



IA/Sciences cognitives quels enjeux ?

par Eric Braine

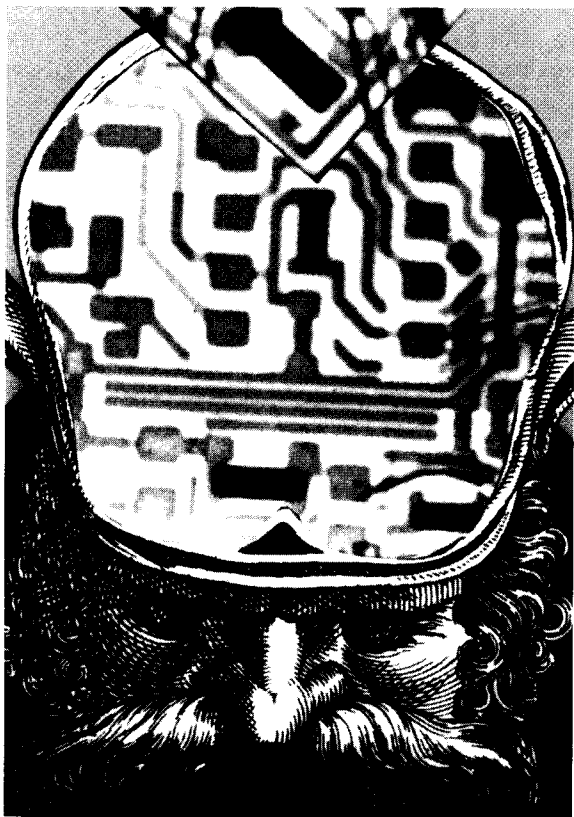
En 1956, des mathématiciens, informaticiens et psychologues réunis en conférence au "Darmouth College", inventent l'intelligence artificielle : l'I.A. Ils la placent sous le paradigme "computationnaliste" : penser, c'est abstraire, représenter, calculer. A la suite d'Alan Turing, ils affirment la possibilité de simuler artificiellement sur un ordinateur les représentations humaines. Ce faisant, ils ouvrent comme le montre l'article de Joëlle Proust : "l'avenir conflictuel de l'I.A." Ce débat philosophique sur la nature et les conditions de l'intelligence n'a pas fini de produire ses effets...

De proche en proche, les différentes disciplines concernées : psychologie, linguistique, philosophie, biologie, neurologie etc..., vont mettre en cause les fondements logico-symbo-

liques de l'I.A. et sa prétention à unifier sous un paradigme mathématique, les différents champs de connaissance. La voie est ainsi ouverte pour des hypothèses naturalistes - biologistes. Désormais, penser ce n'est plus calculer (symboliquement), mais s'adapter, interagir (avec un environnement).

L'article de John Stewart s'inscrit dans cette perspective et souligne le paradoxe de "sciences cognitives" en formation qui ont cependant l'ambition de jouer le rôle hégémonique d'une "interscience" (analogue à celui joué, en son temps, par le structuralisme).

Les enjeux de ces nouvelles sciences du sujet sont importants : - tentative de redéfinition de l'intelligence, donc de l'homme avec ses conséquences, notamment la légitimation des hiérarchies sociales (pouvoir des technos-



L'avenir conflictuel de l'I.A.

PAR JOELLE PROUST

L'intelligence artificielle constitue l'un des axes les plus remarquables du développement de l'informatique des trente dernières années. On connaît la définition classique donnée par Minsky : discipline ayant pour objet de "fabriquer des machines qui font des choses qui, accomplies par l'homme, demanderaient de l'intelligence". Cette définition, et en particulier l'ambiguïté liée à l'expression "des choses qui demanderaient de l'intelligence", contient tout l'avenir conflictuel de l'I.A. : qui va être chargé de délimiter de quelles choses il s'agit, et de caractériser les comportements intelligents ?

ciences) ;- redécoupage du territoire des différentes sciences, nouvelle distribution du savoir, mais au profit de qui ? - rôle accru des sciences humaines et sociales, devenues plus "dures", mais que deviennent alors les cultures particulières, les singularités ? quelles formes de socialité porte ce projet de déconstruction, d'instrumentalisation des subjectivités ? - enfin où se situe la finitude humaine face à cette volonté d'assurer une survie par les machines ?

L'article de Dominique Lestel : "le loup dans la bergerie", finit de mettre en cause la reconstruction artificielle de la cognition, coupée de ses contextes sociaux notamment et introduit une approche écologique de la cognition comprise comme l'intelligence des environnements.

Au congrès de Dartmouth de 1956, fut pour la première fois employé le terme d' "Intelligence Artificielle". Les inventeurs de ce concept convenaient qu'il revient au spécialiste de l'I.A. d'enquêter sur les conditions les plus générales de l'intelligence, parce qu'en vertu de l'hypothèse cognitiviste, il est le mieux placé pour le faire.

L'hypothèse cognitiviste

En effet, l'hypothèse cognitiviste (également appelée parfois l' "hypothèse computationaliste") avance que toute pensée est un calcul sur des symboles, tels que seules y soient pertinentes des considérations de forme. Si l'on exhibe les conditions formelles les plus générales de la cognition, on sera donc en mesure de faire effectuer par un ordinateur les tâches les plus difficiles que des hommes aient à rencontrer.



Voilà le projet de ce qu'on appelle à la suite de John Searle "la version forte de l'I.A." Ce projet, évidemment situé dans le long terme, implique que l'on cherche à obtenir d'un système qu'il ait les connaissances et les mécanismes d'inférence adéquats pour porter des jugements pertinents sur le monde réel et y prendre des décisions rationnelles. Cela suppose, entre autres choses, qu'il soit capable de revoir ses connaissances et d'en acquérir de nouvelles de manière autonome, et ce, sans que le programmeur ait à le protéger des brouillages et de l'encombrement des circuits créés par l'excès de l'information disponible. Dans cette première version, l'I.A. a pour ambition de faire de l'ordinateur bien davantage qu'un simple instrument de la psychologie : elle considère en effet que l'ordinateur convenablement programmé "est un esprit, au sens où les ordinateurs dotés des bons programmes peuvent être dits littéralement comprendre et avoir tous les autres états cognitifs".

Certains praticiens de l'I.A., tout en adhérant pour la plupart à l'hypothèse cognitive, ont souhaité modérer cette ambition un peu utopique des fondateurs. Il ont souhaité subordonner le travail de l'ingénieur en I.A. à celui du psychologue. Il ont en outre voulu restreindre le rôle de l'ordinateur en psychologie à celui qu'il joue dans les sciences de la nature. On ne s'attend pas, dans cette perspective, à ce que l'I.A. produise un analogue fonctionnel de l'esprit humain ; mais on souhaite utiliser l'ordinateur, selon les mots de Searle, "pour formuler et tester les hypothèses d'une façon plus rigoureuse et plus précise" concernant la nature et le fonctionnement de l'esprit (Searle, 1980, p. 417).

Mécanisme faible, mécanisme fort

Les deux interprétations, forte et faible, de l'I.A. peuvent être comparées avec la manière dont, au XVIII^{ème} siècle, on a cru pouvoir tirer parti du mécanisme pour rendre compte des comportements

animaux, voire humains. Le mécanisme est à certains égards une hypothèse analogue au computationalisme : ces deux doctrines suggèrent qu'il existe une théorie scientifique puissante pouvant servir de paradigme causal dans l'explication de phénomènes traditionnellement considérés comme de type supérieur. Le mécanisme, d'origine cartésienne, s'est trouvé ensuite puissamment renforcée par le développement de la physique newtonienne. Le mécanisme suppose en effet qu'une importante classe de faits, jusqu'alors interprétés de manière théologique, peuvent être ramenés à un système de déterminations purement physiques, et expliqués en considérant seulement le mouvement des corps soumis à certaines forces.

Buffon a soutenu, à la suite de Descartes, que le comportement des animaux pouvait s'expliquer en termes matérialistes : "Le vivant et l'animé, au lieu d'être un degré métaphysique des êtres, est une propriété physique de la matière" (1). Les animaux n'ont ainsi qu'une âme "corporelle". Comme le montre bien François Dagognet (2), Buffon fait toutefois du mécanisme un usage partiel et prudent. Il refuse un réductionnisme outrancier qui ne verrait dans l'organisation animale qu'un jeu de roues et de poulies, à la manière de Descartes. Il complète les schémas mécanistes par l'évocation de "forces intérieures", présentes dans les corps organisés, et dont la complexité défie tout espoir de réduction.

La position de Buffon, moniste et matérialiste à propos de l'animal, mais dualiste pour ce qui concerne l'homme, a suscité, en simplifiant, deux types d'objections. D'un côté, certains dualistes, répondant à des inquiétudes théologiques, dénoncent l'application de principes matérialistes (biodynamiques sinon, strictement mécanistes) à ce qui relève du mental, serait-ce l'âme animale. Ils considèrent que si on explique l'intelligence animale en termes naturalistes, c'est la foi en une âme humaine bien distincte du corps qui se trouvera menacée.

De l'autre, les philosophes qui souhaitent au contraire radicaliser l'hypothèse mécaniste, ne se satisfont pas du dualisme que Buffon met en œuvre lorsqu'il dénie aux animaux même les plus élémentaires des capacités cognitives qu'il reconnaît à l'homme. Ces deux courants anticipent de façon assez étroite la divergence entre plusieurs manières de comprendre le projet d'I.A.

Mécanisme et computationalisme

Le paradigme de l'I.A. a plusieurs points communs avec le mécanisme. Tout d'abord, il s'agit dans l'un et l'autre cas d'étendre à une nouvelle classe de faits les acquis d'un certain domaine théorique, et en particulier de comprendre des phénomènes relevant du mental (tels respectivement que le comportement animal ou humain, et de façon générale la cognition) à l'aide de modèles empruntés à une science. Ceux de la mécanique rationnelle, dans le cas du mécanisme des XVII^{ème} et XVIII^{èmes} siècles, ceux de la logique et de l'informatique, dans le cas de l'I.A.

En poursuivant l'analogie avec le mécanisme, comment justifier "l'hypothèse cognitive", c'est-à-dire l'extension de ces travaux initialement logico-

LES THESES DE FREGE

LE TOURNANT DÉCISIF QUI A PERMIS L'ÉLABORATION DU PROJET D'I.A. PEUT ÊTRE DATÉ : 1903, LES LOIS FONDAMENTALES DE L'ARITHMÉTIQUE, DE FREGE, OUVRAGE INAUGURAL DE RECONSTRUCTION EN TERMES LOGIQUES DES THÉORÈMES ARITHMÉTIQUES, SUIVI PAR DE NOMBREUX AUTRES TRAVAUX, (EN PARTICULIER CEUX DE RUSSEL-WHITEHEAD DANS LES PRINCIPIA MATHEMATICA). À PARTIR DES TRAVAUX DE FREGE ET DE RUSSEL, ON DISPOSE D'UNE NOTION PRÉCISE DE SYSTÈME FORMEL, C'EST-À-DIRE D'UN ENSEMBLE D'AXIOMES ET DE RÈGLES DE DÉDUCTION. LES ÉLÉMENTS DE CES SYSTÈMES SONT "FORMELS" EN CE SENS QU'ILS NE SONT MANIPULÉS QUE CONFORMÉMENT AUX RÈGLES RÉGISSANT CHAQUE TYPE DE SYMBOLE. SEULE LA FORME DES SYMBOLES EST PRISE EN COMPTE PENDANT LE CALCUL. CELA NE VEUT PAS DIRE QUE LES SYMBOLES NE PUISSENT RECEVOIR UNE INTERPRÉTATION, C'EST-À-DIRE UN SENS. MAIS UNE FOIS UN SENS ATTRIBUÉ AUX TERMES PRIMITIFS, IL N'A PLUS D'INTERVENTION NOUVELLE EN COURS DE CALCUL. L'IDÉE DE SYSTÈME FORMEL RÉALISE AINSI L'IDÉAL LEIBNIZIEN D'UNE PENSÉE AVEUGLE, C'EST-À-DIRE D'UNE PENSÉE QUI NE PROGRESSE PAS PAR LA RÉFLEXION SUR DES CONTENUS, MAIS PROCÈDE AU MOYEN DE RÈGLES APPLIQUÉES À DES CONFIGURATIONS DE SIGNES.



CONDILLAC

mathématiques en direction de la cognition en général ? Certains travaux sur la calculabilité, par Alan Turing et Alonzo Church, ont permis de donner un sens mathématique précis à l'idée d'une machine capable d'effectuer automatiquement des calculs, aussi bien sur des nombres entiers que sur des symboles primitifs convenablement codés en termes numériques. Ainsi la **"machine de Turing"** est un système formel automatique, un procédé physique qui manipule automatiquement les éléments d'un système formel en suivant les règles de ce système. Le résultat essentiel de Church et Turing, dans la perspective de l'I.A., consiste à démontrer qu'il existe des machines de Turing dites "universelles", capables de simuler n'importe quelle autre machine de Turing, donc de calculer toute fonction calculable par une machine quelconque ou par l'esprit humain. L'hypothèse supplémentaire qui donne corps à la recherche en I.A., consiste à admettre que "penser c'est calculer".

Ce nouveau type de mécanisme ne table plus comme l'ancien sur l'universalité des lois du mouvement, mais sur l'universalité des capacités de calcul des certaines machines. Il s'accompagne d'un nouveau sens du mot de "machine".

L'efficacité de la machine n'est pas à proprement parler d'ordre physique. En effet, les propriétés de "la machine abstraite" (entendue comme description formelle des états successifs d'une machine de Turing) justifient plus le projet d'I.A. que celles de telle ou telle machine concrètement réalisée.

I.A. faible, I.A. forte

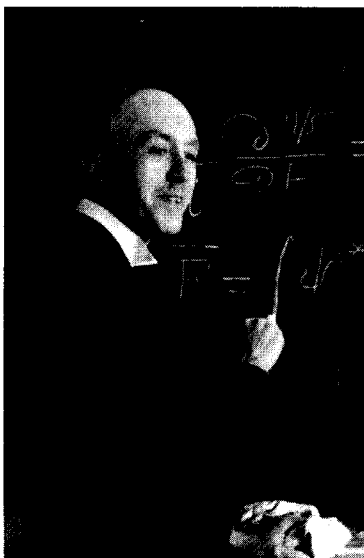
Pour comprendre la portée de la thèse de Church-Turing, revenons aux réactions à une hypothèse de type mécaniste ou matérialiste, en opposant les exemples de Condillac et de La Mettrie.

Condillac, dans le *Traité des Sensations*, s'est rendu célèbre en utilisant l'image d'une statue progressivement animée pour caractériser la genèse progressive du psychisme à partir des sensations. L'originalité de Condillac ne réside pas dans le recours à cette métaphore bien montrée par François Dagognet. Buffon l'avait lui-même utilisée, ainsi que beaucoup d'autres (3). Elle tient à la tentative de concilier deux thèses qui paraissent difficilement compatibles. D'une part, montrer que les animaux ne sont pas des machines, ce qui permet de sauvegarder *a fortiori* le caractère non-machinique du comportement humain. D'autre part, établir non moins fermement la différence entre l'homme et l'animal, différence dont on mesure aisément l'enjeu théologique. Le risque est grand en effet, concédant trop à l'animal, d'avoir des difficultés à caractériser la spécificité humaine, voire la dignité ontologique ou la supériorité supposée du règne humain face au règne animal.

Condillac considère donc que c'est l'existence de phénomènes cognitifs, chez l'animal comme chez l'homme, qui nous contraint à limiter la portée du mécanisme. Le langage mécaniste des "ébranlements" ne parvient pas en effet à rendre compte de

l'adéquation fine des comportements animaux, La connaissance - c'est-à-dire de façon générale les phénomènes mettant en jeu une "liaison des idées" -, marque pour Condillac la limite du mécanisable.

Cependant Buffon n'avait pas fait du mécanisme un usage réductionniste : il continuait à être dualiste.



C'est La Mettrie qui développa avec un certain courage la conséquence philosophique d'un mécanisme rigoureusement généralisé : le matérialisme ; non seulement au sens traditionnel selon lequel seule la matière peut avoir un rôle causal, mais au sens plus précis que le terme reçoit aujourd'hui dans les sciences cognitives : **est matérialiste une théorie qui considère que les états et les processus mentaux se réduisent à des états et des processus physiques.**

Le dilemme

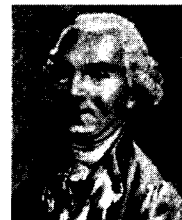
Les chercheurs en I.A. sont confrontés au même type de dilemme philosophique, à la notable différence près qu'il n'est plus question de soustraire l'ensemble de la cognition au domaine du mécanisable, dans la nouvelle acceptation de ce terme, mais simplement de limiter éventuellement la part mécanisable du psychisme.

1) Soit on cherche à donner l'hypothèse computationnelle une extension réduite aux secteurs de la pensée naturelle dans laquelle le calcul et "l'automatisation" suffisent à effectuer les tâches cognitives ; un peu dans l'esprit de Condillac, on admet donc qu'une mécanisation est possible, mais qu'elle est limitée au domaine de ce qu'on appelle l'intelligence opératoire et la pensée formelle, à l'exclusion de tout ce que l'on postule y échapper.

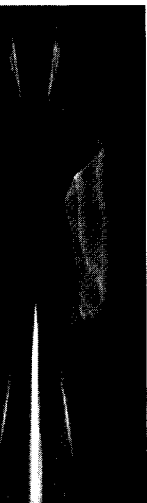
On peut alors, selon les cas, évoquer le rôle d'une intuition transcendant tout mécanisme, de l'expérience globale du monde telle qu'elle est à la fois donnée

LES THÈSES DE CONDILLAC

LA PREMIÈRE THÈSE CONSTITUE L'UN DES POINTS CENTRAUX DU TRAITÉ DES ANIMAUX : "IL EST IMPOSSIBLE DE CONCEVOIR QUE LE MÉCANISME PUISSE SEUL RÉGLER LES ACTIONS DES ANIMAUX". CERTES LEUR COMPORTEMENT PARAÎT AUTOMATIQUE, MAIS IL NE S'AGIT PAS D'UN AUTOMATISME MÉCANIQUE : C'EST UNE HABITUDE ACQUISE PAR L'EXERCICE D'UNE CERTAINE RÉFLEXION. "LA CONNAISSANCE, ÉCRIT-IL, EST ABSOLUMENT NÉCESSAIRE POUR RÉGLER L'ACTION MÊME DU SENS INTÉRIEUR, ET POUR DONNER AU CORPS DES MOUVEMENTS DIFFÉRENTS, SUIVANT LA DIFFÉRENCE DES CIRCONSTANCES" (CH. 4). CONDILLAC ÉTABLIT AINSI LE CARACTÈRE *A FORTIORI* NON MÉCANIQUE DU COMPORTEMENT HUMAIN : LES ANIMAUX NE SONT PAS DE SIMPLES MACHINES, LA STATUE HUMAINE N'EST PAS UN SIMPLE ROBOT. QUANT À LA SECONDE THÈSE, CONDILLAC L'APPUIE SUR LA MANIÈRE DIFFÉRENTE DONT L'ORGANISME HUMAIN ET L'ORGANISME ANIMAL PROGRESSENT DANS L'ONTOGENÈSE : LE PREMIER NAIT PLUS DÉMUNI MAIS PAR LA-MÊME PERFECTIBLE, TANDIS QUE LE SECOND, MIEUX ARMÉ ET PLUS RAPIDEMENT PARVENU À SON POINT DE DÉVELOPPEMENT MAXIMAL ET À LA PLEINE SATISFACTION DE SES BESOINS, N'EST EN REVANCHE QUE MÉDIOCREMENT MALLÉABLE.



BUFFON



et produite par le corps, et de l'intentionnalité, sorte de fluide biodynamique qu'aucune explication mécaniste ne peut adéquatement simuler.

2) Soit on radicalise l'hypothèse computationnelle à la façon de La Mettrie : l'I.A. dans sa version forte vise à simuler l'ensemble des comportements dits intelligents, et tend à considérer que l'homme, l'animal et l'ordinateur sont à certains égards, mais cruciallement, comparables dans leurs capacités de raisonner.

Certains chercheurs ont estimé que la tâche de l'I.A. devait rester très proche de la simulation de la cognition humaine. C'est par exemple le cas de Marvin Minsky et de Roger Schank. D'autres ont pensé que la généralisation à la machine du concept d'intelligence suggérait que l'on dégageât le concept de ses conditions humaines de réalisation ; dans cette optique, l'humain peut utilement être négligé comme un facteur relativement spécifique, mais non nécessaire, du comportement intelligent en général. Une telle radicalisation, présente dès l'origine dans les travaux d'I.A., n'a pas les mêmes implications dans le cas du mécanisme et dans celui du computationalisme. Chez La Mettrie, ce sont les progrès à venir de la science physique, en particulier de la mécanique newtonienne, et plus encore, de la physiologie et de la médecine, qui paraissent susceptibles de fournir une explication de type déterministe (conformément au paradigme concerné) du développement psycho-physiologique humain. Le corps humain est perçu comme machinique dans sa totalité parce que ce sont les lois du mouvement, appliquées aux constituants organiques, qui en expliquent l'organisation. Le détail de la preuve n'est pas à la charge du philosophe. Tandis que les chercheurs en I.A. ne peuvent se contenter durablement d'invoquer des résultats à venir. Il leur faut convertir l'hypothèse en programmes ou, à défaut, en esquisses de programmes.

I.A. forte, logicisme et naturalisme

Deux approches sont actuellement en compétition : l'approche rationaliste et l'approche naturaliste. Les deux approches, soulignons-le, ont ceci de commun qu'elles cherchent à enquêter sur les conditions les plus générales auxquelles on peut parler d'intelligence. Mais deux types distincts de réponses sont possibles, selon la manière dont on comprend la généralité même de la question. Les uns, comme McCarthy, Newell et Simon, impressionnés par le progrès de la technologie informatique, souhaitent développer l'hypothèse cognitiviste en laissant le champ largement ouvert à la reconstruction rationnelle des comportements intelligents. Les autres, tels que Minsky et Schank, estiment en revanche que l'I.A. doit rester étroitement liée à la psychologie, et développer des programmes qui soient en corrélation avec les procédures de la psychologie humaine.

Les partisans de la reconstruction rationnelle, héritiers en ligne directe d'une philosophie de la logique que composent les influences du logicisme de Frege et de Russell et du formalisme de Hilbert, considèrent qu'il convient de donner à la question



Photo B. Faye. Museum d'Histoire Naturelle

toute son universalité, en analysant les conditions formelles *a priori* de tout comportement intelligent. L'avantage de cette approche est que la question posée - "A quelles conditions un système peut-il être dit "intelligent" - reçoit une réponse exhaustive, fondée et donc capable d'orienter la recherche directement dans la direction la plus fructueuse.

Cette façon de procéder n'est pas sans analogie avec la démarche de Kant dans *La critique de la raison pure* : la démarche, dans les deux cas, souhaite dépasser la question de fait - "*comment parvient-on en fait à telle connaissance*", "*décrire tel ou tel comportement intelligent*" -, pour remonter non pas à des conditions plus générales, mais au *système* de ces conditions. Des conditions qui soient *a priori* : en premier lieu indépendantes de l'expérience quoique conditionnant la constitution de l'expérience et deuxièmement nécessaires et suffisantes : tout être humain ou non, pourvu qu'il soit, respectivement, capable de connaître, ou capable d'intelligence, doit mettre en œuvre les moyens *a priori* recherchés, et réciproquement.

C'est l'un des fondateurs de l'Intelligence Artificielle, Allen Newell, qui a développé dans cette direction la réponse la plus détaillée. Evoquons deux aspects essentiels de sa philosophie, centrés sur les concepts clés de *système symbolique physique* et d'*architecture*.

L'idée de système symbolique physique

Le support du comportement intelligent, pour Newell, n'est pas, on l'a vu, l'homme concret, mais l'élément constitutif des capacités opératoires d'un organisme ou d'une machine, donc selon l'hypothèse computationnaliste, le système symbolique physiquement instancié. L'essentiel pour un système symbolique, réside dans les propriétés formelles qui l'habilitent à permettre des calculs. Il doit comporter des *règles* qui déterminent le sens opératoire des symboles ; les éléments constituants doivent pouvoir être ramenés par décomposition à une liste finie de termes primitifs. Enfin, les règles de construction des expressions doivent expliciter les associations légitimes de symboles.

Pour être causalement efficace dans le monde, et ne pas rester une simple "machine abstraite", le système symbolique doit être "physique" : physiquement réalisé dans un matériau quelconque pourvu qu'il instancie les structures formelles de manière adéquate et stable, résistant à l'usure (tant dans la mémorisation que dans l'inférence).

On remarque dans cette approche le vœu de concilier le caractère relativement empirique de la recherche dans son aspect technologique avec la dimension quasi-déductive du fondement en termes de systèmes symboliques physiques. Du concept de système symbolique, tel qu'il est par exemple mis en œuvre dans le langage informatique LISP, on peut en effet dériver/justifier les conditions nécessaires et suffisantes de l'intelligence. Le propos que j'ai nommé "rationaliste" du fait de son souci de parvenir à des conditions nécessaires et suffisantes qui aient la dignité de conditions formelles *a priori*, va de pair avec la prédilection pour les langages de programmation "logiques" par rapport aux langages faisant plus libéralement appel à des représentations.

L'architecture

Un niveau consiste dans un *medium* qui forme le matériau dans lequel s'effectue l'activité caractéristique du niveau (courant électrique d'un certain voltage, au niveau du circuit, bits au niveau logique, ou vecteurs de bits au niveau du transfert de registres, des expressions symboliques au niveau symbolique, des ensembles de données au niveau de la configuration), dans des *composants* qui constituent les termes primitifs du calcul, des *lois de composition* qui permettent d'assembler les composants, et des *lois de comportement* qui déterminent la manière dont le comportement du système dépend du comportement des composants et de la structure du système. La notion d'architecture ainsi comprise est à la base de l'hypothèse de l'I.A. forte.

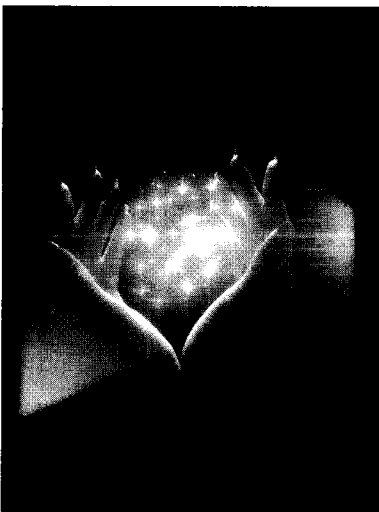
Elle confirme en effet le bien-fondé de la thèse fonctionnaliste : un niveau est *physiquement réalisable* de diverses manières. Ce n'est pas la nature physique du medium qui détermine la performance correcte d'un niveau donné, même si des différences de commodité ou de fiabilité peuvent plaider en faveur de tel ou tel support. Ainsi, on admet que la différence de composition physico-chimique entre le cerveau humain ou animal et l'ordinateur ne suffit pas à réserver au seul cerveau la capacité d'être à l'origine d'états mentaux. Il est aussi légitime d'attribuer de l'intentionnalité à un système automatique ou à un système humain si l'on reconnaît que, dans les deux cas, ce n'est pas l'instanciation physique qui détermine le comportement d'un certain niveau, mais des structures, des lois de composition opérant sur des constituants particuliers à ce niveau.

Deux caractéristiques des niveaux sont pertinentes pour l'idée d'intelligence générale. D'abord, un niveau déterminé peut toujours être décrit et compris de façon autonome, donc sans faire référence à d'autre niveau que lui. Mais par ailleurs on peut toujours réduire un niveau au niveau immédiatement inférieur. On retrouve dans ces deux propriétés cer-

tains aspects des rapports entre le physique et le mental auxquels les philosophes ont été diversement sensibles : que l'on puisse réduire chaque niveau au niveau inférieur n'implique pas qu'un niveau n'ait pas son propre type d'intelligibilité et d'autonomie organisationnelle.

En vertu de cette conception de l'architecture, on peut donc attribuer à un système informatique des états mentaux comme des croyances, des attentes, etc. Newell propose d'appeler "niveau de la connaissance" le niveau supérieur de l'architecture des systèmes d'I.A. Les composants en sont les *buts*, les *actions* et le *corps* qui constituent ensemble un *agent*. Le medium de ce niveau est précisément la *connaissance*. A ce niveau, l'agent traite ses connaissances pour décider de ses actions, la loi de comportement étant ici le *principe de rationalité* (les actions sont choisies en sorte de favoriser l'aboutissement des buts).

Une illustration "concrète" de ce type d'orientation de la recherche est fournie par SOAR, de Laird, Newell et Rosenbloom, présenté dans I.A. en 1987. Soar est un système complexe dont l'objectif explicite est de fournir "l'architecture de l'intelligence générale" ; en d'autres termes, de montrer comment un système est capable d'accomplir toute la gamme des tâches cognitives (3), qu'il s'agisse de travaux de routine ou de problèmes ouverts, d'employer toute la panoplie des méthodes de résolution de problèmes et des représentations corrélatives appropriées, et enfin d'apprendre tout ce qui concerne les divers aspects des tâches, et ce, en utilisant les moyens les plus universels possibles. L'une des originalités de SOAR réside en effet dans l'effort d'"intégrer" l'architecture, et en particulier de faciliter la



Bibliographie

Buffon, Georges-Louis de, **Histoire naturelle**, Paris, ed. Lacépède, 1818.

Condillac, Etienne Bonnot de, **Le Traité des animaux**, introduction par François Dagognet, Paris, Vrin, 1987.

Condillac, Etienne Bonnot de, **Le Traité des sensations**, Paris, Fayard, 1984.

Dagognet, François, "L'animal selon Condillac", introduction au **Traité des Animaux** de Condillac, Paris, Vrin, 1987.

Israel, David, "A short companion to the Naive Physics Manifesto", in **Formal Theories of the commonsense world**, édité par Jerry R. Hobbs et Robert C. Moore, Norwood, Ablex, 1985.

La Mettrie, Julien Offroy de, **L'homme-machine**, Introd. par P.-L. Assoun, Paris, Denoël, 1981.

La Mettrie, Julien Offroy de, **Œuvres Philosophiques**, Paris, Fayard, 1987.

Minsky, Marvin, "A framework for representing knowledge", in **The Psychology of computer vision**, édité par Patrick H. Winston, New York, Mac graw Hill, 1975; reproduit in **Mind design**, édité par J. Haugeland, Cambridge, MIT Press, 1981, 95-128.

Minsky, Marvin, **The society of mind**, New York, Simon and Schuster, 1985. Trad. française par Jacqueline Henry : **La société de l'esprit**, Paris, Inter éditions, 1988.

Newell, Allen, "Physical symbol systems", **Cognitive Science**, 1980,4, 135-183.

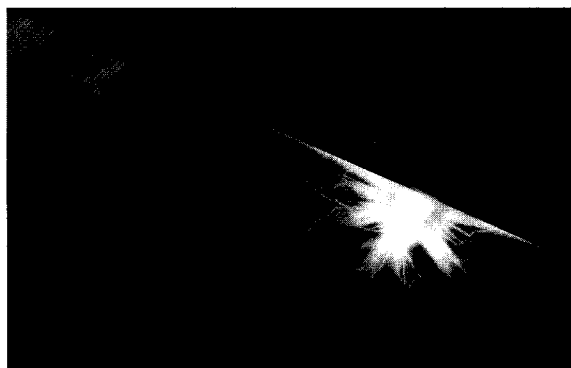
Newell, Allen, "The knowledge level", **Artificial Intelligence**, 1982, 18, 87-127.

Newell, Allen et Simon, Herbert A., "Gomputer Science as empirical inquiry : symbols and search", in **Mind Design**, 1981, 35-66.

communication entre représentation et inférence : la représentation des connaissances est aussi peu "pré-organisée" que possible en domaines disjoints. L'ensemble des connaissances ont une seule et même représentation, sous forme déclarative. L'ensemble des buts est également représenté sous une forme unique, qui est celle des "espaces de problèmes". Soar montre en outre que la production de buts peut se faire de manière automatique, grâce à la façon dont l'architecture est organisée.

L'I.A. naturaliste

Cette seconde façon de comprendre le projet de l'I.A. se distingue de l'I.A. forte de la "reconstruction". Elle tente de restreindre la question des conditions les plus générales de l'intelligence au domaine des faits, en refusant par conséquent d'engager la question vers des conditions générales *a priori* qui seraient indépendantes de la cognition humaine. Cette position, développée dans Minsky (1975), vise principalement l'école logiciste, celle de la reconstruction rationnelle avant qu'elle fasse intervenir les structures appelées "espaces de problèmes". Minsky reproche surtout à cette école de privilégier des considérations formelles sur les exigences liées à l'heuristique. Ainsi, il ne croit pas à la possibilité d'appliquer à des systèmes cognitifs réels, comme l'est l'esprit humain, la distinction nettement faite dans cette école entre, d'une part, les connaissances, et d'autre part, les structures de connaissances. Pour



reprendre les termes de David Israel, "*on ne peut pas dissocier - dans l'esprit - ses connaissances et la manière dont les connaissances sont intégrées à son organisation adaptative globale et à ses principes de fonctionnement*" (Israel, 1985, 440).

Cette idée de structurer les connaissances en fonction du mode d'acquisition et du mode de mobilisation de celles-ci a inspiré à Minsky le concept de "frame" : un cadre cognitif dans lequel s'organisent les concepts subordonnés à un certain but. Elle a été ultérieurement développée par Schank au moyen d'une hiérarchie de structures organisant la mémoire selon des fins. Elle a été si fructueuse que les tenants de la "reconstruction rationnelle" s'en sont inspirés à leur tour, par exemple avec l' "espace de problème" chez Newell.

Un travail plus récent de Minsky, exposé dans *La société de l'esprit*, reprend l'idée de structures à la fois associatives et hiérarchisées, les "K-lines". Philosophiquement importante, elle rapporte autant que possible la structure de la connaissance aux circonstances dynamiques de son acquisition et de sa modification.

Version logiciste et version naturaliste de l'I.A. ont aujourd'hui tendance à fusionner, devant les problèmes rencontrés par la formalisation de la connaissance de sens commun (cf. Israel, 1985). Soulignons que les deux versions de l'I.A. forte défendent une conception matérialiste du mental. Comme dans le matérialisme d'un La Mettrie, la position défendue est moniste : l'homme est un être physique, et toutes ses propriétés, même les plus remarquables, comme l'intelligence ou la volition, doivent pouvoir s'expliquer physiquement. Mais ce en quoi il se distingue du matérialisme à la façon de La Mettrie, consiste à ce que seule la description symbolique des comportements humains lui paraît en mesure de manifester le caractère opératoire, donc computationnel, de l'intelligence.



Comment se passer du nucléaire ? En relançant les énergies renouvelables comme dans tous les autres pays d'Europe. Utopique ? Les exemples concrets sont là pour prouver le potentiel des énergies renouvelables.

56 pages - 50 photos - 30 F franco de port

Commande et règlement :

Numéro Hors-Série

Silence, 4 rue Bodin 69001 Lyon

1. Histoire des animaux. Comparaison des animaux et des végétaux, in Histoire naturelle, t. IV, p. 423 ; cité in (Dagognet, 1987) p. 59

2. Cf. (Dagognet 1987), p.59 sqq.

3. Soar "sait" par exemple comparer des formes et des objets, former des syllogismes élémentaires, chercher les racines d'équations du quatrième degré. Il contient à titre de composantes un certain nombre des systèmes experts les plus performants, entre autre MYCIN, DESIGNER, et même une partie du système R1 qui configure les ordinateur VAX.