

L'esprit est-il mécanisable ?

Jean-François Lambert

Maître de conférences honoraires en Neurosciences à l'Université Paris 8

Le Président : Cher collègue,

Je dis collègue, car vous avez rejoint formellement l'AES en 2015 après y être déjà intervenu en 2008 et avoir assisté régulièrement aux séances depuis plus de 10 ans.

Vous avez été Maître de Conférences en Neurosciences à Paris 6 (Pierre et Marie Curie) et Paris 8 (Vincennes Saint-Denis) et membre de nombreux et divers Conseils universitaires.

Vous êtes toujours enseignant à l'IPC et Associé au laboratoire CHArt (Cognition Humaine et Artificielle) de Paris 8 / Cité des Sciences

Vous avez travaillé d'abord sur les interactions entre sommeil, mémoire et environnement, puis sur les corrélats neurobiologiques de certains états mentaux associés à la dépression, et diverses formes de psychothérapie corporelle (dans le cadre du paradigme dit de la « cognition incarnée ») avec toujours une réflexion sur les enjeux anthropologiques de ces pratiques.

Par ailleurs, vous vous êtes particulièrement intéressé aux travaux de Benjamin Libet sur le libre arbitre et continuez de participer aux débats qu'ils suscitent, dans le domaine du droit notamment.

Vous accompagnez plusieurs recherches utilisant le petit robot Nao sur des applications auprès des enfants en difficulté et auprès des personnes âgées.

Parmi les nombreux « penseurs » et « passeurs » de connaissances et de valeurs qui ont marqués

vos parcours, vous faites une place particulière à André Soulairac, psychiatre et titulaire de la première chaire de psychophysiologie à la Sorbonne, Henri Laborit, chirurgien militaire, biologiste inclassable, « inventeur » du premier neuroleptique et votre cher ami Jacques Arsac dont votre exposé portera fortement la marque. Et puis, parmi les nombreux prêtres et théologiens que vous avez connus, Gustave Martelet, dont vous n'avez cependant pas toujours partagé la lecture qu'il faisait de Teilhard mais qui vous a convaincu de l'intérêt d'une approche ternaire (corps, âme, esprit) de la personne humaine.

Vous êtes issu d'une famille catholique marquée par le scoutisme et les mouvements d'action catholique, comme le CCFD. Vous avez également été très engagé à Amnesty International, dont vous avez été président de la section française en 1979 et que vous quittez en 1982 pour divergence de vue sur l'évolution idéologique de l'association.

Vous vous recentrez alors davantage sur les enjeux épistémologiques et anthropologiques des neurosciences tout en cherchant toujours à prendre en compte à la fois les réalités psychosociales des personnes et les données, incontournables, de notre fonctionnement organique, en priorité celui du cerveau.

Je suis heureux de vous laisser la parole pour cette cinquième séance de notre cycle « Dépasser l'humain » qui traitera de « l'esprit est-il mécanisable ? »

Prologue

L'exposé qui suit est précédé de la présentation d'exemples d'interaction avec différents types de robots. D'abord un « dialogue » entre un enfant et le robot Nao (le jeu dit de la « course à 20 »). Nao demande à l'enfant s'il veut jouer avec lui, et lui explique ensuite qu'il doit choisir un nombre entre 1 et 3. Le robot répète alors le choix de l'enfant, ajoute lui-même un nombre entre 1 et 3, fait la somme et annonce à l'enfant que c'est à lui de proposer à nouveau un nombre entre 1 et 3. Le cycle recommence : le robot répète, propose un nombre, fait la somme et relance l'enfant. Le gagnant est celui qui arrive le premier à 20. L'enfant doit donc faire preuve d'un minimum de stratégie, en tenant compte des réponses du robot, pour atteindre l'objectif. Il y a plusieurs niveaux de programmation du robot : on peut le faire tricher ou distraire l'attention de l'enfant. Ici, l'enfant a gagné et s'est entendu dire par Nao « Bravo ! tu as gagné... tu es très fort ». La démonstration est très spectaculaire.

Ensuite, l'exemple de robots « dessinateurs » a complété la présentation. Pour les uns, il s'agit de mobiles se déplaçant et traçant des courbes de différentes couleurs, de façon semi-aléatoire. Pour les autres, du couplage entre un « œil » (une caméra) et un bras articulé dont la main tient un crayon et dessine le « modèle » posé devant eux (exposition Artistes et Robots, Paris, Grand Palais, 2018).

Enfin, un « dialogue de sourd » entre un enfant et le robot Pepper censé guider les visiteurs à l'entrée de la Cité des Sciences, à la Villette, illustre les limites d'un tel « dialogue ». Le robot détecte la présence de l'enfant et l'accueille en lui disant qu'il peut lui présenter les activités adaptées à son âge. L'enfant répond « oui » et dit « j'ai quatre ans ». Le robot exécute quelques mouvements de la tête et des bras et répond « fais un effort ». L'enfant et sa mère répètent plusieurs fois « j'ai quatre ans ». Le robot répète, à son tour, « je n'ai pas bien compris », puis, après que l'enfant ait répété « j'ai quatre ans », il s'entend répondre « je crois que tu es parti... à bientôt... sympa de te voir ». Pepper a encore des progrès à faire !

Jean-François Lambert :

« L'homme occidental n'a pu se constituer à ses propres yeux comme objet de science (...) qu'en référence à sa propre destruction ».
(Michel Foucault, *Naissance de la clinique*)

« Seul l'esprit, s'il souffle sur la glaise, peut créer l'homme ».
(A. de Saint-Exupéry *Terre des hommes*)

L'esprit est-il mécanisable ? Est-ce bien la question ? Il conviendrait, certes, de préciser de quel genre d'esprit et de quel genre de mécanisme il s'agit. Mais, par ailleurs, de nombreux neuroscientifiques et philosophes sont aujourd'hui convaincus que ce que nous continuons d'appeler l'esprit est bien le produit d'un « mécanisme », un mécanisme neurochimique et fruit de l'évolution, certes, mais totalement réductible aux lois de la physique¹. Pour beaucoup de scientifiques (sinon la majorité), l'esprit est déjà « mécanisé ». La question qui nous occupe concernerait alors seulement la nature du mécanisme et non son

principe. Quel est le rapport de l'esprit au substrat physique dont il serait le produit ? Le support neurochimique naturel est-il le seul susceptible de « produire » de l'esprit, ou ce dernier est-il sous-déterminé² par la nature du substrat physicochimique ? Si, en effet, l'esprit ne dépend que de la structure formelle des opérations de traitement de l'information, la nature physique du système qui effectue ces opérations est indifférente. Notre question d'origine devient alors celle du possible passage d'un substrat (naturel, le cerveau) à un autre (artificiel, l'ordinateur). Certes, cette question du rapport des opérations au substrat est tout-à-fait importante (et loin d'être résolue) mais ce n'est pas ici l'essentiel. En effet, avant de s'interroger sur l'éventualité de la production d'un « esprit » par un mécanisme différent du cerveau, encore faudrait-il que l'esprit soit bien, naturellement, une « production » dudit

¹ Selon le physiologiste Georges Cabanis, le cerveau « digère » les impressions comme l'estomac digère les aliments, et opère ainsi la sécrétion de la pensée : « le cerveau secrète la pensée comme le foie, la bile » (voir Changeux, p.25). Beaucoup pensent aujourd'hui que « l'homme n'a dès lors, plus rien à faire de l'Esprit, il lui suffit d'être un Homme Neuronal » (Changeux, p.227).

² C'est, plus généralement, la question du rapport entre structure et fonction qui est posée ici. Le concepteur du premier avion, Clément Ader (selon les français !), dota son engin d'ailes copiées sur celles des oiseaux (plus exactement des chauves-souris). Or, le vol d'un avion n'a rien à voir avec celui d'un oiseau et, par conséquent, avec la forme de leurs ailes respectives. De fait, pour faire voler un avion, il ne faut pas copier le vol de l'oiseau. De même, un sablier et une horloge permettent tous deux une mesure du temps. Pourtant, simuler une horloge avec un sablier ne fait pas de ce dernier une horloge. L'identité fonctionnelle n'implique pas celle des substrats.

cerveau ! C'est donc principalement à la question de savoir si l'esprit est, naturellement, déjà le produit du mécanisme cérébral que nous nous attacherons particulièrement.

Notre propos s'articule en trois parties. Nous montrerons d'abord comment les questions soulevées aujourd'hui par la théorie de l'information s'inscrivent dans un processus de « réduction » de la pensée, déjà revendiqué par le nominalisme. Nous reviendrons ensuite sur le paradigme cognitiviste pour montrer les impasses auxquelles il conduit (le résultat du calcul ne vaut que pour quelqu'un) et que finalement, rien ne pense, ni dans une machine, ni dans le cerveau. Nous envisagerons enfin comment l'anthropologie biblique permet de prendre pleinement en compte les conditionnements naturels de la personne sans que celle-ci se réduise à ceux-là.

1 - DU NOMINALISME AUX SCIENCES COGNITIVES

Information et signification

Retenons déjà de nos premières réflexions que l'essentiel du débat ne porte plus, principalement, sur les propriétés intrinsèques de la matière-énergie, mais sur l'information qu'elle supporte et qui l'organise. Il convient donc de ne pas se méprendre sur la nature et la signification de ce terme (l'information) qui conditionne toute la problématique des sciences cognitives et plus généralement de la convergence NBIC et du transhumanisme. Je veux profiter de l'occasion qui m'est ici offerte de rendre hommage à Jacques Arsac³ dont le discernement, en matière d'information, est exemplaire. « *Je réalisai tout à coup - écrit-il - qu'une bataille s'était engagée autour de l'intelligence artificielle, à laquelle je ne pouvais rester indifférent, parce qu'elle met en jeu une vision de l'homme : si une intelligence artificielle est possible, alors je suis une machine. Tout mon système de valeurs s'oppose à une telle idée. Il fallait que je le dise* » (p.8). Pour lui, « *l'intelligence artificielle fait problème parce que (...) certains affirment que toutes les activités intellectuelles de l'homme peuvent être transférées aux machines qui les feront mieux que nous* » (p.10). Certes, insiste-t-il, « *les problèmes posés par cette idée d'intelligence artificielle ne sont pas nouveaux (...) Les mots ont-ils un sens ? Nos raisonnements sont-ils réductibles à des calculs (...) Si les questions sont anciennes, peut-être la façon de les considérer a-t-elle changé* » (p.10). L'essentiel, pour Arsac, est de bien

comprendre que « *L'information est un texte susceptible d'apporter une connaissance. Par là - précise-t-il - nous séparons radicalement l'information de la connaissance qu'elle peut éventuellement apporter. Nous séparons même le texte du sens qu'il ne possède que pour certains. Le texte existe pour tous, le sens pour quelques-uns* » (p.44). La définition que l'Académie a donnée de l'informatique confirme cette séparation entre le sens et la forme qui le porte : « *Science du traitement rationnel de l'information considérée comme le support des connaissances* » (p.41). Arsac insiste à de nombreuses reprises sur le fait que « *L'information numérique a donc ceci de très particulier qu'elle est un contenant sans contenu, une forme sans signification* » (p.55) « *l'informatique n'accède qu'à l'information, pas à la connaissance (...) cette connaissance est extérieure au traitement qui est fait : c'est nous qui savons, pas la machine* » (p.57). En conséquence, « *L'informatique ne traite pas les connaissances, mais seulement les informations* » (p.65).

Soulignons encore que cette position n'est pas celle d'un informaticien isolé, réfractaire au progrès, mais qu'elle est partagée par beaucoup parmi les spécialistes de la théorie de l'information. Ainsi, pour Henri Atlan « *la théorie de l'information ne cherche qu'à mesurer l'information que nous apporte un événement du seul fait qu'il s'est produit, c'est-à-dire qu'on ne tiendra pas compte de la signification que cet événement peut avoir, ou ne pas avoir, pour nous* » (p.7). *L'information apportée par un événement est d'autant plus grande que sa probabilité de survenue était faible (...). Information (...) ne doit pas être confondu ici avec signification d'un message* » (p.8). De même, pour le philosophe Hubert Dreyfus, « *Information ne doit pas être confondue avec signification d'un message* » et « *les aspects sémantiques de la communication n'entrent pas en ligne de compte* » (p.204).

L'information et la mondialisation

Même si cela devrait être davantage argumenté que je ne peux le faire maintenant, je voudrais souligner ici combien le projet de naturalisation de l'esprit, cher aux sciences cognitives, s'inscrit dans l'histoire de la pensée occidentale et contribue à l'idéologie mondialiste⁴. Il existe, on le sait, un lien étroit entre pouvoir et langage, et cela d'autant plus qu'un pouvoir universel suppose un langage universel. Ce langage universel est aujourd'hui disponible : c'est celui de

³ Pour tous les auteurs cités, voir bibliographie.

⁴ Je reprends ici, en partie, *La revanche de Babel*, La Nef (Lambert, 1994).

l'informatique. La logique de l'informatique et celle du marché vont de pair car elles sont toutes deux le produit de cette fracture majeure qu'a constitué, au sein de la pensée occidentale, la querelle des universaux qui se solda, on le sait, par l'abandon du réalisme modéré défendu par Thomas d'Aquin au profit du nominalisme revendiqué par Guillaume d'Occam.

Les conceptions de Guillaume d'Occam sont reprises par les nominalistes contemporains pour évacuer l'idée même de sens. Le mot est un signe (un « symbole ») et la proposition se réduit à une suite de signes considérés comme de simples caractères. La pensée se ramène alors à une manipulation de symboles à laquelle la notion de sens est totalement étrangère. On est bien ici au cœur des idées contemporaines sur l'intelligence artificielle et la psychologie cognitive qui conçoivent l'esprit sur le modèle d'une machine capable d'accomplir certaines opérations (calculs), c'est-à-dire d'appliquer des règles de transformation à des chaînes de caractères, indépendamment de leur éventuelle signification (voir *infra*).

L'élaboration d'un langage artificiel procède d'une double nécessité : seul un langage artificiel est capable de s'imposer à tous comme le langage universel dont l'idéologie libérale a besoin (l'artificialité garantit l'universalité) et seul un langage artificiel est susceptible d'être complètement maîtrisé (l'artificialité garantit la calculabilité). L'ambition du libéralisme est bien celle de l'universalité. Le marché mondial est une réalité qui doit déboucher sur une organisation elle-même mondiale. Le discours et la rationalité libérale exigent, au-delà des échanges quotidiens, une règle admise par tous. Or la régulation du jeu de la concurrence suppose une transparence qui n'est pas possible dans la diversité langagière.

En fait, le libéralisme ne cantonne pas ses ambitions à la création et à l'organisation d'un marché mondial, mais vise à constituer une société civile universelle, une communauté mondiale d'échanges assujettie à un droit civil universel fondé sur l'autonomie des individus inspirant une morale - celle des Droits de l'Homme - qui n'est rien d'autre que la règle du jeu de ce marché mondial. La Déclaration des Droits de l'Homme fonde son universalité sur les lois du marché, mais un marché tel qu'il soit susceptible d'exprimer tous les échanges entre les hommes. Il faut alors organiser l'échange verbal de telle sorte qu'il permette de maîtriser le discours universel lui-même.

La recherche d'un langage universel capable d'assurer la cohérence des échanges dans le cadre d'un marché mondial est un prolongement nécessaire au discours libéral. Pour qu'un langage véritablement universel soit possible, il faut que la totalité de la logique humaine - et du discours - soit mathématisable. Il faut que le langage soit réductible au calcul, c'est-à-dire à la manipulation de symboles. Ces conditions sont précisément aujourd'hui remplies par les sciences cognitivistes, dernier avatar du nominalisme.

La singularité humaine contestée

Il convient encore de rappeler que la science s'est imposée en occident aux dépens d'une triple mise en question de la position singulière de l'homme dans la nature. Au décentrage cosmologique de la révolution copernicienne a succédé le décentrage biologique et phylétique du darwinisme puis le décentrage psychologique de l'inconscient. Freud, on le sait, parle à ce propos d'une triple humiliation infligée par la science à l'orgueil et au narcissisme humains. Il existe, en effet, un lien étroit entre la notion d'inconscient et la conception darwinienne de l'évolution qui ont, ensemble, contribué à disqualifier l'idée d'une nature humaine irréductible à ses conditionnements biologiques⁵. En fait, cette volonté de banalisation de l'espèce humaine s'inscrit dans un processus plus vaste de contestation de l'ethnocentrisme inhérent à la conception judéo-chrétienne de la nature comme création, conception qui accorde à la conscience humaine un statut privilégié et introduit une discontinuité dans l'évolution. Cette notion de discontinuité fondatrice justifie par ailleurs les distinctions classiquement admises entre homme et animal, nature et culture, homme et femme, normal et pathologique, naturel et artificiel. Or, ce sont bien toutes ces distinctions qui sont aujourd'hui contestées⁶ et cette contestation trouve dans le darwinisme et l'inconscient ses principales justifications.

Ainsi, de façons certes fort différentes, Darwin et Freud, le libéralisme et le marché, le relativisme

⁵ Voir Marcel Gauchet, *L'inconscient cérébral*. Seuil, 1992.

⁶ « Nous voulons brouiller les frontières établies entre hommes et femmes, humains et machines, humains et animaux, et expérimenter des changements de subjectivité » (Kina, performeuse du groupe espagnol Quimera rosa), *Libération, Hors Série : l'IA au cœur de l'humain*, déc. 2018, p.46. Notons également qu'au Parlement Européen, le lobby LGBT soutient activement l'idée d'accorder aux robots un statut de personne (électronique) ! On sait, par ailleurs, que le Dalaï Lama approuve et encourage le projet « Avatar 2045 » (2045's Avatar Project) dont le but est d'atteindre, en 2045, « l'immortalité cybernétique » par la fusion de l'homme et de la machine.

et le naturalisme, sont des produits du nominalisme que la notion d'information permet désormais de fédérer à travers les SC et ses prolongements transhumanistes.

2 - LE CERVEAU NE PENSE PAS

Retour sur quelques définitions

Bien que souvent utilisées de manière indifférenciée, les notions de psychisme, d'esprit, de pensée ou de conscience doivent être distinguées. Ce n'est pas l'objet principal de cette communication, mais il convient cependant d'y prêter attention.

Le psychisme peut être considéré comme inhérent à tout être vivant. Au sens large, ce terme désigne le fait, pour tout organisme (vivant) d'être le siège de processus (pour l'essentiel, inconscients) dont la finalité est la conservation de son identité (de sa structure). Le psychisme, ainsi entendu, peut être identifié à l'âme (végétative, sensitive, intellectuelle) considérée comme détentrice et/ou génératrice de l'information (de ce qui donne la forme). Bien que non-réductible à la matière-énergie qu'elle « anime », l'âme (*psukhê*) n'a pas ici de connotation spirituelle : il s'agit d'une propriété « naturelle » des organismes vivants, à laquelle on peut rattacher la notion d'homéostasie (Laborit, Damasio) ou celle d'autopoïèse (Varela). Le psychisme est « incorporé », il fait corps avec l'organisme.

Quant à l'esprit, le même mot désigne, on le sait, à la fois le mental, la cognition (souvent réduite à la computation), c'est-à-dire le *mind*, et le principe même de l'être connaissant, le *pneuma* ou *spirit*. Pour compliquer les choses, l'esprit dans sa première acception (*mind*) peut désigner soit l'opérateur (le système de traitement), soit les opérations (la mise en œuvre de règles d'inférence), soit encore le résultat (le contenu) de ces opérations (les connaissances). Pour les sciences cognitives l'esprit s'identifie strictement au mental, lui-même, en principe, réductible à la computation. Doit-on alors considérer que toute forme de computation est une forme d'esprit ? Ou bien existe-t-il une forme d'esprit irréductible au mental, lui-même irréductible à la computation ? Nous savons bien qu'il existe de la computation sans esprit conscient et qu'il n'est pas besoin de revendiquer l'existence d'un opérateur métaphysique pour rendre compte du traitement de l'information par un système physique ou biologique. En revanche, le résultat ne vaut pas en lui-même mais pour quelqu'un qui l'assume comme tel (nous y reviendrons).

Quant à la conscience, elle est souvent confondue, par les physiologistes, avec le fait d'être éveillé ; elle serait ce que l'on perd lorsque l'on s'endort le soir et ce que l'on retrouve le lendemain matin au réveil. Il vaudrait mieux ne parler ici que de vigilance et non de conscience pour désigner le seul fait d'être plus ou moins « éveillé » en fonction d'une hiérarchie de niveaux d'activation cérébrale, mesurables, du coma aux épisodes paroxystiques. En fait, vigilance et conscience ne sont pas strictement corrélés (par exemple dans certains états consécutifs au coma).

La conscience proprement dite désigne d'abord l'*intuition* (pur ressenti) qu'a le sujet de sa propre expérience subjective, indépendamment de toute connaissance réflexive. On parle ici de conscience *primaire* (Edelman, 1992), de conscience *noyau* (Damasio, 1995) ou encore de conscience *phénoménale* (Ned Block, 1980). Notons que chercher à définir la conscience est déjà un acte de conscience : l'énoncé du problème implique les termes de sa solution (voir Bitbol).

Dans sa seconde acception, la conscience exprime la *connaissance* qu'a le sujet de cette expérience subjective : pensée, sentiment, perception, rêve, raisonnement. Ne dit-on pas que l'on perd connaissance lors d'une syncope ? On désigne cette seconde dimension comme conscience *secondaire* ou *d'ordre supérieur* (Edelman), conscience *étendue* (Damasio) ou conscience *d'accès* (Ned Block). Son étude objective suppose de prendre en compte le témoignage des sujets et donc de recourir à l'introspection.

La conscience désigne enfin la conscience de soi dans son rapport à l'autre et renvoie au sens ultime de l'action lorsqu'il s'agit de la conscience morale. La normativité est une dimension nécessaire de cette conscience morale telle qu'elle s'exprime dans le jugement éthique. Il s'agit d'une réalité axiologique qui dépasse largement le cadre des sciences de la nature, bien que la psychologie évolutionniste et la neuroéthique prétendent désormais fournir une justification purement naturaliste des valeurs morales.

Cerveau et conscience

Tout le monde (ou presque) s'accorde, aujourd'hui, à reconnaître qu'il n'existe pas de « centre » de la conscience mais qu'elle est associée au fonctionnement d'un vaste réseau de neurones impliquant différentes régions du cerveau (d'aucuns diront qu'elle émerge *de*, sinon qu'elle est produite *par*, ce réseau). Une telle perspective accorde une importance accrue aux

connexions entre les structures (connectome) et non seulement au fonctionnement exclusif de chacune. Selon les auteurs, on parle de liage perceptif, de coalitions de groupes neuronaux, de synchronie globale, de couplage oscillatoire d'assemblées de neurones distants. Le cortex frontal joue, certes, un rôle crucial au sein de telles coalitions, mais il n'est pas le support exclusif de l'expérience consciente. Le débat reste ouvert entre ceux qui revendiquent l'existence d'un (ou plusieurs) réseau(x) spécifique(s) de la conscience et ceux qui considèrent qu'elle résulte de la synchronisation d'ensembles cérébraux distants qui ne lui sont pas spécifiquement dédiés. Ainsi, selon la théorie dite de « l'espace de travail global conscient » développée par J.P Changeux (voir Dehaene), alors que les informations sensorielles sont traitées de manière inconsciente par de multiples circuits cérébraux distincts, l'accès conscient à ces informations active non pas un circuit ou une région spécifique mais un réseau de neurones beaucoup plus vaste. A partir de ses travaux sur le coma, Steven Laureys défend au contraire que notre conscience n'est pas répartie de manière diffuse dans tout le cerveau, mais dans des zones corticales spécifiques reliées entre elles. Il décrit ainsi un «réseau global de la conscience» constitué du thalamus, des aires associatives préfrontales et temporo-pariétales, du précunéus et du cortex cingulaire (la conscience suppose un minimum d'activité du réseau en question).

En revanche, certains continuent de refuser l'idée selon laquelle la pensée ou la conscience ne seraient qu'une production du cerveau. Selon le philosophe américain Alva Noë, par exemple, la conscience n'est pas du tout produite par notre cerveau. À la différence de la digestion, elle n'est pas le produit d'un organe, mais plutôt le résultat de l'interaction entre plusieurs cerveaux. Comme le souligne également le psychologue Guy Tiberghien (voir Guillaume *et. al.*), ce qui se passe dans nos cerveaux n'a pas pour seule origine notre cerveau. Ce n'est pas un cerveau qui pense mais une personne. Ce qui s'y passe vient aussi de l'environnement et se construit à travers une histoire. Ainsi, la pensée ne peut être comprise hors de tout contexte personnel, culturel, historique et social (nous y reviendrons).

La conscience et l'esprit ne sont pas synonymes, bien que, de fait, fortement corrélés. En tant que telle, la conscience n'est pas un acte mental mais une partie du processus privé que nous appelons esprit. Elle est bien ce qui permet à l'esprit d'exercer son emprise mais elle n'en est pas la substance. Dans la perspective cognitiviste, les

processus de traitement sont non seulement *cognitivement impénétrables* mais nécessairement non accessibles au soi conscient. Autrement dit, je ne peux pas être à la fois, « conscient de » et conscient du mécanisme qui me permet d'être « conscient de ». Nous n'avons pas accès au traitement des symboles mais seulement au résultat. Il n'existe donc pas de connexion nécessaire entre l'esprit (au sens computationnel de traitement de l'information) et la conscience (au sens de connaissance du résultat).

Le paradigme des sciences cognitives

Les Sciences Cognitives (SC) constituent - de fait - la tentative la plus récente - et certainement la plus accomplie - de naturalisation de l'esprit. Cela se traduit par le fait qu'elles abordent la pensée et la connaissance comme des phénomènes dont l'étude relève exclusivement des sciences de la nature de type physico-mathématique. Ce programme scientifique est « révolutionnaire » dans la mesure où, depuis Descartes, l'esprit était considéré comme non mathématisable, donc non mécanisable. Une telle naturalisation de l'esprit procède d'une triple réduction : réduction de l'esprit à la cognition, réduction de la cognition au calcul, réduction du calcul à un mécanisme (et donc réduction de l'esprit au mécanisme).

La première étape (réduction de la pensée au calcul) a été franchie depuis longtemps puisque Leibniz pensait « avoir conçu un élégant artifice en vertu duquel certaines relations peuvent être représentées et désignées de manière numérique » et pour Hobbes « le raisonnement n'est rien d'autre qu'un calcul ». Pour l'un et l'autre la pensée se ramène bien à un processus calculatoire. Ce désir de formalisation de la pensée est une constante de la rationalité occidentale dont on peut déjà trouver la trace chez Platon⁷.

La seconde étape n'a vraiment été franchie que récemment avec l'invention de la machine de Turing⁸ (formulation théorique de l'ordinateur) qui a permis de définir mathématiquement la notion de calcul comme un processus mécanique fini abstrait (sans oublier Pascal ou Lady Lovelace). La découverte d'une telle notion mathématique

⁷ Notamment dans le *Ménon*. Voir Dreyfus, p. 219.

⁸ La machine de Turing universelle (1937) est un dispositif automatique de balayage, de lecture et d'impression associé à une bande de longueur infinie divisée en cases. La machine fonctionne par une succession d'étapes. A chaque étape elle lit un symbole sur une case, imprime un nouveau symbole (ou le même) sur cette case et se déplace d'une case à droite ou à gauche, ou bien s'arrête. La machine (formelle) est ainsi définie par une succession d'états. Un programme spécifique pour chaque état et chaque symbole lu sur la bande quel doit être l'état suivant et l'action sur la bande.

rendait donc possible en pratique le projet de naturalisation de l'esprit, pour peu que l'on accepte d'identifier ce dernier au calcul (première étape). L'approche symbolique ou computationnelle postule, en effet, l'existence d'un langage représentationnel et assimile les processus cognitifs aux opérations d'une telle machine de Turing. L'esprit n'est alors rien d'autre qu'un système particulier de traitement de l'information (d'où l'importance de saisir précisément la nature de cette dernière, comme a cherché à le faire Jacques Arsac).

Parmi les influences dont le paradigme des sciences cognitives porte la marque (particulièrement celle du Cercle de Vienne et de la philosophie analytique), il convient de souligner l'importance de celle du « premier » Wittgenstein. Le *Tractatus* a, en effet, largement contribué à l'essor de la théorie de l'information et des conceptions neurocybernétiques de la pensée en offrant une configuration ontologique parfaitement adaptée à leur sujet (voir Ladrière). Le monde de « tout ce qui arrive », le monde des événements, est bien celui de la théorie de l'information, car celle-ci n'a pas de rapport avec la nature des symboles, mais seulement avec leur occurrence. Une science articulée par la théorie de l'information présuppose, en effet, l'existence d'un univers d'événements descriptibles et traduisibles en symboles discrets, indépendamment de leur nature et plus encore de leur éventuelle signification⁹.

Deux versions du paradigme cognitiviste : le modèle computo-symbolique et le modèle connexionniste ou neuromimétique

Pour l'essentiel, les SC se fondent sur la prise en compte du fait que, contrairement au paradigme behavioriste¹⁰, les comportements et les processus

⁹ Comme le fait remarquer Jean Ladrière, le réductionnisme n'a retenu de Wittgenstein que son apologie de la représentation alors qu'elle ne constitue pour lui qu'un passage obligé vers ce qui ne peut être dit, et qu'il appelle « l'élément mystique ».

¹⁰ La thèse empiriste a longtemps dominé la recherche en physiologie nerveuse. Déjà pour Willis vers 1660, tout mouvement provient d'une sensation et n'est que la restitution de l'énergie reçue de l'extérieur. Pour Setchenov, le maître de Pavlov, tous les actes psychiques, quels qu'ils soient, se développent de la même manière que les actions réflexes. La cause initiale de toute activité humaine se trouve *hors* de l'homme. Avec la notion de *réflexe conditionné* et le schéma Stimulus/Réponse, les travaux de Pavlov constituent l'aboutissement logique de la tradition empiriste - sensualiste. Le stimulus se *réfléchit* dans le mouvement qui restitue au milieu l'énergie de la sensation. Quant au behaviorisme américain (Watson), il se distingue principalement de la réflexologie pavlovienne par le fait qu'il refuse toute référence à un quelconque état interne. Il s'agit d'un

mentaux ne se réduisent pas à l'enchaînement réflexe de stimulus et de réponses, mais qu'il y a lieu de reconnaître entre les deux une médiation. Une telle médiation se traduit non seulement par l'existence d'un délai entre les entrées sensorielles et la réponse motrice mais également par un certain jeu, une certaine incertitude, dans le couplage sensori-moteur permettant la réintroduction des concepts d'image mentale ou d'intentionnalité qui étaient proscrits par le behaviorisme. Ce délai correspond à un processus susceptible d'une double description : en tant que processus physico-chimique, d'une part, et en tant que processus formel, informationnel, cognitif, d'autre part. La question de la pertinence de ce double niveau de description, et celle de la plus ou moins grande subordination de l'un à l'autre, constituent le principal enjeu épistémologique des sciences cognitives.

L'approche symbolique ou computationnelle postule l'existence d'un langage représentationnel et assimile, nous l'avons dit, les processus cognitifs aux opérations d'une machine de Turing (revoir note *supra*). Dans cette perspective le niveau formel est relativement indépendant du substrat qui n'impose que des contraintes limitées (dualité matériel/logiciel). L'approche connexionniste¹¹ ou neuromimétique considère, au contraire, que le raisonnement en termes de symboles est totalement inadapté à la problématique esprit-cerveau. Le neuro-calcul privilégie fortement la nature du substrat (réseau de neurones) et sa structure fonctionnelle (connectivité, poids synaptique...). Le recours aux modèles connexionnistes met en évidence une caractéristique essentielle du cerveau humain : la plasticité de sa structure dont l'organisation est susceptible d'être modifiée par le résultat des opérations qu'elle exécute. On peut simuler cette propriété à l'aide d'un réseau d'automates dont la connectivité (la force des liaisons entre éléments,

objectivisme et d'un positivisme absolus. Les sciences de la vie et singulièrement les sciences humaines restent encore largement dominées par une telle conception empiriste-mécaniste du vivant et de la *psyché*.

¹¹ Dans un réseau connexionniste on a affaire à la résultante instantanée de la coopération d'un ensemble d'automates fonctionnant simultanément, en parallèle, si bien qu'une représentation n'est plus attachée à un symbole donné : la représentation appartient au réseau tout entier. Ce que les machines connexionnistes procurent en fait c'est une autre façon de calculer. Ces réseaux sont, en effet, le siège de calculs non pas symboliques mais numériques qui correspondent à la simulation de systèmes dynamiques. On peut parler ici de cognitivisme physicaliste. En fait le connexionnisme ne constitue qu'un autre versant du même paradigme fondateur. Ce n'est jamais qu'un autre procédé de calcul opérant à partir de composants formels, eux-mêmes simulés sur ordinateur.

ou poids synaptique) se modifie pas à pas en fonction de l'écart mesuré entre le résultat attendu et le résultat effectif grâce à des propriétés de rétropropagation (*deep learning*). Le traitement s'opère ici à un niveau sub-symbolique. Ces dispositifs sont très efficaces dans la reconnaissance des formes y compris les formes linguistiques (par exemple la conjugaison des verbes) ou les objets mathématiques. L'idéologie cognitiviste trouve dans le connexionnisme matière à justifier sa tentation de ramener l'explication de l'univers à un système auto-suffisant dont le principe rationnel unique gouvernerait le niveau cognitif aussi bien que le niveau matériel et leurs interactions mutuelles. Une telle explication justifierait le présupposé de ceux qui la soutiennent, c'est-à-dire que l'homme n'est qu'un objet « naturel » comme un autre.

Quelle que soit la nature des neuro-calculs dans le cerveau, la question de savoir comment et pour qui le résultat fait sens (comment et pour qui l'information devient une connaissance) reste posée. C'est le même processus de réduction formelle qui est partout à l'oeuvre. Cependant, comme le souligne le philosophe Michel Bitbol (p.272), « Ce qui est réduit dans la machine de Turing c'est la séquence objectivée des étapes de l'inférence symbolique et non la pensée qui englobe l'expérience de reconnaître une conclusion comme vraie ».

Les limites du cognitivisme : rien ne voit ni ne pense dans le cerveau

Il existe, en effet, de nombreux exemples illustrant l'impossibilité sémantique absolue de rendre compte de quoi que ce soit en matière de perception, de mémoire, d'affects, de cognition ou d'action sans avoir recours à quelque pronom personnel désignant un sujet agissant et conscient. Ainsi, la connaissance, même ultra-précise de l'état d'activité du cerveau ne donne en réalité aucun accès à l'expérience subjective elle-même. On sait, par exemple, qu'il existe une organisation hautement spécifique du traitement de l'information visuelle. Les différentes caractéristiques de l'objet sont traitées en parallèle par des canaux distincts (couleur, forme, mouvement, position, contraste...). Or le sujet qui perçoit, celui pour qui cette activation fait sens (signifie une couleur), n'est réductible ni à des aires du cortex visuel, ni à aucune autre région du cerveau susceptible de les scruter car ce sont des couleurs (des objets) dans le monde et non des représentations dans le cerveau que nous percevons. Le cerveau ignore tout des stimulus

qu'il reçoit¹². C'est moi qui perçoit et qui pense et non mon cerveau ! Ainsi la vision n'est localisée ni dans la rétine, ni dans le nerf optique, ni dans les corps genouillés, ni dans aucune aire corticale : le cerveau ne voit pas. « Aucun pont conceptuel ne peut être édifié entre l'activité des neurones et le fait de l'expérience des couleurs » (Bitbol, p.275). Mes connaissances sur le cerveau ne changent en rien l'expérience de mon corps vécu. Il demeurera toujours un hiatus infranchissable entre la description à la troisième personne, la seule accessible à la science, et le vécu à la première personne. C'est quelqu'un qui voit, qui a faim ou se souvient et non son cortex occipital, son hypothalamus ou son hippocampe. Ce n'est ni un cerveau, ni une main qui a rédigé ce texte mais un être doté de tels instruments lui permettant d'exprimer des intentions. Comme le dit très justement, Michael Gazzaniga, nous sommes « des gens, pas des cerveaux » (p. 238).

Le cerveau n'est donc pas conscient. Un neurone n'est, en soi, pas plus conscient qu'une cellule cardiaque et le cerveau lui-même n'est pas plus conscient que le cœur. J'ai besoin d'un cerveau et d'un cœur (entre autres) pour être conscient, mais ni l'un ni l'autre ne le sont (d'eux-mêmes ou de moi). Bien plus, je peux être conscient de mon cœur qui bat alors que je ne peux pas être conscient de l'activité de mon cerveau quand je pense. Cela n'a donc aucun sens de distinguer des régions conscientes et inconscientes dans le cerveau. Il y a bien des régions plus nécessaires que d'autres à l'exercice de la raison, à la mémoire, à la conscience de soi et du monde, mais cela ne signifie évidemment pas que lesdites régions sont dotées de raison, de mémoire ou de conscience. Quelle qu'en soit la forme, attribuer de la conscience aux neurones n'a pas de sens autre que métaphorique. Je peux attribuer des propriétés à un dispositif qui ne les possède pas et décider de faire « comme si » il les possédait. Au demeurant, cela ne signifie évidemment pas qu'il les possède. Tous les mécanismes biologiques sont, par nature, inconscients bien que « moi » je puisse en prendre (partiellement) conscience. Comme nous l'avons déjà souligné, le mécanisme par lequel je peux être « conscient de » ne peut (ne doit) être lui-même qu'inconscient (revoir, *Cerveau et conscience*). C'est pourquoi, dire que le cerveau est « attentif », qu'il « décide », « hésite », que le cortex « cherche », « fait des efforts », « à son mot à dire » est typiquement un abus de langage. Dire qu'une aire corticale calcule ou décide, c'est confondre la partie et le tout en attribuant à une partie (fut-elle le cerveau dans son

¹² Voir Erwin Straus, p. 472.

ensemble) les propriétés de la personne comme totalité. C'est également confondre les conditions d'exercice d'une capacité avec cette capacité elle-même (erreur de catégorisation). L'activation d'une région n'équivaut pas à la détermination d'une fonction. Elle n'en est que l'instrument (la condition et non la cause, voir *infra*).

Impossible de ne pas faire référence à quelqu'un

Un mécanisme peut expliquer (causalement) la réalisation d'un processus *par* quelqu'un, mais il ne peut rendre compte de la raison d'agir *pour* quelqu'un. Nous sommes bien obligés de nous arrêter à un moment ou à un autre de décrire des processus physiques dans le cerveau et dire que telle configuration vaut comme résultat pour quelqu'un. De fait, tous les usages courants du terme conscience impliquent une telle limitation indexée par un nom propre ou un pronom personnel. Le sujet *pour qui* un processus fait sens n'est pas réductible au mécanisme *par qui* ce processus advient. L'enchaînement des processus physiologiques a pour limite une personne (désignée par nom propre ou un pronom personnel). C'est pourquoi, même une connaissance complète de la représentation neuronale ne saurait, sans être validée par le témoignage du sujet, nous dire quelle sensation est en train d'être subjectivement vécue (voir Libet). Il est donc faux de penser qu'on puisse effectivement comprendre la nature des expériences subjectives en étudiant seulement leurs corrélats comportementaux. Vous ne voyez pas ce que je vois quand vous examinez mon activité cérébrale au moment où je vois ce que je vois (Damasio, 1995, 1999). L'accès aux images mentales, sous leur forme consciente, n'est possible qu'à partir du point de vue de la première personne alors qu'au contraire, les configurations neuronales ne sont accessibles que du point de vue d'une tierce personne. Si j'avais la possibilité, grâce à une technologie adaptée, de contempler mes propres configurations neuronales ce serait toujours du point de vue de la troisième personne¹³.

On peut, certes, répéter de façon quasi-incantatoire que les images apparaissent « comme ça » dans le cerveau et conclure que (on ne sait comment), *en fin de compte*, les cartes neurales deviennent des images mentales. Mais c'est bien cet « en fin de compte » qui fait problème. Quelle qu'en soit la forme, la question de savoir comment le sens vient aux symboles reste sans

réponse dans un cadre strictement physicaliste. En fait les SC, comme la psychologie objective, ne peuvent pas exister sans « *un marché noir alimenté avec de la marchandise de contrebande venant de la psychologie existentielle* » (Straus, p.195).

C'est l'interprétation qui donne sens aux symboles. Un symbole symbolise quelque chose pour quelqu'un. Rien n'est intrinsèquement une représentation de quoi que ce soit ; quelque chose est une représentation seulement pour quelqu'un. Ainsi n'importe quelle représentation ou système de représentation requiert au moins un utilisateur et un interprète externes à cette représentation.

Il faut quelqu'un pour faire le calcul

Le terme de symbole désigne ici une configuration d'éléments assimilée à un signifiant dont il reste à préciser qui fournit le signifié. L'analogie avec l'ordinateur montre, certes, comment il est possible de simuler des processus sémantiques par des processus formels, mais elle ne permet pas de comprendre comment le sens vient aux symboles. Comme le souligne le philosophe anglais J. Searle « *La syntaxe (...) n'est pas intrinsèque à la physique du système ; elle est dans l'œil de l'observateur (...) le calcul n'est pas un processus intrinsèque qui se trouverait dans la nature comme la digestion ou la photosynthèse ; il n'existe que relativement à un agent qui donne une interprétation computationnelle à la physique. Il en résulte que le calcul n'est pas intrinsèque à la nature mais relatif à l'observateur ou à l'utilisateur* » (p.27). Il s'ensuit donc que ce dernier ne saurait être lui-même réduit (identifié) à un calcul. Dans tous les cas le système n'interprète pas lui-même les symboles ou les représentations qu'il manipule, ce sont le concepteur et, plus tard, l'utilisateur, tous deux extérieurs au système, qui projettent sur ces symboles ou représentations une interprétation. On voit donc difficilement comment on pourrait accorder à un système informatique une intentionnalité intrinsèque. Son intentionnalité apparaît comme une intentionnalité d'emprunt, dérivée de celle du concepteur et de l'utilisateur, qui lui font « crédit » de sens (Cummins, 1983).

L'esprit accède à des vérités qui échappent au calcul

Selon le logicien Alan Turing, tout calcul effectué par un être humain est susceptible d'être réalisé par une machine. Ainsi, l'homme qui fait un calcul se comporte comme une machine et, réciproquement, on peut alors dire que la machine se comporte comme un humain (faisant un calcul). Cependant, l'identification de l'esprit au calcul

¹³ En fait, il s'agirait de la configuration de la perception de la configuration conduisant à une boucle autoréférentielle.

suppose que l'ensemble des assertions portant sur le calcul soit représentable dans le calcul lui-même. Or, ce n'est pas le cas. Comme le montre notamment le logicien Kurt Gödel, il existe des vérités non réductibles au calcul ; donc l'esprit qui accède à ces vérités ne saurait être lui-même réduit à un calcul, qu'il soit effectué par une « machine » ou par le cerveau. Les résultats de Gödel (1931) indiquent en substance qu'il est logiquement impossible de donner une démonstration métamathématique de la consistance de l'arithmétique et qu'il existe donc des propositions arithmétiques vraies qu'on ne peut pas (mécaniquement) déduire des axiomes. Autrement dit, dans tout système contenant l'arithmétique (l'axiomatique des nombres entiers) il existe des énoncés vrais non démontrables. L'esprit accède donc à des vérités qui échappent au calcul.

Plus fondamentalement, le calcul n'est pas la vie. La mathématisation de la nature aboutit à des paradoxes. Par exemple, le calcul infinitésimal transforme le mouvement en un nombre infini d'états stationnaires et conduit donc à la disparition du mouvement en tant que mouvement (comme l'a bien vu Leibniz). L'explication (la réduction) du mouvement est produite aux dépens de sa compréhension. La démarche réductionniste fonctionne de même avec la chaleur ou la couleur par une mise à l'écart de l'éprouvé dans les qualités à réduire. Ainsi, comme le souligne encore M. Bitbol « Toute réduction suppose de redéfinir ce qui est à réduire de manière à en éliminer le seul vrai noyau irréductible qui est expérientiel. Toute réduction exige de conserver le nom de ce qui est à réduire après en avoir tronqué la signification. La réduction délimite en creux le site de l'incomplétude principielle qui affecte toute démarche de ce genre » (p.271).

Nous ne sommes pas réductibles à ce que nous faisons

Chacun sait que, dans l'Évangile, il nous est demandé d'aimer nos ennemis et de prier pour eux (Mt 5, 43-47 ; Lc 6, 27-35). Certes ! Mais il ne nous est pas demandé d'aimer ce qu'ils font (leur comportement, leur conduite). Nous ne sommes donc pas réductible à ce que nous faisons¹⁴ et, comme le rappelle notamment le pape Jean XXIII dans l'encyclique *Pacem in terris* (n°158), « *C'est justice de distinguer toujours entre l'erreur et ceux qui la commettent...* ». La question de savoir s'il existe ou non quelqu'un derrière les apparences ou si la personne s'identifie à son apparence, est évidemment essentielle.

Mais, à quoi peut-on reconnaître qu'une entité possède un esprit conscient ? Quel est le critère de démarcation de la pensée ou plus exactement quelles sont les conditions de possibilité d'un énoncé tel que : « cette entité pense et ou est consciente » ? La pensée ne s'exprime adéquatement dans aucun exemple d'activité pensante, elle est une sorte de point de vue sous lequel nous acceptons ou refusons de considérer cette activité. Nous n'accédons à l'intériorité d'autrui que par une inférence que nous effectuons à partir de son comportement. Mais, suffit-il qu'une entité se comporte comme un humain pour lui attribuer des qualités propres à ce dernier ?

Selon Hofstadter et Dennett, il ne s'agit pas de savoir, par exemple, si une machine souffre, mais si elle se comporte « comme si » elle souffrait. Et d'ajouter que nous procédons de même entre nous. À propos de son fameux test, le logicien A. Turing soutient également que la question n'est pas de savoir si une machine pense ou est intelligente mais si son comportement est indiscernable de celui d'un humain dans les mêmes conditions. Comment sait-on qu'un être humain pense « réellement » sinon en l'inférant de son comportement manifeste ? Le réalisme de l'imitation et la pertinence du comportement sont-ils ici des critères suffisants ? Pour beaucoup, la réponse est oui, car cette inférence serait tout aussi légitime que celle qui permet de conclure à la présence d'un courant électrique à partir de l'observation de ses effets thermiques ou magnétiques.

Au demeurant, même si un homme et un automate satisfont aux mêmes épreuves ce n'est pas ce que l'on voit mais ce que l'on sait de leur comportement qui nous fait en décider. Aucun comportement ne peut donc être la preuve de la présence d'un esprit et inversement, l'absence de comportement n'implique pas nécessairement l'absence de toute forme de « présence ». Il n'y a pas de critères objectifs de la subjectivité. L'apparence n'est pas la présence et inversement l'absence d'apparence n'est pas une preuve d'absence (statut de la personne en état non répondant).

« Nous devons prendre garde à opérer une distinction entre l'aptitude à détecter un signal et la prise de conscience même de ce signal » (Libet, p. 40). Aucun test ne pourra jamais garantir qu'un robot possède bien *autre chose* qu'une apparence de conscience d'accès (en se comportant de manière appropriée dans des contextes appropriés). « *Autre chose qui n'est aucune chose qui n'est même pas quelque chose au sens le plus*

¹⁴ Revoir la note 10 à propos du behaviorisme.

vague du terme. Autre chose qui est à la fois partout et nulle part, qui recouvre toutes les choses sans s'assimiler à aucune. Le réductionnisme nous montre ce que n'est pas sa zone d'ombre, le connaissant inconnu, la précondition non existante de l'existence» (Bitbol, p. 279).

Simuler n'est pas être

Simuler une fonction biologique ou psychologique par un algorithme mathématique ne signifie donc pas que cette fonction s'identifie à cet algorithme. Ce n'est pas parce qu'on peut mécaniser une règle que celle-ci est « mécanique ». Toute règle peut être considérée comme la description d'un mécanisme. La pensée (qui ne se réduit pas à la pensée formelle) est bien autre chose que le fait de suivre un ensemble de règles. On peut, certes, produire des résultats qui découlent d'un ensemble de règles et qui donnent l'impression d'une « pensée » mais cela ne prouve pas qu'une pensée ait été produite par ledit ensemble de règles. La pensée est en amont, dans la production des règles dont l'implémentation donnera l'impression d'une pensée qui n'est rien d'autre que la restitution de celle dont le producteur a fait crédit au système. Comme nous l'avons déjà souligné, il faut qu'un être humain se tienne à côté de la machine pour interpréter ses schémas comme une description de quelque chose. En fait, si le monde est réductible à un calcul, il est ultimement inaccessible car cette inaccessibilité est la condition même de sa calculabilité. En d'autres termes, ce qui fait « tenir » le discours (le calcul, la représentation...) ne peut pas être « tenu » dans le discours lui-même (Gödel, Wittgenstein, Lacan...).

Une machine n'est pas un organisme

Les êtres vivants ne sont pas des machines. Un mécanisme, par lui-même, n'a pas de but, ne tend vers rien ; il n'a pas de finalité propre. C'est son concepteur qui tend vers quelque chose ; c'est lui qui a arrangé des composants de telle sorte qu'ils servent le but que, lui, le concepteur, s'est fixé.

Les systèmes dits intelligents n'ont pas réellement de but même s'ils se comportent, conformément à l'agencement programmé de leurs composants, « comme si » ils en avaient. Leurs actions « ne sont pas motivées par un état interne ; ils sont programmés et déclenchés par le créateur du robot » (Damasio, 2017, pp. 293-294). Un automate réalise des opérations qui ne l'affectent pas (qui n'affectent pas l'existence même de ses composants) alors que le comportement d'un être vivant a pour but premier

la conservation, l'adaptation et le renouvellement de ses propres constituants (et des individus de sa propre espèce). Une machine n'a pas d'affect puisqu'elle n'est pas affectée par ce qu'elle réalise. Les organes d'un organisme vivant sont parties d'un tout intégré (un soi) alors que les composants d'une machine n'ont qu'un rapport extérieur, les uns par rapport aux autres, et ne constituent pas un tout organique.

Ainsi, selon le biologiste français Henri Laborit, un organisme est avant tout une mémoire agissante en perpétuel enrichissement et qui n'agit que pour être¹⁵. Pour lui, en effet, le système nerveux ne sert qu'à agir. Tout ce qu'il peut faire en plus n'est jamais - toujours selon lui - que d'améliorer cette finalité première. Ce qu'il est convenu d'appeler la pensée n'est alors que le moyen le plus perfectionné, propre à l'homme, de réaliser cette fin (agir pour être). Cette idée selon laquelle un organisme vivant est un système auto-centré, spontanément actif, a été développée parallèlement par différents auteurs, dont le neurobiologiste Francisco Varela (1989), à l'origine du concept d'autopoïèse qui désigne la capacité intrinsèque d'un organisme vivant à s'auto-engendrer¹⁶. La structure produit le comportement qui engendre la structure (on parle ici de « clôture opérationnelle », à ne pas confondre avec la nécessaire ouverture thermodynamique des organismes vivants). Un organisme, nous l'avons dit, est affecté par ce qu'il produit, ce qui n'est le cas d'aucun artéfact, à ce jour. L'être même (la « substance ») du robot le plus sophistiqué ne dépend pas de son comportement (il n'a pas de corps), il n'est donc pas « autonome » (autopoïétique) dans la mesure où ses opérations ne concourent pas à la production, à l'entretien et au renouvellement de

¹⁵ Laborit a défendu et argumenté cette thèse dans de nombreux articles (publiés notamment dans la revue *Agressologie*) repris dans plusieurs ouvrages dont *Les comportements* (Masson, 1973), *La Nouvelle Grille* (Robert Laffont, 1974), *Eloge de la Fuite* (Robert Laffont, 1976) et *L'Inhibition de l'Action* (Masson, 1979).

¹⁶ Il s'agit principalement, pour Varela, de comprendre comment se constitue un système autonome et comment il est couplé à son environnement. Un tel système est caractérisé par sa forte détermination interne. Ainsi, « la cognition n'est plus considérée comme une résolution de problèmes s'appuyant sur des représentations; la cognition dans son sens le plus vaste consiste plutôt en l'enaction ou le faire émerger d'un monde par le biais d'une histoire viable de couplage structurel ». Nous sommes ici en présence d'un processus d'émergence de formes fondé sur la notion de co-détermination. L'unité fonctionnelle émerge d'un environnement qui se constitue (pour elle) en même temps qu'elle en émerge. Le cerveau produit des changements en lui-même, changements par lesquels l'esprit et le monde surgissent en même temps.

sa structure. Son existence est indépendante de son fonctionnement (la structure physique de Nao est indifférente au fait qu'il soit « allumé » ou « éteint »).

Damasio (2017) insiste également sur le fait que les organismes vivants « *ne sont pas des lignes de code, mais des entités matérielles tangibles (...)* En affirmant que les organismes vivants sont des algorithmes, on contribue à perpétuer l'idée fautive selon laquelle les substrats utilisés dans la construction d'un organisme ne sont pas dignes d'intérêt (...) On pourrait en déduire que le fait d'appliquer le même algorithme à des substrats différents et dans des contextes nouveaux produirait des résultats similaires. Or, tout tend à prouver le contraire (...) Le substrat de notre vie est une organisation chimique unique en son genre (...) Et ce substrat semble être une composante essentielle de notre identité profonde » (p. 285). Damasio insiste, à juste titre, sur l'importance du corps. Le cerveau n'est pas séparable du corps (il n'est pas un dispositif de calcul, ajouté au corps). On peut, certes, simuler des émotions, mais seul un esprit incarné peut en éprouver. « *Étrangement, les processus purement intellectuels se prêtent bien à l'interprétation algorithmique et ne semble pas dépendre du substrat corporel (...). En revanche, rien n'indique à ce jour que les processus intellectuels puissent à eux seuls constituer la base de ce qui nous distingue en tant qu'humain bien au contraire* » (p. 287).

Le savoir de la Vie

Ainsi, comme le souligne le philosophe Michel Henry, il convient alors de « *réinsérer le savoir dans le champ de la vie* ». Il dénonce « *L'oubli de l'enracinement de la science dans l'humain qui la fonde* » (La barbarie, 1986). Cet oubli s'origine, selon lui, dans la réduction galiléenne de la nature à des idéalités géométriques. Au-delà des qualités sensibles se découvre l'univers abstrait de l'objectivité mathématique. Ces idéalités ne tiennent justement leur être que de ce « *monde sensible, subjectif et relatif dans lequel se déroule notre activité quotidienne* » et non l'inverse. À l'opposé du savoir de la science, le savoir de la vie ne relève nullement d'un voir. La vie se caractérise par une « *sensibilité* » au sens transcendantal de s'éprouver soi-même, « *auto-révélation* » avec laquelle commence et finit la vie. On ne peut accéder au savoir du contenu d'un livre de biologie qu'en sachant mouvoir les mains pour en tourner les pages (ou désormais « cliquer » ou « glisser ») et surtout, il faut bouger les yeux pour lire ! Selon Henry, la vie est pure subjectivité, pure épreuve de soi et son savoir

réside dans le pathos de cette épreuve qui est justement le savoir de la vie. La sensibilité est la « *condition inaperçue* » de la science elle-même (voir Dufour-Kowalska, 2003).

L'esprit est le grand absent du cerveau.

Cependant, comme le souligne P. Ricoeur¹⁷, aucun vécu ne correspond à mon cerveau. Je prends avec mes mains n'est pas identique à je pense avec mon cerveau. Le cortex ne sera jamais dans le discours du corps propre. Comme nous l'avons déjà dit nos connaissances sur le cerveau ne changent en rien l'expérience du corps vécu. En fait, toutes les théories réductionnistes présupposent à leur insu ce qu'elles prétendent éliminer. Dire que le cerveau « *pense* » procède de l'amalgame de deux discours qui ne cessent d'être corrélés mais qui restent irréductibles l'un à l'autre. Comme Paul Ricoeur le fait remarquer à J.P. Changeux « *Nous comprenons soit un discours psychique, soit un discours neuronal, mais leur relation fait problème parce que nous n'arrivons pas à inscrire leur lien à l'intérieur de l'un ou l'autre. Nous manquons du discours tiers* » (p. 84). Le cerveau est bien le substrat, la condition de la pensée mais il ne pense pas au sens d'une pensée qui se pense (p. 54). Cependant, tandis que je pense il se passe toujours quelque chose dans mon cerveau. Or, on ne sait pas indiquer à quoi (à quel troisième terme) la conscience phénoménale et les états représentationnels neuronaux font référence (p. 66).

La conscience, la pensée, l'esprit, ne sont donc pas, par nature, objectivables. L'esprit désigne précisément ce que je dois d'emblée admettre pour que la connaissance soit possible. C'est donc seulement « *en creux* », sur le mode de l'absence, qu'il peut se laisser appréhender empiriquement, mais, de ce fait, on pourra toujours en nier l'existence en tant « *qu'objet* ». Comme le souligne encore M. Bitbol : « *Quelque chose n'est pas réductible, mais quoi exactement ? S'agit-il d'une chose ou précisément de rien de tel ? Peut-on nommer ce qui échappe à tout raisonnement ou est-il voué à rester innommable ? (...)* Le réductionnisme montre en négatif ce que ne peut pas être ce qui lui file entre les doigts » (Bitbol, p. 265).

On peut donc seulement dire que, dans la description physicaliste, quelque chose a été oublié mais l'objet même de l'oubli fait défaut, du moins fait-il défaut en tant qu'objet. Une telle présence dans l'absence n'est pas du domaine de la *preuve* mais de l'*épreuve* c'est-à-dire à la fois de

¹⁷ Voir Changeux et Ricoeur (1998).

ce qui s'éprouve, de ce dont on fait soi-même l'expérience, et de ce par quoi l'on est affecté. C'est seulement aux marges de la description neurophysiologique que peut se montrer celui qui dit je. Le procès du sens, du sujet, de l'esprit, ne peut se solder, du point de vue empirique, que par un non-lieu. Il n'y a pas lieu, en effet, de continuer à chercher la représentation de ce qui n'est pas représentable et qui donc ne saurait être circonscrit en quelque lieu que ce soit. Toute tentative de décrire ce que je suis laisse « quelque chose » de côté. Ainsi, l'intériorité du sujet ne peut se manifester dans le monde objectif que sur le mode de l'absence, comme une sorte d'exception non figurable, d'angle mort de la connaissance, de tache aveugle de la représentation. Comme, pour les chrétiens orthodoxes, une icône ne représente pas ce qu'elle signifie, elle le montre, le tableau de la science ne contient évidemment pas, sinon sur le mode de l'absence, ce qu'il manifeste.

« Il y a en dans le moi une sorte de commencement absolu dont l'origine est inexplicable d'un point de vue purement physique. Chacun de nous est, au milieu du monde, un point de vue unique et irremplaçable qui fait de l'univers entier le décor de son moi. L'esprit situe tout par rapport à soi. Mais, et c'est bien là le paradoxe de notre finitude, le même esprit qui situe et (me) pose est lui-même situé sans avoir l'initiative de la situation. Il est posé sans s'être posé lui-même. Je situe tout de mon point de vue mais je n'ai pas choisi ce point de vue qui est justement le mien. Je peux, jusqu'à un certain point, déterminer toute chose mais je ne détermine pas la position contingente à partir de laquelle je détermine ou envisage de déterminer toute chose » (Léonard, p. 59).

Une absence qui fait signe

Nous voyons bien que ce « quelque chose » (ou plutôt, ce quelqu'un) résiste à la réduction et ce qui résiste c'est précisément ce qui fonde (l'origine). Le vécu auquel se heurte l'objectivation n'est autre que sa condition. Le fondement, l'origine, la cause de tout système est nécessairement indiscernable dans le système, sinon sous la forme d'un manque. Du fait même que la science objective exclut délibérément, par principe, toute forme de transcendance de son champ d'investigation, elle ne saurait ni la rencontrer, ni l'exclure : ce n'est pas de sa compétence. Que l'expérience consciente soit associée à des activités neuronales ne signifie évidemment pas qu'elle s'y réduise. Entre décrire et vivre, un gouffre explicatif demeure. Aucun

tour de passe-passe ne peut permettre à la science de retrouver la subjectivité vécue qu'elle a dû nier pour exister (Bitbol).

Le cerveau conditionne la pensée

La notion de cause est une notion ambiguë. Il ne fait aucun doute que les décharges de certaines cellules hypothalamiques sont spécifiquement activées lorsqu'on éprouve de la rage, de l'excitation sexuelle ou de l'appétit. De plus, si ces cellules sont détruites ces états sont altérés. Il est non moins évident que la lecture de ce texte tout comme l'évocation d'un souvenir particulier suppose l'intégrité de certaines structures cérébrales (notamment l'aire de Wernicke pour la compréhension de ce texte et l'hippocampe pour l'évocation d'un souvenir). Mais ce n'est pas pour autant que ces décharges sont la cause des comportements en question, même si elles sont susceptibles de servir de référent à ces comportements. La science appelle donc cause ce qui est en fait la condition du phénomène observé. Chacun sait que la structure « fusible » conditionne la fonction « froid » du réfrigérateur sans que quiconque ait envie d'en déduire que le fusible soit la cause, l'origine ou le « siège » du froid. De même la roue d'une automobile conditionne sa vitesse sans en être aucunement « la cause ». L'art de Rostropovitch ne se ramène pas à la qualité de son violoncelle, tout en dépendant de lui. Dans l'ordre de la causalité, il existe un rapport direct et de même nature entre l'antécédent et le conséquent (l'abus d'alcool cause l'ivresse). Dans l'ordre du conditionnement, au contraire, le résultat déborde la condition qui l'a rendu possible. Ainsi, il n'y a pas lieu de chercher où et comment l'esprit agit sur le cerveau. Les effets physico-chimiques ont des causes physico-chimiques, mais ces causes ne se « contiennent pas ».

3 - LA PERSONNE : CORPS, AME ET ESPRIT.

Comme le souligne le philosophe américain Philip Clayton, l'existence d'un sujet personnel dans le monde ne peut simplement se justifier par l'ensemble des causes finies ; on doit en effet mentionner le fondement ontologique de l'existence finie en tant que telle. « *L'ordre des causes finies ne procure pas les ressources conceptuelles permettant d'introduire un sujet ou un esprit durable mais uniquement des prédicats s'apparentant à un sujet ou à un esprit* » (p.226). Une « personne » n'est donc jamais réductible à un produit de la nature, même si, organiquement elle en procède. Une personne est d'emblée une

personne : elle « n'émerge » pas. Le caractère ontologiquement unique de l'être personnel, son unité en tant qu'esprit impliquent « quelque chose de plus » - au plutôt « d'autre » - que l'enchaînement causal des déterminismes naturels. L'existence d'un être personnel doté d'un esprit « durable » suppose donc un fondement - une origine (que l'on ne confondra pas avec son commencement) - ontologiquement différents de la causalité naturelle. Il convient alors de reconnaître, indépendamment du cérébral et du mental (de l'homme psychique), l'existence d'une autre « composante » de la personne humaine, de nature spirituelle, qui la constitue précisément en tant que personne appelée à entrer en dialogue avec Celui dont elle procède.

Contrairement à la conception commune qui voit dans le corps et l'âme (matière et esprit ; cerveau et mental) deux composantes de la personne, pour l'anthropologie biblique¹⁸ l'être entier de l'homme comporte l'esprit, l'âme et le corps. « *Il est âme en tant qu'animé par l'esprit de vie; la chair montre en lui une créature périssable; l'esprit signifie son ouverture à Dieu; le corps enfin l'exprime au dehors* » (Vocabulaire de Théologie Biblique, p. 539). À la dichotomie habituelle âme - corps, commune à tous les êtres vivants, l'esprit (au sens spirituel et non pas cognitif du terme) s'ajoute pour l'homme d'une manière distinctive¹⁹. La nature à elle seule ne saurait donc rendre compte de l'émergence de la personne. Pour l'anthropologie tripartite, l'homme est l'être que l'évolution biologique a porté à un stade suffisant pour pouvoir répondre à l'appel du Créateur et recevoir à cet effet les dons de l'Esprit. La réponse à un tel appel est au-delà de la seule nature ; elle

¹⁸ La tradition biblique et l'eschatologie chrétienne expriment fortement la complémentarité d'une diversité de point de vue sur l'unique réalité qu'est la *personne* irréductible à l'interaction de deux principes séparés. La *personne* unifiée et unifiante s'exprime de manière différente à travers ses divers attributs que l'on désigne par des concepts comme la chair, le cœur, l'âme ou l'esprit. La bible hébraïque emploie trois termes différents pour désigner les composantes de l'être humain : *néphesh* (âme), *basar* (chair) et *rouah* (esprit). Le premier de ces termes désignait d'abord la gorge, puis, par extension, la respiration, la vie et finalement l'être vivant envisagé comme un tout. Ce n'est que tardivement (dans le Nouveau Testament) que l'âme devient l'élément essentiel de la *personne* humaine. Le terme *basar* signifie davantage le corps vécu (l'homme tout entier) que le corps organique. Le terme *rouah* désigne le souffle, le vent, porteur de la parole créatrice, mais aussi le souffle de vie qui anime tout être vivant. L'esprit de l'homme apparaît comme une ouverture de la *néphesh* à l'esprit de Dieu. Les pères de l'Église, au contact du monde grec, vont abandonner cette conception tripartite au profit de la dichotomie corps et âme qui restera dominante, surtout en occident.

¹⁹ Voir Lambert, 2015.

implique la liberté d'un sujet personnel se situant au-delà des potentialités naturelles.

Dans un tel contexte, le corps et l'âme sont naturels, mais l'esprit est don. À la complémentarité de l'âme et du corps telle qu'elle peut être étudiée par les sciences cognitives s'ajoute le don de l'esprit. L'homme n'est pas détaché de la nature, mais ce don fait de lui un être d'exception. Le « passage à la personne » n'est pas un plus de l'évolution, ne résulte pas de l'émergence d'un module cérébral supplémentaire ; la personne ne peut que se recevoir d'un Autre par le don de l'esprit. Mais ce don n'est pas contraignant et c'est dans le respect de la liberté de l'être auquel il s'adresse que ce don est fait. Il s'ensuit donc qu'il peut n'être pas reconnu comme tel : l'homme peut prétendre tenir sa dignité seulement de la nature (de l'évolution) ou de lui-même.

Cette conception tripartite s'accorde d'ailleurs avec les données scientifiques qui ne montrent que fort peu de différence biologique entre le chimpanzé et l'homme. La singularité humaine, avec le passage à la conscience réflexive et à l'éthique ne peut s'expliquer par le seul triplement du volume cérébral ou de la surface corticale de l'homo sapiens. Ce n'est donc pas dans l'empirique objectivable qu'il convient de chercher l'irréductible spécificité humaine mais dans le fait que, seul de tous les vivants, l'homme est capable de donner un nom aux êtres et aux choses, de donner un sens à ce qui est, de revendiquer des valeurs. En clair, de telles capacités sont celles généralement reconnues à un esprit tel que l'entendent les philosophes : ce qui a de la volonté, des projets, une mémoire, une conscience. Or, pour un matérialiste conséquent, la matière, est ce qui n'est pas esprit, c'est-à-dire qui ne possède aucune de ces caractéristiques de l'humain. Comment donc une telle matière pourrait-elle produire ce qu'elle n'est pas ? Ce n'est donc pas de la seule matière que l'homme peut tenir ce qui le fait proprement humain²⁰.

²⁰ Rappelons également que pour Thomas d'Aquin, si l'âme humaine n'est pas produite par la matière et ses réaménagements, une certaine organisation de la matière, un certain corps, sont nécessaires pour que l'âme rationnelle puisse s'y exprimer. L'évolution peut donc être comprise comme une préparation de la matière jusqu'à l'apparition d'une forme vivante adéquate à l'apparition de l'âme rationnelle. Saint Thomas conçoit une telle préparation pour chaque personne humaine, puisque la matière qui constitue l'embryon est informée jusqu'à être rendue apte à recevoir une âme raisonnable, immédiatement créée par Dieu. La matière est d'abord en puissance de la forme susceptible de recevoir l'âme rationnelle. Il faut alors admettre l'existence d'un principe actif inhérent à la matière et la rendant apte à

Le don de l'esprit

Tout se passe comme si, en effet, une nouvelle source d'information était nécessaire à l'humanisation. L'extraordinaire surgissement des potentialités de l'esprit échappe à la nature. Comment y a-t-il de la « personne » dans l'univers? La science ne peut apporter de réponse à cette question puisque sa méthode suppose l'objectivation de ce dont elle traite. *« Dans l'homme, l'information n'est pas seulement biologique. Tout en étant de parenté animale, le corps humain n'est pas seulement doué d'un psychisme fonctionnellement dominé par les nécessités de la vie; il est encore au service de l'esprit qui exprime par lui son originalité. Mais loin donc de couper l'homme en deux, l'esprit s'y montre le principe d'une unité nouvelle »* (Martelet, p.87).

On ne saurait accéder à une véritable objectivité en niant notre subjectivité. On ne recompose pas quelqu'un, un soi, un sujet de l'extérieur, en son absence. Multiplier ou augmenter les éclairages n'augmente guère les chances de retrouver les clés perdues ailleurs que là où on les cherche. Reconnaître l'existence d'une instance irréductible à ses manifestations objectives n'a rien de dérisoire *« car ce n'est pas une échappatoire que d'exprimer avec des mots anciens et quelque peu philosophiques les deux bouts d'une chaîne qu'il faut désespérément tenir l'un et l'autre »* (Costabel, p.261). Il convient donc de reconnaître que

« L'explication mécaniste s'arrête devant l'unité non fonctionnelle du sujet pensant et cette difficulté ne tient ni à des préjugés religieux ni à des raisons théologiques (...) L'explication mécaniste s'arrête devant l'unité non-fonctionnelle du sujet pensant et cette difficulté ne tient ni à des préjugés religieux ni à des raisons théologiques (...) Qui -poursuit Lévy - voit donc trembler dans la lumière la poudre d'or du mimosa? Ce ne sera ni cette machine (le cerveau) ni une autre, mais quelque chose d'autre qu'une machine (...) Pourra-t-on désigner scientifiquement celui qui perçoit, celui pour qui le monde existe? Nous pouvons décrire exactement la manière dont le flux photonique tombe des étoiles, on ne sait

produire un être (l'homme psychique) susceptible de répondre à l'appel de Celui qui en fait une personne (l'homme spirituel). Pour Thomas, l'âme humaine est, en quelque sorte, à double face : l'une en contact avec « la chair », l'autre avec « l'esprit ». Elle constitue « l'interface » entre le corps « psychique » (naturel) et l'esprit (surnaturel).

toujours pas comment la lumière jaillit du regard » (Pierre Lévy, pp 197, 201).

Bibliographie

- Arsac, J. (1987). *Les machines à penser. Des ordinateurs et des hommes*. Seuil.
- Atlan, H. (1972). *L'organisation biologique et la théorie de l'information*. Hermann.
- Bitbol, M. (2014). *La conscience a-t-elle une origine ? Des neurosciences à la pleine conscience : une nouvelle approche de l'esprit*. Flammarion.
- Block, N. (1980). What is fonctionnalism ? in N. Block (ed.), *Readings in Philosophy of Psychology*, vol.1, Havard University Press.
- Cabanis, P. (1824). *Rapports du physique et du moral de l'Homme*. Paris : Béchet.
- Changeux, J.-P. (1983). *L'Homme Neuronal*. Fayard.
- Changeux, J.-P. et Ricoeur, P. (1998). *La nature et la règle*. Odile Jacob.
- Clayton, P. (2012). *Les origines de la liberté*. Salvator.
- Costabel, P. (1990). Les notions de corps et de soi à la fin du XVII^e siècle in J. Bernard, M. Bessis et C. Debru (eds.) *Soi et non-soi*, Seuil, 251-261.
- Cummings, A.R. (1983). *The Nature of Psychological Explanation*. Bradford Books.
- Damasio, A. (1995). *L'erreur de Descartes. La raison des émotions*. Odile Jacob.
- Damasio, A.R. (1999). *Le sentiment même de soi. Corps, émotions, conscience*. Odile Jacob, Paris.
- Damasio, A. (2003). *Spinoza avait raison. Joie et tristesse, le cerveau des émotions*. Odile Jacob.
- Damasio, A. (2017). *L'Ordre étrange des choses. La vie, les sentiments et la fabrique de la culture*. Odile Jacob.
- Dehaene, S. (2014). *Le code de la conscience*. Odile Jacob.
- Dreyfus, H. (1984). *Intelligence Artificielle, mythes et limites*. Flammarion.
- Dufour-Kowalska, G. (2003). *Michel Henry. Passion et magnificence de la vie*. Beauchesne.
- Edelman, G. (1992). *Biologie de la conscience*. Odile Jacob.
- Freud, S. (1978). *Essais de psychanalyse appliquée*. Gallimard, 141 et 146.
- Gauchet, M. (1992). *L'inconscient cérébral*. Seuil.
- Gazzaniga, M. S. (2013). *Le libre arbitre et la science du cerveau*. Odile Jacob.
- Gödel, K. (1989). Sur les propositions formellement indécidables des *Principia Mathematica* des systèmes apparentés (1931). *Le théorème de Gödel*, 107-143.
- Guillaume, F., Tiberghien, G. & Baudouin, J.-Y. (2013). *Le cerveau n'est pas ce que vous pensez. Images et mirages du cerveau*. PUG.
- Henry, M. (1987). *La barbarie*. Grasset.
- Hofstadter, D. et Dennett, D. (1987). *Vues de l'Esprit*. InterEditions, Paris.
- Lacan, J. (1966). *Ecrits*. Seuil, Paris.
- Ladrière, J. (1982). Wittgenstein et la philosophie analytique. *Science et foi*, Desclée International, Paris.
- Lambert, J.F. (1992). L'épreuve du sens. Science et incomplétude. *Les Cahiers Jean Scot Erigène*, vol. 3, *Connaissance traditionnelle, connaissance rationnelle*, Guy Trédaniel, 135-222.
- Lambert, J.-F. (1994). *La revanche de Babel*, La Nef, 45 : 28-30.
- Lambert, J.-F. (2010). Peut-on encore parler de sciences humaines ? in AEES (ed.), *Qu'est-ce que l'homme ?* F-X de Guibert, 59-99.
- Lambert, J.-F. (2015). Qui est celui qui dit JE ? in Association des Scientifiques Chrétiens, *Liberté et cerveau*, Parole et Silence, 33-47.
- Lambert, J.-F. (2015). La personne, sujet du miracle. Corps, âme, esprit. De la bible aux sciences cognitives in Projet Nouveau Regard (ed.), *Enquête sur les miracles. Pour la nouvelle Évangélisation*, Editions du Jubilé, 273-283.
- Lambert, J.-F. (2016). Les neurosciences et la justice. L'imagerie cérébrale au tribunal in Ch. Tijus et C. Puigelier (eds.) *L'Esprit au-delà du Droit*, mare & martin, 159-190.
- Laureys, S. (2015). *Un si brillant cerveau*, Odile Jacob.
- Léonard, P. (1987). *Les raisons de croire*, Fayard-Communio.
- Lévy, P. (1987). *La machine univers. Création, cognition et culture informatique*, La Découverte, 197 et 201.
- Libet, B. (2012). *L'Esprit au-delà des Neurones. Une exploration de la conscience et de la liberté*, Dervy.
- Martelet, G. (1992). Actes du Colloque BENA 5..
- Noë, A. (2009). *Out of our Heads. Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*, Hill and Wang.
- Searle, J.R. (1999). *Le mystère de la conscience*, Odile Jacob.
- Straus, E. (1989). *Du sens des sens*, Jérôme Million.
- Varela, F. J. (1989). *Autonomie et connaissance. Essai sur le vivant*. Seuil.
- Wittgenstein, L. (1961). *Tractatus logico-philosophicus et Investigations philosophiques*. Gallimard.
- Wittgenstein, L. (1992). Conversation sur Freud et Conférence sur l'Éthique. *Leçons et conversations*. Gallimard.

Échange de vues

Nicolas Aumonier : En 2012, une équipe japonaise, prolongeant des expériences d'optogénétique, a réussi à allumer et éteindre des neurones spécifiques, notamment ceux de l'hippocampe, donc de la mémoire, et a réussi ainsi, en activant les neurones de la mémoire, à fabriquer des faux souvenirs à des souris grâce à de la fibre optique.

Jean-François Lambert : Grâce surtout au génie génétique qui modifie les neurones, en leur introduisant des récepteurs que les souris n'ont pas naturellement (voir note « Manipuler la mémoire »).

Nicolas Aumonier : Quand même, le résultat est là, on fabrique des faux souvenirs à des souris qui ne les avaient pas. Est-ce que ce n'est pas *a stone* dans votre *garden* ?

Jean-François Lambert : On fabrique de nouvelles connexions dans le cerveau de la souris, que nous interprétons comme des souvenirs. Évidemment, du point de vue de l'observation empirique, vous avez raison. On manipule ce que nous appelons des souvenirs mais, ce qu'on manipule effectivement, c'est le comportement de la souris. Elle se comporte, dans le compartiment où elle n'a jamais reçu de choc électrique, comme si elle en avait reçu un. C'est ce que l'on constate. Maintenant, est-ce que c'est réellement manipuler les souvenirs ? Il conviendrait certainement de distinguer la mémoire des souvenirs. Il y a d'autres techniques très efficaces, sans électrodes ni manipulation génétique, pour manipuler les souvenirs. Les travaux d'Élisabeth Loftus montrent, par exemple, que l'on peut convaincre les gens d'avoir vécu des événements qu'ils n'ont jamais vécu.

Jean-Marie Schmitz : Vous dites de ce que toute information, tout calcul, doivent être interprétés, ce qui semble relever du bon sens le plus évident. Que font, ou qu'argumentent, ceux qui disent que ce n'est pas le cas ? Comment font-ils, ou que disent-ils ?

Jean-François Lambert : C'est le problème de la délégation de compétence ou du crédit de sens dont j'ai parlé, qui est posé. Dans tous les systèmes informatiques, il y a un dispositif, justement, qu'on a appelé interpréteur (ou un compilateur), un programme permettant de

traduire le langage de programmation (par exemple le basic) en langage machine. Ce n'est pas tout à fait innocent d'avoir appelé ça un « interpréteur ». On sait, par ailleurs, qu'il existe aujourd'hui des robots, qui lisent et interprètent des radiographies mieux que les radiologues et « interprètent » certains patterns de l'image comme indiquant la présence d'une tumeur. Mais, c'est de l'interprétation par délégation parce qu'au début, comme à la fin de la chaîne de traitement, il y a toujours un humain ! Cela ne veut pas dire qu'il faut un humain pour tenir la main du robot tout le temps. Le système-expert qui analyse la présence ou non d'une tumeur, le fait bien mieux « tout seul » que si quelqu'un lui tenait la main, mais au final, le résultat ne fait sens que pour un humain ! Vous pourriez m'objecter qu'il y a maintenant des robots qui donnent l'impression d'avoir compris. Oui, tout se passe « comme si » et c'est toute la question. Comme je l'ai dit, simuler n'est pas être !

Rémi Sentis : Bien sûr je suis en total accord avec Jean-François Lambert. Comme vous l'avez très bien dit, il est toujours nécessaire qu'il y ait un intermédiaire entre le logiciel et l'ordinateur, un interpréteur. En fait il y a un compilateur qui traduit les instructions du langage de programmation en code écrit en langage machine (avec un nombre limité d'instructions simples déclenchant une commande exécutée par le processeur). La majorité des gens oublie cela : il y a toujours un compilateur dans tout ordinateur ou même dans une puce électronique. De plus, ce compilateur a forcément été écrit par un humain. La présence du langage machine est tellement évidente pour un informaticien qu'il n'en parle pas souvent.

Par ailleurs, il y a souvent des problèmes de vocabulaire quand on évoque les logiciels informatiques. Par exemple, on parle de la mémoire, et comme chacun sait, la mémoire d'un ordinateur, ça n'a aucun rapport avec notre mémoire à nous ! C'est uniquement des configurations électromagnétiques d'éléments d'un semi-conducteur ; on pourrait appeler ça un « magasin » ou n'importe quoi d'autre... Il se trouve qu'on a appelé ça « mémoire » ; du coup ça donne l'impression que cela a un rapport avec une mémoire humaine, mais ça n'a absolument aucun rapport !

Jean-François Lambert : Il faudrait peut-être, là encore, distinguer la mémoire (le support) du souvenir, mais c'est la question du rapport au substrat que vous posez. Du point de vue fonctionnel, la mémoire électronique et la mémoire neurochimique, ce n'est pas très différent. C'est ce que j'ai essayé de dire dans mon introduction : quel que soit le support, du point de vue formel, le processus mémoire est de même type mais aussi bien pour l'ordinateur que pour le cerveau, le contenu de la mémoire ne fait pas sens par lui-même et suppose donc un interprète distinct de la mémoire elle-même, qui n'est pas autoréférentielle. On peut, certes, postuler l'existence d'un méta-niveau de traitement mais, encore une fois, le résultat du calcul réalisé à ce méta-niveau, comme la signification d'un symbole, ne sera jamais intrinsèque à la physique du système. On a donc accès à des significations qui ne sont pas données par le calcul. Comme le montrent Gödel et d'autres logiciens, nous avons effectivement accès à des vérités qui ne sont pas calculables. Ça veut bien dire qu'il y a quelque chose en nous, que nous pouvons appeler l'esprit (au sens spirituel du terme) qui résiste à la description naturaliste. Ainsi, dans la description naturaliste, il y a quelque chose qui manque, quelque chose qui est fondamentalement oublié et qu'il est possible d'appeler l'esprit. Et ce qui est oublié, c'est ce qui est donné à l'origine mais qui ne fait pas nombre avec le système.

Marie-Joëlle Guillaume : En vous écoutant, je songeais que votre définition de l'esprit a une tout autre ampleur que ce qu'on veut nous dire lorsqu'on réduit l'esprit au calcul. En fait, cette définition pose la question fondamentale de la liberté. Ce qui m'inquiète le plus dans les visions transhumanistes, qui de façon générale partent du principe que l'esprit serait réductible au calcul, c'est que ce sont des visions qui en définitive font l'impasse sur notre liberté. Souvenons-nous de la vieille formule selon laquelle « la liberté ne se prouve pas, elle s'éprouve ». Formule chère aux classes Terminales, dira-t-on peut-être. Il n'empêche que c'est vrai ! Quand vous dites qu'il y a quelque chose dans l'esprit qui dépasse le calcul, et que finalement on l'éprouve plus qu'on ne peut le prouver vraiment, je dirais : la liberté aussi !

J'ai été très sensible à ce que vous avez mis en évidence tout à l'heure, en notant qu'un pouvoir universel exige un langage universel, donc celui de l'informatique, considéré comme tel. Il est vrai que nous nous trouvons aujourd'hui devant des enjeux de pouvoir mondiaux, qui prennent

l'intelligence artificielle comme le moyen de leur réalisation. De là à parler de « langage universel », pour le coup, je ferais un petit parallélisme avec ce que vous avez dit à propos de l'esprit : je pense que c'est donner beaucoup à l'informatique que de parler d'universel à propos d'elle. Car il s'agit peut-être d'un langage unificateur global ; mais l'universel, c'est quelque chose qui touche à l'humain, au contraire. Vous avez évoqué la querelle des universaux. Les nominalistes ne reconnaissaient pas l'aspect universel d'une notion et la réduisaient à de simples mots, de simples noms. Je pense qu'aujourd'hui on se trouve devant un problème analogue quand on considère le langage informatique. L'universalité, le sens de l'universel, c'est-à-dire le sens de ce qu'est l'homme, qui selon Pascal « passe infiniment l'homme », c'est le sens qui nous permet de reconnaître dans l'esprit quelque chose qui n'est pas mécanisable, me semble-t-il.

Donc en vous écoutant, je me suis sentie confortée dans l'idée que l'enjeu est vraiment de défendre notre liberté. C'est la liberté humaine qui est en cause dans cette remise en question de l'esprit. Vous me direz que ce n'est pas neuf, je crois que c'est à Claude Bernard que l'on attribue la phrase, - je ne sais pas si c'est lui qui l'a dite - : « *Je n'ai pas trouvé l'âme sous mon scalpel* ». C'est un peu le même déni. Mais ce déni a aujourd'hui les moyens d'être particulièrement redoutable.

Jean-François Lambert : Quand on citait la fameuse formule de Claude Bernard, le Père Martelet faisait remarquer que ce n'est pas du côté de la pointe du scalpel (qui pointe vers l'objet) qu'il faut chercher l'âme, mais du côté du manche (tenu par la main d'un sujet) ! La question de la liberté, oui, comme cela a été dit, je me suis beaucoup intéressé aux travaux de Benjamin Libet qui est mort il y a une dizaine d'années maintenant. Ses travaux originaux, qui ont été repris avec de nouveaux moyens, sont largement confirmés. A partir de données d'imagerie cérébrale fonctionnelle (EEG et IRMf), on peut prévoir, plusieurs centaines de millisecondes, voire plusieurs secondes avant qu'il soit initié, l'acte que vous allez réaliser (tel qu'appuyer sur un bouton à gauche ou à droite) en fonction de la configuration cérébrale qui le précède. Je ne peux ici discuter davantage ces résultats mais il convient de distinguer ce que le philosophe J. Searle appelle l'intention préalable et l'intention en acte, cette dernière étant seule en cause dans les expériences de Libet (voir Lambert, postface Libet, 2013 et Lambert, 2015). Mais la liberté peut aussi être vécue comme un fardeau dont il serait bon de se débarrasser. Un nombre

significatif de nos concitoyens seraient prêts – semble-t-il - à accepter l'implantation d'une petite puce électronique dans leur cerveau (opération indolore, prise en charge par l'assurance maladie, pas de trace...) permettant une surveillance continue des paramètres vitaux et une prise en charge complète de leur santé... dont ils perdraient, de fait, totalement le contrôle. Par ailleurs, j'approuve bien sûr, la différence entre prouver et éprouver. Comme je l'ai dit la subjectivité s'éprouve plus qu'elle ne se prouve. Dans la mesure où un robot n'a pas de corps et n'est pas affecté (éprouvé) par ce qu'il fait, il ne saurait donc être doté de subjectivité. A propos de l'universel, attention, le langage informatique, c'est bien de l'universel mais de l'universel par défaut. Je partage la manière dont vous entendez l'universel, mais ici c'est un universel par le bas, par élimination des différences entre les hommes et les femmes, les animaux et les hommes, les hommes et les machines. C'est un universel « 0.0 » et c'est pour cela que cela peut marcher.

Grégoire Duchange : Je suis juriste et je m'intéresse surtout à la question du travail. Si j'ai bien compris votre propos, on peut dire qu'un robot ne peut pas penser. Ma question est alors la suivante : un robot peut-il travailler ? Dit autrement, le travail se réduit-il à un ensemble d'actes matériels, qu'une machine peut reproduire, ou est-il davantage que cela ? Dit encore autrement, pour prolonger le thème de votre intervention : le travail est-il mécanisable ?

Jean-François Lambert : La question du travail a déjà été abordée dans ce cycle (voir l'échange après l'intervention de Jean-Paul Delahaye). Est-ce que les robots vont prendre la place des hommes ? Il se trouve que je viens de terminer la lecture du dernier bouquin de François-Xavier Bellamy, *Demeure*, dans lequel il fait une différence entre le travail, l'emploi et le métier. Aujourd'hui, on revendique – à juste titre - un emploi, mais l'emploi n'est pas le métier. Les emplois – ou plutôt ceux qui les occupent - sont interchangeables, alors que le métier ne l'est pas. Avoir du métier c'est détenir un savoir-faire, un « tour de main », c'est le travail de l'artisan. Et ça, justement, ce n'est pas entièrement « calculable ». Or, avec la société du calcul, on est dans un nivellement (j'ai insisté sur l'annulation des différences), une interchangeabilité, des emplois. On crée des emplois qui, pour certains, ne sont pas vraiment « du travail ». Or, si l'interchangeabilité est la norme entre les hommes, elle est tout aussi possible entre les hommes et les machines. En revanche, le métier, au sens de l'artisan, n'est pas, lui, « interchangeable ». Je

profite de votre question pour souligner l'impact de l'IA dans le domaine juridique. Le laboratoire auquel j'appartiens (CHArt) a organisé récemment plusieurs colloques sur ce thème à la Cité des Sciences (voir Lambert, 2016). L'accès aux bases de données et la robotisation des procédures facilitent évidemment le travail de l'avocat ; même avec beaucoup de culture, il ne peut pas connaître toute la jurisprudence. Dans l'échange qui a suivi mon exposé, ici même, en décembre 2008 (« Peut-on encore parler de sciences humaines »), j'insistais déjà (à la fin de l'échange) sur le fait qu'avec la théorie de l'information, le droit, la médecine, la psychologie, la psychiatrie, pouvaient être réduits à de la manipulation formelle de données. C'est particulièrement le cas en psychiatrie avec le fameux DSM (Manuel diagnostique et statistique des troubles mentaux). On n'est plus du tout dans la logique du « colloque singulier », celle de la rencontre de deux personnes, mais dans la logique quantitative des questionnaires et des tests statistiques. Si le questionnaire comporte, par exemple, 25 items, et que vous répondez positivement à 17, vous êtes dépressif, entre 12 et 17 vous êtes entre les deux, à moins de 12 vous reviendrez une prochaine fois... On peut même avoir une consultation automatique (il en existe sur internet) et, pourquoi pas, la commande directe des antidépresseurs sur Amazon ! Je caricature à peine. J'ai eu récemment l'occasion de voir une vidéo montrant comment l'équivalent de « Ooshop » (Carrefour en ligne) en Chine, peut traiter 200 000 commandes par jour avec 4 humains seulement. La robotisation y est très sophistiquée et il y a presque un côté esthétique dans le déplacement en tous sens de ces chariots-robots qui montent et descendent le long des plans de stockage, prennent des objets, les mettent en paquets, trient les colis... Comme j'ai essayé de le montrer, je pense que les enjeux ne sont pas ici de nature technologique. La question de fond est bien celle de savoir si on peut tout ramener à du calcul.

Rémi Sentis : En fait il y a un concept extrêmement important sur ce sujet-là, c'est le concept de responsabilité. On voit cela par exemple quand on parle de la personnalité juridique des robots ; ainsi ceux qui prônent ladite personnalité juridique voudraient que ce soit le robot qui soit responsable en cas de défaillance ou en cas d'accident causé par celui-ci. Ils veulent évacuer un vrai problème - difficile à résoudre. En effet quand il y a un accident ou une défaillance, c'est parce qu'il y a eu une défaillance soit de l'utilisateur, soit du programmeur, soit de celui qui a fait l'interface entre la machine et l'environnement, soit de celui qui a fabriqué le

capteur ou qui l'a installé, etc. Et donc si on dit que le robot est responsable, cela évite de chercher les responsabilités entre les différents

intervenants. Cela permet d'évacuer élégamment la responsabilité des individus (et de continuer à faire du business !).

Séance du 7 février 2019

Manipuler la mémoire

Une protéine sensible à la lumière, la channelrhodopsine, découverte dans la membrane cellulaire d'une espèce particulière d'algue, peut être introduite dans des neurones modifiés génétiquement. Une fraction de cette protéine (le rétinal) change de configuration lorsqu'elle est éclairée par une lumière bleue et forme un canal à travers lequel des ions entrent rapidement dans la cellule. Ces canaux se fixent sur la membrane des neurones génétiquement modifiés, qui lorsqu'ils sont éclairés, via une fibre optique implantée dans le cerveau, changent de configuration. L'équilibre ionique est modifié et les neurones émettent un signal électrique (des variantes de la channelrhodopsine, sensibles à d'autres longueurs d'ondes lumineuses, permettent de contrôler plusieurs types d'ions).

Par cette technique, il est possible de réactiver chez un animal un souvenir désagréable dans un contexte différent de celui dans lequel il s'est formé, en stimulant spécifiquement les neurones associés à la mémorisation de l'environnement désagréable. Il est même possible de substituer un souvenir à un autre, voir d'introduire de faux souvenirs. Les neurones de l'hippocampe (gyrus denté) sont modifiés génétiquement pour être rendus sensibles à des stimuli lumineux dès qu'ils sont activés par un nouvel environnement. Si l'on place des souris ainsi « modifiées » dans un environnement donné, certains neurones de l'hippocampe réagissent à ce contexte particulier et leur activation entraîne l'expression de la channelrhodopsine quelques heures plus tard. Ils deviennent alors sensibles à la lumière et peuvent donc être réactivés à volonté. On place ensuite ces souris dans un dispositif muni d'une grille permettant de leur délivrer de légers chocs électriques dans les pattes et on active (par une stimulation lumineuse) les neurones associés à l'environnement précédent à chaque fois qu'elles sursautent. Replacées dans le contexte initial, ces souris se figent comme si elles se rappelaient y avoir reçu le choc électrique (alors que c'est bien dans le second environnement qu'il leur a été administré). Il y a coexistence de deux souvenirs désagréables : un vrai (dans le second contexte où les souris ont effectivement reçu le choc électrique) et un faux (dans le premier contexte). Dans l'un et l'autre cas, les mêmes structures nerveuses sont impliquées. Les faux souvenirs sont bien des souvenirs et l'évocation de certains événements peut générer des faux souvenirs (fragilité du témoignage, limites de la psychanalyse...).