Peut-on comparer le cerveau avec une machine qui pense ?

Il est vrai que Descartes (1596-1650) et Leibniz (1746-1716) oppose étendue et pensée comme deux régions radicalement opposées et rejettent la machine du côté du mouvement mécanique de l'étendue ; Ils opposent plus précisément la démarche même de la pensée comme réflexion aux fonctions ordinairement attribuées au psychisme et qui seraient à ses yeux explicables par les seuls mouvement mécaniques de l'étendue : comme le chien qui a soif, qui possède donc un signal biologique interne sur le fonctionnement de son corps, il ne sait pas qu'il a soif, ne juge pas qu'il a soif. Pourtant la mise au point de machines capables d'effectuer des calculs infiniment compliqués mais aussi de simuler les comportements humains a remis au goût du jour l'idée d'une machine à penser et permis d'envisager la possibilité de construire une conscience artificielle.

Le débat fait rage pour savoir si l'investigation des fonctions de l'esprit a l'aide de programmes d'ordinateurs est plus qu'une simulation. Est-elle une véritable explication des fonctions de la pensée ou bien simplement un modèle artificiel qui peut approcher certains comportements humains qui reste qualitativement différent: un robot programmé pour écouter la messe, se mettre a genoux au bon moment, distribuer des aumônes aux pauvres et faire ses prières serait-il un bon chrétien ? On peut en douter : il lui manque des pêchés a confesser. ...

Il y a en effet deux différences essentielles entre la machine la plus complexe et l'esprit humain. D'abord un ordinateur exécute un programme (c'est le principe même de l'intelligence artificielle) mais ne pense pas. Si penser, c'est traiter des informations, alors l'esprit humain ne se distinguerait pas fondamentalement d'un ordinateur qui calcule. La pensée serait alors au cerveau ce que le logiciel informatique est à la machine, c'est-à-dire un programme, une série d'instructions, traité sous forme de symboles liés entre eux par des règles logiques. La pensée serait donc une suite de calculs, d'opération logico-mathématiques inconscientes, sur des symboles dans ce langage. Mais alors comment comprendre le sens ? En effet, un programme informatique n'a pas de sens — ou alors il n'a que le sens que lui ont mis les programmateurs. Certes la machine parvient dans ses opérations à imiter la pensée humaine. Or imiter, simuler n'implique pas de reproduire le sens. C'est ce que tend à démontrer une expérience de pensée proposée par John R. Searle : la chambre chinoise. Je suis enfermé dans une chambre close et par une fente du mur, on me passe des questions écrites en chinois. Je n'y comprends rien, mais j'ai un manuel de réponses toutes prêtes et des règles pour associer un idéogramme de question à un idéogramme de réponse. Il est probable que je donne des réponses sensées sans pour autant connaître un seul mot de chinois. Mais rien ne fait sens pour moi. Bref, j'imite la pensée en suivant des algorithmes mais je ne pense pas. Un ordinateur peut en effet « parler » japonais, mais non « penser » japonais.

Autre différence essentielle et qui découle de la première, la machine ne sait ce qu'elle fait. L'ordinateur ne le sait pas pour la raison simple qu'il ne distingue pas ses états internes d'états externes correspondants. L'ordinateur peut a partir d'informations dresser des cartes de paysages de Bretagne, il peut les décrire, afficher "paysage de Bretagne", il ne fait pas référence a la Bretagne, qu'il ne peut identifier. Le pourrait-il d'ailleurs, qu'il ne pourrait distinguer ce qu'il y a sur son écran, de ce qui est hors de lui, c'est-à-dire distinguer les états internes et les états externes, même s'il était muni de capteurs pour "percevoir" les paysages. L'ordinateur ne pourra jamais viser intentionnellement ce qu'il décrit. L'intentionnalité c'est le fait que la conscience est en relation avec l'extérieur et à conscience de cette relation. Il faut distinguer le calcul qui est une suite d'opérations logiques et la pensée qui est intentionnelle. Au mieux donc l'ordinateur serait dans la situation du cerveau dans la cuve de Putnam qui pourrait tout penser et percevoir comme nous, sauf qu'il est un cerveau dans une cuve. Il imagine un cerveau détaché du reste de l'organisme, plongé dans un bain qui le nourrit et qui reçoit toutes sortes d'informations qui simulent la perception effective des choses. La question est simple: le cerveau peut-il distinguer cette sorte d'hallucination permanente d'avec l'expérience courante perceptive qui est la notre?

La réponse de Putnam c'est que le cerveau pourrait tout penser, sauf qu'il est un cerveau dans une cuve. Donc il ne saurait séparer ses informations comme états internes des états externes du monde qui pour nous y correspondent. Le cerveau ne pourrait faire la différence entre image d'arbre et arbre réel. Il ne peut faire la différence entre l'idée qu'il est dans la cuve et le fait d'y être. Le cerveau immerge est dans la situation de l'ordinateur.

Mais du coup l'ordinateur n'est pas vraiment autoreferentiel. Il peut bien afficher "connexion modem inopérante" ou "recharger l'imprimante", quand il est en panne, il ne lit pas son état et ne lie pas son état avec le fait qu'il l'affiche. Il affiche en panne parce qu'Il l'est, non parce qu'il sait qu'il l'est. En quoi consisterait l'auto-referentialite de la machine si elle réfléchissait? A se citer, a mettre ce qu'elle fait entre guillemets. Ce serait marquer ses actes comme des citations de soi. La machine peut bien manipuler des symboles généraux, éventuellement modifier partiellement ses programmes face a un échec, elle ne les cite pas, n'en fait pas la métalangue. L'échiquier électronique va inférer sur les situations, appliquer bien sûr les règles, trancher entre deux exigences, temps contre pièce par exemple, il ne fait en même temps pas la théorie du jeu d'échec. S'il la fait, c'est par un programme supplémentaire spécial. Et les deux ne coïncident pas. Alors que la pensée humaine ne coupe pas langue et métalangue, théorie et meta-théorie. Le lien entre référentialité et autoréférentialité s'éclaire: cette dernière consiste a rapporter le processus de la pensée au locuteur.

C’est donc l’intentionnalité qui constitue la structure de la conscience, ce rapport à soi et cette capacité à distinguer ses états internes de ces états externes. Le fait qu’elle porte sur autre chose qu’elle même et qu’elle en a conscience.

Il n’est donc pas possible de réduire l’esprit à des processus mécaniques ou à des microcircuits électriques et considérer, comme le fait les neurosciences, le corps humain comme une simple enveloppe externe. La réduction des faits mentaux à des processus neuronaux, des a conduit les neurosciences à identifier le cerveau humain à une « machine qui pense ». Cette identification suppose que l’on évacue l’aspect subjectif des actes mentaux au profit de processus cérébraux et de mécanismes logiques. L’esprit ou la conscience pourrait ainsi exister indépendamment du cerveau ou du corps humain.

Or nous nous avons montré qu’il y a quelque chose d'irréductible dans la conscience : irréductible à une approche matérialiste qui tenterait d’expliquer objectivement l’origine et le fonctionnement de la pensée (comme si la pensée était localisée dans le cerveau) ; irréductible également à une approche spiritualiste qui tenterait d’étudier les faits mentaux et intentionnels indépendamment de leur relation au corps (comme si la conscience ou l’esprit pouvait exister sans le cerveau).