

Intelligence artificielle et signification. À propos des limites et des possibilités des sciences cognitives

W. P. Mendonça

Volume 17, numéro 1, printemps 1990

URI : <https://id.erudit.org/iderudit/027102ar>

DOI : <https://doi.org/10.7202/027102ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Mendonça, W. P. (1990). Intelligence artificielle et signification. À propos des limites et des possibilités des sciences cognitives. *Philosophiques*, 17(1), 3–19. <https://doi.org/10.7202/027102ar>

Résumé de l'article

L'auteur distingue, dans les travaux sur l'intelligence artificielle, deux approches : l'approche technologique et l'approche cognitive. Il montre que les rapprochements faits, dans l'approche cognitive, entre l'intelligence humaine et l'intelligence artificielle, ne vont pas de soi, et que les thèses sur l'intelligence artificielle sont largement tributaires de certaines spéculations rationalistes et empiristes de la philosophie classique. Il expose la principale difficulté que rencontre alors une compréhension de l'intelligence humaine à partir de l'approche cognitive, à savoir la nécessité d'oblitérer la dimension sémantique et de professer un solipsisme radical, et il plaide en faveur d'une compréhension de l'intelligence humaine selon ce qu'elle est dans la culture plutôt que dans les machines.

ARTICLES

INTELLIGENCE ARTIFICIELLE
ET SIGNIFICATION.
À PROPOS DES LIMITES
ET DES POSSIBILITÉS
DES SCIENCES COGNITIVES

par W.P. Mendonça

RÉSUMÉ. L'auteur distingue, dans les travaux sur l'intelligence artificielle, deux approches : l'approche technologique et l'approche cognitive. Il montre que les rapprochements faits, dans l'approche cognitive, entre l'intelligence humaine et l'intelligence artificielle, ne vont pas de soi, et que les thèses sur l'intelligence artificielle sont largement tributaires de certaines spéculations rationalistes et empiristes de la philosophie classique. Il expose la principale difficulté que rencontre alors une compréhension de l'intelligence humaine à partir de l'approche cognitive, à savoir la nécessité d'oblitérer la dimension sémantique et de professer un solipsisme radical, et il plaide en faveur d'une compréhension de l'intelligence humaine selon ce qu'elle est dans la culture plutôt que dans les machines.

ABSTRACT. In the literature on artificial intelligence, the author distinguishes two approaches : the technological approach and the cognitivist approach. He shows that the parallels made between human intelligence and artificial intelligence within the cognitivist approach do not go without saying, and that the theses on artificial intelligence are largely dependent upon certain rationalist and empiricist speculations of classical philosophy. He exposes the principal problem that then encounters an understanding of human intelligence in the light of the cognitivist approach, namely the necessity to obliterate the semantic dimension and to profess a radical solipsism, and he pleads for an understanding of human intelligence commensurable with what it is in culture rather than in machines.

I

On constate que les philosophes adoptent des positions parfois assez divergentes en ce qui concerne la science de l'intelligence artificielle. Mais, en règle générale, ils s'intéressent uniquement à la branche « cognitiviste » et non pas à la branche simplement technique de cette science nouvelle. On peut notamment s'occuper de l'intelligence artificielle dans la perspective d'une *finalité technique*, par exemple, en construisant un reconnaiseur automatique de voix ou en programmant une machine qui, d'après certaines données, fournit des diagnostics médicaux. Ne serait-ce qu'à cause de leur compétence, qui laisse peut-être à désirer, les philosophes n'ont rien à dire à propos de la recherche sur l'intelligence artificielle *dans cette perspective*. Cependant, la considération de l'intelligence artificielle comme science technique, comme *science de l'informatique*, ne semble pas tenir compte des « intérêts surtout psychologiques des chercheurs de l'intelligence artificielle ». Ceux-ci se fixent davantage comme tâche « le développement d'une théorie systématique des processus intellectuels ». ¹ Dans cette perspective, ordinateurs et programmes sont importants dans la mesure où une telle théorie, selon une opinion très répandue, doit avoir *la forme d'un programme d'ordinateur*.

C'est au moment où l'on fait abstraction de sa nature spécifiquement technologique et qu'on lui attribue un rôle de premier ordre dans la tentative de comprendre et d'expliquer l'activité humaine intelligente que l'intelligence artificielle a, dans un sens qui reste à expliquer, de l'importance pour l'anthropologie philosophique. En effet, l'intelligence artificielle est couramment utilisée pour fonder une anthropologie matérialiste. Dans ces circonstances, elle est souvent associée à l'idée que la compréhension de la manière dont un être humain fait quelque chose « intelligemment » peut être atteinte par la construction d'une machine qui puisse le remplacer. En outre, les défenseurs de l'interprétation « cognitiviste », c'est-à-dire non simplement technologique de l'intelligence artificielle, opèrent sous la supposition que l'identification théorique entre l'homme et un système de traitement de l'information est une hypothèse scientifique qui peut résoudre à satisfaction quelques

1. BODEN, Margaret, *Artificial Intelligence and Natural Man*. Hassocks : The Harvester Press, 1977, p. 4.

problèmes classiques formulés originellement à l'intérieur de la tradition philosophique post-cartésienne. Ainsi, l'intelligence artificielle, dans son interprétation cognitiviste, est conçue comme la nouvelle forme de la psychologie rationnelle, comme une *top-down psychology* qui cherche à formuler des explications scientifiques du comportement humain *en tant que* comportement intelligent.

Depuis le célèbre article de Alan Turing, « Computing Machinery and Intelligence »², ce but cognitiviste de la recherche dans le domaine de l'intelligence artificielle est propagé par l'idée qu'un ordinateur programmé adéquatement est en fait un être intelligent, doué d'esprit, si bien que l'on peut dire littéralement en parlant de lui qu'il perçoit les choses autour de lui, qu'il comprend le langage articulé, qu'il est intelligent et qu'il possède également les états cognitifs nécessaires que nous pouvons constater chez les êtres humains. Les programmes d'ordinateur seraient dans ce cas les explications psychologiques « profondes » des facultés intelligentes.

La conception selon laquelle nous pouvons construire des systèmes d'élaboration de l'information qui font preuve d'intelligence générale, ou bien à l'inverse, que chaque être intelligent, pourvu qu'il soit soumis à une analyse « plus précise », s'avère être un tel système, est souvent traitée dans la littérature courante comme une *thèse empirique*, et ceci non seulement par les défenseurs *mais aussi* par les critiques du projet cognitiviste de l'intelligence artificielle. Les premiers affirment souvent que cette thèse, dans ses traits essentiels (ou même dans sa totalité), peut être considérée comme étant suffisamment confirmée par le développement de la recherche actuelle dans les laboratoires de l'intelligence artificielle. L'exemple typique dans ce cas est Marvin Minsky.³ Par contre les critiques, comme Hubert Dreyfus,⁴ font remarquer que les programmes existants réellement ne justifient nullement l'enthousiasme avec lequel ils sont présentés. Considérés avec sobriété, ces programmes restent hautement limités dans l'extension de leur capacité et on a trop rapidement prétendu à leur sujet qu'ils sont intelligents dans un sens non métaphorique.

2. *Mind*, 59 (1950), 434-450.

3. MINSKY, M. *Semantic Information Processing*. Cambridge, Massachusetts : The MIT Press, 1968. Voir aussi Minsky, M. *The Society of Mind*. New York : Simon and Schuster, 1985.

4. DREYFUS, H. *What Computers Can't Do*. New York : Harper Colophon Books, 1979.

En ce qui concerne le statut de la thèse mentionnée, le consensus semble dominer largement. La question concernant l'existence possible d'un système de traitement de l'information qui, en considération de l'intelligence, montre le même comportement *input-output* que l'homme, serait donc une question empirique dont la solution ne serait possible que par des expérimentations.

Cette problématique nécessite sans doute une *clarification linguistique critique*. Car en fait, il ne va pas de soi que l'esprit, comme John Haugeland formule la thèse en question,⁵ soit un système d'élaboration de l'information, ou bien que nous soyons ne serait-ce qu'à un niveau « abstrait » — des « ordinateurs naturels ». Le simple fait que, pour évaluer « le degré d'intelligence » des machines, nous prenions pour base notre conception de l'intelligence des personnes, montre à l'évidence qu'il faut d'abord que nous fassions la lumière sur l'intelligence humaine, afin de pouvoir enfin comprendre ce que cela *signifie* de dire : l'ordinateur *fait* quelque chose, *résout* un problème d'une façon intelligente. Il ne s'agit plus dans ce cas de questions auxquelles on pourrait répondre d'une façon empirique, mais plutôt de la *compréhension* adéquate de la *logique* de notre démarche lors de la planification et de la construction des systèmes automatiques. Dans cette perspective, l'expression selon laquelle l'ordinateur pourrait montrer le même comportement *input-output* que l'être humain induit en erreur. Car *comment* interprétons-nous l'*output* d'un ordinateur ? Et *comment* interprétons-nous le comportement (les actions et les gestes) d'un être humain ? Pour répondre à ces questions — c'est ce que j'aimerais expliquer plus amplement — nous n'avons nullement besoin d'expérience scientifique et d'expérimentations : il suffit de recourir à notre expérience de vie et à notre faculté de jugement.

En prenant en considération les distinctions logiques (conceptuelles) d'importance à cet égard, il deviendra clair qu'une recherche significative dans le domaine de l'intelligence artificielle ne peut avoir essentiellement que des buts techniques. L'interprétation cognitiviste de l'intelligence artificielle comme modèle explicatif du « fonctionnement » de l'intelligence humaine, par contre, se

5. HAUGELAND, J. « The Nature and Plausibility of Cognitivism », *The Behavioral and Brain Sciences*, 2 (1978), 235-260.

montre problématique dès le départ. Elle remplace la compréhension de notre pratique de vie par une métaphore trop simple, plus ou moins impropre. De ce fait, quelques techniques créées par nous avec un propos bien défini sont hautement stylisées en « explications » métaphysiques des facultés de l'intelligence.

Les analyses suivantes veulent démontrer que quelques idées centrales sur l'intelligence artificielle sont chargées de *difficultés logiques* apparaissant déjà dans la théorie de la connaissance de la modernité (à la suite de Descartes et de Locke), et conduisant à des *conclusions relativistes et solipsistes*. On ne peut ensuite se soustraire à celles-ci qu'à l'aide de doctrines métaphysiques dogmatiques. La thèse principale est donc que certains développements théoriques dans le domaine de l'intelligence artificielle répètent quelque *spéculation rationalistes et empiristes de la philosophie classique*.

II

L'intelligence artificielle met, aujourd'hui, à la disposition du théoricien cognitiviste le vocabulaire et l'appareil conceptuel pour la description ou bien « l'analyse » des « processus cognitifs ». On entend par là, par exemple, les processus de la reconnaissance des formes sensibles et de la compréhension linguistique. La théorie cognitive actuelle conçoit ces processus comme des processus informatiques (*computational processes*). Elle essaie de ramener la signification des expressions avec lesquelles nous décrivons et évaluons le comportement intelligent dans le langage courant à des régularités qui servent plutôt à la description du *software* d'un ordinateur. C'est pourquoi Margaret Boden⁶ prétend que l'intelligence artificielle ne s'occupe pas essentiellement d'ordinateurs, mais plutôt de « l'intelligence dans la pensée et l'action ».

Cette thèse s'appuie sur l'idée que le comportement intelligent est le résultat d'un calcul exécuté d'après des règles et des principes rigides. La recherche sur l'intelligence artificielle s'efforce de découvrir ces règles. (Il faut que ce soit des règles *rigides* qui constituent une *procédure effective*, car seulement de telles règles peuvent être « suivies » par une machine.) En adoptant l'hypothèse

6. *Artificial Intelligence and Natural Man*, *op. cit.*, p. XI.

que les « processus de pensée » sont des processus d'élaboration de l'information ou des procédures informatiques, la *simulation* dans un ordinateur d'une faculté intelligente est en même temps son *explication*. Ainsi, dans le cadre de la sémantique procédurale,⁷ on prétend que la description d'un compilateur effectif pour une langue est la théorie sémantique qui correspond à la réalité psychologique de ses sujets parlants, si bien qu'un être humain peut saisir la signification des phrases de cette langue uniquement parce qu'il dispose « dans son stockage » d'une quantité effective de critères algorithmiques avec lesquels les conditions de vérité des phrases concernées peuvent être déterminées, quoique « inconsciemment ».

Cette conception aboutit à ce que des processus cognitifs (cela veut dire dans notre cas : comprendre, penser, percevoir etc.) soient conçus comme des opérations formelles (« syntaxiques ») avec des « représentations internes ». En effet, d'après Allen Newell et Herbert Simon,⁸ un système d'élaboration de l'information est un système symbolique physique (*a physical symbol system*), dans lequel des configurations physiques sont produites et transformées d'après des procédures algorithmiques. Ces procédures sont purement *configuratives*, c'est-à-dire, elles prennent en considération seulement la *forme interne* des structures à élaborer et font abstraction de leurs *interprétations externes* possibles. L'application du modèle de l'ordinateur à notre vie intelligente part donc du principe selon lequel nos « états mentaux » se développent à partir du *stimulus input* à travers des processus cognitifs, ces processus étant conçus comme des transformations de représentations internes (c'est-à-dire en dernier lieu : des configurations physiques réalisées physiologiquement). Ainsi, on essaie de montrer dans le cadre de la théorie cognitiviste de la perception comment des représentations internes de plus en plus complexes sont construites d'après des opérations formelles à partir des propriétés distinctives des images rétinienne, par exemple. Par conséquent,

7. Voir, par exemple, JOHNSON-LAIRD, P., « Procedural Semantics », *Cognition*, 5 (1977), 189-214.

8. NEWELL, A. et Simon, H. « Computer Science as Empirical Inquiry : Symbols and Search », *Communications of the Association for Computing Machinery*, 9 (1976), 113-126.

on met au même niveau la « dernière » représentation de cette construction avec la perception comprise et articulée (*percept*).

De ce fait, notre expérience du monde est rendue si subjective (« internalisée ») que la concordance de nos jugements courants de perception et leur relation avec le monde objectif deviennent une énigme impénétrable. À l'aide de la réflexion suivante, nous pouvons reconnaître que nous sommes dans ce cas en présence d'une difficulté qui se trouve au centre même de l'approche cognitiviste.

La tentative de décrire les « processus mentaux » d'après le modèle de l'ordinateur considère ces processus, comme nous l'avons vu, en tant que séquences de manipulation de représentations internes, dont les propriétés « sémantiques » (vérité, référence, signification) sont nécessairement négligées. C'est ce que Jerry Fodor⁹ appelle *formality condition*. Selon ce modèle les terminaisons des organes des sens se chargent de l'unique liaison entre la première représentation interne et le « monde extérieur » — tout le reste se déroule « intérieurement ». Dans le cadre de l'analogie avec l'ordinateur, on compare les terminaisons des organes des sens avec des transducteurs par lesquels les événements du « monde extérieur » sont enregistrés pour être traités. Quand on travaille avec une machine de Turing, cela correspond à l'impression de certains signes sur la bande d'entrée. Le système de traitement de l'information lui-même (en l'occurrence : l'esprit) n'a pas le moindre contrôle sur la relation (sémantique) entre les données enregistrées et le monde « au-delà » des transducteurs (ou des terminaisons des organes des sens). Son accès au monde s'effectue à *travers* des données livrées par les transducteurs (à travers des « stimuli proximaux »). Dans cette mesure, le système ne peut pas non plus « juger » s'il existe réellement un monde « au-delà » des transducteurs. De toute façon, la question concernant ce « monde extérieur » est sans importance pour la « vie intérieure » du système de traitement de l'information. Pour pouvoir effectuer ses processus, celui-ci n'a besoin que de quelques données dans sa « mémoire » (quelques signes dans la bande d'entrée (*input-tape*)). Savoir comment ces données sont parvenues jusque-là et quelles relations

9. FODOR, J. « Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology », In Fodor, J. *Representations*. Brighton : The Harvester Press, 1981.

elles peuvent avoir avec les « événements extérieurs », n'a aucune importance dans ce contexte. (Le meilleur exemple de cela est peut-être le célèbre programme SHRDLU¹⁰ qui opère avec des « objets » qui existent uniquement dans son « monde intérieur ». Les « propriétés » des « objets » « manipulés » par SHRDLU sont toutes spécifiées directement par Terry Winograd.)

Les implications relativistes et sceptiques de cette position sont évidentes. En fait, une théorie qui procède strictement selon la *formality condition* ne peut pas donner de réponse à la question concernant les propriétés sémantiques des représentations internes dont elle décrit la configuration formelle (syntaxique). Elle peut définir les états et processus mentaux uniquement sur la base des représentations internes, sans prendre en considération le fait qu'elles peuvent être des représentations de *quelque chose d'extérieur*. « The formality condition prohibits taxonomising psychological states by reference to the semantic properties of mental representations. »¹¹ Cela correspond, comme Fodor le sait d'ailleurs, exactement au projet épistémologique toujours entrepris (et toujours échoué) depuis Locke, c'est-à-dire, le projet qui consiste à réduire nos perceptions à des sensations ou à des impressions. On remplace simplement « l'association » des empiristes (Hume, par exemple) par des « processus d'élaboration de l'information », les « idées » par des « représentations ». Quiconque ayant des connaissances historiques pertinentes peut reconnaître tout d'abord que cela conduit à un relativisme et à un solipsisme à l'égard de la perception.¹² Toutes les tentatives de mener à terme ce projet épistémologique ont notoirement abouti à l'impasse d'un *ignoramus* total.

10. WINOGRAD, T. « A Procedural Model of Language Understanding ». In Schank, R. et Colby, K. (eds.) *Computer Models of Thought and Language*. San Francisco : W. H. Freeman, 1973.

11. FODOR, J. « Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology », *op. cit.*, p. 239.

12. FODOR (« Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology », *op. cit.*, p. 242) en a conscience. Il fait remarquer que l'attachement à la *formality condition* mène au solipsisme méthodologique, si bien que celui-ci devient la seule stratégie pour la psychologie cognitive, si elle veut encore se présenter comme une théorie représentationnaliste de la perception.

L'essentiel de l'argumentation présentée jusqu'ici vise à mettre en relief que les programmes d'ordinateurs ne peuvent pas déterminer *à eux seuls* les relations entre les symboles manipulés et leurs significations.¹³ La supposition qui constitue l'objet de recherche de la branche cognitiviste de l'intelligence artificielle, à savoir, qu'on considère des processus mentaux comme des opérations formelles avec des symboles, a justement pour conséquence que les théories psychologiques obtenues de cette façon et représentées sous forme de programmes d'ordinateurs ou de procédures informatiques ne peuvent pas répondre à la question concernant l'attribution des significations aux symboles manipulés. *Le programme ne nous dit pas comment interpréter les symboles, comment leur attribuer une signification.* Cela doit être fait plutôt par le programmeur pour ainsi dire « de l'extérieur ». En outre, ceci est également valable pour le cas, induisant particulièrement en erreur, dans lequel les symboles manipulés *ressemblent* à des mots. L'emploi (externe) de ces « mots » et non pas leur traitement (interne au système) leur donne une signification.

Ainsi, il en résulte que l'algorithme lui-même ne peut pas fixer une fois pour toutes les significations des symboles qu'il contient. L'ordinateur n'opère pas avec des symboles qui ont un contenu sémantique en soi. Un programme d'ordinateur décrit purement le rapport formel des représentations internes, détermine, tout au plus, leurs significations *internes*. C'est-à-dire : les significations externes de ces représentations ne sont pas encore suffisamment déterminées par la structure logique que le programme décrit, mais seulement par une interprétation que nous devons faire en tant qu'*usagers externes*. Dans ce sens, la signification des symboles élaborés par l'ordinateur est toujours secondaire, dérivée, relative à nos intérêts et à nos critères globaux de signification ou d'importance pratique. C'est *de cette façon* qu'on interprète l'*output* d'un ordinateur. C'est seulement en vertu d'une coordination externe au système que les structures symboliques reçoivent une

13. Des systèmes formels et, par conséquent, des programmes d'ordinateur imposent, évidemment, quelques restrictions à l'interprétation de leurs éléments : toute interprétation n'est pas possible. Cette détermination préalable du domaine des modèles interprétatifs nous permet de parler, avec pertinence, de significations formelles. Cependant, il est vrai que les systèmes formels, aussi restreints soient-ils, admettent une multiplicité de modèles dont la sélection définitive tient, en dernier lieu, à ceux qui utilisent de tels systèmes.

signification. Uniquement *par cela* elles deviennent des représentations (internes) de quelque chose (d'externe). Pour atteindre ce but, les théories et les modèles de traitement de l'information *ne* suffisent *pas*.

Les modèles basés sur l'ordinateur ne peuvent donc pas nous *expliquer* ce que signifie percevoir ou comprendre. La compréhension de tels modèles *présuppose* plutôt que nous sachions déjà, par exemple, ce qu'est la perception et comment ses caractéristiques principales doivent être décrites. Finalement, quand il s'agit de faire la coordination qui confère des significations aux représentations internes, nous dépendons de l'exercice des habilités acquises de la perception. L'usage des modèles basés sur l'ordinateur *présuppose* aussi la faculté de la reconnaissance certaine des symboles utilisés et la compréhension correcte, acquise et exercée en commun, de l'emploi des règles algorithmiques.

En outre, notre perception n'est pas une représentation *en* nous-mêmes qui pourrait être réduite, du point de vue formel, à des processus physiques se déroulant sur la rétine ou dans le tympan. Il est possible qu'une telle représentation, réalisée de façon organique, lui corresponde. Mais on ne peut parler d'une *perception* que si, au-delà de la structure formelle de cette représentation, sa signification (c'est-à-dire l'état de choses qu'elle représente [ou même *doit* représenter] d'après une coordination *effectuée dans une pratique culturelle normative*) est prise en considération. Voir quelque chose (comme une chose), par exemple, n'est pas un processus *en* soi-même, mais je ne peux voir que ce qui est objectivement à voir et peut également être vu par d'autres, qui ont acquis les concepts pertinents pour l'articulation significative du monde visible ; sinon, j'ai tout au plus des hallucinations. Et même l'objet de l'hallucination (en tant que *type*) doit pouvoir être perçu, en principe, par d'autres.

L'exécution *cohérente* du programme cognitiviste de l'intelligence artificielle implique donc l'abandon de la dimension *sémantique*, de l'interrogation sur la signification des représentations internes.¹⁴ Dans ce cas, il est indispensable de prendre le point de

14. Voir SEARLE, J. « Minds, Brains, and Programs », *The Behavioral and Brain Sciences*, 3 (1980), 417-457.

vue du solipsisme méthodologique. C'est-à-dire : il faut essayer de se tirer d'affaire sans significations, de façon purement « syntaxique », en *identifiant* les états mentaux avec les rapports formels des représentations internes. Cependant, ceci implique la *dissociation conceptuelle* de la perception et de la compréhension des significations auxquelles elles ont *conventionnellement* été associées, et donc également de leur *base épistémique* (ou « critérielle ») *publique*. La perception et la compréhension deviennent ainsi des *processus intérieurs* par lesquels une configuration de stimulation donnée est transformée, de façon « syntaxique », en d'autres configurations, internes celles-là. Percevoir et comprendre ne sont plus conçus dans la perspective de l'appréhension de quelque chose d'objectif, linguistiquement articulable, mais, au contraire, en tant que processus dont il faut chercher les critères uniquement *à l'intérieur d'un sujet*. Au mieux, on peut considérer ces critères seulement du point de vue d'un *observateur*, non sous la perspective d'un *participant* à une pratique commune. (Cependant, comme le soulignent Donald Davidson et Hilary Putnam,¹⁵ c'est précisément la perspective *holiste, impossible à formaliser*, du participant qui s'avère indispensable si nous voulons tenir compte de l'élément *normatif* de la rationalité qui est inscrit dans les significations des expressions psychologiques qui s'appliquent au comportement intelligent.)

L'interprétation cognitiviste de l'intelligence artificielle aboutit ici aux mêmes difficultés que la théorie représentationnaliste classique de l'esprit. Si notre perception n'est qu'une structure formelle *en nous-mêmes*, il ne nous reste qu'à *espérer*, en effet, qu'elle *représente* quelque chose de réel, d'objectif. Car une description de la construction formelle de la représentation interne, pour plus précise qu'elle soit, ne peut rien dire de ce rapport sémantique. En d'autres termes : si l'objet de l'expérience est construit selon le solipsisme méthodologique, il ne nous reste qu'à espérer qu'un objet réel lui corresponde dans le monde objectif, que nous nous entendions vraiment, etc. Dans ce cas, nous espérons, comme les classiques de

15. DAVIDSON, D. « Mental Events », In Forster, L. et Swanson, J. (eds.) *Experience and Theory*. Amherst ; The University of Massachusetts Press, 1970. Voir aussi DAVIDSON, D. « The Material Mind ». In Haugeland, J. (ed.) *Mind Design*. Cambridge, Massachusetts : The MIT Press, 1981 ; PUTNAM, H. « Computational Psychology and Interpretation Theory », In Born, R. (ed.) *Artificial Intelligence. The Case Against*. London : Croom Helm, 1987.

la théorie moderne de la connaissance, qu'il existe une harmonie préétablie entre le monde subjectif de nos représentations internes et le monde objectif des choses externes. (En réalité, cependant, nous n'abandonnons pas cette harmonie au hasard, mais nous essayons de l'établir en enseignant et en apprenant, c'est-à-dire, en agissant.)

La « réalité » de notre connaissance, la compréhension intersubjective et la possibilité (ou le *fait*) de la communication sont, pour les classiques de l'épistémologie, ou bien de vrais énigmes, dont la solution exige, comme chez Descartes, Malebranche, Locke, Berkeley et Pascal, la prise en considération de la bonté et de la sagesse divines, ou bien, et c'est le cas de Hume, des énigmes de la raison humaine qui se dissout, à ce point, en un scepticisme radical. Des représentants plus modernes du solipsisme méthodologique, par exemple, Ernst Mach, Bertrand Russell, Rudolf Carnap, essaient d'échapper aux difficultés solipsistes et sceptiques non plus à travers le recours à la grandeur divine, mais à travers l'observation emphatique que leur solipsisme n'est que « méthodologique ». La totalité de nos concepts et de nos expériences — ainsi l'affirme le solipsiste méthodologique de notre siècle — est une construction logique sur la base d'idées, d'impressions ou de données des sens (*sense data*) d'un *Moi*. Pour éviter un solipsisme tout court, le solipsiste méthodologique étend alors ses « privilèges » à n'importe qui. *N'importe qui* parmi nous pourrait, selon le solipsiste méthodologique, prendre la place de ce *Moi*, de sorte que nous serions *tous* des solipsistes méthodologiques.

Cependant, ceci occasionne une contradiction performative.¹⁶ Le *Moi* de mes expériences et de mes perceptions ne peut pas être, du point de vue solipsiste, au même niveau que l'Autre, à qui le solipsiste méthodologique veut aussi attribuer des expériences comme les miennes. La raison en est que l'Autre — si la théorie du solipsiste méthodologique est maintenue — ne peut être rien de plus qu'une construction à partir de *mes* expériences. En voulant admettre l'existence d'un Autre qui jouerait à l'intérieur de sa construction le même rôle que je joue dans *mon* système, le

16. Je reproduis ici, de manière abrégée, un argument contenu dans : PUTNAM, H. « Was ist Epistemologie ? » In Henrich, D. (ed.) *Kant oder Hegel. Stuttgarter Hegel-Kongress 1981*. Stuttgart : Klett-Cotta, 1983.

solipsiste méthodologique est amené à abandonner (à « transcender ») sa théorie car, d'après celle-ci, tout, y compris les autres et leurs « expériences », doit être réduit à une construction dans mon système. La reconnaissance, même implicite, de cette difficulté a dû être, d'ailleurs, la raison principale du recours à la sagesse et à la bonté de Dieu de la part des représentants classiques de l'épistémologie moderne.

En règle générale, les partisans de l'intelligence artificielle semblent simplement ignorer les problèmes posés par la doctrine du solipsisme méthodologique.¹⁷ Mais il s'avère que l'assimilation théorique entre les états mentaux et les rapports formels des représentations internes a pour conséquence que nous sommes alors privés des critères pour la distinction entre des illusions (ou des hallucinations) et de vraies perceptions. Les partisans de l'intelligence artificielle sont donc obligés, quand ils se rendent compte de cette difficulté, de prendre une mesure particulière pour empêcher un sceptique (du genre de Hume) de transformer ce solipsisme *méthodologique* en un *vrai* solipsisme. Et elle n'est pas moins arbitraire que les mesures proposées par les classiques du solipsisme méthodologique. Ce qui doit empêcher l'écroulement de la réalité en « mondes » solipsistes, c'est, selon Fodor, le *Dasein*. Cependant, il ajoute aussitôt, sans autre commentaire : « I don't know what *Dasein* is, but I'm sure that there's lots of it around, and I'm sure that you and I and Cincinnati have all got it. What more do you want ? »¹⁸ Évidemment, c'est une nouvelle variante de la vieille solution à l'embarras par laquelle on essaie de sauver la théorie représentationnaliste de l'esprit en se rapportant à un

-
17. Haugeland (*Artificial Intelligence : The Very Idea*. Cambridge, Massachusetts : The MIT Press, 1985, p. 33f.) semble avoir, tout d'abord, conscience de ce problème. Ainsi, il montre que le « scepticisme épistémologique », c'est-à-dire, la thèse que la totalité de nos pensées et de nos perceptions puisse n'être qu'une hallucination sans rapport avec le « monde externe », est un corollaire de la conception cartésienne, reprise par l'intelligence artificielle, des pensées comme de simples symboles dans un système notationnel, qui pourraient être combinées de manière à gérer l'illusion de sens. De telles symboles pourraient représenter un état de choses ou un autre, ou même n'en représenter aucun. Comme le montre Haugeland, cela serait indifférent pour le système de la pensée. Toutefois, Haugeland écarte, à ce point, le problème comme une digression dans l'histoire de la philosophie, d'un intérêt à peine périphérique pour le programme cognitiviste de l'intelligence artificielle.
18. FODOR, J. « Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology », *op. cit.*, p. 253.

créateur infiniment sage de l'harmonie préétablie de nos jugements de perception.

III

Parmi les conséquences de ces réflexions, il y en a une qui s'impose spécialement : si nous tenons à une compréhension adéquate des actions intelligentes, nous devons prendre au sérieux la question de la signification comme quelque chose d'irréductible.¹⁹

La stratégie de l'intelligence artificielle, au contraire, est celle d'un réductionnisme formel. Elle essaie d'établir, pour chaque système, une interprétation globale qui tient compte de ses entrées et de ses sorties dans beaucoup de cas, en cherchant à y discerner un certain pattern, par exemple, celui d'un joueur d'échecs. Comme Haugeland l'a montré de manière exemplaire,²⁰ les éléments qui composent les entrées et les sorties d'un système d'élaboration de l'information doivent être des *tokens* de *types* qui forment un ensemble déterminable de manière univoque (*uniquely determinable*). Cela veut dire : pour chaque *token*, on doit déterminer de manière univoque à quel *type* il appartient. En outre, chaque *token* ne peut appartenir qu'à un *type*. Il n'y a donc pas d'ambiguïtés dans la corrélation *token-type*. Haugeland fait encore la distinction entre les *types* simples et les *types* complets. Les *tokens* de *types* complets consistent en des *tokens* de *types* simples combinés selon certaines règles bien définies. Une telle interprétation établit premièrement, de manière univoque, une corrélation entre les *types* simples d'une part et, par exemple, certaines figures du jeu d'échecs ou certaines positions de l'échiquier d'autre part. Les *tokens* de *types* simples acquièrent ainsi des significations : ils *signifient* des figures ou des positions du jeu d'échecs. À travers quelques procédés récursifs qui exploitent les propriétés des règles

19. Voir von WRIGHT, G. H. *Explanation and Understanding*. London : Routledge & Kegan Paul, 1971, p. 114ss. Voir aussi TAYLOR, C. « Interpretation and the Sciences of Man ». *Review of Metaphysics*, 25 (1971), 3-51 ; PUTNAM, H. *Meaning and the Moral Sciences*. London : Routledge & Kegan Paul, 1978 ; HUGHES, J. *The Philosophy of Social Research*. London : Longman, 1980 ; TAYLOR, C. « The Significance of Significance : The Case of Cognitive Psychology ». In Mitchell, S. et Rosen, M. (eds.) *The Need for Interpretation*. London : The Athlone Press, 1983.

20. HAUGELAND, J. « The Nature and Plausibility of Cognitivism ». *The Behavioral and Brain Sciences*, 2 (1978), 215-260.

de combinaison des *tokens* de *types* simples, les *tokens* de *types* complets reçoivent alors des significations. (J'ometts ici des détails plus techniques.)

Cette méthode conduit à un résultat *si* les entrées et les sorties du système pour lequel on veut construire une interprétation sont formées d'éléments essentiellement indépendants du contexte (*context-free elements*). Ceci est apparemment plausible dans le cas des symboles produits dans une notation standard du jeu d'échecs. Dans ce cas, les significations des *tokens* de *types* simples sont spécifiées selon une fonction élémentaire, une liste qui établit une corrélation entre des *types* simples et des « objets » du jeu d'échecs. Et les significations des *tokens* de *types* complets sont déterminées uniquement à travers la manière dont ceux-ci sont formés à partir des *tokens* de *types* simples, c'est-à-dire : selon le système de règles récursives de la notation du jeu d'échecs. Une interprétation dans ce sens est donc un schéma général régulier (*a regular general scheme*) basé sur des stipulations fonctionnelles et des règles « rigides » qui définissent un procédé « effectif ». Une interprétation réussie selon ce schéma transforme les entrées et les sorties d'un système en « représentations quasi-linguistiques », et le système (comme un tout) en un véritable système d'élaboration de l'information.

La thèse radicale de l'intelligence artificielle dit que l'être humain est un système de ce genre, qu'on peut, en d'autres mots, décrire et expliquer le comportement humain intelligent comme la sortie d'un système d'élaboration de l'information.

Ce qui me semble correct dans tout cela, c'est l'idée que le comportement d'un « système » n'a de signification que dans le cadre d'une interprétation globale qui lui fait place dans l'ensemble des autres comportements possibles et réels. Une forme de comportement ne peut véhiculer, faire comprendre quelque chose, signifier quelque chose, qu'en rapport avec un arrière-plan et donc relativement à des considérations plus ou moins holistes. Ceci est valable pour les machines aussi bien que pour les hommes. Dans ce sens restreint, les hommes et les machines sont des « systèmes » qui doivent être constamment interprétés.

Cependant — et les défenseurs de l'intelligence artificielle ne semblent pas s'en apercevoir — l'interprétation globale du comportement des êtres humains est, en règle générale, complètement

différente de l'interprétation qui fait d'une machine un système d'élaboration de l'information. Ce qu'on interprète chez les êtres humains sont leurs propos linguistiques, leurs gestes et leurs expressions corporelles, en un mot : les manifestations vitales dont parlait Dilthey. Dans ces circonstances on retrouvera difficilement ces éléments indépendants du contexte qui caractérisent les systèmes d'élaboration de l'information et que Haugeland met à la base de sa stratégie interprétative.

Ceci est évident surtout dans le cas des gestes et des expressions corporelles. En effet, on n'a pas affaire, dans ce domaine, à des *tokens* essentiellement non ambigus de *types* déterminables de manière univoque. Il n'existe pas une syntaxe formelle des gestes et des expressions corporelles. Et qui serait prêt à se passer des gestes et des expressions corporelles dans la tentative d'interpréter et de comprendre les actions et les manifestations verbales humaines ? Il ne nous reste ici que l'exercice des habilités herméneutiques fondamentales que nous avons appris à maîtriser, non pas selon des schémas récurifs, mais dans la mesure où nous avons été introduits dans une pratique spécifique, caractérisée, entre autres, par le développement de quelques attitudes particulières face aux hommes, voyant les autres hommes et étant vus par eux comme des personnes, et pas comme de simples systèmes qu'on interprète « de l'extérieur ».

Mais aussi dans le cas de pures expressions verbales, séparées des autres manifestations vitales — comme dans la lecture d'un livre — notre compréhension diffère radicalement de la construction d'un système d'élaboration de l'information. Les manifestations verbales ne sont pas des événements indépendants du contexte qu'on pourrait incorporer dans un système effectif. Les phrases de notre langage ne sont pas des structures dont la signification puisse être obtenue selon un schéma. On donne un sens aux propos de ceux qui nous entourent dans le langage quotidien. Et le langage quotidien, à l'opposé de la notation du jeu d'échecs, n'est pas un système rigide. Il faut résister à la tentation de penser « that if anyone utters a sentence and *means* or *understands* it he is operating a calculus according to definite rules ». ²¹

21. WITTGENSTEIN, L. *Philosophical Investigations*. New York : The Macmillan Company, 1953, § 81.

Selon la suggestion de Haugeland, la signification d'un *token* d'un *type* complet est déterminée uniquement à travers sa composition à partir de *tokens* de *types* simples. Ceci n'est pas plausible ou même absolument faux sur le plan du langage normal, où le contexte de l'énoncé verbal détermine sa signification de manière non algorithmique. On est ici très loin des « représentations quasi-linguistiques » de Haugeland. La détermination herméneutique de la signification d'une manifestation verbale dépend des circonstances dans lesquelles elle est située. En effet, dans ce cas, parler d'une manière absolument générale de *types* simples et complets, cela ne veut rien dire de précis. Toute distinction entre simple et non-simple doit être fixée « localement », ayant en vue l'articulation et, éventuellement, la résolution de quelque problème concret, qui doit toujours être expliqué. Cela importe particulièrement si on ne veut pas que notre discours dégénère en des spéculations incontrôlables et en des affirmations dogmatiques qu'on ne peut justifier seulement en les formulant dans un langage apparemment exact issu des sciences de l'informatique.

Tout cela semble confirmer que, si nous devons prendre au sérieux la question de la signification dans nos efforts pour comprendre les actions humaines, nous ne devons pas nous occuper de cette question d'après le modèle solipsiste, ou du point de vue d'un observateur « objectif » qui cherche à construire une interprétation formelle, car cela nous mène seulement à une impasse. Au lieu de cela, nous devons, en suivant l'exemple de Wittgenstein, partir du lien étroit qui existe entre la signification et les expériences personnelles dans une pratique culturelle commune. Ceci nous libérerait de la tentation cognitiviste de l'intelligence artificielle, et nous mènerait plutôt à ne plus chercher le « secret » de l'intelligence auprès des ordinateurs et des programmes que nous devons, en tant qu'êtres intelligents, développer et utiliser raisonnablement, mais précisément là où l'intelligence a son siège « naturel » : dans la culture humaine.

*Departamento de Filosofia,
Pontificia Universidade Católica do Rio de Janeiro*