

Jean Guy Meunier

Meunier, J. G. "La machine humaine et l'information", in *Informatique en perspective*, K. Fall et G. Vignault, (eds) PUQ, 1989 p 121 - 141

La machine humaine et l'information.

1. Introduction

Un des thèmes intégrateurs contemporains pour penser tant le social que l'individuel est celui d'information. Diverses théories portant sur des institutions sociales ou économiques introduisent des catégories appartenant classiquement aux discours sur l'information. Il est, par exemple habituel aujourd'hui de décrire une entreprise en termes de gérance, de circulation et de contrôle d'information. Les théories classiques de la communication utilisent évidemment abondamment ce concept d'information.

Mais on commence à retrouver ce concept dans des lieux qui normalement résistent à la traduction de leur problématique en des termes aussi réducteurs. Ainsi, on voit surgir ce concept d'information dans les théories de la représentation et même dans les théories de la conscience. L'humain est de plus en plus pensé en termes d'information et plus spécifiquement comme une machine à traiter, classer, emmagasiner, etc., l'information. Pour employer un langage plus suggestif, on assimile de plus en plus "l'humain" à un ordinateur, à une machine biomécanique.

Or qu'en est-il de cette conception qui pense l'homme comme un robot cybernétique? Dans la présente recherche nous tenterons d'explorer cette métaphore pour identifier quels concepts elle met en place et comment, au delà de son allure science fiction ces concepts traversent nos conceptions les plus acceptées de l'humain.

2. La question de l'information

Pour de nombreuses théories contemporaines ainsi que pour plusieurs vulgarisations, la pensée n'est qu'un mode de traitement d'information parmi d'autres.

Les romans de science-fiction et les textes de neurophysiologie les plus techniques avanceront sans difficulté une telle thèse. Supportée physiquement par cette machine biologique qu'est le cerveau, la pensée perçoit, se représente, met en mémoire, rappelle, compare, etc., l'information. Bref la pensée est au cerveau ce que l'information est à l'ordinateur. Comme le disaient récemment Rumelhart et McLelland: "Minds are what brains do best". Pour plusieurs penseurs contemporains cette manière de concevoir la pensée n'est plus une métaphore ou une analogie; elle n'est même plus un modèle mais plutôt le mode réel de son fonctionnement:

The general prevailing notion that the functions of the mind could not be described with reductive methods - in opposition to the deductive methods - as for a computer, and hence also for models of the mind, does not hold true any more.

(Taranto, E., 1980:195)

Since the trial and error method was introduced by computers, the reduction method has become part of information theory, and can be utilised in model of the mind and mind processes.

(Taranto, E., 1980: Ibid.)

Pourtant dans cette thèse réductionniste que l'on prétend non-métaphorique, le concept qui fait problème n'est pas celui d'"ordinateur" - ni même, je dirais, celui de "pensée" - mais plutôt celui-là même qui semble le plus innocent, à savoir celui d'"information". En effet, l'analyse critique même la plus simple semble tellement bien comprendre ce concept d'information qu'il passe inaperçu dans la présentation de la thèse. En effet, qui d'entre nous ne sait pas ce que c'est que de l'information?

Cependant, il faut frayé un peu dans le champ disciplinaire des diverses sciences de l'information pour s'apercevoir que ce concept si contemporain demeure des plus problématiques, à tel point que le consensus sur une définition est loin d'être atteint.

2.1. Les divers concepts d'information

En effet, ce n'est pas d'hier que les sciences de l'information sont à la recherche d'une définition rigoureuse de leur objet. Dès les années 50, suite aux critiques de Bar-Hillel et de Carnap sur l'assémantisme des définitions de Shannon et Weaver, de nombreux chercheurs ont tenté de définir cette dite "information" en intégrant les dimensions que l'intuition populaire y associait. C'est ainsi que, dans les années 65 à 75, les diverses théories de l'information se sont associées à la théorie générale des systèmes et de la cybernétique et ont alors intégré les questions : du traitement général de tout ce qui est porteur de message (code et signaux), des structures de décisions (théorie des jeux), du support et du traitement informatique de données et même de la documentation et de bibliothéconomie.

Mais depuis 75 la théorie de l'information s'est surtout identifiée aux théories de la computation et de l'intelligence artificielle, aux sciences cognitives et même à la logique épistémique. Plus récemment encore, depuis 1985 environ, elle s'est presque identifiée aux théories néo-connectionnistes de la cognition.

Ainsi, la question de l'information est devenue aujourd'hui un topos d'interrogations communes à de nombreuses disciplines. Loin d'être un corpus monolithique de réponses et de méthodes, les sciences de l'information sont devenues un carrefour interdisciplinaire aux vecteurs les plus divers. De l'intelligence artificielle à la philosophie, en passant par les théories de gestion administrative et la neurobiologie, tous intègrent d'une manière ou d'une autre ce concept d'information.

2.2. les divers champs disciplinaires de l'information

Cependant, bien que dans ces diverses disciplines et discours, l'"information" apparaisse souvent comme

un concept clair et limpide, on demeure des plus surpris que la littérature technique sur la question ne le définit presque jamais. C'est là d'ailleurs une plainte que font plusieurs chercheurs. Il est par exemple intéressant de constater qu'un ouvrage aussi important que *Human Information Processing* (Lindsay et Norman, 1973), l'utilise sans jamais en donner une définition. et que même après 15 ans, la situation n'a guère changée pour ce qui est des livres techniques tant de biologie, de psychologie (Rumelhart, Mc Lelland (1987)) que de neurophilosophie (Churchland(1986)).

Personne ne semble réussir à donner une définition englobante et synthétique de ce concept. Certes, de nombreuses disciplines s'en donnent une conception opératoire mais celles-ci ont souvent un caractère tellement opératoire qu'il est impossible de les utiliser hors du domaine en question. Pensons ici à certaines définitions de l'information issues des sciences documentaires ou des chercheurs russes l'ont identifiée aux documents scientifiques spécialisés. Certes la définition est opératoire et ce titre indiscutable. Mais si telle est la définition de l'information alors il faut en conclure que les journaux, qui ne sont pas des documents scientifiques, ne contiennent jamais d'information dans ce sens. De ce fait, de nombreuses définitions sont si opératoires qu'on en vient à ne pas pouvoir les employer hors du domaine technique ou elles sont utilisées.

Par ailleurs, dans d'autres cas, la définition devient si générale qu'elle inclue tous les phénomènes de communication. Je pense à cette définition récente de J.C. Burch :

"Information is data that have been put into meaningful and useful context and communicated to a recipient who uses it to make decisions." (Burch J.C. 1986, p. 3.)

Enfin, lorsqu'on fait le tour de toutes ces définitions, même dans les domaines les plus proches de la question de l'esprit ou de la pensée comme cela est le cas pour les théories cognitives, on découvre que l'analogie de l'esprit avec l'ordinateur ne présente plus la limpidité qu'on croyait.

En effet, lorsqu'on dit que la pensée humaine traite de l'information, on ne veut certainement pas dire qu'elle traite des documents scientifiques, ou encore des signaux sans signification. Mais alors voudrait-on dire quelque chose d'aussi général et classique que "la pensée traite des données signifiantes utiles dans les décisions"? Que veut-on dire au juste par cette analogie de l'ordinateur et de la pensée?

2.3. les paradigmes implicites des définitions de l'information

Pour comprendre davantage l'analogie, il faut, je crois, analyser davantage ces diverses définitions. Or, malgré leur allure un peu disparate, certains chercheurs comme Brookes, en 1975, Belkin et Robertson en 1976, Dretske en 1984, ont pu remarquer qu'elles fonctionnaient toutes dans un paradigme commun.

Ces auteurs ont en effet montré que dans ces définitions, la question de l'information opérait, sans que cela soit dit, dans un paradigme de connaissance. De Shannon à Dretske, de Newell à Pylyshyn, de Minsky à Rumelhart, de Yovits à Belkin, l'information fonctionne toujours comme un concept cadré dans l'horizon d'une théorie implicite de la connaissance.

Toutes ces définitions cernent en effet différents aspects du processus de la connaissance. On distinguera par exemple des définitions infra-cognitives (genèse, incertitude, perception, systématisation de la connaissance, etc.) et des définitions liées au processus individuel de connaissance (conceptualisation-communication, représentation de l'information). Certaines touchent par ailleurs les dimensions de connaissance sociale (Concepts sociaux, idéologies, symbolisation) et de métacognition (connaissance formalisée). C'est ainsi par exemple que la définition de Shannon inclut une dimension de "certitude et incertitude", celle de Dretske la présence d'un "believer", ou celle de Bush la "signification".

Dans toutes ces définitions de l'information, on trouve une psychologisation primitive de ce concept. L'"information" est ici quelque chose qui est "attendu", "perçu", "choisi", "retenu", "reconnu", "comparé", "décidé", "communiqué", etc., par un système - que ce système soit de type neurologique ou informatique. Hors du système qui l'intègre, il n'y a pas d'information. Celle-ci est toujours relative à un système qui, "cognitivement", l'envoie ou le reçoit.

Il est très important, il me semble, de reconnaître le paradigme sous-jacent à ces définitions car on voit apparaître sous un nouveau jour la difficulté de la métaphore et de la thèse réductionniste qu'elle soutient. En effet, il semble de plus en plus évident que la métaphore de l'information est loin d'aller dans la direction d'une conception de l'homme comme une machine. Elle me semble aller dans l'autre sens c'est-à-dire d'une conception de l'ordinateur comme un

humain. C'est comme si la métaphore s'était prise au piège de l'anthropomorphisation de l'ordinateur dans son processus de traitement de l'information et qu'elle n'avait pas vu le paradigme cognitif sous-jacent qui la traversait.

Il m'apparaît évident que les théories tant probabiliste que décisionnelle, sémantique ou documentaire de l'information reposent sur ce transfert des concepts d'une psycho-philosophie primitive à une machine soit formelle soit computationnelle.

C'est pour cette raison qu'on aurait distingué divers concepts de l'information:

- l'information du point de vue de la probabilité du signal (théorie de Shannon). C'est là une théorie de l'incertitude "du message" (sic)
- l'information du point de vue de la décision à prendre (Decision theory)
- l'information du point de vue du contenu (théorie sémantique et sémiotique)
- l'information retenue (archivage, documentation).
- l'information transmise et partagée (théories sociale, situationnelle et écologique de l'information).

Nos informaticiens et neurophysiologistes, lorsqu'ils philosophent, semblent, du moins en un premier niveau de lecture de la problématique, des Aristote qui s'ignorent. En effet, si on analyse ces diverses conceptions, on ne peut s'empêcher de retrouver dans leur définitions, les grandes catégories de la psychologie populaire issues de la vieille tradition philosophique traditionnelle telles que la mémoire, l'exécution, l'interprétation, l'intelligence, la décision, et même la représentation etc. L'information nomme le résultat complexe d'une machine qui exécute, décide, retient, interprète etc. C'est même là quelque chose de normal comme le pense McCarthy (1979):

"To ascribe certain beliefs, knowledge, free will, intentions, consciousness, abilities or wants to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person" (McCarthy, J. 1979: 2).

S'il en est ainsi, on comprendra alors pourquoi l'analogie est si efficace. Il n'est pas surprenant que l'esprit humain soit pensable comme un ordinateur. Nous avons projeté sur la machine les catégories et concepts par lesquels nous nous pensons nous-mêmes depuis des siècles. Nous nommons les résultats des opérations de cette machine dans le même vocabulaire avec lequel nous identifions nos

propres actions. Le contenu même de cette nouvelle machine qu'est l'ordinateur, qui sur des points particuliers nous simule nous-mêmes a été, je dirais, "philosophisé" à travers les catégories des plus classiques par lesquelles nous avons toujours pensé l'esprit humain. En Occident, depuis Aristote en passant par Saint Augustin, Descartes, Kant et j'oserais même dire Hegel l' "esprit humain" est toujours un "anima" qui perçoit, choisit, décide, croît, se représente, retient, réfléchit, etc. Doit-on se rappeler que pour Aristote et tous ses commentateurs médiévaux, la raison, la "ratio" était un agent -opérateur extrayant des formes de la matière? On donnait à cet opérateur, je le rappelle, le nom d' "intellect agent" et l' "âme" - le "récepteur qui recevait ce qui était extrait de la matière par le processus dit d' "abstraction" - était "informée" au sens strict, c'est-à-dire au sens d'une mise en "forme".

Il n'est donc pas surprenant que les théories contemporaines de l'information et les modèles de l'esprit (Mind) qu'elles proposent reposent sur une longue tradition occidentale de compréhension de la " machine " humaine comme un agent de mise "en forme" . Information, après tout, vient de " In-forma".

Ainsi, dans les discours de vulgarisation, des théories informatiques et des sciences de l'information, l'homme -machine -ordinateur n'est pas un robot mécanique. Au contraire, c'est la machine qui s'est humanisée via des catégories des plus classiques par lesquelles nous nous sommes traditionnellement pensés.

3. La théorie computationnelle de l'esprit.

3.1. Des concepts originaux.

Il nous faut maintenant analyser plus en profondeur cette métaphore de l'ordinateur et de l'esprit. Bien que dans une première approche, on puisse déceler une anthropomorphisation de la machine, il faut cependant, je crois, voir qu'au-delà de cette métaphore, s'instaure une description spécifique du fonctionnement de l'esprit.

En effet, la thèse de l'esprit ordinateur propose bien d'autres concepts pour décrire le fonctionnement de l'esprit, concepts qui lui sont originaux et que la pensée philosophique classique est loin d'avoir utilisés. Pensons ici aux concepts de compilation, de computation, de récursion, de parallélisme, etc.

Tous ces concepts en viennent à générer une théorie que nous pourrions appeler théorie computationnelle de l'esprit et qui dans le langage de la philosophie présente la version contemporaine de la théorie causale de l'esprit. Cette théorie met en jeu un plus grand nombre de problèmes que la simple question de l' "intelligence" dite artificielle. Car elle inclut la question de la représentation mentale, des croyances, de l' imagination, des émotions ou plus généralement des divers types d' états mentaux que nos théories populaires et techniques reconnaissent à l'humain.

Cette version contemporaine de la théorie causale de l'esprit est assurément complexe et cette complexité est d'autant plus grande aujourd'hui qu'elle en appelle à des concepts et théories de haute technicité qu'il n'est pas toujours facile d'exposer et d'analyser de façon simple. Aussi la présentation, ici, de cette problématique sera-t-elle restreinte.

Nous n'étudierons ici qu'un seul vecteur de cette analogie à savoir celui qui associe les états mentaux à des états de machines et qui, par inférence, établit des relations causales entre ces états mentaux.

L' objectif, ici, est de dévoiler non pas le modèle informatique sous-jacent aux théories contemporaines de l'esprit mais plutôt de révéler le modèle philosophique de l'esprit à l'oeuvre dans les théories computationnelles de ce même esprit. Et je souligne computationnelle car il s'agit d'un modèle très spécifique défendu par une école particulière de pensée. Pensée qui s'oppose au modèle moniste anomal de Davidson, au modèle néo- connexionniste de Rumelhart, Mc Lelland et même du modèle moniste matérialiste de Churchland.

Le modèle dont je parlerai sera celui de Fodor, Pylyshyn, Jackendoff, et pour une part celui de Chomsky et même d' Armstrong.

3.2 nature de la théorie causale de l'esprit.

-La théorie causale et la nature configurationnelle des états mentaux.

Une première sous-thèse de cette théorie computationnelle de l'esprit consiste à considérer l'esprit non pas comme un agrégat matériel de molécules ou d'atomes biologiques, mais bien comme une machine abstraite au sens de la théorie générale des automates et plus précisément d'une machine de Turing et encore plus spécifiquement d'une machine virtuelle de Von Neumann. Pour parler comme Pylyshyn (1984), l'esprit (mind) est un état de l'architecture fonctionnelle de l'organisme biologique humain.

Plus précisément encore, cette thèse consiste à dire qu'un état mental de l'organisme humain est similaire à un état de la machine. L'analogie n'est donc plus entre l'organisme biologique humain ou le cerveau et la machine physique qu'est l'ordinateur (ce qui serait une thèse moniste réductionniste des événements mentaux avec les événements physiques) mais entre la structure d'un état de l'esprit et une structure d'état de la machine. Il est bien important de voir que l'analogie s'instaure via le concept d'état et non le concept de matière. Le cerveau se possède dans des états mentaux comme n'importe quelle machine se trouve dans des états physiques.

C'est ainsi que pour D. Lewis, ce concept d'état de machine définit bien ce que la psychologie populaire entend par "état d'esprit".

"The concept of a system of states that together more or less realize the pattern of generalizations set forth in commonsense psychology" (Lewis D. 1980p. 124)

Boyd (1980) définit cet état directement en ces termes:

"Mental states and processes are computational states" (Boyd, 1980, p. 96). On trouve diverses versions de cette thèse.

Par exemple la douleur, le plaisir, l'imagination, et la pensée, sont en général des états spécifiques dans lesquels l'organisme humain et plus spécifiquement le cerveau se trouvent à un moment ou l'autre.

En d'autres termes, cette version de la théorie causale de l'esprit établit une relation causale entre un état mental comme la douleur, la conscience, etc., Ces expressions sont des expressions descriptives de configurations particulières ou d'états des constituants de l'organisme. Être en douleur c'est, pour un organisme, être dans un certain état corporel. "Être en douleur" c'est pour le cerveau et tout le corps être

dans une configuration spécifique des constituants neurologiques ou de ce que l'on appelle plus communément le système nerveux.

Pour bien saisir la nature de cette première sous-thèse je la traduirai et l'illustrerai par une autre analogie, portant cette fois non pas sur une machine computationnelle mais plutôt sur une machine purement mécanique : la lessiveuse.

La lessiveuse est une machine physique. Il en existe de différentes marques. Cependant quel que soit leur "hardware", les lessiveuses modernes possèdent toutes des cycles de "lavage", de "rinçage" et d'"essorage". Ces termes définissent non une structure macro-moléculaire particulière mais un état abstrait du système d'une machine. Le linge se rince parce que la machine est dans un état particulier de son fonctionnement interne. "Rinçage" est une expression du langage quotidien qui nomme un état de la machine. Cet état est certes causé par une configuration particulière de la mécanique, mais elle n'est pas identique à la matière même de cette mécanique : plusieurs mécaniques matériellement différentes peuvent être dans un tel état.

Certes, le "rinçage" n'est pas un état cognitif ou perceptuel, comme le sont la "douleur", le "plaisir", la "pensée", etc. Mais de telles expressions, pour cette théorie de l'esprit, sont des noms donnés à des états sophistiqués de l'organisme qu'est une machine qu'elle soit humaine ou computationnelle.

Du point de vue des théories des machines abstraites, ces deux types de machines, lessiveuses et organismes humains génèrent des états et des transitions d'états. On peut dire, peut-être en exagérant, que la "pensée" comme concept théorique est un état de la machine humaine comme le "rinçage" est un état de la lessiveuse. Ces deux concepts d'"esprit" et de "rinçage" dans cette thèse définissent donc des états systémiques du type de machine auxquelles ils sont rattachés. Ces états ne s'identifient pas au physique de ces machines.

Il faut bien voir le statut épistémologique de cette théorie causale de l'esprit. Elle n'est pas une théorie moniste c'est-à-dire une théorie qui identifie la nature du fonctionnement de l'esprit avec la matière elle-même, en l'occurrence la dimension physique (hardware) de l'ordinateur biologique que serait le cerveau. Il ne s'agit donc pas de la théorie matérialiste moniste.

*"The causal theory of mind permits and encourages, without actually entailing, the identification of the mental with the physical" .
(Armstrong, 1984, p.120)*

La distinction est ici importante. Car elle met en jeu deux thèses différentes mais souvent confondues. La théorie causale de l'esprit ne s'intéresse pas formellement aux conditions matérielles d'existence de cet esprit mais, dans un premier temps, à la structure configurationnelle de l'état de cet organisme. Ce qui est "causal" est ici lié à la modalité de configuration que prend un état, et non au statut physique qui porte cet état. Une même configuration peut se réaliser dans diverses structures matérielles. La causalité dont il est question est donc d'ordre structural. Structural dans le sens où chaque état se définit par une organisation spécifique ou, en termes techniques : une configuration spécifique des constituants de son état. Un état est avant tout une configuration spécifique de la machine à un moment donné de sa dynamique et non une caractérisation de la physique de cette machine.

Formulée peut-être plus simplement, cette thèse ne soutient pas la proposition qu'une pensée ou une émotion est une simple décharge neurologique et chimique du corps et qu'elle se localise dans une aire cérébrale spécifique. La thèse de la causalité du fonctionnement de l'esprit, dans cette première dimension que j'appelle ici configurationnelle, n'implique donc pas nécessairement la thèse réductionniste matérialiste. L'existence matérielle de nos pensées est une condition nécessaire pour que nous ayons des émotions et des pensées. Sans corps, pas d'esprit. Mais cela ne fait aucunement de cette existence une condition suffisante.

La théorie causale et la nature transitionnelle des états mentaux.

Cette théorie causale de l'esprit met en place une deuxième sous-thèse importante. Dans cette sous-thèse, l'analogie de l'esprit avec la machine-ordinateur ne porte pas uniquement sur les états mais aussi sur les transitions entre les états. Une machine n'est pas qu'un agrégat d'états mais une configuration de transitions entre ces états. Chaque état entraînant d'autres états selon un certain ordre. Ordre qui peut être réglé ou non, déterministe ou non et même auto-régulateur ou non.

Dans cette perspective, cette théorie soutient que le mental aussi passe par des transitions.

Transitions qu'elle thématise d'ailleurs souvent en termes de causalité. Par exemple, un état mental de douleur peut causer un état mental de tristesse et de souffrance ; un état mental de plaisir peut causer un état mental de satisfaction et de contentement. Dans le domaine cognitif et mental de quelqu'un, il y a par exemple une différence entre dire qu'un comportement est causé par une crainte et dire qu'il est causé par une sécrétion hormonale. Affirmer qu'un comportement obsessionnel est causé par une représentation mythique de la pureté c'est autre chose que de dire qu'il est causé par un acide affectant le réseau des neurones de l'hypothalamus. Dans cette perspective, on peut dire que la théorie freudienne est une théorie causale de l'esprit (théorie, que, incidemment, Freud considérait à l'origine comme neuromatérialiste!)

On arrive donc par cette dernière analyse à introduire une deuxième caractéristique de cette théorie computationnelle. Elle est transitionnelle en ce qu'un état peut se transformer en un autre état par une transformation de ses configurations. Nous disons bien ici configurationnelle et non représentationnelle. En effet, ces transitions peuvent selon les théories impliquer ou non des représentations (cf le débat Fodor-Rumelhart).

Encore ici, il faut voir le statut épistémologique de cette dimension de la théorie causale de l'esprit. Si dans certaines machines élémentaires comme notre lessiveuse de tout à l'heure, on peut concevoir que la lessiveuse comme une machine déterministe, rien théoriquement ne nous permet de considérer que les transitions entre les états d'un organisme plus complexes sont déterministes.

Au contraire même sur le plan formel certaines machines, comme on le sait sont non-déterministes. Qui plus est rien ne nous permet de penser que les modèles de transitions au sein d'une machine doivent être purement linéaires. Des modèles parallélistes (Rumelhart McLelland), chaotique et autorégulateur (Freeman) existent. Plus encore, rien sur le plan formel ne nous permet de penser que les transitions sont sérielles. Il est possible de penser formellement des modèles formels récursifs, c'est-à-dire opérant à divers niveaux hiérarchiques mutuellement imbriqués.

Ce point est important car cette sous-thèse et ses divers modèles de réalisation situent cette théorie hors d'une problématique strictement déterministe à la La Mettrie et même à la Skinner. Au contraire, elle instaure dans le modèle explicatif une dimension historique. En effet, qu'elle prenne une forme

computationnelle (à la Pylyshyn) ou générative (à la Fodor-Chomsky), chaotique et autorégulatrice (à la Freeman)cette théorie met en jeu des suites de transitions non déterministes mais déterminables d'états qui se déroulent dans le temps et qui donc, chacun selon leur cas, ont une histoire .

Le problème qui vient alors ici est de comprendre comment des états mentaux instaurant entre eux des relations transitionnelles peuvent-ils être compris en termes de rapport de causalité mécanique . Ou plus spécifiquement de quelle nature sont ces rapports de causalité pour qu'ils permettent ces transitions et ces transformations de configurations?

La théorie computationnelle et la nature représentationnelle des états mentaux.

La réponse aux dernières questions se trouve dans une troisième sous-thèse spécifique à la version computationnelle de cette théorie. C'est probablement une idée des plus répandues que le modèle dit computationnel des états mentaux (c'est-à-dire l'intelligence artificielle) est purement physicaliste et réductionniste. Pourtant , l'histoire de cette théorie montre qu'il en est tout autrement. En effet, Allen Newell(1983), dans une étude du développement historique de l'intelligence artificielle montre comment cette discipline a cheminé à travers divers débats théoriques pour enfin arriver à définir son propre territoire. Depuis ses origines, vers 1950-55, l'intelligence artificielle a dû tour à tour se distinguer de la cybernétique, de la théorie des systèmes , de la reconnaissance des formes, de la calculabilité numérique , de la théorie des programmes, de l'ingénierie électronique , et finalement même de la logique formelle .L'une des étapes par lesquelles l'intelligence artificielle s'est dissociée des diverses disciplines mathématiques et informatiques dominantes fut, de l'avis même de Newell, l'identification de l'ordinateur comme une machine (au sein abstrait du terme) traitant essentiellement des objets de nature symbolique comme l'est par exemple un système algébrique.

"As soon as 1965, scientists in AI saw computers as machines that manipulated symbols.The great things was, they said that everything could be encoded into symbols, even numbers." Newell 1983: p 196.)

La théorie computationnelle

Cette compréhension de l'intelligence artificielle comme système de traitement de symboles est renchéri par les ajustements récents des théories de la computation.

En effet cette théorie qui , à l'origine, ne considérait que les nombres, c'est à dire des unités numériques, s'est progressivement orientée vers une définition qui la conçoit presque uniquement comme un théorie de la manipulation des symboles ayant syntaxe et sémantique. J'en prendrai pour exemple la définition relativement récente de Pylyshyn: la théorie computationnelle est...

"the rule governed transformation of formal expression viewed as interpreted symbolic codes" (Pylyshyn 1984: 55)

ou encore

" A computational process is one whose behavior is viewed as depending on the representational or semantic content of its states. This occurs by virtue of there being another level of structure- variously called the " symbol level" of the syntax or logical form - which possesses the following two essential properties. One, the formal syntactic structure of particular occurrences (tokens) of symbolic expressions corresponding to real physical differences in the system, differences that affect the relevant features of the systems behavior. Two, the formal symbol structures mirror all relevant semantic distinctions to which the system is supposed to respond and continue to do so when certain semantically interpreted rules are applied to them , transforming them into new symbol structures. For every apparent, functionally relevant semantic distinction there is a corresponding syntactic distinction."(Pylyshyn 1983: 74)

Comme on le voit cette thèse de la computationnalité n'est pas une thèse sur la nature physique d'un processus mais au contraire sur la dimension symbolique en jeu.

En effet la thèse conçoit la computation comme une fonction qui prend comme intrant non des objets physiques mais des symboles atomiques ou de base décrivant une quelconque réalité, c'est-à-dire interprétables. L'essentiel de la thèse dit qu'il est possible d'effectuer des opérations sur des symboles purs, non interprétés et d'en arriver à générer des formes symboliques transformées recevant à leur tour une interprétation. La computation est une manipulation de symboles et uniquement de symboles.

Exemple : soit la liste { #, @ } et la computation suivante : inverser les types de symboles. On obtiendra alors : { @, # }. On remarquera que la règle utilisée dans ce cas ne tient aucunement compte de l'interprétation que l'on peut associer aux symboles d'entrée et de sortie. Pour manipuler # et @, point n'est besoin de savoir à quoi ces symboles réfèrent.

Ceci par ailleurs ne veut pas dire qu'une règle dans ses constituants ne peut pas tenir compte d'une interprétation éventuelle. Or au contraire, c'est en raison d'une interprétation possible que certaines règles seront choisies plutôt que d'autres. La computation se fait certes sur des symboles non interprétés - mais ceux-ci devront être éventuellement interprétés. Ceci contraint alors le choix des règles. Ce point est important car il fait précisément le lieu de la dispute Searle-Newell sur la nature sémantique ou non-sémantique de l'intelligence artificielle.

-La question causale et la représentation.

Ainsi donc, la théorie causale de type computationnel (l'"intelligence artificielle") ne nous met pas en présence d'une problématique physicaliste et réductionniste de la pensée. Elle n'est pas une théorie moniste. Mais du moins dans cette version elle nous instaure dans problématique représentationnelle. Dans cette perspective, les états mentaux et surtout les relations entre ces états mentaux traitent et transforment des symboles. En soi une machine qui traitent et transforment des symboles peut être réalisée dans divers modes matériels. Mais dans sa nature fonctionnelle, elle est essentiellement sémiotique.

Une représentation ne cause pas (au sens classique de ce terme) un comportement. Une relation causale n'existe qu'entre des objets physiques. Ainsi l'analyse un peu plus serrée du modèle causal

computationnel soutient que ce à quoi nous sommes confrontés appartient à l'ordre sémiotique et sémantique et non à l'ordre causal physique. On peut même se demander si la théorie computationnelle des états mentaux est encore causale.

Cette nature représentationnelle des états mentaux et de leur transitions n'est pas une thèse acceptée par tous. Au contraire, le débat sur le statut de la représentation est maintenant enclenchée. La thèse étant qu'il n'est pas nécessaire que chacun des états par lesquels la machine ou l'organisme passe prennent la forme de "représentations" qui plus est il n'est pas nécessaire que la transitions entre ces états soient réglées et encore nomiques. Une théorie représentationnelle et normale de l'esprit d'une part implique une sémantique dont la fonction est extrêmement difficile à circonscrire (Searle, 1984) et d'autre part surtout ne réussit pas à expliquer les nombreux phénomènes cognitifs qui sont tantôt adaptatifs, transitoires, et soumis à des contraintes multiples. (Rumelhart et McClelland 1987) C'est là cependant un point que nous ne pouvons ici discuter car il nous entraînerait au delà de nos préoccupations immédiates.

4. Portée de la métaphore.

Qu'en est-il maintenant de cette métaphore de l'homme machine? L'homme est-il toujours ce robot mécanique? La version contemporaine de cette conception de la pensée est plus complexe qu'en présente la métaphore quotidienne de l'homme-ordinateur. Elle ne le réduit pas à un mécanisme simpliste. L'Homme-machine contemporain est ici conçu comme un système complexe de traitement d'information où jouent configurations, représentations, transformations et transitions entre représentations d'information. Et une telle conception ne peut être que fonctionnelle au sens mathématique du terme. Elle s'avère donc plus complexe que ne le laisse croire l'image vulgarisante de nos discours quotidiens qui ne voient dans cette métaphore que le retour de l'homme-machine des siècles derniers.

Mais il y a plus dans cette conception. Cette métaphore séduit. Et je crois que sa séduction et son succès malgré des apparences contraires, relève de sa grande similitude avec les conceptions dites philosophiques et humanistes de nos discours contemporains sur l'humain. En effet de nombreux discours critiques, sociologiques, anthropologiques ou littéraires monnaient de plus en plus cette conception fonctionnaliste de l'homme comme machine traitant des représentations et des relations entre ces représentations. Notre conception moderne de l'humain avec sa "conscience", son "inconscient", ses "pulsions", ses "idéologies", ses "phantasmes", est, je

crois, notre modalité contemporaine de transformer l'automate de l'homme-machine de La Mettrie. L'humain d'aujourd'hui n'est certes plus l' "anima-teur" central d'Aristote. Il est devenu une "machine désirante" aux phantasmes les plus pulsionnels, aux déterminismes biologiques et sociaux les plus fonctionnels.

Je doute que même la critique éthique la plus féroce de cette conception retrouvera sous de nouveaux lexèmes les vecteurs qu'elle avait elle-même mis dans la compréhension de son objet. Tant que notre conception de l'humain restera du type fonctionnaliste il sera difficile, je crois, de se défaire de cette métaphore de l'homme-machine même dans sa version computationnelle ou néo-connexionniste contemporaine.

Bibliographie et références.

Dreske F.I. Knowledge and the flow of information. MIT Press , 1982.

Davidson D., Essays on Actions and Events, Oxford University Press. 1982

Armstrong, D.M, A Materialist Theory of the Mind, Londres, Routledge & Kegan Paul, (1968) .

Armstrong D.M.,& Malcolm, N., Consciousness & Causality. Blackwell,1984.

Bar Hillel, Y., "An examination of Information Theory", Philosophy of Science, vol. 22 ,1955, pp. 86-105.

Bar Hillel, Y. Carnap R. " Semantic Information":The British Journal of the Philosophy of Science, 4,(14): 1953: 144-1157.

Belkin ,N.J., Towards a definition of information for informatics" Informatics #2, 1976 p.50-56.

Brookes,B.C. Information science ,In British Librarians and information science ,1966-70 H.A. Whatley (ed) London: The Library Association, 1972, p137-49.

Block, N., "Troubles with Functionalism" in Block , Readings in the Philosophy of psychology. Methuen, 1980.

Block, N., "Are Absent qualia Impossible" in Philosophical Review ,89 1980,

Boyd,R., "Materialism without Reductionism : What Physicalism Does Not Entail" In Block.(1980)

Burch J.C. Information systems , John Wiley , New York, 1979, 1986.

Cherry, E.C., "A history of the theory of Information", Proceeding of the Institute of Electrical Engineers, vol. 98, (III) . 1951) p. 383-393.

Chomsky, N., "Human Language and Other Semiotic systems," Semiotica, 25. 1/2 31-44 ,1979.

Chomsky,N., Rules and Representations , New York,Columbia U.P.1980.

Churchland ,P., Matter and Consciousness, MIT Press, 1984.

Churchland-Smith,P. Neurophilosophy .Cambridge:MIT Press, 1986.

Freeman W.,J., & Skarda,C., "How brains make chaos in order to make sense of the world" Behavioral and the Brain Sciences, 10: 161-173, 1987

Follesdal,D "Causation and Explanation: A problem in Davidson's View of Action and Mind" : In Lepore,E., et B McLaughlin (eds), page 311-323.

Fodor, J.A . Modularity of Mind .MIT Press .1983

Fodor, J., " Special Sciences, or The Disunity of Science as a Working Hypothesis" in Block,(1980) p. 125.

Fodor,J., Representations. MIT Press, 1981.

Fodor, J. and N. Block, "What Psychological States are not" in Block, vol 1

Jackendoff,R., Semantic and Cognition, MIT Press, Cambridge, 1983.

Lepore,E., et B McLaughlin , (eds),Actions and Events; Oxford ,Blackwell, 1985.

La Mettrie, O, L'Homme-machine, Paris, Pauvert, (1748).

Lewis, D., "Mad Pain and Martian Pain" , in Block (1980) vol 1.

Lindsay & Norman D., Human Information Processing. New York: Academic Press. 1972.

Rumelhart D., Mc Lelland,J., Parallel distributed processing, Vol I et II, MIT Press, 1987

Newell, A., "Physical Symbol systems "Cognitive science. 4:2 135-183.

Neumann J von, The Computer and the Brain, New Haven, Conn, Yale University Press, 1958.

Pylyshyn, Zenon W., Computation and Cognition. Towards a Foundation for Cognitive Science, MIT Press, 1984.

Searle J., Mind Brains and Science, Harvard University Press,1984

Taranto ,R., The Mechanics of semiotics and of the " human mind", Cybernetica, 23, (3) 1980, pp. 195-219

Turing ,A.M. "Computing Machine and intelligence" in Anderson, A.R. (eds) Minds and Machines,Englewood Cliffs, N.J. Prentice Hall. 1964.

Wilkes, K,V.Physicalism, Routledge & Kegan Paul, 1978.

Wilkes K.V., " Functionalism, Psychology and the Philosophy of Mind" in Philosophical Topics, 12, 1981, p. 165.

Yovits,M.C., A theoretical framework for the development of information science: In: Problems of information science (FID 530) Moscow VINITI, 1975, pp 90 -114.

J.G . Meunier : La machine humaine et l'information_____