Hubert DREYFUS

LES ORDINATEURS PEUVENT-ILS ETRE VRAIMENT INTELLIGENTS ?*

Même les personnes qui ont un esprit scientiste ont le sentiment que bien qu'elles soient faites de matière, elles ne sont pas des machines ; cependant, récemment, il leur est de plus en plus répété, comme si cela était évident, que « chaque être humain est un ordinateur superbement construit » (1) et que les ordinateurs se comporteront éventuellement aussi intelligemment que les humains. Certains savants disent que les ordinateurs, comme Hal dans le film *« 2001 : Une odyssée de l'espace »*, seront exactement comme les humains ; d'autres prétendent que les machines intelligentes seront supérieures aux êtres humains, puisqu'elles n'auront pas à subir la fatigue, les émotions, les crises de doute, et l'illusion qu'elles ne sont pas des machines. Chacune de ces prédictions est accompagnée de son propre scénario pour un désastre : l'ordinateur émotif s'affolle et détruit tout le monde dans son effort irraisonné pour sauver la mission ; l'ordinateur purement intellectuel rend calmement la société un enfer rationnel convenant uniquement aux robots. Puisque répandre la bonne nouvelle de l'imminence de l'intelligence artificielle aussi bien que prophétiser un désastre inévitable sont devenus une nouvelle industrie des média, il est grand temps de songer à nouveau à notre tranquille certitude que nous ne sommes pas des ordinateurs et que prétendre qu'ils peuvent être intelligents doit être une absurdité.

Deux des « succès » les plus popularisés des ordinateurs, qui semblent donner quelque fondement à l'idée que les savants sont en train de faire des progrès réguliers vers la création de machines intelligentes, sont le programme des blocs (SHRDLU)

(*) Les sujets traités dans cet article, ainsi que plusieurs autres programmes d'ordinateur du même genre et leurs problèmes, sont analysés plus en détail dans *What Computers Can't Do, A Critique of Artificial Intelligence*, par Hubert L. DREYFUS, Harper and Row, 2e éd., 1979.

(1) C. SAGAN, « In Defense of Robots », dans *Broca's Brain*, Random House, New York, 1974.

REVUE CRITIQUE, Août-septembre 1980, N° 399-400

de Terry Winograd (2) et les performances impressionnantes des récents programmes pour jouer aux échecs.

Quand il fut dévoilé pour la première fois, il y a dix ans, le programme de Winograd parut être, vraiment, une enjambée importante vers la création de machines intelligentes. SHDRLU simule le bras d'un robot qui peut faire bouger un ensemble de blocs de formes différentes et permet à quelqu'un d'engager un dialogue avec l'ordinateur, en posant des questions, en faisant des déclarations, et en donnant des ordres à propos de ce monde simple de blocs mobiles. Ceux qui travaillent dans le domaine de l'IA (Intelligence artificielle) n'ont pas essayé de cacher le fait que c'est grâce au domaine restreint de SHRDLU qu'une apparente compréhension est possible. Ils avaient même un nom pour désigner la façon dont Winograd réduit le champ du discours. Celui-ci traitait avec un « micro-monde ». Marvin Minsky et Seymour Papert, co-directeurs du « projet-robot » au Massachusetts Institute of Technology (M.I.T.), expliquent :

> Chaque modèle — ou « micro-monde » comme nous l'appellerons — est très schématique ; il s'agit d'un monde imaginaire dans lequel les choses sont si simplifiées que presque chacune des déclarations faites à leur sujet seraient littéralement fausses si elles étaient prononcées à propos du monde réel (3).

Mais ils ajoutent immédiatement :

> Cependant, nous avons le sentiment qu'ils (les micro-mondes) sont si importants que nous sommes en train de consacrer une grande partie de notre effort au développement d'une série de ces micro-mondes et nous cherchons comment utiliser les puissances de suggestion et de prédiction des modèles sans être incommodés par leur incompatibilité avec la vérité.

Ce qui caractérise la période du début des années soixante-dix, et qui fait paraître SHRDLU comme un pas en avant vers une intelligence générale, c'est ce concept apparemment scientifique d'un micro-monde. Etant donné le caractère artificiel et arbitraire des micro-mondes, pourquoi est-ce que Minsky et

(2) T. WINOGRAD, « A Procedural Model of Language Understanding », dans *Computer Models of Thought and Language*, édité par R. Schank et K. Colby, W.H. Freeman Press, San Francisco, 1973. [Le sigle SHRDLU n'a aucune signification. Winograd l'a emprunté au magazine *Mad* qui utilise souvent cette erreur typographique pour nommer les monstres mythiques et autres créatures de ce genre.]

(3) M. MINSKY et S. PAPERT, Projet d'une proposition à l'*Advance Research Projects Administration* (*ARPA*) pour des recherches sur l'Intelligence artificielle au M.I.T. de juillet 1970, publié par M.I.T. en 1970/71.

Papert ont pensé qu'ils fournissent une direction de recherche intéressante ?

Pour trouver la réponse, nous devons suivre les remarques pénétrantes de Minsky et Papert sur la compréhension de la narration et leurs moins pénétrantes conclusions :

> ... Dans la fable familière, c'est par la ruse que l'astucieux Renard arrive à ce que le vain Corbeau lâche le fromage en lui demandant de chanter. Le test habituel de compréhension est la capacité de l'enfant à répondre à des questions comme celle-ci : « Est-ce que le Renard pense que le Corbeau a une jolie voix ? »
>
> Le sujet est quelquefois classifié comme « manipulation du langage naturel » ou comme « logique déductive », etc. Ces descriptions sont mal choisies. Car le vrai problème n'est pas de comprendre l'anglais ; c'est de *comprendre* tout simplement. La difficulté pour obtenir de la machine une bonne réponse ne dépend absolument pas de la « disambiguation » des mots (tout au moins pas dans le sens habituel de choisir une « signification » parmi un ensemble discret de « significations »). La difficulté ne réside pas non plus dans la nécessité d'avoir un appareil logique inhabituellement puissant. Le principal problème est que personne n'ait construit les éléments d'un ensemble de connaissances à propos de telles matières qui soit adéquat pour comprendre l'anecdote. Voyons ce que cela entraîne.
>
> Pour commencer, il n'y a jamais une solution unique à ce genre de problème ; aussi nous ne demandons pas ce que l'Entendeur *doit* connaître. Mais il aurait avantage à posséder le concept de *flatterie*. Pour procurer cette connaissance, nous imaginons une « micro-théorie » de la flatterie — un ensemble extensible de faits ou de procédures qui décrit les conditions dans lesquelles on peut s'attendre à trouver la flatterie, quelles formes elle prend, quelles sont ses conséquences, etc. La complexité de cette théorie dépend de ce qui est présupposé. Aussi serait-il très difficile de décrire la flatterie à notre Entendeur [homme ou machine] si il/elle ne savait pas déjà que des propos peuvent servir à d'autres buts que de convoyer une information factuelle littéralement correcte. Cela serait pratiquement impossible s'il/elle n'avait même pas un concept comme *but* ou *intention* (4).

Le coup surprenant ici est qu'ils conclurent qu'il *pourrait* exister une « micro-théorie » circonscrite de la flatterie — en quelque sorte intelligible en dehors du reste de la vie humaine — tout en montrant que comprendre la flatterie entraînerait une pénétration toujours plus poussée dans la compréhension du reste de notre vie quotidienne, avec ses buts et ses intentions complexes.

La notion d'un micro-monde, comme un domaine pouvant être traité en isolation, vient d'une fausse analogie avec la physique. En effet, durant notre vie quotidienne nous sommes engagés dans une multitude de « sous-mondes » tels que le monde du théâtre, des affaires, ou des mathématiques, mais

(4) *Idem.*, p. 42-44.

chacun de ceux-ci est une « modalité » du monde quotidien que nous partageons (5). Les sous-mondes ne sont donc pas reliés au monde comme les systèmes physiques isolés sont liés à de plus grands systèmes dont ils sont les *composantes ;* ils sont plutôt les élaborations locales d'un tout qu'ils *présupposent.*

C'est seulement récemment que l'illusion que l'on pouvait généraliser les travaux faits dans des domaines étroitement circonscrits, a été repérée et rejetée par Winograd lui-même :

> Les programmes d'IA des dernières années soixante et du début des années soixante-dix prennent toutes choses au pied de la lettre. Ils traitent la signification comme si elle était une structure à construire avec des briques et du mortier fournis par les mots... Cela leur donne un caractère « friable », capable de fonctionner seulement dans les domaines à la signification très spécifique, comme une conversation formelle artificielle. De même, ils sont pauvres s'il s'agit de traiter les énoncés naturels, pleins de morceaux et de fragments, de métaphores continues (non décelables), et de références à des domaines moins aisément formalisés (6).

Tandis que les vulgarisateurs sont toujours pleins de louanges pour SHRDLU, il est maintenant généralement reconnu que le recours au micro-monde pour programmer l'intelligence quotidienne est une impasse.

Tandis que la vie quotidienne est la totalité des activités reliées les unes aux autres, les jeux sont juste le genre de micro-mondes totalement circonscrits dans lesquels les ordinateurs excellent. Ainsi, dans la mesure ou l'on peut s'attendre à des échecs lorsque l'on travaille sur le langage humain, on pourrait espérer de grands succès avec les programmes de jeux. Mais nous devons faire attention à ne pas attribuer ces succès à une capacité ressemblant à l'intelligence humaine.

Les échecs, par exemple, sont un parfait micro-monde dans lequel ce qui est important est réduit au domaine étroit défini par le genre de pièce (pion, chevalier, etc.), sa couleur, et la position qu'elle occupe sur l'échiquier. La taille, le poids et la température de la pièce ne sont jamais pertinents. Mais tandis que le caractère circonscrit du jeu rend possible, en principe, un programme pour jouer aux échecs au niveau du championnat du monde, il y a beaucoup d'indications que les êtres humains jouent aux échecs d'une façon tout à fait différente des ordinateurs. En effet, les ordinateurs n'utilisent pas de stratégie

(5) Cette conclusion est démontrée dans *Sein und Zeit* de M. HEIDEGGER. (Voir particulièrement la section 18.)

(6) T. WINOGRAD, « Artificial Intelligence and Language Comprehension », dans *Artificial Intelligence and Language Comprehension*, National Institute of Education, Washington, 1976, p. 17.

à long terme, n'apprennent rien par expérience, ou même ne se souviennent pas des tours précédents.

Pour comprendre quelle est la différence entre le jeu humain et celui d'un ordinateur, nous devons d'abord comprendre comment fonctionne un programme d'échecs. Un programme d'échecs utilise des règles qui relient une situation à une action. Une situation est caractérisée en termes d'attributs indépendants du contexte ; par exemple, la position et la couleur de chaque pièce sur l'échiquier. Toutes les manœuvres légales et les positions qui en résultent sont alors définies en termes de ces attributs. Pour évaluer et comparer les positions, des règles sont fournies pour calculer les scores des attributs tels qu'« équilibre matériel » (où une valeur numérique est désignée pour chaque pièce de l'échiquier et le score total est calculé pour chaque joueur) ou « contrôle du centre » (où le nombre de pièces qui dominent le centre de l'échiquier sont comptées). Finalement, il doit y avoir une formule basée sur ces scores pour déterminer la valeur de chaque position possible. En utilisant cette approche et en considérant environ trois millions de positions possibles à chaque tour, CHESS 4.5 a récemment gagné le 84e tournoi du Minnesota, mais un maître envisage les résultats de moins de cent coups possibles à chaque tour et cependant a un jeu bien meilleur. Comment cela se fait-il ?

Il semble qu'en rejouant les jeux publiés et qu'en participant à des tournois, les maîtres développent une capacité pour reconnaître les positions du jeu actuel comme étant similaires aux positions qui se sont produites dans les jeux classiques. Ces positions-là ont déja été analysées en fonction de leurs aspects significatifs. Les aspects d'une position dans un jeu d'échecs comprennent des caractéristiques globales telles que « contrôle de la situation » (le degré dans lequel les coups du joueur opposé peuvent être forcés par une manœuvre menaçante), « une situation comprimée » (ce qui reste de la possibilité de manœuvre pour chaque joueur), ou « trop grand éparpillement » (le fait que, bien que la position soit superficiellement forte, on n'ait pas suffisamment le contrôle de la situation pour poursuivre et que, avec un jeu correct de l'adversaire, une retraite massive soit obligatoire). Les positions déjà analysées dont il se souvient permettent au joueur d'éviter un grand nombre de calculs et de concentrer son attention sur les zones critiques avant de commencer à compter un à un chaque coup.

La distinction entre attributs et aspects est centrale ici. Dans une analyse du jeu des joueurs humains, les *aspects* jouent un rôle similaire à ceux des *attributs* dans un modèle d'ordinateur, mais il y a une grande différence. Dans le modèle d'ordinateur, *la situation est définie en fonction des attributs,*

tandis que pour les joueurs humains *la compréhension de la situation précède la détermination des aspects.* Par exemple, la valeur numérique d'un attribut tel que l'équilibre matériel peut être calculée indépendamment de toute compréhension du jeu, tandis qu'un aspect comme un trop grand éparpillement ne peut être calculé seulement en fonction de la position des pièces, puisqu'une même position sur l'échiquier peut avoir des aspects différents selon sa place dans la stratégie à long terme du jeu.

On ne peut pas dire que la capacité du maître à utiliser une expérience passée pour concentrer son attention sur les éléments importants qui émergent du jeu présent tient de la correspondance, basée sur les attributs, entre la position présente et des positions précédentes placées dans sa mémoire comme des livres sur les rayons d'une bibliothèque. Il est extrêmement improbable que deux positions soient jamais *identiques ;* aussi faut-il comparer des positions *similaires.* Mais on ne peut pas définir la similarité comme le fait d'avoir un grand nombre de pièces sur des carrés identiques. Deux positions qui sont identiques sauf pour un pion placé sur un carré adjacent peuvent être totalement différentes, tandis que deux positions peuvent être similaires bien qu'aucune pièce ne soit sur le même carré dans chacune. Ainsi la similarité dépend du sentiment qu'a le joueur de ce que sont les issues en question, pas simplement de la position des pièces. Pour voir que deux positions sont similaires, il faut précisément avoir une compréhension profonde du jeu. En structurant ainsi la situation courante en fonction des aspects des situations similaires dont il se souvient, le joueur humain est capable d'éviter la masse de calculs que doit faire un ordinateur qui peut seulement « reconnaître » des positions caractérisées en fonction d'attributs indépendants.

Donc l'intelligence humaine, même dans les jeux, demande l'utilisation d'un fond de connaissances pratiques ; dans la vie quotidienne ce fond est fait de la compréhension commune, que nous partageons avec les autres êtres humains, à savoir faire les choses. Les travaux récents en intelligence artificielle ont été forcés de traiter directement avec ce fond quotidien. Confrontés à cette nécessité, les chercheurs ont implicitement essayé de traiter le fond comme un ensemble de faits reliés par des règles — quelquefois appelé un « système de croyances ». Ce préjugé, que le fond des connaissances pratiques peut être traité simplement comme un autre objet, sert de base à la prétention que tous les êtres humains ne sont que des ordinateurs très compliqués. Cette conviction est ancrée profondément dans notre entière tradition philosophique. Suivant Martin Heidegger, qui est le premier à avoir identifié et critiqué cette vue, je l'appellerai le préjugé métaphysique.

La question évidente à poser est : Y a-t-il des raisons de croire qu'en dehors des difficultés persistantes et du passé de promesses non remplies de l'IA, le préjugé métaphysique soit injustifié ? Y a-t-il aucune défense contre cette version subtile de la philosophie mécaniste ? Le meilleur argument, je crois, c'est que chaque fois que la conduite humaine est analysée en termes de faits reliés par des règles, ces règles doivent toujours contenir une condition *ceteris paribus*, c'est-à-dire qu'elles s'appliquent, « toutes choses étant égales », et ce que « toutes choses » et « égales » signifient dans une situation donnée ne peut jamais être complètement expliqué. De plus, cette condition *ceteris paribus* n'est pas un inconvénient montrant que l'analyse n'est pas encore complète et peut être une « tâche infinie ». La condition *ceteris paribus* indique plutôt le fond des pratiques comme condition de la possibilité de toute activité qui peut être décrite par des règles. Lorsque nous expliquons nos actions nous devons toujours, à un moment ou à un autre, recourir à nos pratiques quotidiennes et simplement dire « ceci est ce que nous faisons » ou « être un être humain, c'est ça ». Ainsi en dernière analyse tout intelligibilité et toute conduite intelligente a sa source dans le sentiment de ce que nous *sommes*, qui est quelque chose que nous ne pouvons jamais *connaître* d'une façon explicite.

Pour rendre cette prétention plus plausible on peut se servir comme exemple d'un programme fait au M.I.T. pour comprendre des anecdotes. Considérez l'extrait suivant d'une anecdote :

> Aujourd'hui, c'était l'anniversaire de Jacques. Penny et Jeannette sont allées au magasin. Elles sont allées acheter des cadeaux. Jeannette décida d'acheter un cerf-volant. « Ne le fais pas », dit Penny. « Jacques a un cerf-volant. Il te *le* fera retourner. » (7)

Le but est de construire une théorie qui explique comment le lecteur comprend que « le » se rapporte au nouveau cerf-volant, et non pas à celui que Jacques possède déjà. Des astuces grammaticales (telles que « le » se rapportant au dernier substantif mentionné) sont clairement inadéquates, parce que le résultat serait de commettre l'erreur de comprendre la dernière phrase de l'anecdote comme signifiant que Jacques demandera à Jeannette de retourner le cerf-volant qu'*il possède déjà*. Il est aisé de voir qu'on ne peut savoir que « le » se rapporte au nouveau cerf-volant si l'on ne sait pas quelles sont les habitudes commerciales de notre société. On pourrait imaginer un monde

(7) I. GOLDSTEIN et S. PAPERT, Laboratoire d'IA du M.I.T., IA memo n° 337, juillet 1975, revu en mars 1976, « Artificial Intelligence, Language and the Study of Knowledge », p. 29-31.

différent dans lequel les objets que l'on vient d'acheter ne sont jamais retournés au magasin, mais que les anciens le sont.

L'approche de l'IA dictée par le préjugé métaphysique est, évidemment, d'essayer de rendre le fond des pratiques impliqué dans la compréhension de cette anecdote explicite comme étant un ensemble de croyances. Mais une fois que les jeux et les micro-mondes sont laissés de côté, un abîme géant menace d'engouffrer ceux qui essayent d'accomplir un tel programme. Comme Papert le note :

> ... l'anecdote n'inclut pas d'une façon explicite tous les faits importants. Regardons à nouveau l'anecdote. Certains lecteurs seront surpris de noter que le texte lui-même ne dit pas (a) que les cadeaux achetés par Penny et Jeannette étaient pour *Jacques*, (b) que le [cerf-volant] acheté par Jeannette était en fait un cadeau, et (c) qu'avoir un objet implique que l'on ne veut pas en avoir un autre (8).

Notre exemple repose sur la question : Comment est-ce que l'on met en mémoire d'ordinateur les « faits » mentionnés en (c) ci-dessus à propos du retour des cadeaux ? Pour commencer, il y a peut-être un nombre indéfini de raisons pour retourner un cadeau. Ce cadeau n'a peut-être pas la bonne taille, fonctionne sur un voltage différent, est cancérigène, fait trop de bruit, est considéré comme étant trop enfantin, trop féminin, trop masculin, trop américain, etc. Et pour comprendre chacun de ces faits, il en faut d'autres. Mais nous nous concentrerons sur la raison mentionnée en (c) : que normalement, c'est-à-dire *toutes choses étant égales*, si l'on a un objet on n'en veut pas un autre semblable. Bien sûr, cela ne peut pas être acquis simplement comme une vraie proposition. Cela n'est pas vrai pour des billets de banques, des biscuits, ou des billes. (Il n'est même pas évident que cela soit vrai pour des cerf-volants). Papert pourrait répondre que, bien sûr, une fois que l'on parle de ce qui est normal il faut être prêt à traiter les exceptions.

Mais là commencent les clins d'œil désespérés, car le texte n'a pas besoin de mentionner de façon explicite les exceptions. Si le cadeau était des billes ou des biscuits, le texte sûrement n'aurait pas à mentionner que ce sont des exceptions à la règle générale, que posséder un unique exemplaire est suffisant. Aussi la base des données devrait contenir *une description de toutes les exceptions possibles* pour compléter le texte, à supposer que de penser à cela comme une liste définitive ait même un sens. Pire encore, même si on avait la liste de tous les cas exceptionnels où quelqu'un serait content de posséder plus qu'un seul spécimen d'un certain genre d'objet, il y a des situations qui permettent une exception à cette exception :

(8) *Idem.*, p. 33.

avoir déjà un biscuit est plus que suffisant si le biscuit en question a un mètre de diamètre ; un millier de billes est plus que ce qu'un enfant normal peut manipuler. Devons-nous donc établir une liste des situations qui font que l'on doit s'attendre à avoir des exceptions aux exceptions ? Mais ces exceptions aussi peuvent être outrepassées dans le cas, par exemple, d'un goinfre ou d'un collectionneur de billes, et ainsi de suite. Le programmeur qui écrit un programme pour comprendre les anecdotes doit essayer d'énumérer toutes les informations utiles, et une fois que cette information fait appel au *normal* ou au *typique* il n'y a aucun moyen d'éviter que s'ajoutent des séries sans fin de conditions aux conditions d'application de ce savoir à une situation spécifique.

La seule « réponse » qu'offre Papert est le préjugé métaphysique qui veut que le fond des connaissances de la vie quotidienne soit un ensemble de situations rigidement définies, dans lesquelles les faits pertinents sont clairs et circonscrits comme dans un jeu :

> Les prémices fondamentales de la théorie des cadres c'est la thèse qui veut que... la plupart des situations dans lesquelles les gens se trouvent *ont suffisamment en commun* avec des situations rencontrées précédemment pour que les traits saillants soient *pré-analysés* et mis en mémoire sous la forme « situation-spécifique » (9).

Mais cette « solution » est intenable pour deux raisons :

1. Même si la situation présente est, en effet, *similaire* à une situation pré-analysée, nous avons toujours le problème de décider à quelle situation elle ressemble. Nous avons déjà vu que, même pour les jeux tels que les échecs, il est peu probable que deux positions soient identiques ; aussi une profonde compréhension de ce qui se passe est-elle nécessaire pour décider ce qui compte comme deux positions similaires dans deux jeux donnés. Ceci devrait être encore plus évident dans le cas où le problème est de décider à quelle situation pré-analysée une situation réelle donnée ressemble le plus : par exemple, si une situation dans laquelle il y a des bébés bien habillés et des jouets neufs que l'on offre est plus en rapport avec un anniversaire ou un concours de beauté.

2. Même si toutes nos vies *étaient* vécues comme des situations stéréotypées, nous venons juste de voir que tout cadre du monde réel doit être décrit en fonction du normal, et que faire appel au normal entraîne nécessairement une régression. Car lorsque nous essayons d'énumérer explicitement les conditions nécessaires et suffisantes de la normalité, nous sommes entraînés dans le marécage des exceptions aux exceptions. Seul le sentiment général que nous avons de ce qui est typique nous permet de décider de ce qui est normal dans un cas particulier, et *cette* compréhension du fond ne peut pas être par définition « spécifique à la situation ».

(9) *Idem.*, p. 30-31. (Souligné par moi.)

Le sens que nous avons de notre situation est déterminé par nos humeurs changeantes, nos intérêts et projets courants, le projet global de notre vie, et aussi probablement par nos facultés sensori-motrices à manier les objets et à vivre avec les gens — facultés que nous développons par la pratique sans jamais avoir à nous représenter à nous-mêmes notre corps comme étant un objet, notre culture comme étant un ensemble de croyances, et nos habiletés comme étant produites par des règles liant les situations aux actions. Toutes ces capacités uniquement humaines donnent une « richesse » et une « épaisseur » à notre façon d'être-dans-le-monde et ainsi semblent jouer un rôle essentiel dans le fait que nous sommes en situation, qui à son tour soutient toute conduite intelligente.

Etant donné que le fait d'être en situation est essentiel à l'intelligence humaine, ceux qui désirent modeler l'esprit d'après un ordinateur, qui est essentiellement hors de toute situation, sont confrontés à un dilemme. Ou bien ils doivent essayer de formaliser le fond entier des pratiques quotidiennes humaines. Terry Winograd, par exemple, reconnaît maintenant que l'intelligence humaine est « holistique » et que la signification dépend de « l'ensemble entier formé par les buts et la connaissance ». Si cette approche était poursuivie d'une façon consistante, elle obligerait les travailleurs de l'intelligence artificielle à capturer les humeurs, les intérêts, et les pratiques incarnées dans un réseau formel de croyances (tentative qui est sûrement condamnée à échouer). Ou bien, quand ils auront un aperçu de l'énormité de la tâche, les travailleurs de l'IA peuvent essayer d'isoler l'intelligence du fond de la compréhension quotidienne. A ce stade presque tous les travailleurs de l'IA et tous ceux qui croient que l'esprit fonctionne en traitant l'information sont plus ou moins lucidement condamnés à la seconde alternative de ce dilemme. Ils présument que les aspects non-cognitifs de l'esprit peuvent être ignorés en toute sécurité — que la cognition consiste simplement dans les manipulations d'ensembles complexes de symboles sans signification qui correspondent aux attributs du monde réel. Winograd est entraîné dans cette voie lorsqu'il écrit : « L'IA est l'étude générale des aspects de la cognition qui sont communs à tout système de symboles physiques, y compris les humains et les ordinateurs. (10) » Mais cela simplement définit le domaine de l'intelligence dé-située ; cela

(10) T. WINOGRAD, Panel on Natural Language Processing, IJCAI-77, *Proceedings*, p. 1008.

ne montre en aucune façon qu'une telle chose existe. L'anecdote du cerf-volant, qui débouche sur la totalité des pratiques humaines, suggère fortement que cette approche, basée sur un désir désespéré plutôt qu'une estimation réaliste du phénomène, est également vouée à l'échec.

Cependant, à ce dilemme les chercheurs de l'IA peuvent répondre d'une façon plausible : « Quelque soit le fond des intérêts, sentiments et pratiques partagés, nécessaires pour comprendre des situations spécifiques, ce savoir *doit* être d'une manière ou d'une autre dans l'esprit des êtres humains qui ont cette compréhension. Et comment représenter autrement ce savoir sinon comme un ensemble explicite de faits et de croyances ? » En effet, le genre de programmation d'ordinateur accepté par tous les travailleurs de l'IA nécessiterait une telle structure des données, et une telle structure est aussi présupposée par les philosophes qui soutiennent que toute connaissance doit être représentée explicitement dans nos esprits. Mais il y a deux alternatives qui, en évitant l'idée que tout ce que nous savons doit exister dans nos esprits sous la forme d'une description explicite, éviteraient les contradictions inhérentes au modèle du traitement de l'information.

Une réponse, partagée par des phénoménologues existentiels comme Maurice Merleau-Ponty (11) et des philosophes du langage ordinaire tels que Ludwig Wittgenstein, est de dire qu'un tel « savoir » des intérêts et pratiques humains n'a pas besoin d'être représenté. Comme Wittgenstein le dit dans *De la certitude* : « Les enfants n'apprennent pas que les livres existent, que les fauteuils existent, etc., — ils apprennent à aller chercher les livres, à s'asseoir dans les fauteuils, etc. » (12). Juste comme il semble plausible que je puisse apprendre à nager en m'exerçant jusqu'à ce que je développe la bonne forme d'activité qui se déclenchera automatiquement sans que je décrive jamais à moi-même mon corps et les mouvements musculaires, ainsi également ce que je « sais » des pratiques culturelles qui me rendent capable de reconnaître et d'agir dans des situations spécifiques a été graduellement acquis par entraînement — dans un environnement déjà signifiant — bien que personne n'ait ou ne puisse jamais rendre explicite ce qui était appris.

Une autre description possible aurait une place pour les représentations, tout au moins dans certains cas où je dois m'arrêter et réfléchir, mais une telle position soulignerait le fait que ces représentations ne sont que rarement des descriptions explicites mais sont plutôt des images au moyen desquelles j'explore ce que je *suis*, non pas ce que je *sais*. D'après cette vue je ne me représente pas que j'ai des désirs, ou bien que pour rester debout il faut être en équilibre, etc. Quand cela aide, cependant, comme pour comprendre une anecdote, je peux m'imaginer dans une situation particulière et me demander ce que je ferais ou quels seraient mes sentiments (si j'étais à la place de Jacques, comment réagirais-je si l'on m'offrait un second cerf-volant ?) sans avoir à rendre explicite tout ce qu'il faudrait dire à un ordinateur pour qu'il puisse parvenir à la même conclusion. Ainsi nous faisons appel à des représentations *concrètes* (images ou souvenirs) basées sur notre propre expérience sans avoir à rendre explicite les règles strictes supposées et expliquer complètement (si cela a même un sens) leurs conditions *ceteris paribus* qui seront nécessaires pour construire des descriptions symboliques *abstraites*.

(11) M. MERLEAU-PONTY, *Phénoménologie de la Perception*, Gallimard, Paris, 1945.

(12) L. WITTGENSTEIN, *On Certainty*, Harper Torch Book, New York, 1972, p. 62.

En effet, il est difficile de voir comment les diverses façons subtiles dont les choses nous concernent peuvent être entièrement énumérées. Nous pouvons anticiper et comprendre la réaction de Jacques parce que nous nous rappelons ce que c'est d'être amusé, stupéfié, incrédule, déçu, contrarié, attristé, ennuyé, dégoûté, bouleversé, en colère, furieux, scandalisé, etc. et nous reconnaissons ce que ces divers sentiments nous inciteraient à faire. Pour programmer l'ordinateur on aura besoin de lui fournir une description de chaque nuance des sentiments aussi bien que l'occurrence normale et le résultat probable de chacun.

L'idée que les sentiments, les souvenirs, et les images *doivent* être la partie consciente d'une description explicite inconsciente se trouve confrontée à la fois par l'évidence *prima facie* et le problème d'explication des conditions *ceteris paribus*. De plus, ce préjugé mécaniste n'est supporté par aucune miette d'évidence scientifique venant de la neurophysiologie ou de la psychologie, ou des succès passés de l'IA, dont les échecs répétés ont motivé auparavant l'appel à cette même métaphysique. Quand les travailleurs de l'IA feront finalement face et analyseront leurs échecs, il se peut très bien que ce qu'ils auront à rejeter se trouvera être le préjugé métaphysique/mécaniste.

Si l'on passe en revue les vingt dernières années de recherche de l'IA nous pouvons dire que le point essentiel qui émerge est que *puisque l'intelligence doit être située elle ne peut pas être séparée du reste de la vie humaine*. Cependant, le refus obstiné de ce fait apparemment évident ne peut pas être mis sur le compte de l'IA. Il commence avec Platon qui a séparé l'intellect ou âme rationnelle du corps avec ses

compétences, ses émotions, et ses appétits. Aristote a continué cette peu vraisemblable dichotomie quand il sépara le théorique du pratique, et définit l'homme comme un animal rationnel — comme si l'on pouvait séparer la raison de l'homme de ses besoins et appétits animaux. Si on pense à l'importance des facultés sensori-motrices dans le développement de notre capacité à reconnaître et manier les objets, ou au rôle des besoins et des désirs dans la structuration de toutes les situations sociales, ou, en fin de compte, à l'ensemble du fond culturel d'interprétation de la vie humaine implicite dans nos actes quotidiens, quand nous identifions et utilisons des chaises, l'idée que nous pouvons ignorer ce savoir-faire, tout en formalisant notre compréhension intellectuelle comme un système complexe de faits et de règles, est hautement improbable.

Pourtant incroyable, cette dichotomie douteuse imprègne maintenant notre façon de penser à propos de tout, y compris les ordinateurs. Dans la série télévisée *Star Trek*, l'épisode intitulé « Le Retour des Archons » raconte ce qui arriva après que le sage homme d'Etat Landru ait programmé un ordinateur pour que celui-ci dirige la société après sa mort. Malheureusement, comme il n'a pu donner à l'ordinateur que son intelligence abstraite, non sa sagesse concrète, l'ordinateur changea cette société en un enfer planifié. Personne ne s'arrête un moment pour se demander comment, sans les compétences, les sentiments, et les intérêts propres à Landru, l'ordinateur pouvait vraiment comprendre les situations quotidiennes et ainsi du tout diriger la société.

Les vrais artistes ont toujours pressenti la vérité, obstinément niée à la fois par les philosophes et les techniciens, que, précisément parce que l'homme a un corps, il ne peut pas posséder la clarté caractéristique de l'ordinateur. Les artistes sentent que la base de la compréhension humaine ne peut pas être isolée et comprise explicitement. Dans son livre *Moby Dick*, Herman Melville écrit à propos du sauvage tatoué, Queequeg, qu'il a « inscrit sur son corps une théorie complète des cieux et de la terre, et un traité mystique sur l'art d'atteindre la vérité. Ainsi Queequeg dans sa propre personne était une devinette à résoudre, une œuvre merveilleuse en un volume ; mais les mystères de cette œuvre, même pas lui-même ne pouvait les lire... » Yeats exprime d'une façon encore plus succincte le respect du poète pour les limites que pose notre incarnation : « J'ai trouvé ce que je voulais — pour le dire en une phrase, je dis : « L'homme peut incarner la vérité, mais il ne peut pas la connaître ». »

HUBERT L. DREYFUS.

(traduction de Maurice Boissier.)

LA PHILOSOPHIE ANALYTIQUE ET L'HISTOIRE DE LA PHILOSOPHIE

L'histoire de la philosophie compte-t-elle encore comme philosophie ? Personne ne soutiendrait que l'histoire de l'art est de l'art ou que faire des recherches sur l'histoire de la cuisine c'est cuisiner. Si l'on soutient que toute recherche sur l'histoire de la philosophie est philosopher, il s'agit là d'une affirmation d'un type particulier qui demande à être justifiée. Dire que la connaissance de l'histoire de la philosophie constitue un arrière-plan nécessaire à la pensée philosophique est une chose, affirmer qu'une telle connaissance est partie intégrante de la philosophie en est une autre.

1. De nombreuses recherches sur l'histoire de la philosophie n'ont de toute évidence rien à voir avec la philosophie en tant que telle. Ainsi l'attribution d'une date ou d'un auteur à un ouvrage philosophique au moyen de l'étude de l'occurrence des termes ou des références faites à des ouvrages contemporains est un travail de détective où la philosophie n'est nullement impliquée. (Il existe bien sûr des travaux de chronologie qui requièrent essentiellement l'intuition et l'argumentation philosophiques. Un bon exemple en est le travail de G.E.L. Owen (1), professeur à Cambridge, qui a rétabli la chronologie du *Timée* de Platon). Y a-t-il d'autres aspects de l'histoire de la philosophie qui la rendent essentielle à la philosophie ? Parmi les philosophes d'orientation analytique, il y a sans doute plus de désaccord sinon de scepticisme au sujet de cette question, que dans la tradition continentale.

2. Hegel a dit dans ses *Leçons sur l'histoire de la philosophie* que l'histoire de la philosophie n'est pas une succession de points de vue, mais la saisie du développement d'une *idée*. Pour lui, l'enchaînement historique des systèmes philosophiques est analogue à la déduction logique de la détermination

(1) G.E.L. OWEN, « The Place of the *Timaeus* in the Order of Plato's Dialogues », *Classical Quarterly*, 1953.