

1 LE MECANISME INDEXICAL

1.1 Philosophies mécanistes

1.1.1 Différentes sortes de Mécanismes

Je distingue les hypothèses mécanistes suivantes :

MECANISME BEHAVIORISTE *Des machines peuvent avoir un comportement d'êtres pensants (vivants, conscients, etc.) (MEC-BEH)*

MECANISME FORT *Des machines peuvent penser (être vivantes, conscientes, avoir un vécu privé, etc.) (MEC-FORT)*

MECANISME INDEXICAL *Je suis une machine (ou vous êtes une machine, ou encore les êtres humains sont des machines) (MEC-IND)*

En remplaçant machine par machine digitale, on obtient les thèses digitales correspondantes.

Le mécanisme behavioriste digitale MEC-DIG-BEH correspond grossièrement à celui que Turing expose dans son article de 1950. Le mécanisme digital fort MEC-DIG-FORT correspond de même à ce qui est appelé dans la littérature la thèse forte de l'intelligence artificielle (strong AI thesis).

Dans ce travail je m'intéresse exclusivement au mécanisme indexical et digital (MEC-DIG-IND ou encore MDI). La digitalité nécessite la thèse de

Church , c'est pourquoi l'aspect digital est relayé dans la deuxième partie. J'y montrerai comment un procédé, essentiellement dû à Gödel, permet de traiter l'indexicalité chez les machines digitales.

Proposition :

MEC-IND => MEC-FORT => MEC-BEH, et
MEC-BEH ≠> MEC-FORT ≠> MEC-IND.
(avec ou sans l'hypothèse de digitalité)

justification : On admet que les êtres humains savent penser, (être conscient, avoir un vécu privé, etc.). Dès lors MEC-IND entraîne MEC-FORT et MEC-FORT entraîne MEC-BEH. Que MEC-BEH n'entraîne pas MEC-FORT est justifié par Weizenbaum (1976) (voir aussi Gunderson¹ 1971). MEC-FORT n'entraîne pas MEC-IND, car le fait que les machines puissent penser n'entraîne pas qu'elles soient les seules à penser. Il est concevable que les machines puissent penser sans que nous soyons nous-mêmes des machines. Wang (1974) présente une justification similaire. Mais de nombreux philosophes utilisent implicitement l'opinion contraire : MEC-FORT => MEC-IND, voir par exemple Arzac 1987.

La thèse mécaniste, d'ordre pragmatique, suivante ne sera pas abordée :

MECANISME METAPHORIQUE *Tels organes biologiques peuvent être avantageusement comparés à telles machines (ex; le coeur comparé à une pompe, le cerveau à un réseau électrique régulé chimiquement, etc.)*

Nous rencontrerons occasionnellement l'hypothèse suivante :

PANMECANISME *L'univers est une machine* ou encore *tout est machine* (PM) (on aura MEC-IND -> PM est indécidable)

et accessoirement :

ULTRA-PANMECANISME *L'univers est une machine globalement finie.* (UPM). Globalement signifie que la machine est finie ainsi que son espace de travail : elle ne peut pas croître indéfiniment.

¹ Gunderson 1971 critique le test de Turing. Le test de Turing est un test pour MEC-BEH. En gros, une machine (cachée) passe le test si elle parvient à se faire passer pour un être humain lors d'une "conversation" par l'intermédiaire d'un télétranscripteur.

1.1.2 Philosophie mécaniste : bref historique

La philosophie mécaniste digitale contemporaine est due en grande partie à Descartes et à Hobbes² (voir Rogow 1986, Bernhardt 1989). Descartes voulait distinguer l'homme de l'animal. Il argumente que l'animal, ainsi que le *corps* de l'homme (cerveau compris), est une machine. Il entendait par là un assemblage fini de pièces matérielles déterminant de manière univoque la fonction et le comportement de l'assemblage. Descartes estime que l'âme n'est pas mécaniste. En séparant alors l'âme du corps, et donc l'esprit de la matière, il est à l'origine de la position dualiste diversement appréciée en philosophie de l'esprit. On parle de *dualisme cartésien*.

Voilà trois arguments présentés par Descartes en faveur de sa distinction entre l'homme et l'animal-machine (Notons que la distinction entraîne la négation de MEC-IND) :

1) les animaux ne sont pas doués de raison et ne pratiquent pas le langage.

Cet argument a perdu en force puisque le langage et le raisonnement semblent plus accessibles aux machines d'aujourd'hui que, par exemple, l'émotion qui est communément admise chez certains animaux (voir par exemple Lévy 1987). Descartes reprend ici la position d'Aristote qui fait de l'homme un animal rationnel.

2) Les machines sont des êtres finis. Un être fini ne peut pas concevoir l'infini. Or je sais (dixit Descartes) concevoir l'infini. Donc je ne suis pas une machine.

Cet argument contre MEC-IND met en évidence deux questions fondamentales :

- a) L'homme sait-il concevoir l'infini ?
- b) La machine sait-elle concevoir l'infini ?

La question a) distingue Hobbes de Descartes. Hobbes estime en effet qu'il n'est pas possible de concevoir l'infini.

3) Une machine ne peut effectuer que des tâches particulières, en l'occurrence, les tâches pour lesquelles elle a été construite. Descartes dit en effet :

²On peut dégager des affirmations ou des interrogations mécanistes chez les philosophes (pré et post-Socratiques, pas nécessairement matérialistes) de l'antiquité grecque (cf le *Timée* de Platon, voir aussi Odifreddi 1989), chez des philosophes chinois, par exemple dans Lié-tseu, un moine est admiré pour avoir pu faire passer ses *automates* serviteurs pour des êtres de chair et d'os, chez des philosophes Hindoux, par exemple, dans les *questions au roi Milinda*, le corps *humain* est comparé au char, et l'esprit humain est comparé à l'*assemblage* des pièces du char, à la façon dont Hume (1739) aborde le problème de l'identité avec son bateau. La tentation de construire des artefacts à l'image de l'homme est aussi une composante de quelques mythologies, (par ex. le Golem dans la culture juive, voir par exemple Breton 1990). Il n'est pas exagéré de prétendre que l'idée Mécaniste est apparue partout où des machines se sont développées.

Car, au lieu que la raison est un instrument universel qui peut servir en toutes sortes de rencontres, ces organes ont besoin de quelque particulière disposition pour chaque action particulière; d'où vient qu'il est moralement impossible qu'il y en ait assez de divers en une machine, pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir (Descartes, "1953", page 165).

L'idée d'une machine universelle avait pourtant déjà traversé l'esprit de Raymond Lulle (1302) que Descartes avait étudié. Cette idée réapparaîtra chez Leibnitz, et culminera dans les travaux de Turing, comme cela sera expliqué dans la deuxième partie.

La Mettrie reprendra le mécanisme-animal de Descartes pour l'étendre à l'homme (La Mettrie 1748, voir aussi Gunderson 1971).

Parallèlement à Descartes, Hobbes développe lui aussi l'hypothèse mécaniste (Rogow 1986). On peut situer l'origine de la motivation de Hobbes pour le mécanisme lors de sa découverte de la géométrie. Après avoir été particulièrement impressionné par le fait qu'il ait pu être *convaincu* par une *communication finie* basée sur un raisonnement logique en géométrie, Hobbes conçoit le caractère mécanisable de la pensée. Il pense alors qu'il devrait être possible de réduire la pensée à l'addition et à la soustraction. (voir Webb 1980). Il est proche ainsi de la position *fonctionnaliste* en philosophie de l'esprit : que les additions et les multiplications soient réalisées par un *réseau télégraphique, un système hydraulique, un dispositif électromagnétique, ou même un moulin, une catapulte ou un ordinateur*, pour citer une énumération due à Searle (Searle 1984), la pensée est réduite à des opérations ne dépendant pas nécessairement de l'outillage, et de la matière qui le constitue, utilisé pour les réaliser. La Mettrie, à sa façon, argumente dans le même sens :

Ainsi une Ame de boue, qui découvreroit, comme d'un coup d'oeil, les rapports et les suites d'une infinité d'idées, difficiles à saisir, seroit évidemment préférable à une Ame sote et stupide, qui seroit faite des Elémens les plus précieux. (La Mettrie 1748).

De même Lafitte nous apprend au sujet de Babbage, précurseur de l'informatique au 19ème siècle, sur qui nous reviendrons dans la deuxième partie,

Pour Babbage, toute machine étant un composé d'organes différenciés reliés entre eux d'une manière complexe, ce qu'il importe de fixer, c'est moins la forme même des organes que l'enchaînement des fonctions, qui tient aux liaisons organiques et qui fait le fonctionnement de l'ensemble. (Lafitte 1930)

A la différence de Descartes, Hobbes estime qu'il n'est pas possible que l'homme, être qu'il considère comme étant fini, conçoive l'infini. La motivation, finitiste et indexicale (la *pensée humaine* est mécanisable) de Hobbes est donc opposée au mécanisme animal de Descartes et est, en ce sens, plus proche de la motivation contemporaine pour l'intelligence artificielle. Je reviendrai bientôt sur la relation qui existe entre le mécanisme et le fonctionnalisme.

1.1.3 Qu'est-ce qu'une machine ?

Etant donné les connotations familières du mot machine - locomotives, poteaux électriques, automobiles, ordinateurs, microscopes, lave-vaisselle, machines à coudre, chaudière à mazout, horloge, le mécanisme peut paraître grotesque.

Même si la machine est considérée comme un artefact de construction exclusivement humaine, c'est-à-dire un être artificiel, la machine n'est pas si facile à définir, Lafitte, en 1911, argumente qu'une telle définition est vaine :

Prétendre définir la machine c'est supposer que la science des machines est parvenue, ou peut parvenir quelque jour, à son entière perfection. Outre que c'est, chimériquement, assigner des bornes au développement des formes mécaniques, c'est supposer, d'abord, une connaissance entière et parfaite de tous les caractères de tous les individus mécaniques présents et à venir, puis la perfection d'un instrument de mesure situant, dans une classification définitive, tous ces individus d'après l'ensemble de leurs caractères. Mais c'est encore admettre, implicitement, que l'ensemble de ces individus forme, réellement, une masse fortement tranchée, aux contours nettement saisissables et ne présentant aucun lien avec les autres corps. (voir aussi plus loin 2.3)

De même La Mettrie, dans "L'Homme-machine" écrit :

L'Homme est une Machine si composée, qu'il est impossible de s'en faire d'abord une idée claire, et conséquemment de la définir.

Ce qu'il y a de commun dans la conception de Hobbes et de Descartes est que la machine est un être localement fini. Son comportement global est déterminé par le comportement de ses composantes élémentaires lesquelles sont à chaque instant en nombre fini (c'est l'aspect digital). Le nombre de composants peut cependant croître avec le travail de la machine.

Une philosophie appelée *philosophie mécanique* et qui s'était développée sous l'influence des travaux de Newton, mais aussi de Boyle (voir de Brouckère 1982), était plus matérialiste et déterministe (dans le sens vague contemporain) que finitiste. L'action à distance qui semble exister entre les corps matériels dans la mécanique de Newton, et qui inquiétait (par ailleurs) Hobbes, semblait exclure la composante finitiste (voir Metaxopoulos 1986). De même Searle, que je ne classerais pas parmi les

philosophes mécanistes, argumente que seule une machine, comme l'est (dit-il) le cerveau humain, est capable de penser, mais Searle n'est pas clair sur la nature des identifications considérées. La difficulté de définir les machines est reflétée dans la difficulté de cerner la philosophie mécaniste.

Gandy (1980) a isolé 4 principes qui capturent l'idée intuitive de machine (digitale ou digitalisable en fait) :

1) *Principe de détermination* : une machine peut être décrite par la donnée d'une structure hiérarchique S , ainsi que d'une fonction F telle que l'itération $F(S)$, $F(F(S))$, $F(F(F(S)))$, ... donne l'évolution de la machine.

2) *Principe de la hiérarchie limitée* : la structure S est hiérarchisée en un nombre fini de niveaux.

3) *Principe du réassemblage unique* : la structure S est décomposable en parties limitées et recomposable de façon unique à partir de ces parties et d'un plan (fini lui aussi).

4) *Principe de localité* : la structure hiérarchique de la machine admet une description topologique telle que l'état d'une partie de la machine $F^n(S)$ lors de son évolution ne dépend que de l'état des parties appartenant à son voisinage dans $F^{n-1}(S)$ (pas d' "actions à distance").

Gandy et Shepherdson (1988) donnent une formulation beaucoup plus précise en termes d'ensembles héréditairement finis. Gandy démontre, (et Shepherdson généralise), l'équivalence entre cette définition et la définition conceptuelle de Turing (avec oracle), si bien que la thèse de Church (généralisée) nous permettra, dans la deuxième partie, de nous passer des principes explicites de Gandy. J'ai néanmoins informellement exposé ces principes car ils résument les conceptions de Descartes et de Hobbes, et éclairent les motivations biologiques du mécanisme (digital) ainsi que les doutes qui viennent de la chimie.

1.1.4 Les motivations qui viennent de la biologie

Les motivations qui viennent de la biologie concernent directement le mécanisme indexical puisqu'elles proviennent d'une observation de nous-mêmes, à la différence des motivations qui viennent de la contemplation du travail de l'ingénieur (voir par exemple Lafitte 1930, Vésale 1543, Ambroselli 1987).

En voulant montrer que les animaux sont des machines, Descartes se heurte à un problème qu'il ne parvient pas à résoudre.

1°) Le problème de Descartes

Comment un animal-machine est-il capable de se reproduire ? Comment une machine pourrait-elle construire une machine identique à elle-même ? Elle devrait contenir en elle-même une description complète d'elle-même, et cela semble impossible³. Le fait que la division cellulaire a été comprise dans ses grandes lignes au niveau moléculaire (Cairns, Watson et Crick, (voir Watson (1965, 1989) pour les références détaillées) constitue une motivation par la biologie contemporaine pour la mécanique.

2°) Le problème de Driesch

De même lorsque Driesch réalise les premières expériences d'embryologie, la différenciation cellulaire lui semblera être un phénomène à ce point surprenant qu'il l'utilisera comme argument décisif en faveur du vitalisme, et donc d'une conception contra-mécaniste⁴ de la vie⁵. De même comment expliquer la souplesse dont font preuve les cellules lors du phénomène analogue de régénération cellulaire. Pensons au clonage de la grenouille (voir Watson & Al. 1989 page 835), ou encore au stupéfiant exemple de la planaire, petit vers plat, long d'environ 1 cm, qui habite certaines eaux douces (voir le dessin page suivante) et qui est une sorte de champion de la régénération des tissus parmi les animaux qui possèdent un système nerveux central (Buchsbaum 1938, Buchsbaum et al. 1987). Les phénomènes de régulations génétiques sont éloquentes à l'égard du mécanisme (Jacob et Monod 1961, Thomas 1978, Thomas et Van ham 1974).

Voilà encore ce que disait Diderot dans son entretien avec d'Alembert, confrontant le mécanisme Cartésien et le développement de l'embryon :

Voyez-vous cet oeuf ? c'est avec cela qu'on renverse toutes les écoles de théologie et tous les temples de la terre. Qu'est ce que cet oeuf ? une masse insensible avant que le germe y soit introduit ; et après que le germe y est introduit, qu'est-ce encore ? une masse insensible, car ce germe n'est lui-même qu'un fluide inerte et grossier. Comment cette masse passera-t-elle à une autre organisation, à la sensibilité, à la vie ? par la chaleur. Qui produira la chaleur ? le mouvement. Quels seront les effets successifs du mouvements ? Au lieu de me répondre, asseyez-vous, et suivons-les de l'oeil de moment en moment. D'abord c'est un point qui oscille, un filet qui s'étend et se colore, de la chair qui se forme ; un bec, des bouts d'ailes, des yeux, des pattes qui paraissent ; une matière jaunâtre qui se dévide et produit des intestins ; c'est un animal. Cet animal se meut, s'agite, crie ; j'entends

³ L'argument est encore utilisé de temps en temps, ainsi Cossa en 1955 : "si les machines se trouvaient, par impossible hypothèse, dotées du pouvoir de se reproduire, ces irrégularités, ces malfaçons, au départ minimes et sans conséquences fonctionnelles, subiraient de génération en génération, une amplification telle que la machine cesserait vite de tourner"

⁴ La différence entre le non-mécaniste et le contra-mécaniste est la même qu'entre l'agnostique et l'athée.

⁵ Par réaction, Helmholtz, qui proposera une théorie mécaniste de la perception que nous allons en quelque sorte retrouver, proposera un pacte anti-vitaliste (de Brouckère 1982).

ses cris à travers la coque ; il se couvre de duvet ; il voit. La pesanteur de sa tête, qui oscille, porte sans cesse son bec contre la paroi intérieure de sa prison ; la voilée brisée; il en sort, il marche, il vole, il s'irrite, il fuit, il approche, il se plaint, il souffre, il aime, il désire, il jouit ; il a toutes vos affections ; toutes vos actions il les fait. Prétendez-vous, avec Descartes, que c'est une pure machine imitative ? Mais les petits enfants se moqueront de vous, et les philosophes vous répliqueront que si c'est là une machine vous en êtes une autre⁶.

Le biologiste contemporain peut estimer, que *relativement aux lois de la chimie*, le problème de la reproduction biologique est résolu. La découverte par les biochimistes et les généticiens moléculaire du plan de la cellule, et de la façon dont ce plan est chimiquement représenté, décodé, exécuté au sein de l'organisme constitue une motivation pour l'application du principe de réassemblage unique, le principe de détermination, et le principe de la hiérarchisation limitée (cela apparaît déjà avec la génétique classique, voir Cuny 1969). De même que la découverte plus ancienne de l'importance des échanges de particules avec le milieu ou entre organismes, comme lors de la respiration, de la digestion, de la fécondation, est en faveur de l'application du principe de localité (Van Helmont, Mendel, Lavoisier, Vesale pour citer les plus connus, voir de Brouckère 1982, Ambroselli et Al. 1987, Vesale 1543).

1.1.5 Le doute qui vient de la chimie

Comme dit Watson "la cellule obéit aux lois de la chimie", et les motivations précédentes justifient peut-être une croyance dans un mécanisme indexical ***relatif aux lois de la chimie***. Si celles-ci se révélaient non-mécanisables, le mécanisme se verrait affaibli, peut-être réfuté, certainement relativisé.

Cette suggestion est d'autant plus fondée que les lois de la chimie sont capturées par la mécanique quantique. Celle-ci, en dépit de son nom (*mécanique* est utilisé dans le sens Newtonien), attire les philosophes et les théologiens qui verraient dans les faits décrits (et jusqu'à présent confirmés) de cette théorie une justification empirique de la nature non-mécaniste du monde et/ou de la conscience⁷.

Les arguments anti-mécanistes fondés sur la mécanique quantique sont variés. Examinons-en brièvement quelques-uns :

⁶ Remarquons l'esprit moderne de Diderot qui place l'animal sur le même pied que l'humain, qui refuse ainsi la distinction de Descartes. En général, à l'exception notable de La Mettrie, le mécanisme sera mal reçu. On se rappelle de l'argument de Pascal. Ce genre d' "argument" n'est pas loin de celui que Turing qualifie d'objection de l'autruche et pour laquelle il estime une consolation plus adéquate qu'une réfutation. (Turing 1950)

⁷ Letovski 1987 résume une (trop ?) rare rencontre entre les cognitivistes ouverts aux approches computationnelles de la conscience et les neuro-chercheurs ouverts à l'utilisation de la mécanique quantique pour résoudre le saut subsistant entre le cerveau et l'esprit.

a) Le plus ancien argument : la MQ met en évidence un indéterminisme intrinsèque dans le monde (ou plus précisément concernant les relations entre l'observateur et le monde). Le mécanisme est déterministe. Donc notre relation au monde n'est pas mécaniste.

Ceux qui usent de cet argument sont tenter d'"expliquer" le libre-arbitre au moyen de cet indéterminisme. L'argument a déjà été réfuté par Carnap ou McKay ou Schrödinger. De plus, je vais montrer que le mécanisme *n'est pas* déterministe.

b) Le plus récent argument : la MQ rend possible des matériaux très particuliers, comme par exemple les quasi-cristaux de Penrose et Shechtman (voir Penrose 1989). Penrose suggère, sans vraiment se convaincre lui-même semble-t-il, que le cerveau pourrait être une *sorte* de quasi-cristal. De même Margeneau 1984 et Squires 1990 cherchent à utiliser la MQ pour développer une théorie dualiste et non mécaniste de l'esprit (voir aussi Stapp 1993).

Les arguments suivants méritent d'être examinés plus en détail parce que l'hypothèse mécaniste (indexicale) les éclaire considérablement. A cette fin j'expose le minimum qu'il faut avoir à l'esprit sur la mécanique quantique pour suivre l'argument.

Newton concevait la matière et la lumière comme constituée de particules interagissant les unes avec les autres. Huygens quant à lui réserve cette façon de voir pour la matière exclusivement. Il développe une théorie ondulatoire de la lumière qui rend compte avec succès de nombreux phénomènes lumineux. Einstein mettra en évidence, dans son travail sur l'effet photo-électrique, un aspect corpusculaire de la lumière, sans détrôner pour autant la théorie ondulatoire. Il fonde ainsi la théorie quantique de la lumière. De Broglie étend cet aspect onde-corpuscule de la lumière à la matière. Cela permet de rendre compte du comportement des électrons dans les atomes décrits par Bohr et cela signifie la naissance de la théorie quantique de la matière. Le tableau suivant résume l'évolution du concept de lumière et de matière :

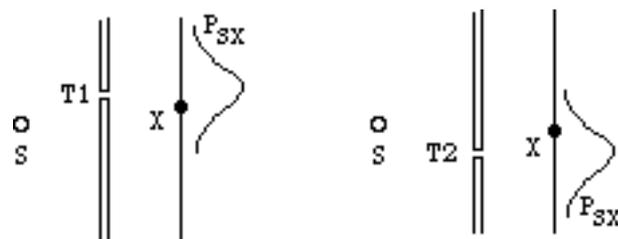
<i>Newton</i>	lumière:	corpuscule
	matière:	corpuscule
<i>Huygens</i>	lumière:	onde
	matière:	corpuscule
<i>Einstein</i>	lumière:	corpuscule (& onde)
<i>de Broglie</i>	matière:	onde (& corpuscule).

Born, à son tour, donne une interprétation probabiliste de l'onde accompagnant la particule. Il associe donc une grandeur qu'il appelle *amplitude de probabilité* à la particule. C'est cette amplitude qui oscille et est responsable de la présence des phénomènes quantiques d'interférence ondulatoire. A présent, lorsqu'on décide d'observer la position de la particule, le résultat sera une position bien précise, résultat prédit par l'amplitude de probabilité de la valeur de l'onde en cette position mise au carré.

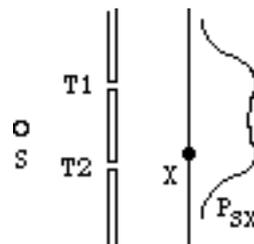
Illustration :

Supposons qu'une particule ait le choix entre deux chemins pour aller d'une source S vers un point X d'un écran, en traversant par exemple une plaque comportant deux trous T1 et T2. Dans le cas classique, la probabilité $P_{SX/T1}$ (resp $P_{SX/T2}$) d'arriver en X en passant par le trou T1 (resp T2) est égale au produit des probabilités d'aller de S à T1 (resp T2) avec la probabilité d'aller de T1 (resp T2) à X :

$$P_{SX/T1} = P_{ST1} * P_{T1X} \quad (\text{resp } P_{SX/T2} = P_{ST2} * P_{T2X}).$$

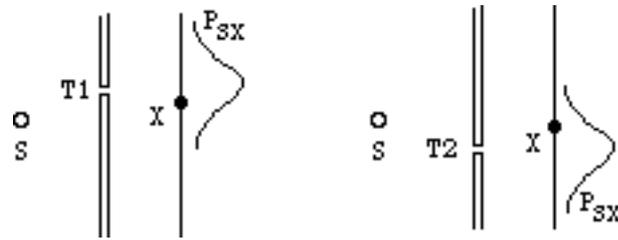


La probabilité d'aller de S en X, quel que soit le trou emprunté est alors égale à la somme de ces deux probabilités: $P_{SX} = P_{ST1} * P_{T1X} + P_{ST2} * P_{T2X}$



Quantiquement, le raisonnement est le même excepté que, tant qu'aucune mesure n'est effectuée, on additionne et on multiplie les *amplitudes* ondulatoires. Pour trouver la probabilité du résultat de la mesure *finale* de la position de la particule sur l'écran, on élève l'amplitude *finale* au carré. Si A_{XY} représente l'amplitude pour aller de X en Y, la probabilité d'aller de X en T1, $P_{SX/T1}$, est égale à $(A_{ST1} * A_{T1X})^2$, et de même

$P_{SX/T2} = (A_{ST2} * A_{T2X})^2$. Avec un seul trou ouvert on retrouve la situation classique⁸,

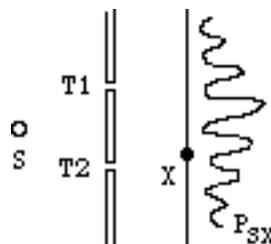


mais avec les deux trous ouverts, l'amplitude finale est égale à la somme des amplitudes correspondant aux alternatives : $A_{SX} = (A_{ST1} * A_{T1X}) + (A_{ST2} * A_{T2X})$, si bien que la probabilité pour aller de S en X : =

$$P_{SX} = ((A_{ST1} * A_{T1X}) + (A_{ST2} * A_{T2X}))^2$$

$$= P_{ST1} * P_{T1X} + P_{ST2} * P_{T2X} + 2(A_{ST1} * A_{T1X}) * (A_{ST2} * A_{T2X})$$

Dans le cas quantique, un terme supplémentaire doit donc être pris en compte. Comme l'amplitude est ondulatoire, c-à-d que $A_{XY} = Ae^{iquelque\ chose}$ (avec le *quelque chose* dépendant, d'une façon générale du temps et de l'espace X, Y), la valeur de la probabilité finale sera oscillante le long de l'écran :



Si on avait mesuré par quel trou passe la particule (par exemple en mesurant l'impulsion de l'écran), on aurait pu additionner directement les amplitudes au carré des chemins alternatifs, et le terme d'interférence aurait été annulé.

La suppression du terme d'interférence est-il dû à la perturbation physique de la mesure ?

Bohr admet déjà dans ses discussions avec Einstein que la perturbation qu'il invoque au sujet de l'observation n'est pas physique au sens usuelle du terme. Einstein le précise avec son paradoxe d'Einstein (voir de Broglie 1957), et cela apparaît clairement avec le paradoxe d'Einstein Podolski Rosen (EPR) en admettant qu'une perturbation physique ordinaire (causale) doit

⁸ en négligeant les effets quantiques dus au bord du trou.

être locale. Le fait est qu'avant la mesure l'amplitude est non nulle dans une vaste région de l'espace. Après la mesure elle est partout nulle sauf à l'endroit où la particule est détectée. Lorsque deux particules interagissent le système des deux particules est décrit en mécanique quantique par *une seule* fonction d'onde, si bien que malgré qu'elles puissent s'éloigner l'une de l'autre, la mesure sur une des deux particules supprime instantanément des possibles termes d'interférence concernant des mesures possibles sur l'autre particule. Cette non-localité, mise en évidence par Einstein et qui a pu être testée expérimentalement grâce aux travaux de Bell (voir annexe 5) ne permet pas d'envoyer de l'information à distance, et ne viole pas le principe de localité de Gandy parce qu'elle exhibe seulement une corrélation statistique. Il n'est pas possible de transmettre de l'information avec deux dés qui seraient corrélés puisqu'il n'y a pas moyen de prédire ou de choisir le résultat du lancement du dé. Localement, le principe de localité de Gandy est respecté bien qu'il faille ajouter localement un oracle (une donnée infinie) aléatoire pure. Dans 1.3 et 3.3 une phénoménologie mécaniste d'un tel oracle est exhibée.

Axiomes de la mécanique quantique : la mécanique quantique décrit deux types d'évolution d'un système physique :

1) le **développement continu et déterminé de l'onde** (décrit par l'équation différentielle de Schrödinger dans le cas non relativiste par exemple). La grandeur oscillante est appelée *amplitude*.

2) la **réduction abrupte de l'onde lors d'une observation**. Le résultat de l'observation peut être prédit avec une probabilité calculable à partir de la forme de l'onde, (l'amplitude de l'onde au carré). C'est le principe de réduction.

Nous pouvons aborder l'argument suivant. Il est à la base d'une approche contra-mécaniste en philosophie de l'esprit :

c) *L'argument classique* : il constitue la théorie de la mesure de von Neumann (1932), reprise par London et Bauer (1939), Wigner (1967) etc. Puisque le comportement de tout ce qui est physique y compris mon corps et mon cerveau, en l'absence d'observation semble être décrit par le développement déterminé et continu de l'onde, c'est que la réduction de l'onde est opérée par ce qui est non physique en moi, c'est-à-dire, selon von Neumann, la conscience (parfois appelée ici *spectateur ultime de la chaîne de von Neumann*, la chaîne étant constituée de l'objet observé, l'oeil, la rétine, le nerf optique, les neurones du cortex visuel, etc.)

Cette théorie soulève des difficultés logiques et physiques considérables. Imaginez une mesure quantique faite par un aveugle qui se contente de

photographier les résultats (sans les connaître, et donc sans prendre conscience de ces résultats, et donc sans réduire l'onde), et d'envoyer ces résultats par la poste à un ami physicien. Avec la théorie de von Neumann, le physicien aveugle est décrit par une superposition ondulatoire d'états incompatibles tant que son ami n'a pas pris conscience du résultat (voir Shimony 1963, ou encore 1989 pour des analyses plus détaillées de cet argument).

Avec l'hypothèse mécaniste, l'observateur-machine ne peut pas être privilégié par rapport à l'objet observé. L'observateur doit donc être décrit, comme l'objet, par l'équation continue et déterministe de l'onde. Il y a ici un plongement du sujet dans l'objet. On reviendra en 3.3 sur la portée philosophique des plongements de ce genre. Everett, en 1957, montre qu'il est encore possible d'assigner des états aux systèmes observés à condition de relativiser la notion d'état. Comme l'aveugle décrit plus haut l'observateur-machine se retrouvera dans une superposition d'états incompatibles. Mais comme le montreront indépendamment Graham et Hartle, chacun des observateurs-machines, multiplié par les interactions, placera dans sa mémoire un résultat bien précis, cohérent pour chacun des observateurs avec la statistique attaché au postulat de réduction de l'onde. Everett inaugure ainsi une des premières formulations de MQ sans réduction de l'onde. Il construit à partir de l'équation d'onde appliquée au système complet [observateur + objet-observé] une phénoménologie de la réduction en utilisant l'hypothèse que l'observateur est une machine et qu'elle obéit dès lors aux lois de la physique.

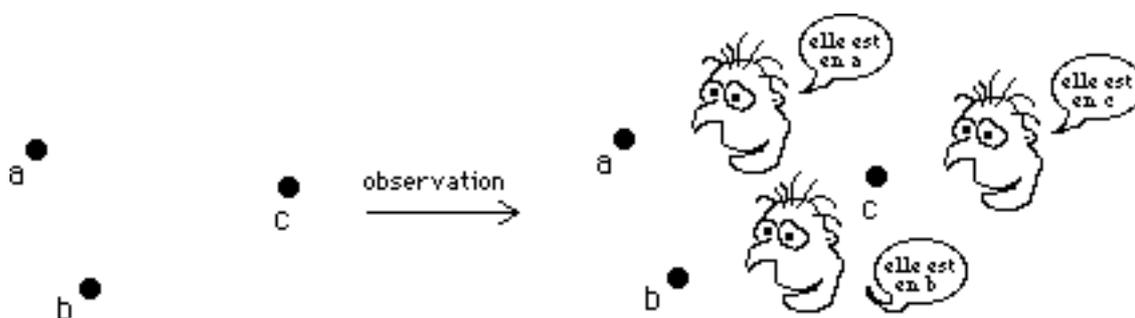
A fin que cela soit plus claire voici une présentation plus schématique des positions respectives de von Neumann et d'Everett :

Une particule, tant qu'elle n'est pas observée, se manifeste comme si elle était en plusieurs endroits à la fois par l'intermédiaire de son amplitude ondulante. Des observations indirectes permettent de confirmer cette prédiction (ou plutôt rétrodiction) de la mécanique quantique, mais si on observe la position de la particule directement, l'observateur positionnera la particule à un endroit précis :



C'est la réduction de l'onde. Comme le cerveau de l'observateur est lui-même un système de particules, von Neumann a conclu que la réduction est opérée par la conscience -explicitement non physique, et dissociée ainsi du cerveau- de l'observateur.

Avec l'hypothèse mécaniste, on est obligé d'admettre que l'observateur a le même statut que la particule non observée. Dans ce cas il n'y a tout simplement pas de réduction: :



L'observateur se multiplie comme la particule observée. Mais chacun des nouveaux observateurs pense localiser la particule en un endroit précis : pour chaque observateur on obtient *une phénoménologie* de la réduction du paquet d'onde.

Everett a montré que le formalisme de la mécanique quantique permet de justifier pourquoi l'observateur ne perçoit pas sa propre multiplication ni la présence de ses "sosies". De plus cette interprétation est nécessaire lorsqu'on désire appliquer la mécanique quantique en cosmologie où les systèmes sont gigantesques et incluent d'office les observateurs.

L'interprétation ou la formulation d'Everett de la mécanique quantique est pour le moins contre-intuitive puisque l'observateur est multiplié, sans qu'il ne puisse d'ailleurs le remarquer directement. Cette interprétation est controversée. Cependant, **toutes** les interprétations de la mécanique quantique sont contre-intuitives et controversées.

Tout argument en faveur de l'interprétation d'Everett, ou plus généralement en faveur des interprétations sans réduction du paquet d'onde, fait de la mécanique quantique une compagne pour la philosophie mécaniste de l'esprit plutôt qu'une plausible concurrente. **L'argument** le plus simple pour défendre Everett est peut-être donné par l'application d'une version *conceptuelle* du rasoir d'Occam : pour le philosophe mécaniste la théorie

équation de Schrödinger

est non seulement plus courte, mais est surtout conceptuellement plus simple que la théorie

équation de Schrödinger + principe de réduction

D'autant plus que le principe de réduction se réfère de façon arbitraire à l'observateur et à ses décisions.

Nous aurons l'occasion de constater (en 1.3, 3.3) que la multiplication du sujet est un phénomène plus intrinsèquement *mécaniste* que *quantique*. Contentons-nous momentanément de retenir que les faits quantiques non seulement ne contredisent pas l'hypothèse mécaniste, mais semble plutôt la confirmer. L'interprétation d'Everett mue le doute qui vient de la chimie en motivation pour le mécanisme.

1.1.6 Niveaux d'explications et indépendance a priori

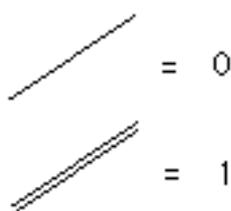
Nous avons vu que la biologie donnait des motivations pour un mécanisme indexical relativement aux lois de la chimie. Afin de clarifier cette notion de dépendance, je présente ici un petit ordinateur⁹ (de type von Neumann) qui va nous introduire à la notion de niveaux, et d'indépendance de niveaux, d'explication. Cela permettra aussi d'établir une relation entre le mécanisme et le fonctionnalisme en philosophie de l'esprit (et de présenter, si c'était nécessaire, un embryon de la logique propositionnelle classique). Ce petit ordinateur sera aussi réutilisé dans la troisième partie pour exposer le paradoxe du graphe filmé.

1°) Un petit ordinateur

L'ordinateur que je présente est un graphe, on y rencontre 4 sortes de sommets : que je désigne et appelle le "&" (*et*), le "V" (*ou*), le "¬" (*non*), et le branchement ou *split* (sommet noire).

Lorsque l'ordinateur est en activité, les arcs sont parcourus par "quelque chose" qui dépendra de la réalisation physique du graphe. Pour l'instant on peut imaginer qu'il s'agit de courant électrique.

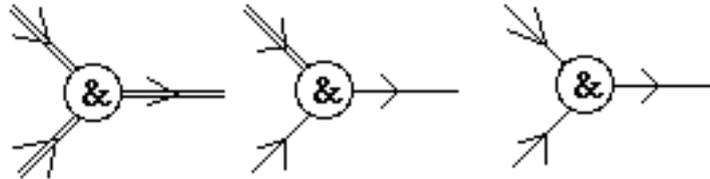
On marque l'arc par 1 ou 0 selon qu'il est actif ou inactif. Ce qui est représenté graphiquement ainsi :



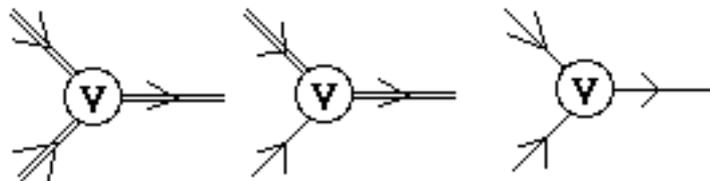
⁹ Je résume un exposé que j'ai tenu à Arlon à l'invitation du professeur Papy (Marchal 1983)

Les sommets obéissent aux règles suivantes :

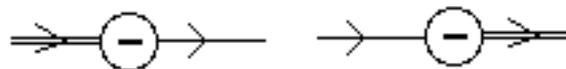
-Règle du *et* :



-Règle du *ou* :



-Règle du *non* :

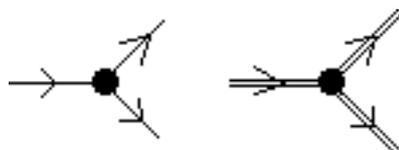


Si l'entrée des sommets est interprétée par des propositions, de telle façon que l'activité de l'arc correspond à vérité de la proposition, on voit que le sommet "&" calcule la table de vérité du "&", de même pour le "V" et le non.

p	q	p&q	p	q	pvq
1	1	1	1	1	1
1	0	0	1	0	1
0	1	0	0	1	1
0	0	0	0	0	0

Remarquons que le "V" est inclusif, il correspond au latin *vel*.

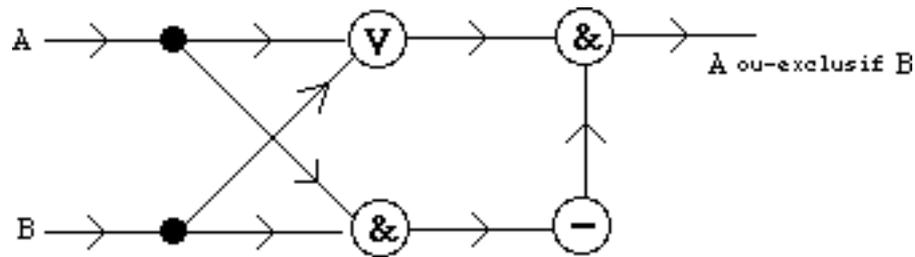
Règle du *split*



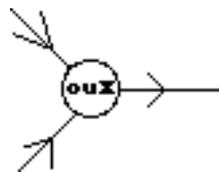
Le split n'est pas un connecteur de la logique classique habituelle¹⁰, il sert ici à dupliquer physiquement l'information.

En combinant de façon appropriée les sommets, on peut construire un graphe capable d'effectuer un calcul.

1°) **Le ou exclusif** (en latin *aut*), en français, c'est le *ou bien ... ou bien*.
A ou-exclusif B est équivalent à $(A \vee B) \& \neg(A \& B)$
 il est donc réalisé par le graphe suivant :



nous disposons ainsi d'un nouveau *module*



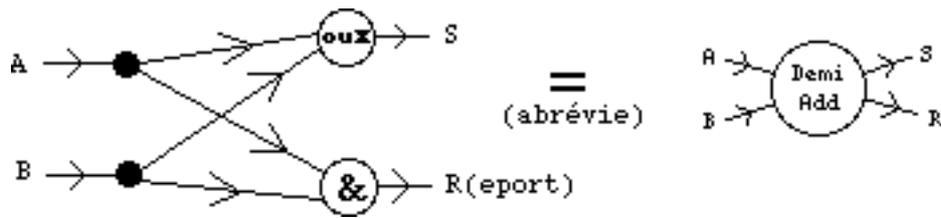
2°) Addition en binaire

En binaire $1 + 1 = 10$. Pour additionner deux nombres, par exemple $111 + 11$, l'algorithme traditionnel ne dépendant pas de la base, on additionnera les premiers chiffres du rang des unités, on dira par exemple : $1 + 1 = 10$, j'écris 0 et je retiens 1. Pourquoi retenir 1 ? Parce qu'il appartient au rang des deuzaines et qu'il faut le reporter sur la somme des deuzaines, j'additionne donc les deuzaines $1+1 = 10$, auquel j'ajoute le report, $10 + 1 = 11$, j'écris 1, et je reporte Le 1 qui appartient au rang des "centaines", je passe au "centaines", $1+0 = 1$, additionné au report 1 cela donne 10 et je l'écris.

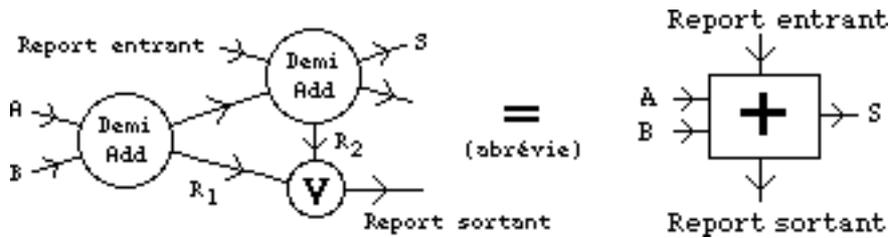
$$\begin{array}{r} 1^1 1^1 1 \\ + 1 1 \\ \hline 1 0 1 1 \end{array}$$

Dans $a + b = rs$, le résultat de l'addition s , que j'écris, est réalisé par le ou-exclusif. Le report r est réalisé par le "&". Cette demi-addition est donc réalisée par le graphe suivant :

¹⁰ on pourrait cependant l'introduire comme tel en affaiblissant les règles (structurales) de la logique usuelle.

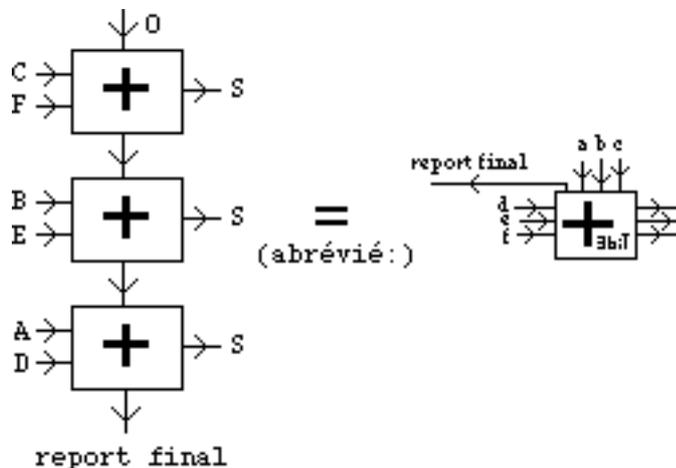


Il faut encore tenir compte du report entrant,



Le report final étant est donné par l'application d'un ou "v" sur les deux reports. On dispose ainsi d'un module capable d'effectuer le travail sur une colonne

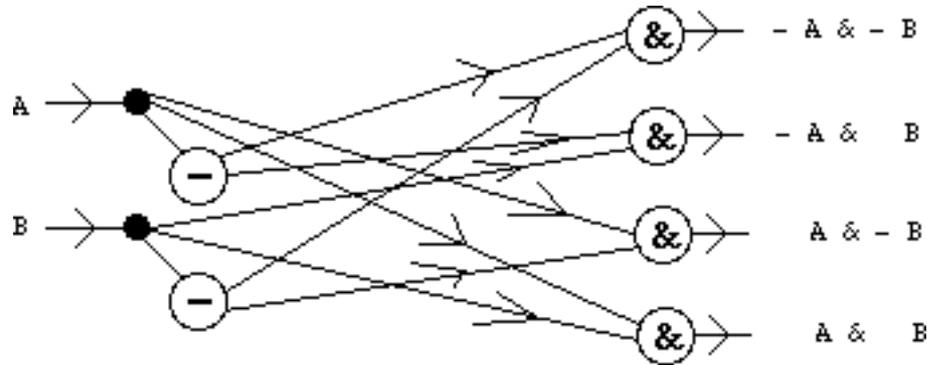
Pour additionner des nombres de n bit (c'est à dire de longueur n), il suffit d'accoler suffisamment de tels additionneurs. Le calcul de ABC + DEF, où A, B, C, D, F représente 1 ou 0 est réalisé par le graphe suivant :



Un tel graphe ne calcule qu'une fonction (et encore sur des nombres très petits, notre exemple "111 + 11" donne un message d'erreur). Existe-t-il un graphe universel, dans le sens qu'il calculerait n'importe quel fonction calculable par une machine ? Répondre "oui" à cette question revient à accepter la thèse de Church, nous y reviendrons. L'idée de von Neumann pour construire un graphe universel nécessite quelques organes supplémentaires :

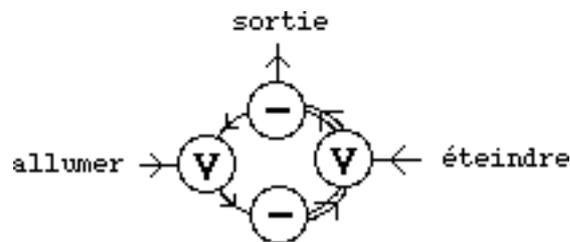
3°) Le décodeur

Le décodeur permet avec un mot de n bit (une suite de 0 et de 1 de longueur n) de sélectionner une sortie parmi 2^n ($= 2$ fois 2 fois 2 ..., n fois¹¹). Par exemple, un décodeur opérant sur des mots de longueur 64 permet la sélection d'une sortie sur 2^{64} , soit 18 446 744 073 709 551 616 (environ 18,5 milliards de milliards) J'illustre le principe en exhibant un décodeur qui opère sur des mots de longueur 2. Il permet de sélectionner une sortie sur 4.



4°) La mémoire

Un bit, un mot de longueur 1, peut être capturé par le graphe suivant (flip-flop) :

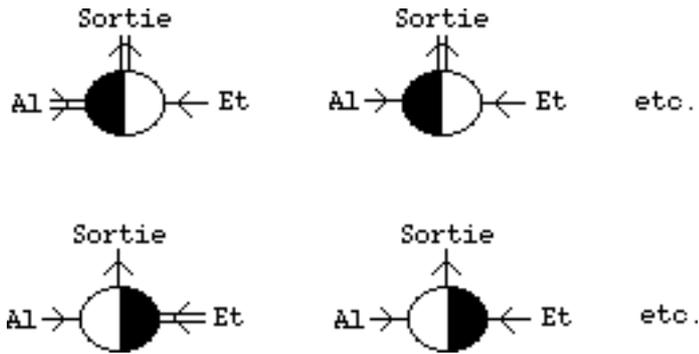


que l'on peut représenter :

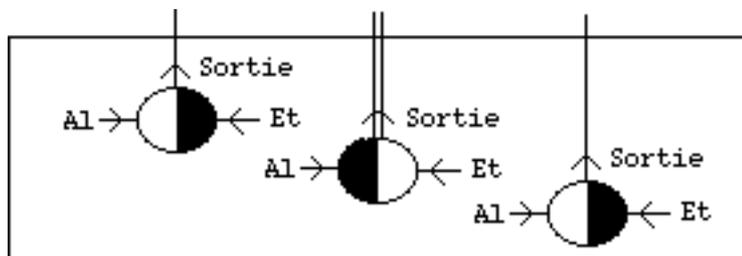


En activant la flèche *allumer* pendant un laps de temps, on passe d'un circuit stable à un autre circuit stable, ensuite allumer n'a plus d'action sur le circuit. De même pour éteindre, on active (pendant un laps de temps) la flèche *éteindre*.

¹¹ On présume la compréhension intuitive des nombres naturels.

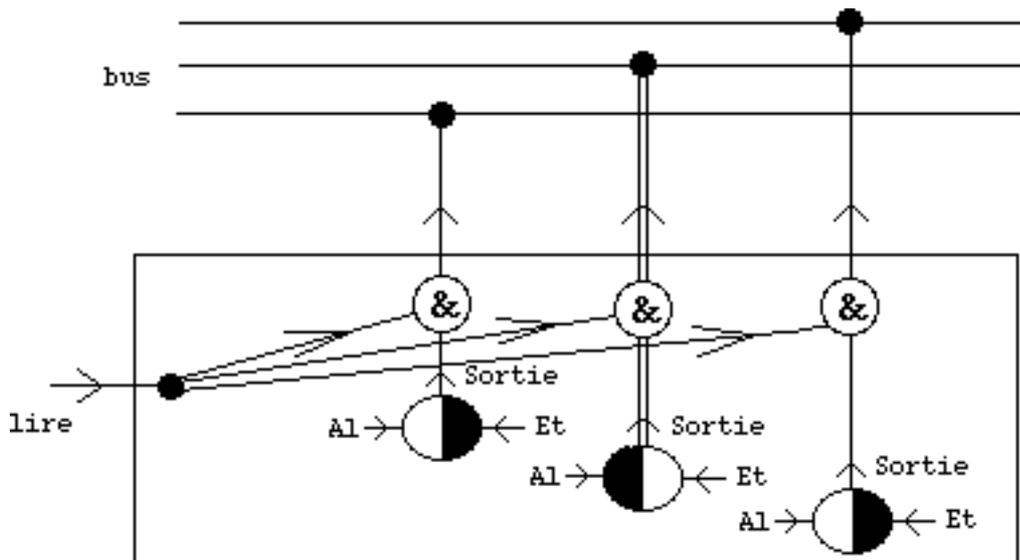


Pour mémoriser un mot de n bit, on utilisera un registre de n bits, c-à-d une collection de n flip-flop. Voici un registre de 3 bits :



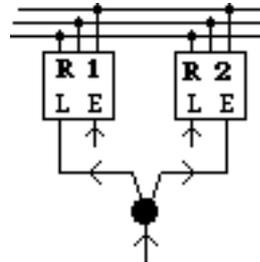
5°) Le bus

Pour utiliser le contenu du registre de 3 bits, on aimerait pouvoir lire son contenu et le transporter là où on en a besoin. 3 flèches du graphe, constituant le bus, suffisent. Si le bus est préalablement désactivé, 3 "&" suffisent pour permettre au moment opportun (quand la flèche *Lire* est activée) de transporter le contenu (déterminé par l'activité et la non-activité des flip-flop du registre) sur le bus :



Un registre est encore plus utile si on parvient à y inscrire de l'information. Une entrée "*écrire*" doit être ajoutée permettant de faire

passer le contenu du bus dans un registre. Je laisse le soin au lecteur de compléter ce module. On dispose à présent de registres dans lesquels on peut lire et écrire. Par exemple, voici un dispositif qui recopie le contenu d'un registre R1 dans un registre R2



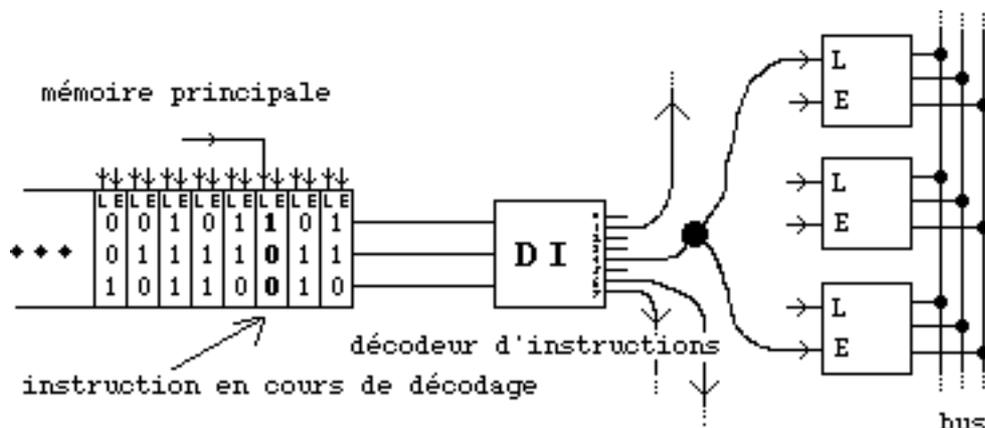
Il permet de réaliser l'instruction :

recopier le contenu du registre R1 dans le registre R2

6°) Le principe du fonctionnement

Le principe de fonctionnement de la machine de von Neuman est le suivant :

Pour satisfaire l'universalité (espérée), la machine doit pouvoir calculer une fonction calculable par une machine quelconque. L'idée de von Neumann est de placer la description de la machine (fonction) à imiter dans la mémoire, et de successivement décoder cette description en une activité qui correspond à celle de la machine qu'on veut imiter. La description est une suite d'instructions codées en binaire. Le décodage d'instructions est réalisé par un décodeur, appelé décodeur d'instructions :



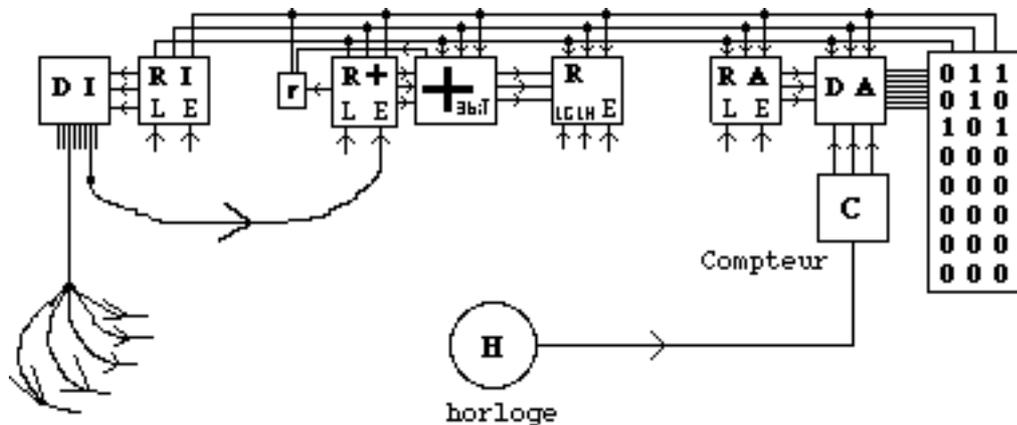
Exemple : calcul de la factorielle de n ($= n*(n-1)*...*1$)

- 1) mettre n dans le registre R1
- 2) recopier le contenu du registre R1 dans le registre R2

- 3) soustraire 1 du contenu de R1 (le résultat reste en R1)
- 4) si le contenu de R1 = 0, sortir le contenu de R2 (sortir = stop)
- 5) multiplier le contenu de R1 dans R2, et le laisser dans R2
- 6) aller en 3.

Le fait que le calcul de toutes les fonctions en principe calculables peut être décomposé en une suite finie de calculs de fonctions élémentaires (comme le non, le &, etc.) est remarquable et peu évident. C'est ce fait qu'énonce la thèse de Church, et les arguments pour cette thèse seront abordés ultérieurement.

Voici le graphe dans son entièreté, où je n'ai cependant pas placé toutes les sorties du décodeur d'instructions afin de rendre le schéma plus lisible :



2°) Niveaux d'indépendances d'explications

Considérons la machine possédant le code de la factorielle fixé une fois pour toute dans sa mémoire centrale. Ainsi, si on lui donne 5 (en binaire ici) elle répond 120 :



Considérons deux observateurs, n'ayant a priori aucune connaissance de la machine, et essayant de construire une théorie de cette entité particulière. Le premier observateur regarde l'entité d'assez haut et fini par *inférer* qu'elle calcule la factorielle. Sa théorie T1 est donc (une description du calcul de) la factorielle. Le second observateur regarde l'entité d'assez près et *infère* que le graphe est constitué des "&", des "v", et des "non". Sa théorie T2 est constituée de la description intégrale du graphe (comprenant le code de la factorielle dans la mémoire centrale). Nous, pour avoir construit cette machine, savons que les deux théories sont correctes et compatibles bien que

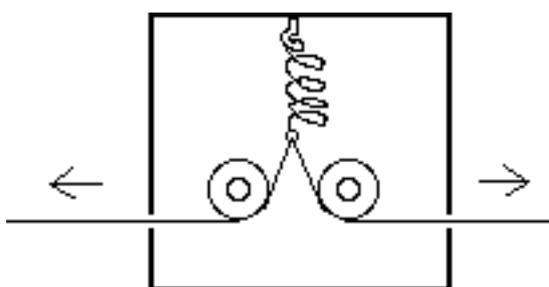
très différentes. Informellement T1 est de plus haut niveau que T2. Par construction nous savons aussi que T1 est indépendante de T2. Choisir entre T1 et T2 revient à choisir un niveau de description. Remarquons que *nous ne savons pas*, de la même façon, que les lois de la biologie sont indépendantes des lois de la chimie. Il est néanmoins possible d'inférer empiriquement l'indépendance de quelques niveaux.

L'indépendance de T1 par rapport à T2 apparaît peut-être plus clairement si on réalise que T1 ne dépend pas de la réalisation physique (l'incarnation) de T2. Ce qu'on va regarder à présent :

3°) Incarnation du graphe booléen

Seul l'implémentation, où l'incarnation de la machine lui permet de se manifester concrètement relativement à un environnement concret. Turing avait déjà remarqué la qualité pédagogique de la machine de Babbage. Purement mécanique -à la façon d'une horloge- elle permet de démystifier le rôle de l'électricité en informatique et en philosophie mécaniste de l'esprit. Comme la théorie T1 est indépendante de la théorie T2, cette dernière peut être remplacée par une autre pourvu qu'on change adéquatement la représentation de T1. Les organes de l'ordinateur décrits plus haut peuvent ainsi être réalisés électriquement ou électroniquement ou quantiquement, etc., sans que cela ne change rien à la fonctionnalité des théories de plus haut niveau.

Dewdney présente par exemple, une incarnation en terme de cordes, poulies et ressorts. A titre illustratif voici comment est réalisé le *non* :



On se débrouille pour que les cordes soient toujours suffisamment tendues. L'équivalent de l'intensité électrique est l'augmentation de la tension dans une corde *au-dessus d'un certain seuil*. On constate sur la figure que si on tire à gauche, la tension diminue à droite, et si on relâche à gauche, la tension augmente à droite. Ce système de poulie réalise donc la règle du *non*. Dewdney explique comment réaliser de façon similaire le *&*, le *∨*, et, par combinaison de ces systèmes élémentaires, les divers organes d'un ordinateur.

On peut de la même façon réaliser des ordinateurs hydrauliques, pneumatiques, etc.

Parmi les ordinateurs non-standards, mentionnons les circuits ingénieux de Fredkin et Toffoli (1982), ainsi que leurs incarnations utilisant une table de billard, des boules de billard et des plaques de diverses formes. De tels systèmes constituent des machines dont la dynamique est en principe réversible et conservative¹².

Une incarnation fantaisiste et souvent utilisée en philosophie de l'esprit : la population actuelle de la planète, avec un service postal convenable. Un réalisateur du *non*, par exemple, est une personne qui, lorsqu'elle reçoit une lettre sur laquelle est dessiné un grand UN (resp. un grand ZERO) envoie à un ou plusieurs destinataires fixés une lettre avec un grand ZERO (resp. un grand UN), etc.

Les systèmes de régulation génétique sont capables d'imiter les ordinateurs. Thomas a notamment programmé une boucle biologique par mutation d'un chromosome de bactériophage. Celui-ci itère successivement une insertion et une sortie du chromosome viral dans le chromosome de la bactérie hôte (voir aussi Richelle 1974).

Remarques :

1) Machine finie et machine infinie

On a remarqué, je suppose, la très petite capacité de mémoire de l'ordinateur présenté plus haut. Le dessin serait devenu inutilement illisible si j'avais décrit une machine possédant un bus plus raisonnable de 8 bits, avec sa capacité de décoder $2^8 = 256$ instructions. (que dire si j'avais utilisé un bus de 64 bits !). Mais même avec un bus de 64 bits, la quantité d'instructions est finie, ainsi que la mémoire accessible. Une telle machine comporte un nombre fini d'états accessibles. Si elle ne devait pas s'arrêter, elle finirait nécessairement par retomber sur un état par lequel elle est déjà passée. En particulier, il ne lui serait pas possible de calculer un développement décimal non périodique (comme celui de la racine de deux¹³). Les machines qui vont nous intéresser seront supposées disposé d'un stock de mémoire aussi grand qu'elles le désirent, on adjoint donc les circuits nécessaires à la gestion d'une mémoire virtuelle, sur disque, etc. Il n'est pas interdit à la machine d'écrire sur les murs! J'appellerai et justifierai l'appellation "environnement localement universelle" pour une machine qui *serait* universelle si elle avait assez de mémoire.

¹² De telles systèmes ont été utilisés par Drescher pour construire une version computationnelle de la mécanique quantique dans la formulation d'Everett. (Drescher 1991)

¹³ En fait, Minski montre qu'une machine finie n'est même pas capable d'être programmée pour multiplier deux nombres naturels quelconques. Il montre aussi qu'elle est cependant capable d'être programmée pour additionner deux naturels quelconques (voir Minski 1967).

2) Les principes de Gandy sont automatiquement vérifiés sur ce petit ordinateur. Il est facile de se convaincre qu'ils restent vérifiés lorsqu'on combine un nombre fini de telles machines pour obtenir un réseau. Si on fait abstraction du temps de calcul et de l'espace mémoire utilisé, (ce qui ne nous concernera seulement d'une façon relative puisque nous allons étudier le comportement des machines au voisinage de l'infini), un tel réseau de machines peut être imité parfaitement par une seule machine¹⁴.

1.1.7 Behaviorisme, fonctionnalisme et mécanisme

1°) Behaviorisme

Le behaviorisme est la doctrine qui réduit directement la pensée au comportement, et réduit ainsi la psychologie à l'étude du comportement.

Des behavioristes comme Watson (psychologue américain le fondateur de la doctrine 1878-1958), ou comme Skinner, dénie explicitement l'existence des expériences subjectives. Ils sont à l'origine du développement considérable de la psychologie expérimentale mais leur influence fut dévastatrice en philosophie de l'esprit, tout discours sur le mental étant considéré comme non scientifique. Beaucoup parmi les psychologues expérimentaux ont adopté un behaviorisme méthodologique : ils se limitaient à l'étude du comportement sans souscrire au point de vue philosophique de Watson ou Skinner.

Behaviorisme et extensionnalité

Le point de vue behavioriste sur une entité (machine) consiste à ne regarder que le comportement de cette entité (machine). Ce comportement est défini pour une machine par ses entrées-sorties. L'ensemble des entrées-sorties constituent ce qu'on appelle l'extension de la machine. On peut éventuellement étendre l'usage de ce vocabulaire pour une entité quelconque pour autant qu'on sache définir de façon pas trop ambiguë le *bord* de cette entité, c'est-à-dire une notion d'entrée et de sortie.

Faiblesse du behaviorisme

Comme doctrine philosophique elle n'est pas capable de distinguer la différence entre la doctrine et le behaviorisme méthodologique.

¹⁴On peut faire un réseau neuronal, mais on peut aussi simuler le réseau neuronal avec la petite machine de Von Neumann. (Pour un neurone Booléen voir C. de Caestecker 1984), on peut encore simuler une collection finie de telle machines, simuler une collection infinie de telle machines etc. Le parallélisme est émule par le séquentiel au "temps réel" près. Ceci sera illustré par le dovetelleur en 3.3.

Le behaviorisme est facilement mis en difficulté avec l'indexical¹⁵:

Exemple 1 : une entreprise *E & Inc.* qui occupe cent employés. Supposons qu'on remplace un employé de *E & Inc.* par un nouveau venu avec les mêmes compétences. Pour un pur fonctionnaliste rien n'a changé dans l'univers, mais si on interroge l'employé *x*, (le point de vue intérieur de *x*), il y a eu un changement, il a perdu son emploi.

Exemple 2 : un behavioriste-positiviste ne peut pas prendre en considération l'univers. En effet celui-ci n'a, par définition ni entrées, ni sorties si bien que du point de vue purement extensionnel l'univers est équivalent¹⁶ à la fonction vide. Cela peut ne pas être faux mais cela ne peut avoir de sens que si une justification phénoménologique de la plénitude accompagne la réduction au vide. Notons que *Univers = notre Univers*, le problème provient une fois de plus du caractère indexical. Le behaviorisme pur et dur a pratiquement disparu tant cette position est extravagante. Selon Ullmo (1969), les premiers behavioristes s'interdisaient même l'usage du microscope et du télescope.

Exemple 3. le behaviorisme met en évidence la ressemblance entre le *rêveur* et l'*univers* : vu l'absence de comportement de l'un comme de l'autre. Le *rêveur* comme l'*univers* se réduit (extensionnellement) au vide.

2°) Le fonctionnalisme

Le fonctionnalisme est un progrès par rapport au behaviorisme. La notion de comportement (ensemble d'entrées-sorties) reste la notion fondamentale mais n'est plus restreinte à l'enveloppe extérieure d'un sujet-cobaye (au bord forcément arbitraire). Le cerveau (par exemple) n'est plus regardé comme une boîte noire. Il est considéré comme un ensemble organisé d'entités connectées entre elles et agissant les unes sur les autres en obéissant au principe de causalité (ou plus abstraitement avec le mécanisme explicitement digital : à un principe de détermination computationnelle du genre de celui de Gandy. On peut donc le considérer comme une sorte de sous-behaviorisme. La nature matérielle des constituants de ces entités ne joue pas d'autre rôle que de permettre aux fonctions (que l'on peut décrire en terme d'intention, de croyance) de se manifester relativement à un environnement. Il laisse une place à la possibilité d'expériences subjectives. Cela ne signifie pas que le fonctionnalisme explique ces expériences, mais il en tient compte en adoptant un principe d'identité c'est-à-dire en associant les états mentaux à un état fonctionnel du cerveau. A une croyance par exemple, correspond une disposition à agir de telles ou de telles façons dans

¹⁵ De telles analyses existent dans la littérature. Voir Madell 1988 par exemple dans son chapitre "indexical". La critique de Madell est une critique du fonctionnalisme et du mécanisme car Madell utilise l'hypothèse matérialiste. La troisième partie du travail permettra de dissocier davantage le mécanisme et les interprétations matérialistes du fonctionnalisme grâce à l'abandon de l'hypothèse matérialiste.

¹⁶ Extensionnellement.

telles ou telles circonstances¹⁷. Cette disposition est incarnée par la réalisation physique du cerveau (ou du corps ou de l'univers).

Selon qu'on associe l'état mental à l'état dispositionnel (= état computationnel avec MEC-DIG) ou qu'on associe l'état mental avec la réalisation concrète (matérielle, en terme de *token*) correspondant à l'état dispositionnel on dit qu'on use d'un principe faible d'identité ou d'un principe fort. En présence du principe fort, on peut affirmer que c'est l'activité du cerveau (corps, univers, ...) qui véhicule l'état mental (la conscience privée). On dira aussi dans ce cas que la conscience supervène sur l'activité physique du cerveau.

Cette forme de fonctionnalisme, avec principe fort d'identité, sera montrée difficilement compatible avec le mécanisme digital.

Le mécanisme possède une relation privilégiée avec une forme de fonctionnalisme relativisé à un niveau :

FONC(n) :

*Si on isole dans un système A un sous-système B et que l'on remplace B dans A par un sous-système B' **extensionnellement équivalent** à B relativement à un niveau n, alors FONC(n) signifie qu'on obtient un système A' **fonctionnellement équivalent** à A relativement au niveau n.*

Le mot fonction est bien sûr utilisé dans un sens intensionnel plus large (et beaucoup plus difficile à capturer formellement) que le sens extensionnel.

Le mécanisme (digital) peut alors s'énoncer sous la forme : il existe un niveau (finiment descriptible) où le fonctionnalisme est correct (où une description fonctionnelle est correcte)

$$\text{MDI} \Leftrightarrow \exists n \text{ FONC}(n)$$

Pour le mécanisme digital, la formule $\exists n \text{ FONC}(n)$ peut être dérivée à partir des principes de localité, réassemblages uniques, et détermination. Pour une machine digitale construite explicitement, comme le graphe booléen, le niveau est déterminable par le rang de la machine (du système).

Dire que le fonctionnalisme est correcte est malheureusement encore ambigu. On peut lever l'ambiguïté en tenant compte du fait que la doctrine

¹⁷ Putnam, qui fut partisan et initiateur du fonctionnalisme (Putnam 1960), devint plus tard un de ses détracteurs (voir par exemple Putnam 1988). Putnam reste fidèle à l'idée selon laquelle la fonction de la machine ne dépend pas de la nature matérielle de sa constitution. Ce que Putnam abandonne c'est l'identification directe de l'état mental avec le type d'état fonctionnel d'une machine (ordinateur), c-à-d la thèse (faible) de l'identité. Il en vint cependant à critiquer pour cette raison le mécanisme. Pour une analyse critique de cette critique voir Marchal 1990. Je vais montrer dans 3.3 que le mécanisme est incompatible avec les thèses d'identité.

fonctionnaliste est indexicale par nature, en tant que principe de base de la psychologie cognitive computationnelle, elle s'applique à l'être humain.

Réécrivons alors des versions indexicales des trois formes de mécanisme abordées au début de la section

MEC-BEH-ind Des machines peuvent avoir mon comportement ;

MEC-FORT-ind Des machines peuvent avoir mon comportement et être une personne (consciente, vivantes, etc.) ;

MEC-IND-ind Des machines peuvent être moi.

Notons que MEC-IND-ind et MEC-IND sont des propositions *classiquement* équivalentes : si je suis une machine alors il y a une machine qui forcément peut être moi. Toutefois on aura l'occasion de mettre en doute ou de nuancer cette équivalence avec des logiques plus faibles (plus exactement avec l'interprétation intensionnelle (modale) de telles logiques).

Ces trois formes de mécanisme admettent alors une description usant du fonctionnalisme relatif. Cela revient à donner trois sens à l'expression "*le fonctionnalisme est correcte*" :

***MEC-BEH-ind* il existe un niveau tel que des machines, obtenues par substitutions extensionnellement équivalentes à ce niveau peuvent avoir mon comportement.**

***MEC-FORT-ind* il existe un niveau tel que des machines, obtenues par substitutions extensionnellement équivalentes à ce niveau peuvent avoir mon comportement et être une personne (conscientes, vivantes, etc.)**

***MEC-IND-ind* il existe un niveau tel que des machines, obtenues par substitutions extensionnellement équivalentes à ce niveau peuvent être moi.**

"Avoir mon comportement" signifie aussi être capable de se faire passer pour moi. "Être moi" signifie ici que je peux survivre personnellement à la substitution extensionnelle. Cela sera détaillé en 1.3.

On peut opposer de prime abord un mécanisme indexical (éventuellement digital) *de très bas niveau*, où pour survivre il faut dupliquer tout l'univers, et un mécanisme indexical (digital) *de haut niveau* : lorsque le niveau est déterminé, par exemple au moyen d'une description digitale approximative de la composition chimique (sous forme des concentrations de molécules et d'ions, pas sous la forme de la description de l'état quantique de ces molécules) et électrique du cerveau.

Le fonctionnalisme contemporain est un corps vague de doctrine. Le caractère vague provient de l'usage ambigu du mot fonction. Le mécanisme (digital et indexical) est plus précis en adoptant un fonctionnalisme, dans un sens extensionnel clair, à un *certain* niveau. Quel niveau ? Ce sera l'objet du présent travail de montrer que l'indexicalité et la digitalité entraîne le caractère nécessairement indéterminé de ce niveau. En particulier le "∃" de MDI apparaîtra comme nécessairement non constructif.

1.1.8 Bibliographie locale

AMBROSELLI C., FAGOT-LARGEAULT A., SINDING C., 1987, Avant-propos à *La fabrique du corps humain*, d'André Vésale, Editions Actes Sud, Inserm.

ANDERSON A. R. (ed.), 1964, *Minds and Machine*, Prentice Hall inc. New Jersey. Trad. Française : Pensée et machine, Editions du Champ Vallon, 1983.

ARSAC J., 1987, *Les machines à penser*, Editions du Seuil, Paris.

BARWISE J., KEISLER H. J. and KUNEN K. (eds.), 1980, *The Kleene Symposium*, North Holland Pub. Company.

BERNHARDT J., 1989, *Hobbes*, Presses Universitaires de France.

BODEN M. *Minds and Mechanisms: Philosophical Psychology and Computational Models*. The Harvester Press, 1981.

BORN R., (ed), 1987, *Artificial Intelligence : the Case Against*, Croom Helm, London & Sidney.

BENZER S., 1962, *The fine Structure of Gene*, Sci. Am. Janvier.

BENZER S., 1965, *The Elementary Units of Heredity*, paru la première fois en 1957 dans *The Chemical Basis of Heredity* W.D. McElroy et B. Glass (eds), Johns Hopkins Press, Baltimore, Maryland. Repris dans Taylor 1965.

BUCHSBAUM R., 1938, *Animals Without Backbones*, second ed., Penguin Books Ltd, Harmondsworth, Middlesex, England, 1948, (first published in USA, 1938).

BUCHSBAUM R., BUCHSBAUM M., PEARSE J., PEARSE V., 1987, *Animals Without Backbones*, third ed., The University of Chicago Press, Chicago & London.

BRETON P., 1987, *Une histoire de l'informatique*, Seuil 1990.

COLLINS A. W., 1987, *The Nature of Mental Things*, University of Notre Dame Press, Indiana.

COSSA P., 1955, *La cybernétique "du cerveau humain aux cerveaux artificiels"*, Masson & Cie, Paris.

- CUNY H., 1969, *Thomas-Hunt Morgan et la génétique*, Editions Seghers, Paris.
- DE BROGLIE, L., 1957, *La théorie de la mesure en mécanique ondulatoire*, Gauthier Villars, Paris.
- DE BROUCKERE L., 1982, *L'évolution de la pensée scientifique*, Culture Laïque, édité par la fédération des amis de la morale laïque, Maison de la laïcité, Bruxelles.
- DECAESTECKER C., 1983, *Graphes Booliens et Réseaux Neuronaux*, mémoire de license en sciences mathématiques, Université Libre de Bruxelles, année académique 1983-1984.
- DENNETT D. C. and HOFSTADTER D., 1981, (composed and arranged by), *Mind's I*, Basic Books, Inc., Publishers, New-York. (paru en français sous le titre *Vue de l'esprit*, interEditions, Paris, 1987).
- DESCARTES R. *Oeuvres et Lettres*. Bibliothèque de la Pléiade. Gallimard, 1953.
- DEWDNEY A. K., 1990, *The Magic Machine. A Handbook of Computer Sorcery*, W. H. Freeman and Company, New-York.
- DEWITT B. S. and GRAHAM N. (Eds), 1973, *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton Series in Physics NJ.
- DIDEROT D., *Oeuvres*, Bibliothèque de la Pléiade. Gallimard, 1951.
- DRESCHER G. L., 1991, *Demistifying Quantum Mechanics: A Simple Universe with Quantum Uncertainty*, Complex Systems 5, pp. 207-237.
- EVERETT, III, H., 1957, "Relative State" Formulation of Quantum Mechanics, Review of Modern Physics, Vol. 9, N° 3, pp. 454-462. (also in DeWitt and Graham, 1973).
- FREDKIN E. & TOFFOLI T., 1982, *Conservative Logic*, International. Journal of Theoretical. Physics. 21, pp. 219-253.
- GANDY R., 1980, *Church's Thesis and Principles for Mechanisms*, in BARWISE J., KEISLER H.J. and KUNEN K. (eds), 1980.
- GUNDERSON K., 1971, *Mentality and Machines* (2ed) Croom Helm, London & Sidney, 1985.
- GULYAS B. (ed), 1987, *The Brain-Mind Problem. Philosophical and Neurophysiological Approaches*, Leuven University Press.
- GRAHAM N., 1973, *The Measurement of Relative Frequency*, in DeWitt and Graham.
- HARTLE J. B., 1968, *Quantum Mechanics of Individual Systems*, American Journal of Physics, Vol 36, N° 8.
- HERKEN R. (ed), 1988, *The Universal Turing Machine A Half-Century Survey*, Oxford University Press.

- HUME D., 1739, *A Treatise of Human Nature*, Fontana/Collins eds, 1987.
- JACOB F. et MONOD J., 1965, *Genetics Regulatory Mechanisms in the Synthesis of Proteins*, Journal of Molecular Biology, 3, 318, 1961. Repris dans Taylor.
- LAFITTE J., 1932, *Réflexions sur la science des machines*, Vrin, 1972, Paris.
- LA METTRIE, J. O. de, 1748, *L'homme machine*, Leyde. Repris dans *Oeuvres philosophiques 1*, 1751. Réédité par Fayard 1987.
- LETOVSKY S., 1987, *Ecclesiastes, A report from the Battlefields of the Mind-Body Problem*, AI Magazine, vol. 8, n° 3, pp. 63-69.
- LEVY P., 1987, *La Machine Univers*, Editions La Découverte, Paris.
- LIE-TSEU., 1980, *Le Vrai Classique du Vide Parfait*, dans Philosophes Taoïstes Bibliothèque de la Pléiade. Gallimard.
- LONDON F. and BAUER E., 1939, *La théorie de l'observation en mécanique quantique*, Hermann & Cie, Editeurs, Paris.
- LULLE R., 1308, *L'Art bref*, Les éditions du Cerf, 1991, Paris.
- MADELL G., 1988, *Mind & Materialism*, Edinburgh University Press.
- MARCHAL B., 1983, *L'elaboratore e'un grafo, L'insegnamento della matematica e delle scienze integrate*, Vol. 6, n° 2, pp. 43-58.
- MARCHAL B., 1988, *Informatique théorique et philosophie de l'esprit*, Actes du 3ème colloque international de l'ARC, Toulouse.
- MARCHAL B., 1990, *Des fondements théorique pour l'intelligence artificielle et la philosophie de l'esprit*, Revue Internationale de Philosophie, 1, n° 172, pp 104-117.
- MARGENEAU H., 1984, *The Miracle of Existence*, Shambhala, New Sciences Library, Boston & London.
- METAXOPOULOS E., 1986, *Métaphysique mécaniste-matérialiste et empirisme scientifique chez Hobbes : étude d'une explication astronomique*, dans *Epistémologie et matérialisme*, Méridiens Klincksieck, Paris.
- MINSKI M., 1967, *Computation: Finite and Infinite Machines*, Prentice Hall.
- ODIFREDDI P., 1989, *Classical Recursion Theory*, North-Holland, Amsterdam.
- PENROSE, R., 1989, *The Emperor's New Mind*, Oxford University Press, Oxford.
- PUTNAM H., 1988, *Representation and Reality*, A Bradford Book, The MIT Press, Cambridge.

RICHELLE J., 1974, Contribution à l'étude théorique de certains aspects de la régulation de l'immunité du bactériophage tempéré λ , mémoire de license en sciences chimiques, Université Libre de Bruxelles.

ROGOW A. A., 1986, Thomas Hobbes, Radical in the service of Reaction, W.W. Norton & Company, NY.

SCHRÖDINGER E., 1951, Science and Humanism, Cambridge University Press.(repris dans Schrödinger 1992)

SCHRÖDINGER E., 1992, Physique quantique et représentation du monde, Seuil.

SEARLE J. R., 1984, Minds, Brains and Sciences. British Broadcasting Corporation. Trad. Franç. chez Hermann, 1985.

SEARLE J. R., 1980, Minds, Brains and Programs dans The Behavioral and Brain Sciences, vol 3, Cambridge University Press. (repris dans DENNETT D.C. and HOFSTADTER D.1981).

SHEPHERDSON J. C., 1988, Mechanisms for Computing Over Arbitrary Structures, in HERKEN 1988.

SQUIRES E., 1990, Conscious Mind in the Physical World, Adam Hilger, Bristol.

SHIMONY A., 1963, Role of the Observer in Quantum Theory, Am. J. of Physics 31, 6, pp. 755-773

SHIMONY A., 1989, Conceptual Foundations of Quantum Mechanics, in The New Physics, Paul Davies (Ed), Cambridge University Press.

STAPP H., P., 1993, Mind, Matter and Quantum Mechanics, Springer-Verlag, Berlin.

TAYLOR G. R., 1963, Histoire Illustrée de la Biologie. (texte français de C. Vendrely), Hachette, 1965. (ed. original : Thames & Hudson, Londres)

TAYLOR J. H., 1965, Selected Papers on Molecular Genetics. Academic Press, New-York & London.

THOMAS R. et VAN HAM P., 1974, Analyse formelle de circuits de régulation génétique : le contrôle de l'immunité chez les bactériophages lambdaïdes, Biochimie, 56, n° 11-12, p. 1529.

THOMAS R., 1978, Logical Analysis of Systems comprising Feedback Loops, ULB, preprint.

TURING A., 1950, Computing Machinery and Intelligence, Mind, Vol. LIX, N° 236. Aussi dans Anderson 1964.

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

VESALE A., 1543, Andrae Vesalii De humani corporis fabrica libri septem. Bâle, J. Oporinus, La fabrique du Corps Humain. Actes Sud et Inserm, 1987.

VON NEUMANN J. 1932, *Mathematische Grundlagen der Quantenmechanik*, Berlin. Trad. Angl. : *Mathematical Foundations of Quantum mechanics*, Princeton University Press 1939, Trad. Franç. A. Broca, librairie Félix Alcan, Paris, 1946, aussi : Gabay, Sceaux 1988.

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

WATSON J.D., HOPKINS N.H., ROBERTS J.W., STEITZ J. A. and WEINER A.L., *Molecular Biology of the Gene*, fourth edition, (1^{er} ed. 1965, 1970, 1976, 1987), The Benjamin/Cummings Publishing Company, Inc. Texte Français élaboré par H. Costinesco, InterEditions, Paris, 1989.

WEBB J. C., 1980, *Mechanism, Mentalism & Metamathematics*. An Essay on Finitism. D. Reidel Pub. Company.

WEIZENBAUM J., 1976, *Computer Power and Human Reason*, W. H. Freeman and Company, San Francisco and London. (traduction française aux Editions d'informatique, 1981).

WIGNER E., 1967, *Symmetries and Reflections*, Indiana University Press, Bloomington.