpha \in W$ que si $F \in$ à tous les mondes β accessibles à partir de α, alors $\Box F \in \alpha$.

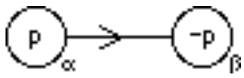
Reste à définir la valuation V : une formule atomique p est ***vraie*** dans un monde si elle ***appartient*** à ce monde. Il n'est plus difficile alors de prouver, par induction sur la complexité des formules F, que F est vraie dans le monde α ssi F appartient à α.

$$\alpha \Vdash F <=> F \in \alpha$$

Il ne reste plus alors qu'a montrer que les modèles canoniques de KT, KT4, ainsi de suite sont réflexifs, transitifs etc. Cela ne pose aucune difficulté (voir par exemple Boolos 79).

Remarque

Le métathéorème de déduction n'est pas valable pour la logique modale. On a p=> $\Box$ p par nécessitation, mais p-> $\Box$ p est déjà contredit par le monde α dans le modèle suivant :



4°) Savoir, Croire et le stratagème

Je résume, motive et complète la théorie de *celui qui sait ou peux savoir*. Je commence par utiliser le mot dans un sens très général, puis je nuance entre diverses significations intuitives à l'aide de formule modale.

Celui qui sait est *in touch* avec la vérité ou la réalité, lorsqu'on sait *p*, *p* est vrai. L'exemple le plus typique est la joie ou la douleur. Losqu'on sait qu'on a mal au dent, ou lorqu'on ressent un maux de dent, on a mal au dent : $\Box$ p->p. On pourrait objecter qu'une lésion accidentelle ou une tumeur cérébrale puisse induire la sensation de maux de dent, sans que les dents ne soient impliquées (comme il semble que cela puisse être le cas, voir Sacks (1985) pour des descriptions de situations analogues, voir aussi Marcel & Bisiach 1988). En ce sens le savoir est le savoir de la sensation. *Mal au dent* ne doit pas être pris pour un diagnostic, mais pour la référence à une sensation dont la connaissance est présupposée chez l'interlocuteur. Plus simplement, quant on sait qu'on souffre, on souffre. On admet l'axiome *savoir p -> p*, et j'utiliserai $\boxdot$ pour les modalités qui vérifie cet axiome.

T $\boxdot$p->p

Cette formule est attribuée à Feys et (indépendamment) von Wright.

L'axiome $\boxdot$p->$\boxdot\boxdot$p, dont le nom est *4*, est adopté pour les raisons décrites plus haut. Je m'intéresse à ce qui peut être su sans limitation de ressources de la part du sujet[5]. C'est la même raison qui nous invite à admettre l'axiome (appelé K pour Kripke) $\boxdot$(p->q)->($\boxdot$p->$\boxdot$q). Par exemple, je sais que si x est un multiple de 4 alors il est un multiple de deux, si j'apprends que tel nombre est un multiple de 4, je saurai qu'il est un multiple de 2. Cet exemple mathématique n'est pas fortuit. Qu'une partie du savoir mathématique peut être *sue* est à l'origine d'une conception solipsiste des mathématiques, due principalement à Brouwer, où les

5 En ce qui concerne les machines, plus tard nous étudierons leurs comportements dans les voisinages de l'infini.

concepts se développent constructivement, sans quitter le cordon ombilical de la vérité intuitive, celle-ci reposant sur une intuition élémentaire des nombres naturels (accompagnée d'une intuition du continu dans les développements plus avancés). Il s'agit des mathématiques intuitionistes. Une formalisation des mathématiques intuitionistes a été entreprise par Heyting, et, à la présentation près, est équivalente à la présentation que j'ai donnée de la logique classique où l'on remplace l'axiome 10 :

$$p \vee \neg p$$

dit du *tiers-exclu*, par un axiome plus faible :

$$\neg p \to (p \to q)$$

On ne dispose plus de la méthode des tables de vérité pour vérifier une *tautologie* intuitioniste. De multiples interprétations ont été proposées sans interruption depuis, (arbre de Beth, ouvert d'un espace topologique, logique interne d'un topos, etc...)

Je propose ici une interprétation informelle due encore à Heyting (voir Heyting 1956). En gros l'assertion *p* de la part d'un intuitioniste revient à l'affirmation qu'il possède une preuve constructive de p. La notion de preuve constructive est alors précisée ainsi :

a) e est une preuve de p&q <=> e est une paire (e_1, e_2) telle que e_1 est une preuve de p, et e_2 est une preuve de q.

b) e est une preuve de pvq <=> e est une paire (e_1, e_2) telle que e_1=0 et e_2 est une preuve de p, ou e_1=1 et e_2 est une preuve de q.

c) e est une preuve de p->q <=> e est un *outil* qui est capable de convertir toute preuve x de p en une preuve e(x) de q.

d) e est une preuve de $\neg$p <=> e est une preuve de p-> $\perp$.

On voit avec cette interprétation que le principe du tiers exclu entraînerait la décidabilité de toutes les formules. On verra comment la théorie de la récursion permet de préciser des sémantiques informelles de ce genre.

Dès 1933, Gödel a annoncé que les axiomes modaux K, T, 4 :

$\boxdot(p \to q) \to (\boxdot p \to \boxdot q)$	K
$\boxdot p \to p$	T
$\boxdot p \to \boxdot\boxdot p$,	4

accompagnés des règles décrites (Nécessitation, Modus Ponens) permet une modélisation par la logique classique (LC) de la logique intuitioniste (LI). Ce système modal est connu sous la dénomination S4 (depuis Lewis) et il constitue notre (embryon de) théorie de la connaissance, ou plutôt du *connaissable* (vu la transitivité). En particulier LI $\vdash$ p <=> S4 $\vdash$ $\boxdot$p. Cela va incidemment nous permettre de profiter d'une relation entre le *doxastisme* (savoir, connaissance) et l'*intuitionisme* de Brouwer.

La croyance (ou la crédibilité), à l'inverse est corrigible, on peut croire du faux, par exemple lorsqu'on se trompe ou lorsqu'on rêve. C'est K (ou K4).

S4 et K4 permettent de décrire des théories de la connaissance vérifiable (éventuellement correcte), et permettent de lier des aspects formels du solipsisme et du positivisme[6]. Je laisse en suspens, à présent, la question de savoir si la communication positiviste est de style crédibilité (K4) ou connaissabilité (S4).

Proposition 1) tous les référentiels réflexifs respectent $\Box$ p->p,
2) Seuls les référentiels réflexifs respectent $\Box$ p->p.

Les référentiels réflexifs caractérisent ainsi la formule modale $\Box$ p->p. Dit de façon plus succincte :

(W,R) respecte $\Box$ p->p <=> R est réflexive

Preuve
1) R est réflexive => (W,R) respecte $\Box$ p->p.
je démontre par l'absurde.
Supposons que R est réflexive et que (W,R) ne respecte pas $\Box$ p->p. Dire que (W,R) ne respecte pas $\Box$ p->p, cela signifie (en vertu de la définition fondamentale) qu'il existe un modèle (W,R,V) comprenant un monde où $\Box$ p->p est fausse. Mais cette implication est fausse seulement si $\Box$ p est vrai et p est fausse dans ce monde. La suite est encore un dessin animé :

Voilà le monde (de W) coupable du non-respect de $\Box$ p->$\Box$ $\Box$ p par (W,R) :

[6]Le lien entre intuitionisme, épistémisme et positivisme est fait explicitement par Grzegorczyk 1964.

mais la relation R est supposée réflexive, donc

ce qui, par Kripke, est contradictoire.

2) (W,R) respecte □ p->p => R est réflexive. Supposons que (W,R) respecte □ p->p et que R ne soit pas réflexive. On va construire une valuation témoignant que (W,R) ne respecte pas □ p->p. Dire que R n'est pas réflexive signifie qu'il existe un monde μ dans le référentiel qui ne possède pas de boucle. Prenons (à nouveau) la valuation qui consiste à mettre p (V(p) = 1) dans tous les mondes auquel μ accède, et p faux (V(p) = 0) partout ailleurs. Que de tels mondes existent ou non, cela rend, par Kripke, □ p vrai dans μ, et, puisque μ n'accède pas à μ, p est faux dans μ par le choix de la valuation. Mais si □ p est vrai dans μ sans que p ne soit vrai, c'est que □ p->p est faux dans μ, et cela montre que le référentiel ne respecte pas □ p->p.

Définition

Il est facile cependant de montrer que si la communication positiviste (le positivement justifiable), (□) est décrite par K4, la communication positiviste rendue correcte ⊡p, c-à-d définie par □ p & p, est décrite par S4. Par exemple l'axiome T devient une tautologie classique :

□ p & p -> p

J'appelerai ***stratagème de Théétète*** ou simplement stratagème (ou encore stratagème ***fort*** par opposition à une version affaiblie du stratagème que l'on va considérer) la définition de la connaissance par la croyance ou la justification vraie[7]. La plausibilité de cette définition de la connaissance fait l'objet d'une discussion approfondie entre Théétète et Socrate dans le Théétète de Platon. Après avoir réfuté l'idée que la sensation constitue la connaissance, Socrate réitère son interrogation :

SOCR. : <...>. Allons ! dis-moi de nouveau ce que c'est que la connaissance. -

---

7 Sémantiquement (avec les logiques normales, c-à-d qui admettent une sémantique de Kripke) cela revient à ajouter des boucles sur tous les mondes. Une démonstration due indépendamment à Boolos et Goldblatt (notamment) est donnée dans 2.3 (ainsi que les références).

> *THEET. : Répondre que c'est tout jugement est impossible, Socrate, puisqu'il y a aussi des jugements qui sont faux ; mais il est bien possible que le jugement vrai constitue la connaissance : et même, que ce soit là ma réponse ! <...>*

Bien entendu Socrate ne sera pas vraiment satisfait de cette réponse. Plus tard Théétète proposera alors de définir la connaissance par la le jugement vrai accompagné d'une justification. Comme j'inclus la justification dans la communication positive le stratagème se réfère plus à cette dernière proposition (malgré qu'elle ne satisfera pas davantage Socrate (ou Platon) qui préferera ne pas conclure).

Le stratagème jouera un certain rôle lorsque j'analyserai une théorie de la connaissance des machines dans les voisinages de l'infini.

La version affaiblie du stratagème, où $\boxdot$ p est défini par $\Box$p & $\Diamond$p, semble jouer un rôle équivalent, dans les voisinages de l'infini, pour une notion de croyance, ou même pour une notion d'*immédiatement observable ou probable*. La motivation originale pour le stratagème affaibli repose cependant sur une idée de la justice. L'axiome choisi est cependant très faible et est souvent proposé aussi pour les croyances et une notion de probabilité (voir par exemple Thomason 1980, Alechina 1994, et Turner 1990).

<u>6°) droit et justice,</u>

Avant d'aborder une tentative vers une théorie de la conscience, je vais décrire un axiome de base pour une théorie du droit. Outre que cela fournit un exemple supplémentaire d'approche modale linguistique-géométrique, cela permet de comparer, d'une certaine façon, la conscience et la justice.

Cette fois-ci $\Box$ p est interprété par *p est obligatoire*. Et dualement -$\Box$-p, noté encore $\Diamond$, est interprété par *il n'est pas obligatoire que non p*, c-à-d *p est permis*.

Une conception pratique de la justice est incarnée dans un texte de loi. Celui-ci peut contenir une série d'obligations et de permissions, par exemple :

$\Box$ service-militaire V $\Box$ service-civil,
$\neg\Diamond$tuer-son-prochain, etc...

Mais la donnée de symboles accompagnée d'une interprétation intuitive ne constitue pas une théorie. Est-il possible de trouver un ou plusieurs axiomes[8] capturant l'idée de justice : voilà D, appelé axiome déontique,

---

[8] Pour un panorama d'applications diverses de la logique modale, voir van Benthem 1985. Pour des applications aux théories de la vérité et de la connaissance, voir Turner 1990.

rencontré dans la littérature[9] (voir Chellas 1980, Turner 1990, Bailhache 1991, Gardies 1979) :

$$\Box A \rightarrow \Diamond A \qquad (D)$$

Cette formule capture l'idée qu'il est juste de permettre ce qui est obligatoire. Un texte de loi qui ne vérifierait pas l'axiome déontique permettrait la présence de lois du genre :

$$\Box p$$
$$\neg \Diamond p,$$

p dénotant une certaine proposition fixée. Si on admet que la société peut punir un individu qui aurait commis une action interdite ou qui n'aurait pas commis une action obligatoire, alors tout le monde peut être punis. L'axiome déontique capture un aspect élémentaire de la justice ou du droit. Il ne constitue nullement une définition de la justice, d'autres (schéma d') axiomes, peut-être une infinité, doivent être introduits pour affiner la théorie.

sémantique

Les lois doivent être telles que les obligations puissent être tenues. Ce n'est certainement pas le cas dans un dernier monde où tout est nécessaire et rien n'est possible, ou, *déontiquement parlant*, tout est obligatoire et rien n'est permis.

A quoi pourrait ressembler un référentiel respectant l'axiome déontique ? Il suffit qu'il ne possède pas de dernier monde.

définition : un **référentiel** <W, R> est **idéal** s'il ne possède pas de dernier monde. Je dirai aussi que R est idéale sur W, ou simplement que R est idéal[10].

définition : un **monde** est **transitoire** ssi il n'est pas un dernier monde.

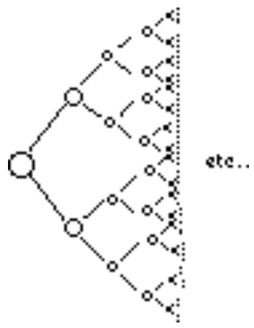
On a : un référentiel est idéal <=> tous ses mondes sont transitoires.

examples

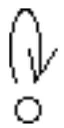
1) l'éventail

[9] Selon Gardies et Kalinowski, *D* remonterait à Leibnitz (voir Gardies 1979). Sa réapparition moderne serait due à Jeremy Bentham (voir Bailhache 1991).

[10] J'emprunte le terme "idéal" pour de tel référentiel chez Fitting 1983.



2) la boucle



Propositions

1) tous les référentiels idéaux respectent □ p->◊p,
2) seuls les référentiels idéaux respectent □ p->◊p.

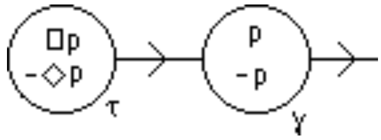
Les référentiels idéaux caractérisent ainsi la formule modale □ p->◊p. Dit de façon plus succincte :

(W,R) respecte □ p->◊p <=> R est idéale

Preuve

1) R est idéale => (W,R) respecte □ p->◊p.

Si R est idéale, tous les mondes sont transitoires, et si (W,R) ne respecte pas □ p->◊p, il existe un monde transitoire t dans lequel □ p est vrai et ◊p est faux. Dans les mondes auxquels τ accède (et il y en a puique τ est transitoire) on a p et on a ¬ p.



Contradiction.

2) (W,R) respecte □ p->◊p => R est idéale. Supposons que (W,R) respecte □ p->◊p et que R ne soit pas idéale. Il existe dès lors un dernier monde δ. Comme δ est un dernier monde, quelle que soit la valuation qu'on pourrait imposer, □ p est vrai dans δ Comme <W,R> respecte □ p->◊p,et que □ p est vrai δ, ◊p est vrai dans δ. Absurde puisque δ est un dernier monde.

T, D et le stratagème affaibli

J'insiste sur le fait que D est plus faible que T. Le schéma d'axiome $\Box p \rightarrow p$, est équivalent, par contraposition propositionnelle classique, au schéma $p \rightarrow \Diamond p$. Donc T => D. La non-réciprocité est sémantiquement évidente : l'éventail sans aucune boucle n'est pas réflexif. Le stratagème affaibli revient à définir un nouveau système modal en garantissant, non plus la *vérité*, mais la *possibilité* pour les formules justifiées. De cette façon l'axiome D est automatiquement entraîné, comme T l'est avec le stratagème fort.

1.2.3 Une théorie de la conscience

Survivre à une expérience nécessite la préservation de la conscience à travers cette expérience.

L'influence du behaviorisme et du positivisme au début du siècle avait pratiquemment banni le terme conscience, non seulement dans le discours philosophique, mais aussi dans les discours psychologiques (à quelques exceptions près, notamment celles des écoles psychanalytiques et des onirologues[11]) au point que Marcel et Bisiach dans leur recueil d'articles sur la conscience compare la réapparition récente de l'intérêt pour celle-ci au retour du fils prodige (Marcel et Bisiach 1988). Ces auteurs font néanmoins remarquer combien le concept de conscience reste ambigu, peu clair, difficile à cerner, et que les approches proposées sont en général inconsistentes.

Le fait que le terme conscience, ainsi que des termes relatés (expérience subjective, vie privée, sentiments, etc...) aient été bannis par les philosophes positivistes, constitue en soi une information intéressante et non négligeable. Cela suggère une approche axiomatique[12] de la conscience, ou plus modestement, une approche axiomatique des discours possibles sur la conscience. On pourrait ainsi, directement à partir du tractatus de Wittgenstein (1918), cerner la conscience par *ce dont on ne peut parler*, ou de façon plus exacte par *ce dont le philosophe positiviste ne peut parler*, ou

[11] Le positivisme, élaboré par Auguste Comte, a été dogmatique dès le départ. L'existence des atomes, la constitution des étoiles, l'unification des lois physiques, et même l'usage du microscope semblent avoir été considérés comme des activités scientifiques déviantes par les premiers positivistes. (cf Ullmo J. 1969).

[12]Les théories mathématiques définissent rarement univoquement les entités auxquelles elles se réfèrent, et depuis Gödel on sait qu'il en est presque toujours *nécessairement ainsi* (ce qui par ailleur sera justifié tout au long de la deuxième partie). Les mathématiciens travaillent alors axiomatiquement en épinglant les propriétés pertinantes des objets visés, et en dérivant des nouvelles propriétés à partir de ces propriétés pertinantes. Une théorie axiomatique de la conscience ne nécessite donc pas de définir univoquement le terme conscience, pas plus qu'une géométrie ne nécessite la définition univoque des notions de points ou de droites. Il se peut, et cela arrive souvent (presque toujours) que de nouveaux objets, contraire à l'intuition de départ, vérifient les axiomes (qui décrivent les propriétés pertinantes choisies avec l'intuition initiale). Dans ce cas, soit on généralise le sens des termes utilisés (en géométrie non euclidienne on décide d'appeler droites des objets qui, en géométrie euclidienne s'appellent cercle!), ou alors on cherche des nouvelles propriétés pertinantes permettant une meilleure approximation de l'objet capturé par l'intuition initiale (comme le postulat d'Euclide). Dans tous les cas de figure on se contente de diverses approximations.

encore de façon plus précise : ce qu'on ne peut justifier de façon positiviste. Une axiomatique analogue usant du solipsisme consisterait à cerner la conscience par ce dont le philosophe solipsiste n'a pas *besoin* de parler.

Ce que j'essaie de capturer axiomatiquement est vague. Conscience, conscience morale, foi, responsabilité, sagesse, relation avec une divinité, intelligence (au sens religieux cf Krishnamurti 1973, a priori différent de compétent ou doué). Voici quelques textes qui confirment l'importance de l'aspect incommunicable, ou ineffable de l'expérience de la conscience. Ces textes sont choisis pour leur élégance et leur diversité.

La stratégie adoptée consiste en l'utilisation de l'approche modale *linguistique d'abord - géométrie après* pour poser une équation dont l'inconnue désignerait la conscience. Ensuite une approche géométrie d'abord - linguistique après sera proposée pour isoler une solution.

Les textes illustrent, chacun à leur façon, la proposition de Wittgenstein comme quoi il existe quelque chose dont il est difficile de parler. (Il s'agit du Wittgenstein mystique, déjà présent dans le tratactus, dont Carnap (un des pères de la philosophie positiviste) se serait méfier. (voir Kerr 1991).

1°) Textes

Voici comment Ouzoulias décrit la conscience dans son petit livre à l'usage des étudiants :

> *Je suis moi, non plus ce sujet grammatical que je mets presque négligemment dans mes discours, mais ce qui, au loin du langage, à couvert des mots, y demeurait tapi. Je suis moi et non un sujet en général. Je suis confronté à ce moi irréductible à toute détermination objective qui serait purement contingente : ce qui peut être déterminé, c'est seulement ma personnalité physique et morale, fruit de l'histoire et d'un enchaînement de causes et d'effets souvent hasardeux qui ne m'appartiennent pas. En revanche, ce moi paraît indéterminable, car il m'apparaît comme la condition par laquelle tout autre personnalité aurait pu m'advenir. Il fallait bien que je fusse avnt de devenir quelqu'un. Il y a dans l'expérience de la conscience celle de l'indéterminablité du moi ou d'une subjectivité pure que j'appellerais le sentiment de l'"égocité". Mais maintenant, cette égocité me paraît innéfable. Le langage, en signifiant, extériorise et objective. Il semble bien que voici avec la conscience une expérience qui ne peut que se vivre sans jamais pouvoir se dire. Au dedans de moi, il me semble maintenant que je suis dans une irrémédiable insularité, comme Robinson isolé sur l'île de Speranza et interrogeant le destin.*

Ouzoulias décrit le mystère de l'*innefable égocité.*

Faisant écho à la critique par Goethe de la théorie des couleurs de Newton, Schrödinger, dans un tout autre style et dans un tout autre contexte, illustre l'irréductible différence entre les *qualia* (sensation vécue, par exemple face à un objet coloré) et une quelconque description "objective" de la relation entre l'observateur *objectivé* et l'objet :

> *On ne peut rendre compte de la sensation de la couleur par le modèle objectif des ondes lumineuses que le physicien possède. Le physiologiste pourrait-il en*

rendre compte, s'il disposait d'un savoir plus complet que celui dont il dispose sur les processus se déroulant dans la rétine, et sur les processus nerveux que ces derniers déclenchent dans le faisceau du nerf otique et dans le cerveau ? Je ne le pense pas. Nous pourrions au mieux connaître objectivement quelles fibres nerveuses sont excitées et en quelle proportion elles le sont, peut-être même pourrions-nous connaître exactement les processus qu'elles engendrent dans certaines cellules cérébrales - lorsque notre esprit enregistre la sensation de jaune dans une direction ou un domaine particulier de notre champ visuel. Mais même une connaissance aussi intime ne nous dirait rien sur la sensation de couleur, et plus particulièrement du jaune dans cette direction. On pourrait concevoir que le même processus physiologique aboutisse à une sensation de goût sucré, ou à n'importe quoi d'autre. Je veux simplement dire que nous pouvons être sûrs qu'il n'y a pas de processus nerveux dont la description objective inclue les caractéristiques "couleur jaune" ou "goût sucré" pas plus que la description objective d'une onde électromagnétique n'inclut l'une de ces caractéristiques.

et comme l'ajoute le traducteur et commentateur M. Bitbol :

...La pertinence de la réflexion philosophique de Schrödinger sur les qualités sensibles n'est en aucune manière affectée par la perte progressive d'actualité du tableau des connaissances physiologiques tel qu'il le brosse dans ce chapitre.

Cette réflexion peut être naturellement étendue en un argument en faveur de l'incommunicabilité (positiviste ou solipsiste) de la perception des qualia. Ceux-ci sont souvent utilisés pour réfuter l'approche fonctionnaliste de l'esprit[13].

Voici Arsac, concluant un colloque sur la science et le sens :

Puis-je dire qu'à titre personnel je crois profondément à l'existence du sens. Je ne suis pas nominaliste : il n'est pas pour moi un mot aussi vide que l'éther. Il est une réalité au plus profond de moi-même, si forte qu'il m'arrive de ne pas trouver les mots pour le dire (ce qui pour moi, est la preuve flagrante de son existence avant et au-delà des mots).(voir Arsac et Sentis 1990).

Le sens est aussi attaché au sentiment de conviction, que celle-ci soit religieuse ou reliée à la preuve dans l'acceptation générale (informelle) du mot. Que ce sentiment précède le langage est défendu avec vigueur par Brouwer, le fondateur de la philosophie intuitioniste et par Bergson, auquel Brouwer se réfère occasionnellement (cf van Stigt 1990). Je cite une traduction anglaise :

Ridiculous is the use of language of so called philosophers or metaphysici discussing morality, God, consciousness, immortality and the free will (Brouwer 1905)

---

[13] Une idée qui se trouve déjà chez Leibnitz (1714) avec son expérience par la pensée qui consiste à être réduit de taille et à se promener ainsi dans un cerveau.

Voilà une citation de l'Islam tiré de " Salam A., Heisenberg W., Dirac P. A. M., 1990, page 72." :

> *Quand tous les arbres de la Terre seraient des roseaux et la Mer un encrier,*
> *Avec sept Mers encore pour l'emplir,*
> *Seigneur, je ne pourrais transcrire toutes Ta parole.*
> *Seigneur Tout-Puissant, infini est Ton savoir.*
>
> *Coran (XXXI, 27)*

Cette présentation de textes et de religions n'est évidemment pas exhaustive.

Si l'on représente la communication positiviste (vérifiable[14]) par un connecteur modal, obéissant au moins à K, la façon la plus simple de capturer un aspect de la conscience, ou de l'expérience sensible ou privée, indépendamment du choix des axiomes pour l'axiomatisation que nous cherchons, est :

$$\neg \Box x,$$

ce qui encore, classiquement (ou intuitionistiquement), est équivalent à :

$$\Box x \rightarrow \bot, \qquad (W)$$

où x peut abrévier *je suis conscient, je suis vivant.*[15],... Pour fixer les idées[16] j'appelle cette formule ***principe de Wittgenstein* (W)**.

A la différence du droit, le candidat axiome pour la conscience est exprimé sous forme d'équation que doit vérifier la conscience. Il ne s'agit donc pas d'étudier les conséquences de W comme si c'était un axiome d'un système modal, mais de choisir, *sur base d'une réflexion linguistique*, un système modal axiomatisant convenablement la communication positiviste ou solipsiste et de résoudre dans celui-ci l'équation (W).

Dans les textes qui suivent une autre formulation se dégage :

$$\Box x \rightarrow \neg x. \qquad (LWV)$$

---

[14] voir Wittgenstein 1911, ou Amstrong & Malcolm (1984) et Gaston Granger (1992).

[15] Pour autant qu'elle ne tombe pas dans la trivialité, je ne craindrai pas la généralité de la théorie axiomatique. A la différence de Griswold 1986, je ne sépare pas a priori le soi de l'épistémologue de celui du théologien, ou même de l'immunologiste. Cela ne signifie pas que ces termes aient les même référents, je tente seulement d'axiomatiser un pattern commun ou une vérité commune les concernant.

[16] Mais aussi parce que je pense que ce principe, qui est pratiquemment contradictoire, joue un rôle de charnière dans l'oeuvre deWittgenstein. C'est un raccourci entre le *Tractatus* (Wittgenstein 1911) et, par exemple *De la certitude* (Wittgenstein 1965).

Je l'appelle principe de *Lao-tseu*, ou *principe de Watts-Valadier* ou encore ***principe de Lao-tseu-Watts-Valadier*** (**LWV**) car ces auteurs l'ont abondamment et méticuleusement illustré (Watts 1951, Watts 1974, Valadier 1990[17]). L'avantage de cette formulation est qu'elle donne une lueur de réponse à la question suivante : s'il faut *taire ce dont on ne peut parler*, que se passe-t-il si on en parle quand même? Question importante quand on réalise que lorsque Wittgenstein dit qu'il faut le taire, il est en train d'en parler! N'est-ce pas contradictoire ? André Frossard écrit :

> *Démontrer Dieu est impossible, et cela ne pourrait d'ailleurs se faire qu'au détriment de la foi (Frossard 1990).*

De même le Tao-te-king de Lao-tseu commence ainsi (je donne plusieurs traductions) :

Le sens qu'on peut exprimer n'est pas le sens éternel (trad R. Wilhelm et E. Perrot)

La voie qui peut s'énoncer n'est pas la voie éternelle (trad F. Houang et P. Leyris)

Le Tao qu'on tente de saisir n'est pas le Tao lui-même (trad Liou Kia-hway)

Le Tao qu'on saurait exprimer n'est pas le Tao de toujours (trad Liou Kia-hway)

Voie qu'on énonce, N'est pas la voie (trad C. Larre)

Le principe qui peut être énoncé, n'est pas celui qui fut toujours (Trad L. Wieger)

Alan Watts, dans son petit livre -*The Wisdom of Insecurity*-- interprète le Tao-te-king et notamment l'énoncé cité, comme un principe de l'effort inversé, ou encore une loi des effets contraires. Alan Watts suggère de nombreux exemples dans la vie quotidienne mais s'attache surtout à illustrer la loi des effets contraires dans la quête de la sécurité psychologique (sociale ou individuelle). Il montre que l'acharnement vers la sécurité et le besoin de fixer des garanties dans ce domaine est une des sources principales de l'insécurité. Il termine en appliquant son principe pour la quête de sécurité "religieuse". Il rappelle par exemple le dicton "qui veut sauver son âme la perdra" et suggère que les dogmes religieux ainsi que l'autoritarisme dont ils sont entourés sont des causes d'affaiblissement de la foi. Lorsque Lao-Tseu affirme que *le sage se tait*, il rencontre la même difficulté que Wittgenstein lorsqu'il dit qu'il y a de l'inexprimable *dont le philosophe ne parle pas*. On peut d'une part se demander si Lao-Tseu et Wittgenstein n'ont pas *raté l'occasion de se taire*[18], d'autre part on évacue difficilement le sentiment qu'ils pointent chacun sur un aspect clé de l'expérience de la

[17] Cette liste est, bien sûr, loin d'être exhaustive, voir Huxley 1946, pour ne citer qu'un autre.

[18] Très instructif dans ce contexte est aussi le petit livre de l'abbé Dinouart : l'art de se taire (Dinouart 1771).

conscience. Le principe de l'effort inversé apparaît plus clairement encore dans les deux 1ères propositions du 2ème discours du Tao-te-king de Lao-Tseu :

Quand chacun sur terre tient le beau pour beau
Cela implique d'emblée la laideur
Quand chacun sur terre tient le bien pour bien
Cela implique d'emblée le mal (trad R. Wilhelm et E. Perrot)

Quand chacun tient le beau pour beau vient la laideur
Quand chacun tient le bon pour bon viennent les maux (trad F. Houang et P. Leyris)

Tout le monde tient le beau pour le beau
C'est en cela que réside sa laideur
Tout le monde tient le bien pour le bien
C'est en cela que réside son mal (trad Liou Kia-hway)

Tout le monde sait que la beauté est belle
Voilà ce qui fait sa laideur
Tout le monde sait que le bien est bien
Voilà ce qui fait son imperfection (trad Liou Kia-hway)

Chacun de par le monde décrète le Beau
Et voici venir le Laid
Chacun de par le monde décrète le Bon
Et voici venir le Mal (trad C. Larre)[19]

Watts, pour illustrer son principe, cite les Upanishad hindoues, mais aussi le chrétien Jean de la Croix. Ils ne craignent pas la franche contradiction (en apparence sans doute, et dans une traduction, je le précise) :

Celui qui pense que Dieu ne peut pas être compris, comprend Dieu, mais celui qui croit que Dieu peut être compris ne le connaît pas. Dieu est inconnu à ceux qui le connaissent, et il est connu par ceux qui ne le connaissent pas du tout. (Upanishad).

L'une des grandes faveurs accordées à l'âme en cette vie passagère est de pouvoir si distinctement et sentir si profondément qu'elle ne peut absolument pas comprendre Dieu. (Jean de la Croix).

Dans un ouvrage plus récent (1990) Paul Valadier, jésuite, examine en profondeur la différence entre la morale qui se veut objective ou qui s'est laissée objectivisée et la morale pratique de celui qui, dans une situation concrète, rencontre ce qu'il rappelle être un cas de *conscience* :

---

[19] Remarquons que Lao-Tseu ne dit pas "quand chacun tient le laid pour laid cela implique d'emblée la beauté". Le tao, le bon, le bien, la beauté, la vertu, ne sont pas symétrisables comme le *long* avec le *court*, le *facile* avec le *difficile*, le *haut* et le *bas*, etc. Certaines traductions (dont celle de Wieger, reprise par Grenier (1973) notamment) oblitèrent cette distinction. Certains passages de Lie-Tseu semble aussi ne pas tenir compte de cette distinction (par exemple le chapitre XVI du *vrai classique du vide parfait* de Lie Tseu "1980"). Cette oblitération réduit alors LWV à un principe de relativisme, bien présent par ailleurs chez les taoïstes, mais qui n'ont pas la forme de LWV et en particulier ne s'applique pas au Tao selon l'interprétation suggérée dans la présente approche.

Négatrice de la pratique, la morale objectiviste risque bien d'être immorale, malgré sa prétention contraire.

... l'objectivisme moral conforte l'immoralisme ambiant...

Valadier met en exergue ce qui se passe lorsqu'on prétend rendre nécessaire ou communiquer objectivement (à la façon dont communiquerait un positiviste ou un néo-positiviste ou simplement un scientifique) une valeur morale ou une expérience religieuse. On obtient, comme l'illustrait Watts, l'effet inverse de l'effet attendu. Face au cas de conscience, la conscience *subjective*, qui par chance n'a pas subi d'endoctrinement lourd, ou de lavage de cerveau, reconnaît, selon Valadier, son inéluctable manque de certitude et la nécessité de son humilité. Valadier écrit :

...jamais une question morale ne surgit autrement que sous cette forme d'un "étonnement" à laisser sourdre, qui déplace les anciennes évidences (Valadier 1990).

Valadier et Watts mettent ici en évidence une caractéristique universelle de l'*intention* (concept abondamment étudié par les philosophes anglo-saxons, voir Searle 1992, Dennett 1988), c'est que la communication (finie), ou même la mise en action *normalisée* de l'intention, *trahit* cette même intention. On n'aboutit pas nécessairement à une "catastrophe" ou au faux, comme avec le principe de Wittgenstein, car si l'esprit (la conscience) est ouverte, capable de reconnaître cette trahison, capable de *laisser sourdre l'étonnement*, alors les *anciennes évidences*, on peut en tout cas l'espérer (à la lecture de Watts et Valadier), sauront se déplacer d'elles-mêmes. Dans ce cas un apprentissage est possible. Je le représente de façon symbolique :

□⊥ ---> □' ⊤

signifiant que le sujet (□) a opéré une transformation (logiquement non-monotonique) sur lui-même : □ ---> □'. Ceci suggère par ailleurs une relation entre la conscience et l'apprentissage (voir aussi Oatley 1988).

Je ne prétends pas que le propos de Watts et les propos de Valadier ou de Lao-tseu, dans leurs ouvrages respectifs cités, soient semblables. Ce que j'appelle Principe de Lao-tseu-Watts-Valadier capture seulement une intuition commune concernant la conscience, intuition que je tente d'expliciter au moins partiellement.

On trouvera aisément d'autres exemples, notamment dans le livre de Martin Gardner (Gardner 1983). Les exemples abondent dans la poésie et les chansons. Les sentiments amoureux sont toujours trahis par les mots et cela d'autant plus qu'on tente un usage direct et littéral, d'où l'intérêt de la poésie en ce domaine. Ils abondent aussi dans l'histoire politique ou l'intention (qui est toujours au départ l'apanage d'une conscience, ou d'une prise de conscience) semble toujours trahie par son institutionnalisation (laquelle est une façon de rendre cette intention nécessaire). On reconnaît encore le dicton populaire : *l'enfer est pavé de bonnes intentions.*

<u>*2°) Principes équationnels*</u>

J'espère avoir assez illustré deux candidats pour une *équation* de la conscience, x est l'inconnue:

$$\Box x \rightarrow \bot \quad \text{ou} \quad \neg \Box x \qquad \text{(W)}$$
$$\Box x \rightarrow \neg x \quad \text{ou} \quad x \rightarrow \neg \Box x \qquad \text{(LWV)}$$

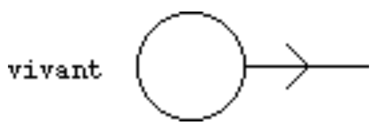
<u>*3°) Recherche d'une solution*</u>

La plus simple des solutions est $x = \bot$. Dans ce cas W semble transmettre une injonction du genre "on ne communiquera (justifiera) pas le faux". Cela ressemble à une injonction morale du genre "tu ne mentiras pas", ou à un principe de consistance "tu ne te tromperas pas". Tout laisse penser cependant que Wittgenstein considérait l'objet de l'incommunicabilité comme une proposition vraie bien que métaphysique. Notons cependant qu'avec le faux W est une instantiation de l'axiome T de la connaissance. Avec LWV, quant à lui, avec $x = \bot$, il devient une tautologie classique et est d'office respecté par tous les référentiels. Le carré modal peut vérifier n'importe quel axiome propositionnel modal reposant sur la logique classique. Ce respect est un peu trivial puisqu'il est dû au fait que, classiquement, le vrai est impliqué par toute proposition.

A présent on va utiliser l'approche modale *géométrie d'abord, linguistique ensuite* pour tenter d'isoler une solution qui ne soit pas équivalente à une proposition fausse. Une telle approche a été proposée dans la brève introduction à la théologie au début de cette section, et l'expérience personnelle (indexicale) y est représentée (géométriquement) par une flèche

Mourir, de façon absolue, consiste à accéder à un dernier monde :

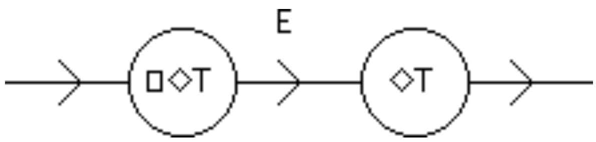
Vivre ou être vivant, ou être conscient, ou être dans un état vivant (conscient), dans ce cas consiste à être dans un monde (état) transitoire :



Afin de ne pas confondre la vie spirituelle dont je parle avec la vie clinique, j'utiliserai exclusivement le terme conscient.

Prétendre survivre après une expérience, prétendre qu'il y a une vie après la mort clinique par exemple, est certainement une chose dont on ne peut parler, si on veut respecter le principe de communicabilité vérifiable de Wittgenstein (voir aussi Amstrong & Malcolm 1984). *Survivre à quoi que ce soit*, y compris une nuit de sommeil, est une proposition métaphysique par excellence. Malheureusement ce principe, s'il est correcte reste incommunicable et je devrai justifier la façon dont je le communique. Il ne peut s'agir d'une communication positiviste sous peine de franche inconsistance. En 1.3 on verra avec des expériences par la pensée comment distinguer W, vrai mais non communicable, de LWV, vrai et communicable. Toute la deuxième partie montrera la même chose, mais sans utiliser d'expériences par la pensée. Il n'en demeure pas moins qu'admettre à présent l'incommunicabilité de la survie permet de motiver pour un référentiel respectant LWV.

Survivre à une expérience nécessite de rester conscient à travers cette expérience. On survit donc à une expérience si on accède par elle à un monde transitoire. Dans le cas linéaire, c-à-d dans le cas où de chaque état part au plus une flèche,



$\Box\Diamond \top$ garantit la survie à l'expérience E. Communiquer de façon positiviste (vérifiable), ou solipsiste (vérifiable et vraie en usant du stratagème) revient à admettre $\Box\Diamond \top$ ou $\boxdot\Diamond \top$.

Dans ce cas une solution est suggérée par la brève introduction à la théologie :

$$x = \Diamond \top,$$

ou plus généralement $x = \Diamond p$, avec p vraie puisque la proposition "je peux survivre à une expérience E et donc accéder à un monde où p est vrai" est tout autant *métaphysique*. On est donc amener aux formules modales suivantes. Les deux premières illustrent le principe de Wittgenstein, les deux suivantes illustrent le principe de Lao-Tseu, Watts, Valadier.

$$\neg \Box\Diamond \top$$
$$\neg \Box\Diamond p$$

$\Box\Diamond\top \to \neg\Diamond\top$
$\Box\Diamond p \to \neg\Diamond p$

Anticipation.

Avec l'hypothèse mécaniste, deux justifications supplémentaires seront apportées pour la solution $\Diamond\top$: par des expériences par la pensée (en 1.3), et par les conséquences du théorème de Gödel avec l'hypothèse du mécanisme digital (c'est l'objet de toute la deuxième partie).

Les principes de Wittgenstein et de Watts-Valadier capturent un aspect de la conscience religieuse, la foi, la morale, et même, semble-t-il des qualia. Par exemple:

$\Box$ beau -> $\neg$ beau (avec $\neg$beau = laid[20])
$\Box$ bon -> $\neg$ bon (avec $\neg$bon = mal)

est une réécriture des énoncés de Lao-tseu où $\Box$, jouant ici le rôle du *décret*, représente à nouveau un essai de capture formelle, essentiellement finie. Notons que $\Box$ ne peut pas représenter la perception du qualia avec ou sans le stratagème S. Intuitivement, cependant, la communication positiviste (éventuellement vérifiable) et le qualia de l'expérience sont liés car une suite de confirmations semble nécessaire pour qu'une foi, éventuellement attachée à des propositions non confirmables ou très indirectement confirmables, puisse se développer. A ce sujet on peut déjà mentionner la théorie de Helmholtz dans laquelle la perception est décrite comme une inférence (donc de type foi) inconsciente ou instinctive. (voir Gregory 1988, Oatley 1988, Barlow 1987). Notons aussi que la communication positiviste est sociale et s'adresse à l'autre, tandis que l'expérience est purement individuelle.

remarque : si on accepte d'entendre par *blasphème* une proposition de type $\Box\Diamond\top$, on peut déduire (informellement) des réflexions de Valadier qu'utiliser l'argument d'autorité dans la religion, est une façon d'institutionnaliser le blasphème. Ce que Locke <ici : lettre sur la tolérance> avait, je pense, déjà saisi.

<u>*4°) Sémantique de Kripke*</u>

1) Wittgenstein : $\neg\Box\Diamond\top$

---

[20] En écrivant (-beau = laid), je néglige l'état "agnostique" de l'indifférence. Je commets donc l'erreur sur laquelle j'attire l'attention depuis le début. On peut la corriger en introduisant une notion de ressource et en interprétant la négation comme une diminution de ressource. Il faut alors indexer les flèches (numériquement ou avec les points d'un treillis). Cela revient à "fuzzifier" la présente approche. J'aurai ultérieurement d'autres raisons de mentionner cette possibilité en (3.3).

Il n'y a pas de sémantique de Kripke pour le principe de Wittgenstein. En effet si $\neg \Box \Diamond \top$ est vrai dans tous les mondes d'un référentiel non vide, alors $\Diamond \neg \Diamond \top$ est vrai dans tous les mondes, et cela montre que tous les mondes sont transitoires. Mais dans ce cas tous les mondes n'accèdent qu'à des mondes eux-mêmes transitoires, et dès lors $\Box \Diamond \top$ est vrai dans tous les mondes. C'est juste la négation de ce que l'on souhaitait. Cela montre qu'il va falloir creuser davantage (ou espérer un cadeau inattendu !).

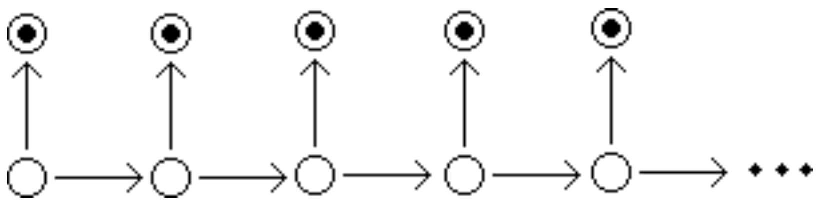
2) Wittgenstein généralisé : $\neg \Box \Diamond p$

Il en va de même pour $\neg \Box \Diamond p$ puisque le respect ne dépend pas de la valuation.

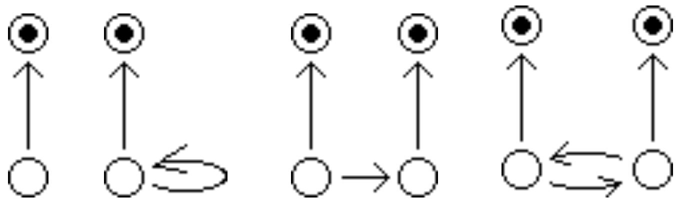
3) Watts-Valadier :

définition un référentiel <W,R> est réaliste si pour tout monde transitoire α appartenant à W il existe un dernier monde δ accessible à partir de α.

Exemple :



autres exemples :



On voit qu'à la différence d'un référentiel idéal où "l'immortalité"[21] est en quelque sorte garantie, dans un référentiel réaliste, bien que l'immortalité est possible (si le référentiel est infini ou possède une boucle) elle n'est jamais garantie. Partout on peut emprunter un chemin (une expérience, ou une suite d'expériences si la relation est transitive) qui aboutit à un dernier monde.

Propositions

[21] notons l'abus de langage

1) tous les référentiels réalistes respectent $\Box\Diamond p \rightarrow \neg\Diamond p$,

2) Seuls les référentiels réalistes respectent $\Box\Diamond p \rightarrow \neg\Diamond p$.

Les référentiels réalistes caractérisent ainsi la formule modale $\Box\Diamond p \rightarrow \neg\Diamond p$. Dit de façon plus succincte :

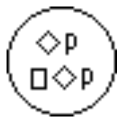
$$(W,R) \text{ respecte } \Box\Diamond p \rightarrow \neg\Diamond p \iff R \text{ est réaliste}$$

Preuve

1) R est réaliste => (W,R) respecte $\Box\Diamond p \rightarrow \neg\Diamond p$.

je démontre par l'absurde.

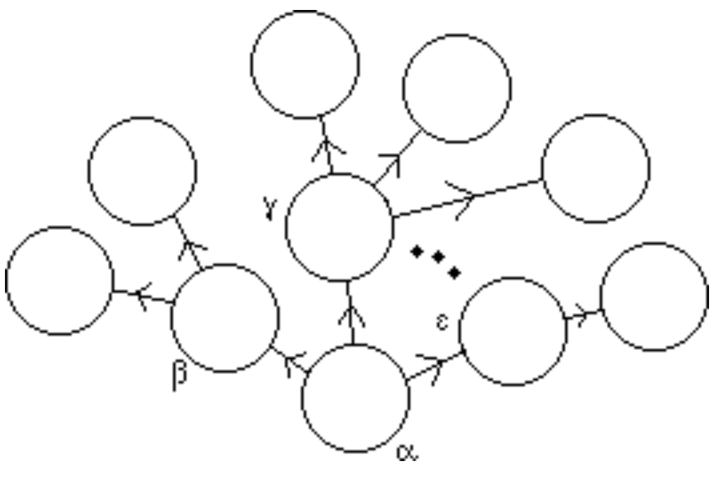
Supposons que R est réaliste et que (W,R) ne respecte pas $\Box\Diamond p \rightarrow \neg\Diamond p$. Il existe donc un modèle (W,R,V) comprenant un monde où $\Box\Diamond p$ est vrai et $\neg\Diamond p$ est fausse, c-à-d $\Diamond p$ est vrai



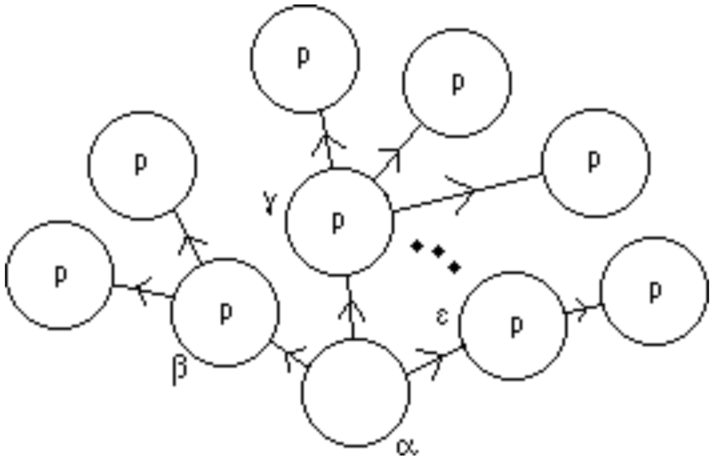
Mais si R est réaliste, ce monde accède à un dernier monde dans lequel on aura $\Diamond p$, ce qui est absurde pour un dernier monde.

2) (W,R) respecte $\Box\Diamond p \rightarrow \neg\Diamond p$ => R est réaliste

Supposons que (W,R) respecte $\Box\Diamond p \rightarrow \neg\Diamond p$ et que R ne soit pas réaliste. On va construire une valuation témoignant que (W,R) ne respecte pas $\Box\Diamond p \rightarrow \neg\Diamond p$. Dire que R n'est pas réaliste signifie qu'il existe un monde α transitoire n'accédant à aucun dernier monde, c-à-d n'accédant qu'à des mondes eux-mêmes transitoires : $\beta\ \chi$ .. ε :



Prenons la valuation qui consiste à mettre p (V(p) = 1) dans tous les mondes auxquels α accède ainsi que dans tous les mondes auxquels ces mondes transitoires eux-mêmes accèdent et p quelconque (V(p) = ce que vous voulez) partout ailleurs. En l'occurence on peut prendre simplement pour p, la constante propositionnelle $\top$.



Il est clair, par Kripke, que $\Diamond$p et $\Box\Diamond$p sont vraies dans α, ce qui contredit notre hypothèse selon laquelle le référentiel respecte $\Box\Diamond p \to \neg\Diamond p$. Remarquons que dans le monde α, on a pas eu besoin de donner de valuations. On peut remplaçant "p" par "$\top$", on voit que le principe de Watts Valadier restreint $\Box\Diamond\top \to \neg\Diamond\top$ est sémantiquement équivalent au principe général.

Complétude (W,R,V) étant le modèle canonique de C (voir plus haut la définition), il suffit de démontrer :

1) C $\vdash$ F <=> F est vraie dans tous les mondes de W,

2) (W,R,V) est réaliste.

1) C $\vdash$ F <=> F est vraie dans tous les mondes de W.

a) C $\vdash$ F => F est vraie dans tous les mondes de W. Supposons que C $\vdash$ F et que F soit fausse dans un monde de W. Il existe alors un ensemble E maximal consistant qui étend {$\neg$F} ce qui contredit le fait que C $\vdash$ $\neg\neg$F et que E soit consistant.

b) F est vraie dans tous les mondes de W => C $\vdash$ F. En effet si C $\nvdash$ F, il existe un ensemble maximal consistant qui étend {$\neg$F}, donc $\neg$F appartient à un monde de W, donc $\neg$F est vraie dans un monde de W et F n'est pas vraie dans tous les mondes de W.

2) (W,R) est réaliste.

Supposons que le modèle (W,R,V) ne soit pas réaliste. Dans ce cas il existe un monde transitoire α aboutissant, par R, exclusivement à des mondes β_i transitoires. Ces mondes β_i aboutissent eux-mêmes à des mondes, qui comme ensembles maximaux consistants de formules contiennent tous $\top$. Grâce au "point délicat", voir plus haut, on a $\Diamond\top$ dans tous les mondes β_i, et on a encore $\Box\Diamond\top$ dans α. Par ailleurs, C $\vdash$ $\Box\Diamond\top\text{->}\neg\Diamond\top$, donc $\alpha \Vdash \Box\Diamond\top\text{->}\neg\Diamond\top$, et donc (les extensions maximales étant fermées pour le modus ponens) on a $\alpha \Vdash \neg\Diamond\top$, donc $\alpha \Vdash \Box\bot$. Donc les β_i (qui existent) $\Vdash \bot$ ce qui contredit leurs natures d'extensions maximales consistantes.

5°) Relation entre référentiels idéals et référentiels réalistes

1) Un référentiel idéal n'est jamais réaliste,

2) Un référentiel réaliste n'est jamais idéal,

3) Il existe des référentiels qui ne sont ni idéal ni réaliste,

4) En se limitant à la collection des référentiels transitifs, on a : un référentiel est non-réaliste si et seulement si il admet un sous-référentiel idéal.

preuve

1) et 2) sont évidentes à partir des définitions.

3) voici un référentiel qui n'est ni idéal ni réaliste :

4) Si un référentiel (transitif ou non) admet un sous-référentiel idéal, il n'est pas réaliste puiqu'il admet un monde transitoire qui n'aboutit pas à un dernier monde. Reste à démontrer la réciproque : un référentiel réaliste, avec R transitive, admet un sous-référentiel idéal. En effet si le référentiel <W,R> n'est pas réaliste il existe un monde transitoire α tel que pour tout β $\alpha R\beta$ -> β transitoire. Soit W' = $\{\xi \text{ tq } \alpha R\xi\}$, et soit R' = R restreinte à W'. Je prétends que <W', R'> est un sous référentiel idéal. En effet si <W', R'> n'était pas idéal, il existerait γ appartenant à W' avec $\gamma R'\delta$ et δ = dernier monde. Mais g $\in$ W' entraîne $\alpha R\gamma$ (par définition de W'), et $\gamma R'\delta$ entraîne $\gamma R\delta$ (puisque R' est une sous relation de R). Si R est transitive cela entraîne $\alpha R\delta$. Cela contredit le fait que α n'aboutit qu'à des mondes transitoires.

Je propose un sujet de réflexion : dans quel sens pourrait-on dire que l'application du stratagème affaibli rend idéal un référentiel réaliste ?

remarques

1) Si on examine chacune des citations qui ont motivé les principes de Wittgenstein et les principes de Watts et Valadier, on remarquera qu'elles

sont toutes sous la forme : "on ne peut pas communiquer ...". J'ai toutefois, par chance, découvert une "définition de la conscience" non seulement positive mais formellement plus proche de la réflexion sur la vie ($\Diamond\top$) que je propose :

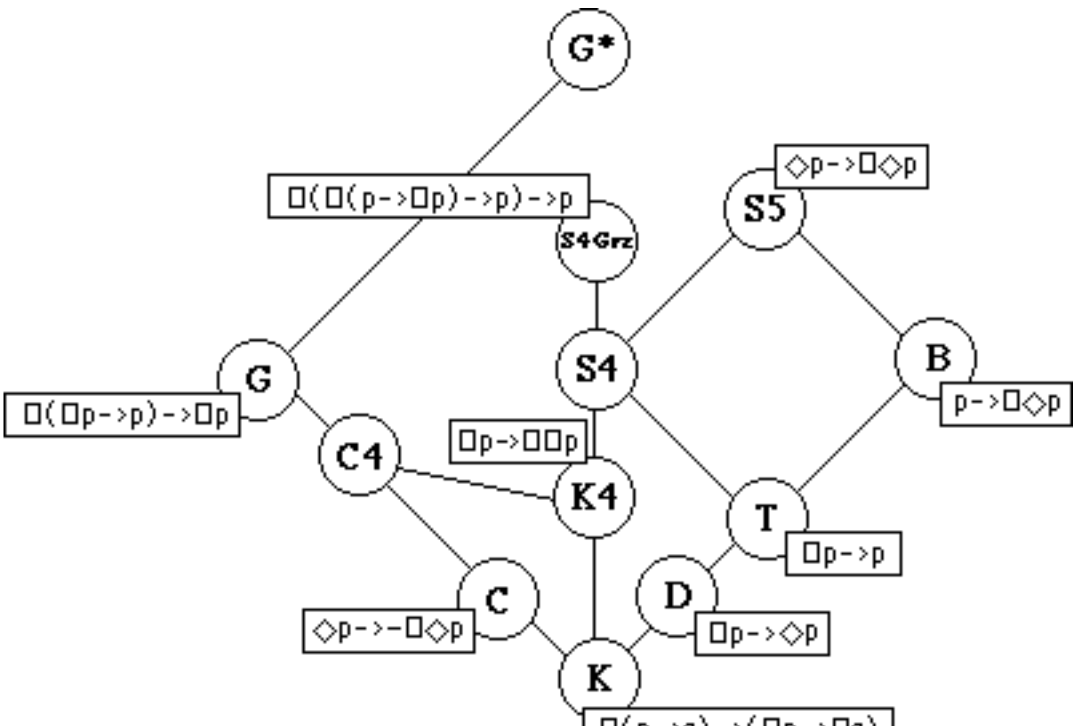
> *la conscience, c'est le pressentiment de la vérité accessible par un homme.*
> Dostoïevski (voir Thibaudat 1989)

2) L'accessibilité à la vérité, $\Diamond\top$, peut bien sûr s'écrire négativement puisque $\Diamond\top$ est équivalent à $\neg \Box \bot$. Si on interprète $\Box$ par la prouvabilité formelle, le principe de Watts-Valadier est formellement semblable au second théorème d'incomplétude de Gödel :

$$\Diamond \top \rightarrow \neg \Box \Diamond \top$$

L'hypothèse du mécanisme et la thèse de Church permettront en effet d'interpréter les phénomènes d'incomplétude Gödeliens dans les termes d'une théorie abstraite de la conscience.

Résumé graphique des modalités rencontrées et à venir : les arcs allant du bas vers le haut représentent l'inclusion (les théories étant vues comme des collections de théorèmes).



A l'exception de G*, je n'ai pas placé sur ce diagramme les logiques non normales que nous rencontrerons aussi. La plupart des logiques non normales résulteront ici de l'application du stratagème affaibli sur une logique normale, ou sur G*.

6°) Ce qu'on peut attendre d'une théorie de la conscience

1) Qu'elle ne tombe pas dans le piège (consistant à communiquer de l'incommunicable), dans lequel Lao-Tseu ou Wittgenstein ont été accusés de tomber. De toutes théories on peut exiger, ou simplement espérer qu'elle ne soit pas inconsistante. Comme on le voit ici le risque est élevé, et, paradoxalement d'autant plus élévé que la théorie est formelle ou susceptible d'être formalisée. En ce qui concerne le principe de Watts Valadier l'existence d'un modèle de Kripke garanti sa consistance, mais ce n'est pas aussi clairement le cas pour le principe de Wittgenstein.

2) Qu'elle réponde aux questions que les philosophes abordent sur la conscience :

a) Qu'est-ce que l'identité personnelle ? (voir 2.2)

b) La conscience a-t-elle un rôle ? A-t-elle plusieurs rôles, lesquelles ? (cf Jackendoff 1987, Barlow 1987, Oatley 1988) (voir 2.3).

c) Quelles relations y a-t-il avec d'autres approches, comme les idées sur la perception de Helmoltz 1896 (voir 2.3).

d) Quelle distinction peut-on faire entre la conscience transitive et la conscience intransitive de Malcolm (Armstrong et Malcolm 1984).

e) Quelle relation peut-il exister entre la théorie de la conscience au sens large présentée ici et l'intelligence artificielle. Quelle relation la conscience entretient-elle avec l'intelligence au sens restreint de compétence, comme l'intelligence au sens de Binet (1911) par exemple, (voir 2.3).

f) Y a-t-il compatibilité avec la neurophychiatrie, théorie des rêves (sommes nous conscients pendant un rêve, etc... (voir 3.1).

g) Comment expliquer l'aspect *supervénience* et *flux de la conscience ?* Comment formuler le problème du corps et de l'esprit ? C'est pratiquement l'objet de toute la troisième partie.

h) Quelles relations y a-t-il entre la conscience et les qualias ? La contreverse entre Goethe et Newton sur les couleurs (cf le texte de Schrödinger plus haut) devrait être clarifiée (voir 3.3).

On devine la grande difficulté d'une telle approche. D'où l'intérêt d'utiliser une hypothèse supplémentaire : comme l'hypothèse mécaniste digitale. Tombera-t-on dans le piège qui consiste à justifier l'injustifiable ? Pas nécessairement car la thèse de Church, analysée dans la deuxième partie, implique l'inexistence d'une théorie formelle capable de décider systématiquement les propositions sur les machines elles-mêmes, elle sert ainsi de protection contre un réductionnisme du mécanisme lui-même. La simple hypothèse du mécanisme indexical, MEC-IND, permet cependant d'affiner considérablement l'approche présente sans utiliser la thèse de Church. L'idée est d'utiliser des expériences par la pensée que l'on peut concevoir avec la thèse indexicale pour traiter la question de la survie à une *expérience particulière* que j'aborde dans la section suivante. Cette section justifiera concrètement l'impossibilité de garantir la survie lors de cette expérience particulière et permet de montrer ainsi pourquoi la vérité de W (avec $x = \bot$, $x= \neg \Box \bot$, etc.) entraîne son incommunicabilité, ce qui justifie LWV. Il s'agit donc d'un approfondissement *mécaniste* de l'approche :*géométrie d'abord, linguistique ensuite.*

1.2.4 Résumé

La théologie est conçue ici comme l'étude des mondes ou des états possibles en relation avec la question de survie de soi (de son âme) ou des autres. La logique modale, extension de la logique classique (le langage comprend un symbole supplémentaire "$\Box$") et la sémantique de Kripke (en terme de collection de mondes) est naturellement appropriée dans ce contexte. L'interprétation de "$\Box p$", due à Kripke, est que $\Box p$ est vrai, dans le monde μ, si p est vrai dans tous les mondes auxquels je peux accéder (à partir de μ). Dualement, l'interprétation de $\Diamond p$ (abréviant $\neg \Box \neg p$) est que $\Diamond p$ est vrai, dans le monde μ, s'il existe un monde accessible, à partir de μ, où p est vrai.

Des référentiels spécifiques permettent de caractériser des formules modales. Ainsi $\Box(p\text{->}q)\text{->}(\Box p\text{->}\Box q)$ est satisfaite dans tous les mondes de tous les référentiels, quelle que soit la nature de la relation d'accessibilité ou la valeur de la proposition p. Un axiome acceptable (linguistiquement) pour la "crédibilité", ou pour la "communicabilité positiviste" (c-à-d $\Box$ = croyable), est $\Box p\text{->}\Box\Box p$. Cette formule est caractérisée par les référentiels transitifs. Pour le "savoir" et le "connaissable", ou pour la "communication positiviste correcte" (c-à-d $\boxdot$ = savoir, sachable) on admet de plus la formule $\boxdot p\text{->}p$, qui est caractérisée par les référentiels réflexifs.

Un stratagème consistant à identifier $\boxdot p$ avec $\Box p$ & p est permis en ce sens que $(\Box p \,\&\, p)\text{->}p$, donc $\boxdot p\text{->}p$, et on peut vérifier la transitivité, ainsi que l'axiome K (voir 2.3 pour plus de détails).

Mentionnons l'axiome déontique $\Box p\text{->}\Diamond p$, il est caractérisé par les référentiels idéaux.

A partir de considérations linguistiques basées notamment sur Wittgenstein (la fin du tractatus), mais aussi Lao-Tseu, Watts et Valadier, j'ai suggéré comme axiome équationnel intuitif pour une théorie de la conscience (celle-ci jouant le rôle de l'inconnue x sous réserve d'une théorie de la communication convaincante ou positiviste) :

¬ □ x (principe de Wittgenstein)
□ x->¬ x (principe de Lao-Tseu, Watts, Valadier)

A partir de considérations plus géométriques reposant sur l'idée que la survie nécessite la préservation de la conscience lorqu'on accède à "un autre monde", la solution x= ◊T a été suggérée. La (méta)théorie de la conscience est alors donnée par l'axiome □◊T->¬◊T, Ou plus généralement par □◊p->¬◊p. Une telle formule est caractérisée par les référentiels que j'appelle "réalistes" : un dernier monde est accessible à partir de tout monde transitoire.

1.2.5 Biblio-locale

ALECHINA N., **1994**, *Logic with Probabilistic Operators*, 1-13.preprint.

ARMSTRONG D. M. & **MALCOM** N., **1984**, *Consciousness & Causality,* Basil Blackwell, Oxford.

ARSAC J et **SENTIS** P., (eds), **1990**, *Sciences et Sens*, Actes du Colloque organisé dans le cadre de l'Académie Meudonnaise. Vrin, Paris.

AUBERT J.-M., **1991**, *Et après... Vie ou néant ?,* Desclée de Brouwer, Paris.

BAILHACHE P., **1991**, *Essai de logique déontique*, Vrin, Paris.

BARLOW H., **1987**, *The Biological Role of Consciousness*, in C. Blakemore & S. Greenfield (Eds.) 1987.

BENOIT P., **1989**, *La théologie au XIIIe siècle : une science pas comme les autres*, in Eléments d'histoire des sciences, Bordas, Paris.

BINET A., **1911**, *Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école*, L'année psychologique, 19, pp. 145-201.

BLAKEMORE C. & **GREENFIELD** S., **1987**, *Mindwaves,* Basil Blackwell, Oxford.

BROUWER L.E.J., **1905**, *Leven, Kunst en Mystiek* (La Vie, l'Art et le Mysticisme), Waltman, Delft. Cité dans van Stigt 1990 page 199.

CANTO-SPERBER M., (Ed.), **1991**, *Les paradoxes de la connaissance, essais sur le Ménon de Platon*, Eds Odile Jacob, Paris.

CARSE J. P., **1986**, *Finite and Infinite Games*, Macmillan Publishing Company, New York, Traduction française, Seuil 1988.

CHELLAS B. F., **1980**, *Modal logic an introduction*, Cambridge University Press, Cambridge.

DENNETT D. C., **1988**, *Intentionality,* 3ème Colloque International de l'ARC, (Association pour la Recherche Cognitive), Cognition et Connaissance, CNRS / Université Paul Sabatier-Toulouse, Toulouse, pp. 193-227.

DINOUART (abbé), **1771**, *L'art de se taire, principalement en matière de religion*, Ed. Jérôme Millon, 1987, France.

FEYS R., **1937**, *Les logiques nouvelles des modalités*, Revue néoscolastique de philosophie, vol. 40, pp. 517-553, + vol. 41 (1938), pp. 217-252.

FITTING M., **1983**, *Proof Methods for Modal and Intuitionistic Logics,* D. Reidel Publishing Company, Dordrecht.

FROSSARD A., **1990**, *Dieu en questions,* Desclée de Brouwer, Stock / Laurence Pernoud.

FORD N. M., **1988**, *When did I begin?***,** Cambridge University Press, Cambridge.

GARDIES J. L., **1979**, *Essai sur la logique des modalités*, Presses Universitaire de France, Paris.

GARDNER M., **1983**, *The Why of a Philosophical Scrivener,* Oxford University Press, 1985, (first published in 1983 by William Morrow and Company, Inc.)

GRANGER G.-G., **1992**, *La vérification*, Ed. Odile Jacob, Paris.

GODEL <pas de table de vérité en logique int, Glivenko ?>

GRZEGORCZYK, A., **1964**, *A Philosophicaly Plausible Formal Interpretation of Intuitionistic Logic*, Indagationes Math. 26, pp. 596-601.

GREGORY R. L., **1988**, *Consciousness in science and philosophy: conscience and conscience*, in A. J. Marcel & E. Bisiach (Eds.) 1988.

GRENIER, J., **1973**, *L'esprit du Tao*, Flammarion, Paris.

GRISWOLD C. L., **1986**, *Self-Knowledge in Plato 's Phaedrus*, Yale University Press. New Haven and London.

HELMOLTZ H., **1896**, *Handbuch der Physiologischen Optik*, Zweite Auflag., Voss, Hamburg & Leipzig.

JACKENDOFF R., **1987**, *Consciousness and the Computational Mind*, The MIT Press, Cambridge.

HEYTING A., **1956**, *Intuitionism, an Introduction*, North-Holland, 3ème éd. 1980.

HUXLEY A., **1946**, *Les mots et la réalité,* dans *Les portes de la perception*, trad. Franç. Jules Castier, Éditions du Rocher, 1979, Monaco, pp. 123-128.

KERR F., **1991**, *La théologie après Wittgenstein*, (traduction), Les éditions du Cerf.

KNEALE W. and **KNEALE** M., **1962**, <u>*The developpement of Logic*,</u> Clarendom Press, Oxford (last reprinting 1988).

KRIPKE, S. A., **1963a**, *Semantical Considerations on Modal Logic*, Acta Philosophica Fennica 16, pp. 83-94.

KRIPKE, S. A., **1963b**, *Semantical Analysis of Modal Logic I : Normal Propositional Calculi*, Zeit. Math. Logik. Grund., 9, pp. 67-96.

KRISHNAMURTI, **1973**, <u>*L'éveil de l'intelligence*</u>, Textes enregistrés aux Etats-Unis, en Inde, en Suisse et en Grande-Bretagne, traduit par Anette Duché, Ed. Française établie par Nadia Kossiakov, Ed. Stock, 1975, 1980.

LEIBNITZ G. W., **1714**, *Monadology*, in <u>Philosophical Writings</u>, édité par G.H.R. Parkinson, trad. Angl. de Mary Morris & G.H.R. Parkinson, Dent, London.

LAO TSEU, **1977**, <u>*Tao Te King, Le livre de la Voie et de la Vertu*,</u> trad. de Claude Larre, Desclée de Brouwer, Bellarmin.

LAO TSEU, **1974**, <u>*Tao-te-king*,</u> version allemande de R. Wilhelm, trad. en français E. Perrot, librairie de Médicis, Paris.

LAO-TSEU, **1967**, <u>*tao tö king*,</u> trad. de Liou Kia-hway, Gallimard.

LAO-TSEU, **1967**, <u>*tao tö king*,</u> trad. de Liou Kia-hway, dans Philosophes Taoïstes, bibliothèque de la pléiade, Nrf, Gallimard.

LAO-TZEU, **1979**, <u>*La Voie et sa vertu*</u> *Tao-tê-king*, texte chinois présenté et traduit par F. Houang et P. Leyris, Editions du Seuil.

LOCKE J., **1667**, *Lettre sur la tolérance*, dans <u>Lettre sur la tolérance et autres textes</u>, trad. française par Jean Le Clerc, Flammarion, 1992, Paris.

MACKAY D. M, **1960**, *On the Logical Indeterminacy of a Free Choice*, Mind 69, n° 273, pp. 31-40.

MARCEL A.J. and **BISIACH** E., (eds), **1988**, *Consciousness in Contemporary Science*, Clarendom Press, Oxford.

MARTIN R. N. D. & **BROWN** S., (eds), **1988**, <u>*G. W. Leibnitz: Discourse on Metaphysics*,</u> Manchester University Press, Manchester and New-York.

OATLEY K., **1988**, *On changing one's mind: a possible function of consciousness*, in A. J. Marcel & E. Bisiach (Eds.).

O'MEARA D. J., **1989**, <u>*Pythagoras Revived Mathematics and Philosophy in Late Antiquity*</u>, Clarendon Paperbacks 1990.

OUZOULIAS A., **1989**, <u>*La conscience*,</u> Editions Quintette, Paris.

PARFIT D., **1984**, <u>*Reasons and Persons*</u>, Clarendom Press, Oxford.

PERRY, J., **1975**, *Personal Identity,* University of California Press, Berkeley.

PLATON, ????, *Théétète ou de la science,* oeuvre de la pléiade, Editions Gallimard, 1950, pp. 83-192.

SACKS O., **1985**, *The Man Who Mistook His Wife for a Hat*, Gerald Duckworth & Co (Pan Books 1986, traduction française chez Seuil 1988).

SALAM A., **HEISENBERG** W., **DIRAC** P. A. M., **1990**, *La Grande Unification,* Seuil, Paris.

SCHRÖDINGER E., **1958**, *Mind and Matter,* Cambridge university Press, Cambridge. Trad. franç.par M. Bitbol: l'esprit et la matière, Editions du Seuil, Paris 1990.

SEARLE J. R., **1992**, *The Rediscovery of Mind*, The MIT Press, Cambridge.

SMULLYAN R., **1977**, *The Tao is Silent,* Harper and Row, New-York.

ROCHAT DE LA VALLEE E, **?**, *Saine Incertitude. Texte, présentation, traduction et commentaire du chapitre 2 du Zhuangzi*, Institut Ricci, Paris.

SOLOVAY R. M., **1976**, *Provability Interpretation of Modal Logic*, Israel Journal of Mathematics, Vol. 25, pp. 287-304.

SWINBURNE R. and **SHOEMAKER** S., **1984**, *Personal Identity*, Basil Blackwell, UK.

THIBAUDAT J. P., **1989**, *De l'autre coté du rideau rouge*, Libération, N° 2, juillet 1989. Dostoïevski est cité par Oleg Tabakov (homme de théatre Soviétique).

THOMASON R. H., **1980**, *A Note on Syntactical Treatment of Modality*, Synthese, 44, pp. 391-395.

TURNER, R., **1990**, *Truth and Modality for Knowledge Representation*, Pitman UK.

TROELSTRA A. S. and **VAN DALEN** D., **1988**, *Constructivism in Mathematics, An Introduction*, (2 volumes) North Holland.

ULLMO J., **1969**, *La pensée scientifique moderne*, Flammarion, Paris.

VALADIER P., **1990**, *Inévitable Morale.* Editions du Seuil.

VAN BENTHEM J., **1985**, *A Manual of Intensional Logic,* Lecture Notes CLSI, Stanford.

VAN STIGT W. P., **1990**, *Brouwer's Intuitionism,* Studies in the history and philosophy of Mathematics, Vol 2, North Holland, Amsterdam.

WANG H, **1974**., *From Mathematics to Philosophy*, Routledge & Kegan Paul, London.

WATTS A., **1951**, *The Wisdom of Insecurity,* Pantheon Books, a Division of Random House, Inc. (trad. Franç. : Bienheureuse insécurité, Stock / Monde ouvert 1977).

WATTS A., **1974**, *The Essence of Alan Watts,* Celestial Arts, Millbrae, California. (trad. Franç. : L'Envers du Néant, Editions Denoël/Gonthier, Paris, 1978).

WIEGER, L., **1950**, *Les pères du système taoïste*, Les belles lettres, Paris.

WITTGENSTEIN L.,**1918**, *Tractatus Logico-Philosophicus,* Gallimard, Paris, 1961.

WITTGENSTEIN L.,**1965**, *De la certitude*, Gallimard, 1976.