HAL
open science

Par quel prisme concevoir la conscience ? Neurosciences / sciences cognitives et philosophie

Francis Vincenti

▶ To cite this version:

Francis Vincenti. Par quel prisme concevoir la conscience ? Neurosciences / sciences cognitives et philosophie. Philosophie. 2021. dumas-03520653

HAL Id: dumas-03520653
https://dumas.ccsd.cnrs.fr/dumas-03520653

Submitted on 11 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE de REIMS
CHAMPAGNE ARDENNE
UFR Lettres et Sciences Humaines

MEMOIRE
de MASTER de recherche, niveau 2
PHILOSOPHIE

Par quel prisme concevoir la conscience ? Neurosciences / sciences cognitives et philosophie

Francis VINCENTI

Sous la direction de
Madame Véronique LE RU
Professeure de philosophie
Directrice du département philosophie de l'URCA

Année 2020 - 2021

REMERCIEMENTS

Je tiens à remercier particulièrement
ma directrice de mémoire,
Véronique LE RU
Directrice du département de philosophie
de l'Université de Reims Champagne Ardenne (URCA)
qui a bien voulu accepter,
avec bienveillance,
de me guider au cours de mes pérégrinations
et de m'apporter une aide précieuse
en 3 « r » : rééquilibrage, resserrement, réécriture ...

Je remercie également
l'ensemble des professeurs
du département de philosophie de l'URCA
pour leur enseignement et leur accompagnement
tout au long de mes années de découverte et d'exploration
en *Terra Incognita*.

Merci enfin à
Mesdames Isabelle Beautrait et Amandine Tong
des services administratifs du SEAD de l'URCA
pour leur accueil toujours chaleureux et amical.

Introduction

1° Le cerveau peut être considéré comme l'organe le plus « déterminant » de l'être humain.
Dans ce sens, il suffit de rappeler que la définition actuelle de la mort énoncée à l'article R 1232-1 du Code de la Santé publique (modifié par le décret n° 2005-949 du 2 août 2005) comprend trois critères cumulatifs dont les deux premiers se rapportent à la conscience et aux fonctions cérébrales[1]. Il en ressort que la mort se caractérise par la disparition complète et totale de toute activité cérébrale (par un encéphalogramme plat).
Le cerveau humain, a une double nature : d'abord biologique car son élément de base, est une cellule : le neurone ; ensuite, socioculturelle car il est le vecteur des interactions entre les individus humains et entre l'individu et son « milieu », selon l'expression de Georges Canguilhem[2].
A cet égard, le cerveau humain, comme tout organe d'un être vivant, est lui-même un produit de l'évolution. En réalité, il s'agit d'une double évolution, biologique et culturelle. En effet, selon Henri Atlan *« L'idée est communément admise, aujourd'hui, que l'évolution biologique dans l'espèce humaine s'est considérablement ralentie et a été relayée par une évolution culturelle »*[3].
Cette approche a conduit à l'élaboration de modèles dits de « coévolution cerveau/culture » dans lesquels

---

[1] Légifrance, Article R1232-1 du Code de la Santé publique (modifié par le décret n° 2005-949 du 2 août 2005) *« Si la personne présente un arrêt cardiaque et respiratoire persistant, le constat de la mort ne peut être établi que si les trois critères cliniques suivants sont simultanément présents :*
1° Absence totale de conscience et d'activité motrice spontanée ;
2° Abolition de tous les réflexes du tronc cérébral ;
3° Absence totale de ventilation spontanée ».

[2] *La connaissance de la vie*, chap III, « Le vivant et son milieu », Paris, Librairie philosophique J. Vrin, 2003, p 165.

[3] Henri Atlan, *Entre le cristal et la fumée*, Paris, Seuil, 1979, p 182.

« l'évolution du cerveau humain (est) *fonction de ses propres productions : le langage, les cultures, la technique, l'organisation sociale …* (C'est) *ce que F. John Odling-Smee* (et d'autres) *appelle une « construction de niche »*[4].

La seconde moitié du XX° siècle notamment a connu une accélération fulgurante et un développement considérable de la science du cerveau. Il est utile d'en rappeler le contexte historique en soulignant, à titre liminaire, que l'avènement de la modernité, à l'époque classique (XVII° siècle) et au temps des Lumières (XVIII°) a conduit à l'édification d'un modèle particulier de la science, celui de la science « moderne ». C'est sur cette base que le champ des connaissances scientifiques n'a cessé de s'étendre à partir de la fin du XVII° siècle. Dans ce cadre, l'objet de la science est d'abord le monde extérieur, la nature. Mais progressivement l'homme devient objet de la science en tant qu'être social au XIX° siècle (sociologie, anthropologie) puis, au cours du XX°, en tant qu'être vivant (biologie, génétique) et en particulier être vivant doté d'un cerveau (neurologie).

On soulignera quelques découvertes récentes de la science du cerveau qui comptent parmi les plus marquantes :

- Les neurones « miroirs ». Il s'agit de neurones qui contrôlent *« … l'exécution non de mouvements mais d'actes finalisés, c'est-à-dire tendus vers un but … ils* (s'activent) *à la fois dans le cerveau* (de celui) *qui agit et* (de) *celui … qui observe. Ces neurones peuvent aussi s'activer quand une action n'est pas visible* (ou audible) *…. et qu'elle peut être imaginée »*[5] ;
- Le réseau du mode par défaut. *« … même lorsque le cerveau est au repos, il fonctionne dans ce « mode par défaut ». Il y a une activité permanente au sein d'un réseau*

---

[4] Achille Weinberg, « L'évolution conjointe du cerveau et de la culture », in *Sciences Humaines*, n° 309, 12/2018, p 53.

[5] Vittorio Gallese, « co-découvreur » des neurones-miroirs avec Giacomo Rizzolatti, Entretien in *Philomagazine*, n° 135, 12/2019 – 01/2020, p 48.

réparti entre (les régions) *du cortex* ... ». A noter que « ... *les fluctuations observées dans le réseau en mode par défaut sont propres à chaque individu. Elles constituent une sorte de carte d'identité cérébrale ... »*[6] ;

- La mise en évidence que l'essentiel des opérations cérébrales sont inconscientes. Ainsi, il apparaît que la part inconsciente de l'activité cérébrale est plus grande que celle de la vie consciente, ce qui conduit à la définition d'un « nouvel inconscient », selon l'expression du neurologiste Lionel Naccache, à côté de celui, plutôt de ceux, au fondement des psychanalyses de Freud et de Jung. On ne peut ignorer que cette avancée des connaissances du cerveau n'a pu être réalisée qu'en raison de la mobilisation de moyens humains, matériels et financiers considérables, dans le cadre d'organismes ou de programmes de recherche publics ou privés. On peut citer, en France, l'Institut du Cerveau et de la Moëlle Epinière, fondation privée reconnue d'utilité publique, créée en 2006, près de l'Hôpital de la Pitié-Salpêtrière à Paris ; aux USA, le Allen Institute for Brain, organisme privé à but non lucratif, créé en 2003 à Seattle par Paul Allen, cofondateur de Microsoft. A noter que son président et directeur scientifique est Christof Kock, un des deux concepteurs de la théorie de la conscience dite de « l'information intégrée ». Il faut aussi évoquer, pour souligner l'importance des moyens mis en œuvre, au moins trois mégaprogrammes de recherche :

- Le Human Connectome Project (HCP) lancé en juillet 2009 par le National Institute of Health aux USA, afin de cartographier le fonctionnement des neurones, c'est-à-dire d'établir une « carte du réseau neuronal » (le connectome humain) ;

- Le Brain Activity Map Project lancé le 2 avril 2013 par l'Administration Obama doté d'un budget de 3 milliards de dollars sur dix ans en vue d'accélérer le

[6] Hervé Chneiwess, Entretien, in *La Recherche*, n° 557, 03/2020, p 70.

développement et les applications de technologies innovatrices dans la connaissance du cerveau humain ;

- Enfin, le Humain Brain Project lancé en 2013 par l'Union Européenne doté d'un budget de 1,2 milliard d'euros sur dix ans, pour étudier le cerveau humain, le soigner, le stimuler et s'en inspirer afin de développer des applications en matière d'« intelligence artificielle » (IA). A noter que compte-tenu de son intégration dans un but de développement de l'IA, ce projet a été contesté et réorienté dans une perspective plus large des neurosciences et sciences cognitives.

Le cerveau humain constitue un système qui présente trois propriétés principales :

- Il est d'une incroyable complexité ;
- C'est un système de circuits et de réseaux d'informations d'où son ultra-« connectivité » ;
- Il est doté d'une propriété étonnante : la « plasticité » c'est-à-dire la capacité de se modifier, se restructurer au cours du temps ; plus précisément, il s'agit, selon Jean-Pierre Changeux, de la *« capacité générale du neurone et de ses synapses à changer de propriétés en fonction de leur état d'activité ... C'est cette propriété qui confère aux réseaux de neurones à la fois flexibilité fonctionnelle, propriété de stockage et capacité d'auto-organisation »*[7] ;
- C'est un organe qui n'est pas indépendant des autres organes du corps, au contraire, c'est un système qui fait partie d'un ensemble, qui est intégré dans un tout. Ainsi, l'étude *« des communications entre le cerveau et le reste du corps ... s'est enrichie par la découverte ... du « cerveau entérique » … « un système d'interactions* (entre les systèmes) *intestinal, immunitaire, nerveux, central et périphérique et endocrinien »*[8] (d'où la naissance d'une nouvelle discipline appelée la neurogastroentérologie »),

[7] Jean-Pierre Changeux, *L'Homme de vérité*, Paris, Odile Jacob, 2002, p 43 et 45.

[8] Henri Atlan, *Cours de philosophie biologique et cognitiviste*, Paris, Odile Jacob, 2018, p 381 et 382.

un deuxième cerveau situé dans les intestins, qui est apparu le premier au cours de l'évolution.
2° Naturellement, la science du cerveau n'a pas achevé la découverte de son objet. Néanmoins, son développement a conduit à l'ouverture d'un nouveau champ d'étude et de recherche scientifique : la conscience.
Certes, cela fait longtemps que la conscience est un objet d'études. Déjà John Locke, s'était interrogé sur la question de la « conscience » (« consciousness ») notamment dans son œuvre majeure *Essay concerning Human Understanding* (*Essai sur l'entendement humain*) de 1694. Mais celle-ci ne devient un objet des neurosciences qu'à partir de l'article co-écrit par Francis Crick (codécouvreur de la structure de l'ADN, reconverti à la neurobiologie) et Christof Koch « Vers une théorie neurobiologique de la conscience » (« Toward a neurobiological theory of consciouness ») dans la revue *Neuroscience* de 1990 vol 2. Dans cet article, les auteurs soulignaient la nécessité d'étudier la nature des « liens » qui peuvent être établis entre activité neuronale et états mentaux. Il s'agit du problème dit des « corrélats neuronaux de la conscience » qui est – toujours – au cœur de cette discipline scientifique. Depuis le début des années 1990, les neurosciences ont elles-mêmes connu un fort développement, et, ont conduit, en particulier, à la formulation de diverses théories de la conscience, plus ou moins étendues.
Figurent aujourd'hui, parmi les grands thèmes de recherche en cette matière, outre les « corrélats neuronaux de la conscience » déjà cités :

- Le « code neural » ou les « signatures de la conscience » c'est-à-dire l'explication du fait que les aires corticales ont une fonction particulière, une fonctionnalité spécifique (exemple : l'aire de Broca et le langage) ;
- Les liens entre le cerveau et son environnement ; à cet égard, *« L'élucidation des relations entre le cerveau, le* (reste du) *corps et l'environnement, notamment … humain*

et culturel, constitue l'une des grandes questions des sciences de la conscience »[9] ;

3° Le développement des connaissances du cerveau, leur extension à celles concernant la conscience a donné naissance aux neurosciences et aux sciences cognitives dont il importe de retracer la formation.

A l'origine, l'analyse des actes, des comportements individuels relevait de la philosophie, dans le cadre d'une de ses branches : la psychologie. A la fin du XIX° siècle, celle-ci s'est éloignée de la philosophie pour se constituer en science autonome, en adoptant la méthode expérimentale mise au point par les sciences de la nature.

A partir du XX° siècle, la définition du neurone, en tant que cellule nerveuse conduit à l'émergence d'un nouveau champ de disciplines scientifiques dans le cadre des sciences du vivant et de la biologie : les neurosciences avec de multiples branches dont la neurobiologie (étude des composants du cerveau) et la neurophysiologie (étude de l'activité cérébrale).

En outre, dans les années 1940 et 1950 sont formulées – et mathématisées - les théories dites de l'information lesquelles vont servir de modèle aussi bien pour refonder la psychologie que pour développer un nouveau champ à la fois intellectuel et technologique connu sous l'appellation d'« intelligence artificielle ».

Enfin, il importe de relever que les études du cerveau puis de la conscience ont été menées en lien avec celles concernant la problématique du langage – donc la linguistique – ou encore l'approche objective de l'être humain et de son évolution c'est-à-dire l'anthropologie. Finalement, c'est de la convergence de ces différents champs de questionnement et de recherche que vont naître les « sciences cognitives ».

A ce stade, une observation s'impose : toute la difficulté, on le sait, est de bien nommer les choses. Il s'avère donc

---

[9] Lionel Naccache, « Observer la conscience », in *Pour la Science*, n° 500, 06/2019, p 80.

utile de définir les termes « neurosciences » et « sciences cognitives »[10], étant entendu que dans les pages qui suivent, on entendra, par neurosciences, les sciences qui ont pour objet le cerveau et ses liens avec la conscience ; et par sciences cognitives, le groupe pluridisciplinaire dont les neurosciences ne sont qu'un élément au côté de la philosophie, la psychologie, l'intelligence artificielle, la linguistique, l'anthropologie bien qu'elles en constituent le « bateau amiral », un ensemble cohérent représenté par l'hexagone figurant en appendice ci-dessous.

De même, que faut-il entendre par l'âme, l'esprit, la pensée, l'idée, la conscience (ou les consciences puisqu'en anglais, il y a deux termes - consciousness et awareness - pour désigner la « chose ») le psychique, la cognition, l'intelligence ? On parlera principalement de conscience – encore faudrait-il distinguer entre les états de conscience : « l'éveil », le(s) sommeil(s), la conscience immédiate, réflexive, intentionnelle ou phénoménale. On parlera également d'intelligence artificielle, (parfois (re)nommée science numérique ou encore science de l'Informatique et de l'Automatique - encore IA) qu'il serait plus juste de nommer « computation artificielle », mais que nous continuerons, dans un souci de simplification, à désigner par l'expression « intelligence artificielle » voire par son acronyme « IA ».

4° Où en est-on aujourd'hui ? Rien n'est encore certain ni définitif s'agissant de la connaissance de la conscience. Celle-ci reste un « mystère » (John Searle). Et d'après

---

[10] Nous proposons les définitions suivantes :

Neurosciences : Ensemble des disciplines biologiques qui ont pour objet l'étude du système nerveux central, son anatomie, sa physiologie et son fonctionnement, le plus souvent en liaison avec des activités mentales particulières (le langage, la mémoire, la vision etc…).

Sciences Cognitives : Ensemble des disciplines qui ont pour objet commun l'étude de la fonction psychique qui assure la réception, la transformation et le traitement des données provenant du monde extérieur, à partir desquelles l'être humain élabore la connaissance du réel.

Jerry Fodor *« … ce que la science cognitive a principalement découvert sur l'esprit, c'est que nous ignorons comment il fonctionne »*[11]. Mais, certains neuroscientifiques soulignent que l'on en est qu'au début de la science de la conscience. Si, force est de constater qu'il n'existe pas, à ce jour, de théorie générale de la conscience, il existe au moins deux théories dominantes. En premier lieu celle dite de « l'espace de travail neuronal global conscient » formulée par Jean-Pierre Changeux et Stanislas Dehaene qui fait apparaitre que la conscience ne constitue pas un tout mais qu'elle émane d'un centre, d'un lieu originaire, d'un niveau d'organisation et d'activité d'une grande complexité. En second lieu, la théorie de l'information intégrée *(« integrated information theory »)* soutenue par Giulio Tononi rejoint par Francis Crick et Christof Koch, pour laquelle le niveau de conscience d'un système (cerveau ou machine) dépend des interconnexions que l'on retrouve dans ce réseau. Autrement dit, tout système complexe interconnecté par des relations causales aura un certain niveau de conscience qui dépendra de sa complexité et du niveau d'intégration dont il est capable.

5° L'hypothèse de travail qui sous-tend cette réflexion est que le développement de ces connaissances ne peut pas ne pas avoir d'impacts sur la philosophie entendue comme un ensemble de « matières à penser » (métaphysique, épistémologie, éthique, esthétique etc.). Pour donner une vue d'ensemble du mémoire, on précisera que sa problématique générale est de savoir quels sont les implications et les enjeux philosophiques de cette extension du champ des connaissances scientifiques concernant le cerveau et la conscience, dans le cadre des neurosciences et des sciences cognitives. En quoi ce fait a-t-il des incidences en philosophie et en quoi la philosophie peut-elle être utile à celles-ci ?

A cet égard, trois questions peuvent être formulées :

---

[11] Jerry Alan Fodor, *L'esprit, ça ne marche pas comme ça*, Paris, Odile Jacob, 2003, p 166.

a) Le développement des connaissances du cerveau permet-il d'expliquer la conscience ? Celle-ci est-elle d'emblée contenue dans le cerveau ou apparaît-elle à partir d'un certain niveau de complexité de celui-ci, qui en est la condition et non la définition ? Cette question porte sur la nature de la conscience considérée comme objet scientifique, elle est en rapport avec l'ontologie et concerne directement un des problèmes majeurs de la philosophie : le problème corps-esprit (« mind-body problem »). Elle conduit à considérer le contexte historique, déjà évoqué, d'édification de la science moderne, dans lequel les neurosciences sont nées et se sont développées. Ceci implique une mise en perspective qui comportera deux volets. Le premier se rapporte à la mise en place de cette problématique dans le cadre de la philosophie de la modernité et du sujet, à partir du XVII° siècle, notamment avec Descartes et Spinoza. (1ère partie) ; le second couvre deux moments essentiels de la philosophie de la 2ème moitié du XX° : d'une part, la confrontation entre deux grands courants, le fonctionnalisme et le naturalisme, d'autre part, l'élaboration d'un autre courant qui a, au moins tenté de renouveler la philosophie de l'esprit, la neurophénoménologie (2ème partie).

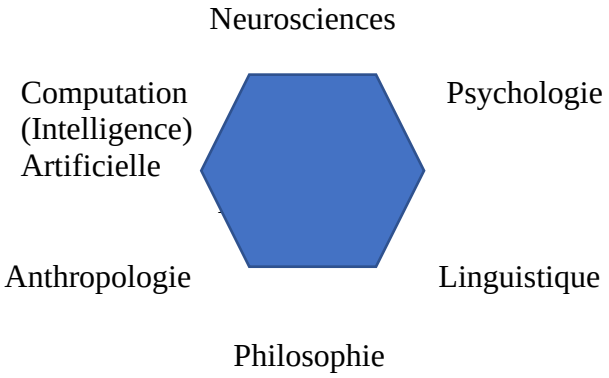
b) Le développement des connaissances du cerveau implique nécessairement l'augmentation du pouvoir, de la puissance d'agir des hommes sur l'être humain lui-même et sur le monde. Dès lors, ne faut-il pas évaluer les impacts et, le cas échéant, fixer des limites aux applications de ces connaissances ? Cette question est en rapport avec l'éthique. Elle conduit à situer les domaines dans lesquels la progression du savoir en sciences cognitives est susceptible d'avoir des effets sur la vie des individus et des sociétés. A cet égard, On peut considérer que le premier champ concerné est celui de l'intelligence artificielle. Dès lors, il importe de bien le délimiter et d'estimer la nature et de mesurer l'étendue des

conséquences qui en découlent sur l'être humain. En fait, cette problématique pose un ensemble de questions touchant à la régulation de ces impacts. En outre, sur un plan plus pratique, ce sujet porte sur la place de la science dans la société, le rôle du citoyen à son égard, et la mission du chercheur. Ces points seront traités dans la 3ème partie.

c) Enfin, le fait que la recherche scientifique ait commencé à aborder le sujet de la conscience conduit à poser une troisième question : peut-on construire une théorie scientifique de la conscience et, dans cette hypothèse, qu'elles en sont les conditions ? Cette question est en rapport avec l'épistémologie. Elle ne sera pas développée dans le cadre de ce mémoire sauf à indiquer ces quelques éléments de réponse : ce sont a priori les conditions générales qui valent pour toute discipline scientifique qui devraient s'appliquer ici, avec une double spécificité qui concerne déjà les sciences humaines et sociales : d'une part, la question de la position du sujet par rapport à l'objet tenant au fait que l'individu humain est à la fois sujet et objet de la discipline ; d'autre part, la prise en compte d'une contrainte qui s'impose de manière incontournable aux neurosciences et plus largement aux sciences cognitives : la nécessité de mener la recherche dans un cadre pluridisciplinaire.

……....

Appendice : Hexagone des sciences cognitives



Ière partie

La mise en place de la problématique de la nature de la conscience comme objet de science au XVII° siècle

Introduction

Cette question s'inscrit, on l'a vu, dans le cadre plus large du problème corps-esprit (« mind-body problem ») récurrent depuis la naissance de la philosophie.
Dès l'Antiquité, cette problématique était l'objet d'une opposition radicale entre d'un côté les Epicuriens et les Stoïciens, initiateurs du matérialisme, et de l'autre côté Platon suivi par Aristote, concepteurs de l'idéalisme.
Dans la tradition philosophique, chez Platon puis dans la pensée chrétienne et scolastique dominante au Moyen-Âge, l'âme et le corps sont les deux parties distinctes et séparées, constitutives de l'être humain. En outre, la tradition philosophique confère à l'âme une supériorité par rapport au corps.
Mais à l'Age classique, au XVII° siècle, dans le contexte marqué par la révolution scientifique et la naissance de la physique moderne, la philosophie, notamment à partir de Descartes et Spinoza, propose et développe une nouvelle conception de l'âme ou de l'esprit définie, en rupture avec la tradition évoquée ci-dessus, comme une activité de nature mentale, ainsi qu'une nouvelle conception du corps régi par des lois qui lui sont propres, de nature mécanique, physique. Cette reconceptualisation des deux termes du problème corps-esprit conduit à reposer la question fondamentale de leur possible relation.
Descartes et Spinoza, ont avancé des philosophies fondées sur la notion générale de substance, englobant les concepts de corps et d'esprit, mais une notion de substance pensée de manière radicalement différente, sinon contradictoire.
Il importe donc dans un premier temps de faire l'analyse de ces deux philosophies classiques, qui expriment cette nouvelle approche, celle de Descartes (Chap. 1) puis de

Spinoza, (Chap. 2) avant de tenter d'établir un lien entre elles (Chap. 3).
En revanche, nous n'aborderons pas un autre volet ouvert par la philosophie classique, à savoir l'empirisme, initié par les philosophes britanniques John Locke et David Hume, qui n'est pas directement lié au problème corps-esprit.

Chapitre 1 : René Descartes

En préalable, il importe de rappeler les liens qui rattachent la philosophie de Descartes (1596 – 1650) à la tradition aristotélico-thomasienne et à l'Ecole Scolastique du Moyen Âge.
Ce sont là les sources de sa pensée, issues de l'éducation qu'il a reçue de huit à seize ans au collège des Jésuites de La Flèche *« l'une des plus célèbres écoles de l'Europe »* [12].
Sa philosophie se situe, dans la continuité de ces courants de pensée, avec le maintien de l'usage de certains concepts, tel celui de la substance. Mais, surtout, Descartes opère une rupture radicale avec la Tradition, par la consécration de la Raison dans son *Discours de la méthode*, par l'invention du sujet avec le *Cogito*. Cette rupture marque l'avènement du courant philosophique et culturel qui nait au XVII° siècle sous l'appellation de « Modernité ». Cette doctrine se présente sous un double aspect : ontologique d'une part, et épistémologique, méthodologique, d'autre part.
Au regard du problème corps-esprit, la pensée de Descartes tend au rejet de la représentation aristotélicienne de l'âme comme principe d'animation du corps.

[12] *Discours de la Méthode*, 1ère partie, 6ème §, (AT VI, 5) Paris, Editions Garnier Frères, 1960, p. 35.

La philosophie cartésienne s'articule autour de deux points fondamentaux : le dualisme ontologique et l'union substantielle de l'âme et du corps.

§ 1 : Le dualisme ontologique
Théorie de la dualité des substances de l'âme et du corps

1° Dans les *Méditations métaphysiques*, la deuxième et surtout la sixième, Descartes souligne que l'âme et le corps sont deux substances distinctes et séparées : idéelle, *« chose pensante » (res cogitans)* pour l'âme et matérielle ou mécanique, *« chose étendue » (res extensa)* pour le corps : *« ... parce que d'un côté j'ai une claire et distincte idée de moi-même, en tant que je suis seulement une chose qui pense et non étendue, et que d'un autre j'ai une idée distincte du corps, en tant qu'il est seulement une chose étendue et qui ne pense point, il est certain que ce moi, c'est-à-dire mon âme ... est entièrement et véritablement distincte de mon corps ... »*[13].
Ces deux substances, du fait de la volonté de Dieu de les séparer, existent donc l'une indépendamment de l'autre, elles n'ont pas besoin l'une de l'autre pour agir.
A noter, en outre, que leur existence est saisie par l'expérience, la démarche empirique.
Il importe de relever leurs différences de nature.
2° L'âme ou l'esprit est « hiérarchiquement » la substance première. Soulignons d'abord que, selon l'intitulé même de la deuxième méditation, « (l'esprit) *est plus aisé à connaître que le corps* ». Mais, surtout, l'esprit est ce *« je suis .... une chose qui pense »*, ce moi constitutif du sujet, qui va servir de fondement à la philosophie du sujet qui sera dominante jusqu'au XIX° siècle.

[13] *Méditations métaphysiques*, sixième, (AT IX, 62) Paris, Flammarion, coll « GF », 2011 édition revue et corrigée, p. 187.

Il n'existe qu'une fonction caractéristique et unique de l'âme : une fonction théorétique ou cogitative, à savoir que c'est par l'âme que j'existe, que j'accède à la conscience, au cogito ainsi qu'à la vérité : *« … c'est, ce me semble, à l'esprit seul, et non point au composé de l'esprit et du corps, qu'il appartient de connaître la vérité de ces choses-là* (qui sont hors de nous) »[14].

En outre, la nature de l'âme est en elle-même une et indivisible, c'est-à-dire qu'*« il ne faut pas imaginer que l'esprit ait des parties »*[15].

Pour autant, l'âme dispose de facultés, de capacités, de ressources. Elle ne se réduit pas aux capacités intellectuelles, à la faculté de penser. En effet *« … entendre, vouloir, imaginer mais aussi sentir, est la même chose que penser »* ou encore *« Nous pouvons aussi concevoir fort distinctement diverses façons de penser, comme entendre, imaginer, se souvenir, vouloir … »* [16].

L'âme recouvre toutes les opérations qui relèvent du psychisme et de ce que l'on appelle aujourd'hui la « cognition », jusqu'aux « *émotions intérieures* » visées à l'article 147 du Traité *Les passions de l'âme*.

Dans la perspective cartésienne, l'âme est également le siège de la volonté libre, c'est-à-dire de la capacité, de la puissance de vouloir, de juger, autrement dit du libre arbitre.

A noter encore que dans la conception de Descartes, l'esprit (et le produit de son activité, à savoir la pensée, les idées) s'analyse comme une suite d'images, de représentations des choses, de lui-même et du monde extérieur. Cette approche est au fondement du modèle

---

[14] *Ibid.* (AT IX 65-66) p. 197.

[15] *Méditations métaphysiques*, Réponses aux cinquièmes objections … contre la sixième Méditation, traduction de Clerselier, 1647, Paris, Flammarion, coll « GF », 2011 édition revue et corrigée, p. 422.

[16] *Les principes de la philosophie*, 1ère partie, § 9 et 65, (AT IX, 28 et 54) Paris, Libraire philosophique J. Vrin, 1989, pp. 56 et 102.

représentationnel de l'esprit. Ce *« moi qui suis une chose qui pense »* se définit dans des termes qui se réfèrent à l'intériorité et à la réflexivité du sujet, comme un processus interne impliquant un moi se percevant et se représentant lui-même ; le moi étant alors son propre spectateur, d'où l'assimilation de l'esprit cartésien à une sorte de « théâtre intérieur ». Cette conception sera vivement critiquée par certains philosophes de l'esprit contemporains.

3° Le corps, de son côté, en tant que substance étendue, matérielle, est ce dont nous avons la perception immédiate en tant que « chose », puis en tant qu'acte ou sensation : *« Je me considérais, premièrement, comme ayant un visage, des mains, des bras et toute cette machine composée d'os et de chair ... laquelle je désignais par le nom de corps ... Je considérais ... que je me nourrissais, que je marchais, que je sentais et que je pensais »*[17].

Mais, on peut distinguer, dans la pensée cartésienne, deux approches de la conception du corps.

A titre principal, le corps est considéré comme un ensemble de parties liées les unes aux autres, un assemblage de membres et d'organes, une composition de multiples pièces et rouages, d'où sa conceptualisation sur le modèle de la *« machine composée d'os et de chair »*, de l'automate, de l'*« horloge composée de roues et de ressorts »*[18].

Contrairement à l'âme, le corps est divisible et il est aussi modifié dès lors que l'assemblage est modifié.

Pour autant, en seconde approche, le corps a une unité au moins relative. Il s'agit d'abord d'une unité propre fondée sur la complémentarité, sur l'indissociabilité plutôt, de ses organes : *« ... il est un et en quelque façon indivisible à raison de la disposition de ses organes qui se rapportent tellement tous l'un à l'autre que, lorsque quelqu'un d'eux*

---

[17] *Méditations métaphysiques*, deuxième, (AT IX, 20) Paris Flammarion coll « GF », 2011, édition revue et corrigée, p. 75.

[18] *Discours de la Méthode*, 5ème partie, 9ème §, (AT VI, 59) Paris, Editions Garnier Frères, 1960, p. 98.

est ôté, cela rend tout le corps défectueux »[19]. Ensuite son unité résulte plus encore du fait de son union avec l'âme laquelle est jointe à tout le corps et non à l'une au l'autre de ses parties. Il s'ensuit que si le corps change, il demeure malgré cela, le même tout le temps qu'il est uni à l'âme.

§ 2 : L'union substantielle de l'âme et du corps et leurs interactions

1° Pour autant, Descartes n'est pas un dualiste radical.
L'âme ou l'esprit et le corps sont, certes, deux substances distinctes et séparées, mais elles sont aussi conjointes, confondues, mélangées, finalement unies ; au point que l'on peut qualifier cette analyse d'union psychophysique.
Celle-ci se trouve explicitement affirmée, à de multiples reprises dans l'œuvre de Descartes. Ainsi dans le *Discours de la méthode* 5ème partie (1637), la Lettre à Hyperaspistes d'août 1641, *Les Principes de la philosophie* 1ère partie § 48 (1644), le traité des *Passions de l'âme* (1649) et notamment, dans la fameuse formule de la sixième des *Méditations Métaphysiques* (1642) : *« La nature m'enseigne ... que je ne suis pas seulement logé dans mon corps, ainsi qu'un pilote en son navire, mais, outre cela, que je lui suis conjoint très étroitement et tellement confondu et mêlé, que je compose comme un seul tout avec lui »*[20]. Mais, c'est particulièrement dans ses lettres à la princesse Elisabeth de Bohême des 21 mai et 28 juin 1643 que Descartes, poussé par sa correspondante, « dans ses derniers retranchements », soutient et développe sa conception.
2° Il convient d'en souligner les caractéristiques essentielles.

---

[19] *Les passions de l'âme*, art 30, (AT XI, 350) Paris, Flammarion, coll « GF », 1996, p. 118-119.
[20] *Méditations métaphysiques*, sixième, (AT IX, 64) Paris, Flammarion, coll « GF », 2011 édition revue et corrigée, p. 193.

L'union de l'esprit et du corps résulte de la puissance absolue et ordonnée de Dieu qui a décidé que les deux ne peuvent jamais exister l'un sans l'autre.
Mais pour l'être humain, cette réalité est une donnée de l'expérience, un enseignement de la vie avec les sensations de douleur et de plaisir, de faim et de soif, les émotions et les passions de l'âme (colère, joie, tristesse, amour).
Cette union recouvre un système d'interactions entre l'esprit et le corps, des mécanismes d'actions réciproques de l'esprit sur le corps et du corps sur l'esprit, dont Descartes va tenter de donner une description physiologique. Elle comporte deux éléments principaux tenant à l'existence, d'une part, de ce qu'il appelle les « esprits animaux », que l'on pourrait rapprocher aujourd'hui de l'« influx nerveux » et, d'autre part, d'une localisation de l'âme dans le corps, laquelle aurait son « siège » dans le cerveau, plus particulièrement, dans la glande pinéale. Cette description est présentée dans la 2ème Partie du *Monde ou Traité de la Lumière,* intitulée *l'Homme* puis, sous forme résumée, dans le *Discours de la Méthode* 5ème partie et dans la *Dioptrique* qui suit le *Discours*, enfin de manière plus développée dans le Traité des *Passions de l'Âme*. Les « esprits animaux » sont « ... *les plus vives et les plus subtiles parties du sang, que la chaleur a raréfiées dans le cœur* (qui) *entrent sans cesse en grande quantité dans les cavités du cerveau ... celles de ses parties qui sont les plus agitées et les plus subtiles ... composent les esprits animaux ... ce que je nomme ici des esprits* »[21]. La glande pinéale quant à elle, est « ... *une petite glande dans le cerveau ... située dans* (son) *milieu ... et tellement suspendue au-dessus du conduit par lequel les esprits de ses cavités antérieures ont communication avec ceux de la postérieure que les moindres mouvements qui sont en elle peuvent beaucoup pour changer* (le) *cours de ces esprits, et réciproquement que les moindres*

[21] *Les passions de l'âme*, art 10, (AT XI, 334) Paris, Flammarion, coll « GF », 1996, p. 105-106.

changements qui arrivent au cours des esprits peuvent beaucoup pour changer les mouvements de cette glande »[22]. La glande pinéale est donc l'organe qui sert d'interface, d'intermédiaire entre l'âme et le corps. Ses mouvements traduisent les échanges entre les deux. En fait, dans l'anatomie cartésienne, elle est le seul organe, dans le cerveau, à ne pas être double. C'est son unicité qui la rendrait compatible avec l'indivisibilité de l'âme. Elle constitue un organe unique où sont fusionnés les stimuli provenant des organes sensoriels ; elle est le *« lieu où les deux images qui viennent par les deux yeux, ou les deux autres impressions qui viennent d'un seul objet par les doubles organes des autres sens, se puissent assembler en une avant qu'elles ne parviennent à l'âme, afin qu'elles ne lui représentent pas deux objets au lieu d'un. Et on peut aisément concevoir que ces images ou autres impressions se réunissent en cette glande par l'entremise des esprits qui remplissent les cavités du cerveau ... »*[23].

Néanmoins, cette notion de l'union substantielle de l'âme et du corps, complétée par celle de leurs interactions, ne permet pas à Descartes d'expliquer clairement et précisément les mécanismes de leurs relations.

C'est bien ce qu'il ressort de ses échanges de correspondance avec la princesse Elisabeth de Bohême.

Ainsi, en réponse à une lettre dans laquelle Elisabeth lui demandait comment l'âme, substance immatérielle peut déterminer le corps, substance matérielle, Descartes se référait, dans sa lettre du 21 mai 1643, à la notion de « pesanteur » de la « physique » scolastique, supposée être « *... une qualité réelle, dont nous n'avons point d'autre connaissance, sinon qu'elle a la force de mouvoir le corps ... vers le centre de la terre ...* (sans) *que cela se fasse par un attouchement réel d'une superficie contre une autre* ». Il affirmait que cette notion *« nous a été donnée pour*

[22] *Ibid.* art 31, (AT XI, 352) p. 119.

[23] *Ibid.* art 32, (AT XI, 353) p. 120.

concevoir la façon dont l'âme meut le corps »[24]. Dans son autre lettre du 28 juin 1643, en réponse à Elisabeth lui proposant plutôt de considérer l'âme comme une chose douée d'étendue, Descartes paraît faire sienne la thèse d'une certaine co-extension de l'âme. On pourrait donc admettre une certaine « corporéité » de l'âme en tant précisément qu'elle est unie, plus précisément, co-étendue au corps. Mais cette « étendue » de l'âme ne serait pas du même ordre que celle du corps, en ce que celui-ci occupe nécessairement un certain lieu. Ainsi, l'extension de l'âme devrait s'entendre au sens d'une co-extension sans lieu.

En dernier ressort, et principalement dans sa lettre du 21 mai 1643, Descartes justifie sa conception en recourant à des *« notions primitives »* : il s'agit de facultés que Dieu a mises en nous, des idées « innées », des vérités qui nous sont connues d'elles-mêmes, *« ... des originaux, sur le patron desquels nous formons toutes nos autres connaissances* ... (que) *nous ne pouvons chercher ... ailleurs qu'en notre âme qui les a toutes en soi par sa nature, mais qui ne les distingue toujours assez les unes des autres, ou bien ne les attribue aux objets auxquels on les doit attribuer »*[25].

Parmi celles-ci *« nous n'avons, pour le corps ... que la notion d'extension, ... pour l'âme seule, nous n'avons que celle de la pensée, ... enfin pour l'âme et le corps ensemble, nous n'avons que celle de leur union, de laquelle dépend celle de la force qu'a l'âme de mouvoir le corps, et le corps d'agir sur l'âme ... »*[26]. Autrement dit, il faut bien distinguer le plan des essences ou substances et le plan de la configuration des existences. Si sur le premier plan (essences) se trouvent bien l'étendue et la pensée, deux substances irrémédiablement distinctes et séparées ;

---

[24] Lettre à Elisabeth 21 mai 1643, (AT III, 667) Paris, Flammarion, coll « GF », édition mise à jour 2018, p. 69-70.

[25] *Ibid.* (AT III 665-666) p. 68-69.

[26] *Ibid.* (AT III 665) p. 68.

l'union de l'âme et du corps ensemble se situe sur le second (existences) et elle n'est entendue qu'au sens de « composé » des deux substances, et nullement comme une troisième substance. Pour autant, cette union constitue la nature humaine.

Il est clair que ces notions primitives ont des allures de postulats métaphysiques. Dès lors, l'explication de l'union substantielle par le recours à un postulat métaphysique fondé sur la jonction du corps et de l'esprit par l'action divine, est l'indice de la difficulté rencontrée par Descartes qui ne parvient pas à expliquer l'union du corps et de l'esprit à partir de leur distinction réelle, empirique et d'une démonstration objective et scientifique. Il importe de souligner que lui-même avoue cet état de fait : *« ... ne me semblant pas que l'esprit humain soit capable de concevoir bien distinctement, et en même temps, la distinction entre l'âme et le corps, et leur union ; à cause qu'il faut, pour cela, les concevoir comme une seule chose, et ensemble les concevoir comme deux, ce qui se contrarie ... »*[27]. Cette difficulté à penser et à justifier à la fois la distinction entre l'esprit et le corps et leur union substantielle constitue une problématique qui traverse toute la philosophie de Descartes. Comment résoudre en effet, l'énigme ou le paradoxe de deux choses radicalement différentes qui ne seraient pourtant qu'une seule ?

On peut considérer que Descartes a tenté de dépasser ce problème en indiquant que la dualité des substances est une donnée résultant des notions primitives, des idées innées, alors que l'union, bien qu'elle soit elle-même une notion primitive, est plutôt perçue à travers l'expérience subjective, le « vécu » personnel dès lors qu'elle *« s'éprouve mieux qu'elle ne se conçoit »*[28].

Force est de constater que les deux éléments du problème ne se situent pas sur le même plan, mais relèvent de deux

---

[27] Lettre à Elisabeth du 28 juin 1643, (AT III, 693) Paris, Flammarion, coll « GF », éd mise à jour 2018, p. 75.

[28] Véronique Le Ru, cours M2, p. 34.

domaines différents : d'une part, celui de la métaphysique, de l'ontologie, s'agissant de la co-existence de deux substances distinctes et séparées et sans interactions possibles, l'une idéelle, sensible (l'esprit) et l'autre matérielle, physique (le corps) ; d'autre part, celui de la réalité, de l'épistémè, concernant leurs relations réciproques. La dualité des substances est une vérité métaphysique, un enseignement de la philosophie, accessible au seul intellect ; l'union du corps et de l'esprit est un enseignement de la vie qui résulte de l'expérience.
Il s'ensuit une différence de mode d'approche : d'un côté, la rationalité philosophique qui révèle la dualité de l'âme et du corps, de l'autre l'expérience vécue qui montre leur union. En introduisant cette différence de plans, Descartes s'éloigne du plan ontologique de la « substance », de la définition de l'être, pour se placer sur le plan épistémique, de l'accès à la connaissance de la nature humaine. A cet égard, L'union substantielle de l'âme et du corps est une construction rationnelle pour mieux connaître et mieux dire ce qu'est l'Homme en prenant en compte leurs interactions.
La philosophie de Descartes touchant au problème du corps et de l'esprit ouvre droit à deux lectures selon que l'on s'appuie, soit sur son dualisme substantiel corps-esprit fondé sur les postulats métaphysiques qui aboutit à l'idéalisme ; soit sur sa théorie de l'union substantielle laquelle peut fournir un modèle causal en vue d'élaborer une théorie des processus mentaux.
C'est à cette dernière interprétation que se réfèrent certains neuroscientifiques ce qui montre que la pensée cartésienne est toujours d'actualité.

Chapitre 2 : Baruch Spinoza

Après avoir développé sa philosophie sur l'origine, la nature et la force des affects, Spinoza (1632 - 1677) se livre, dans la préface de la 5ème partie de son œuvre majeure

l'*Ethique*, à une double critique de la théorie cartésienne relative au problème corps-esprit. En effet, il récuse tant la distinction des substances de l'âme et du corps, que la réalité de leurs interactions. De même, il rejette (implicitement) la conception représentationnelle de l'esprit.

Il mène cette critique à partir d'une analyse de la théorie des rapports entre l'âme et la glande pinéale. *« En vérité je ne puis assez m'étonner qu'un Philosophe ... admette une hypothèse plus occulte que toute qualité occulte. Qu'entend-il, je le demande, par l'union de l'Âme et du Corps ? Quelle conception claire et distincte a-t-il d'une pensée très étroitement liée à une certaine petite portion de l'étendue ? »*[29].

Les reproches de Spinoza à Descartes sont de trois ordres : Sur l'intelligibilité, d'abord. Descartes ayant distingué et séparé l'âme du corps, n'a pas pu établir leur union par une *« cause singulière »* (telle cause et pas une autre). Pour l'expliquer, il recourt à la *« cause de tout l'Univers c'est à dire à Dieu »*[30]. Sur la causalité, ensuite. Descartes n'explique pas comment fonctionnent les rapports entre l'âme et les mouvements et positions de la glande pinéale, comment des éléments psychiques (*« passions »*, *« jugements »*) peuvent être affectés (joints ou disjoints) par des *« causes corporelles »*. Enfin, Spinoza met en cause l'existence même de la glande pinéale.

En réponse, il propose, dans la perspective d'un monisme ontologique, sa théorie du « parallélisme » qui le conduit à concevoir l'identité individuelle de l'homme qui s'exprime à travers ces deux attributs.

§ 1 : Le monisme ontologique
Théorie de l'unité de la substance

---

[29] *Ethique*, V° partie, préface, trad. par Charles Appuhn, Paris, Flammarion, coll « GF », 1965, p. 305.

[30] *Ibid.* p. 305.

1° A titre liminaire, il convient de considérer les définitions, données par Spinoza de trois notions de base : substance, attribut et mode :

a) *« J'entends par substance ce qui est en soi et est conçu par soi : c'est-à-dire ce dont le concept n'a pas besoin du concept d'une autre chose, duquel il doive être formé »*[31]. Spinoza définit la substance comme étant ce qui est cause de soi, sans autre relation qu'à soi-même. Il énonce ensuite le postulat métaphysique au fondement de sa philosophie *« ... il ne peut y avoir qu'une substance unique d'une certaine nature »* ; *« ... il n'existe qu'une seule substance de même nature »*, et *« Toute substance est nécessairement infinie »*[32]. Au sens de Spinoza, infini s'entend sans début ni fin, c'est-à-dire de ce qui n'a pas été créé. Il n'est qu'une seule et même substance, *« Dieu ou la Nature »* (« *Deus sive Natura* »)[33], caractérisée par l'unité mais aussi la totalité, l'infinité, l'intemporalité et dont fait partie l'homme.

b) *« J'entends par attribut ce que l'entendement perçoit d'une substance comme constituant son essence »*[34]. Il y a une infinité d'attributs, chacun étant une expression de l'essence de la substance unique mais deux seulement sont accessibles à notre connaissance : la pensée et l'étendue, (*Cogitatio et_Extensio*). Ainsi, celle-ci se présente à nous sous la forme de ces deux attributs. Pour autant *« ... tout ce qui peut être perçu ... comme constituant*

---

[31] *Ibid.* I° partie, 3ème définition, p. 21.

[32] *Ibid.* I° partie, 8ème proposition, scolie II, p. 27 et p. 28. ; 8ème proposition, p. 26.

[33] *Ibid.* IV° partie, 4ème proposition, démonstration, p. 224.
A noter deux autres traductions : Robert Misrahi, *« Dieu c'est-à-dire la Nature »*, Paris, LGF, Le livre de poche, coll « Les classiques de la philosophie », 2005, p. 287. ; Bernard Pautrat, *« Dieu autrement dit la Nature »*, Paris, Seuil, coll « Points Essais », dernière édition 2010, p. 36.

[34] *Ibid.* I° partie, 4ème définition, p. 21.

une essence de substance, appartient à une substance unique et en conséquence ... substance pensante et substance étendue, c'est une seule et même substance comprise tantôt sous un attribut, tantôt sous l'autre »[35].

c) *« J'entends par mode les affections d'une substance, autrement dit ce qui est dans une autre chose ; par le moyen de laquelle il est conçu »*[36]. Chaque mode relève d'un attribut dont il est une expression particulière. Il importe donc de distinguer deux niveaux de réalité ontologique : celui des attributs d'une part et celui des modes d'autre part, dont relèvent le corps et l'esprit, le premier étant un mode de l'attribut étendue, le second un mode de l'attribut pensée.

2° La distinction entre pensée et étendue n'implique aucune dualité ontologique ou substantielle, elle n'est qu'une distinction conceptuelle. Pour autant, si les deux attributs se conçoivent chacun séparément, l'un de l'autre, cette séparation n'exclut pas qu'il y ait des correspondances entre les deux, notamment que l'attribut pensée puisse correspondre à l'attribut étendue.

Ainsi, en vertu de la thèse ontologique de la substance unique, Spinoza rejette le postulat métaphysique cartésien de la dualité de substances, de l'existence de deux substances, la pensée et l'étendue, comme substances constitutives de l'homme.

§ 2 : L'identité individuelle de l'homme

1° Le monisme ontologique de Spinoza doit être précisé dans sa portée. Ayant posé qu'il existe une seule et même substance qui se présente sous la forme d'une infinité d'attributs dont deux seulement sont connus de l'être humain, Spinoza conclut que ce dernier s'identifie par son intégration dans la substance unique ; le corps et l'esprit

---

[35] *Ibid.* II° partie, 7ème proposition, scolie, p. 76.

[36] *Ibid.* I° partie, 5ème définition, p. 21.

n'étant qu'une seule et même substance, réalité saisie sous la forme de deux attributs, deux aspects homologues, sous deux perspectives distinctes : *« … l'Âme et le Corps sont un seul et même individu* (ou *« une seule et même chose »*) *qui est conçu tantôt sous* l'attribut *de la pensée, tantôt sous celui de l'étendue … »*[37]. Spinoza recourt à la métaphore des deux faces d'une même pièce de monnaie. Il n'y a donc pas d'unité du corps et de l'esprit, au sens de confusion entre les deux.

La 13ème proposition de la 2ème partie de l'*Ethique* énonce des caractéristiques de chacune de ces deux « faces » qu'il convient d'analyser.

2° Ce qui identifie le corps, c'est son appartenance à l'attribut étendue : *« … le corps, c'est-à-dire un certain mode de l'étendue existant en acte »*[38]. Les corps sont une modification particulière de l'attribut étendue, ils possèdent tous à l'origine une communauté d'essence. Mais tous les corps sont en rapport, en interaction les uns avec les autres.

Il suit de là une véritable dialectique des modalités selon lesquelles un corps est affecté par un autre corps en fonction de la nature du corps qui affecte et la nature du corps qui est affecté. *« Toutes les manières dont un corps est affecté par un autre, suivent de la nature du corps affecté et en même temps de celle du corps qui l'affecte ; en sorte qu'un seul et même corps est mû de différentes manières en raison de la diversité des corps qui le meuvent, et qu'en retour différents corps sont mus de différentes manières par un seul et même corps »*[39].

Aussi, Spinoza distingue les corps simples et les corps composés qu'il appelle « individus »[40]. Les corps sont

---

[37] *Ibid.* II° partie, 21ème proposition, scolie, p. 99. et III° partie, 2ème proposition, scolie, p. 137.

[38] *Ibid.* II° partie, 13ème proposition, p. 83.

[39] *Ibid.* II° partie, 13ème proposition, lemme III, axiome I, p. 87.

[40] *Ibid.* II° partie, 13ème proposition, lemme III, axiome II définition, p. 88.

composés d'autre corps qui forment des individus sans cesse soumis à des processus de modification, de remplacement ou de substitution par élimination ou adjonction d'autres corps, sans qu'ils perdent leur identité. Cette composition ou « individuation » des corps peut se faire à l'infini, des plus simples aux plus complexes. Le corps humain ne présente pas, à cet égard, de différence de nature avec les autres, étant observé qu'il s'agit d'un composé, d'un individu dont Spinoza souligne les particularités dans les six postulats énoncés à la fin de la 13ème Proposition de la 2ème partie de l'*Ethique* qui constituent ce que d'aucuns appellent sa « petite physique », son « abrégé de physique », sa « théorie de l'individu » ou encore son « traité des corps ». Spinoza souligne notamment que : *« Le corps humain est composé d'un très grand nombre d'individus (de diverse nature) dont chacun est très composé »* ; que *« ... lui-même est affecté, d'un très grand nombre de manières par les corps extérieurs »* ; que *« Le corps humain a besoin, pour se conserver d'un très grand nombre d'autres corps par lesquels il est continuellement comme régénéré » ;* qu'il *« peut mouvoir d'un très grand nombre de manières et disposer en un très grand nombre de manières les corps extérieurs »*[41].

3° Ce qui identifie l'esprit humain (« mens humana ») – plutôt que l'âme -, selon le monisme ontologique de Spinoza pour qui la substance est identifiée à Dieu ou la Nature, substance unique et infinie ; c'est son appartenance à l'attribut pensée. Plus précisément ce qui spécifie l'esprit, c'est d'être un mode de l'attribut pensée qu'on peut aussi considérer comme l'idée du mode de l'attribut étendue, le corps, son objet. Ainsi : *« L'objet de l'idée constituant l'Âme humaine est le Corps ... »*[42]. Autrement dit, si le corps (l'objet) est le mode de l'attribut

[41] *Ibid.* II° partie, 13ème proposition, postulats I, III, IV et VI p. 91.
[42] *Ibid.* 13ème proposition, p. 83.

étendue existant en acte, l'idée (de l'objet) – l'esprit, la conscience - est le mode de l'attribut pensée. C'est cette définition de l'esprit humain comme idée du corps qui fonde la thèse de l'identité de l'homme puisque le corps, mode de l'étendue, et l'idée de ce mode, elle-même mode de la pensée, sont une seule et même chose. Selon Henri Atlan « la *conception de la nature de l'esprit comme « idée du corps » est l'une des parties les plus originales de* la *philosophie »* de Spinoza[43].

Naturellement, le corps humain étant complexe, l'idée du corps est elle-même complexe. En outre, l'esprit humain perçoit les affections du corps, les évènements qui le modifient. Plus encore, *« les idées des corps extérieurs que nous avons indiquent plutôt l'état de notre propre Corps que la nature des corps extérieurs »*[44]. L'esprit humain ne se connait lui-même qu'en tant qu'il perçoit les idées des affections du corps. Il en résulte que nul ne peut comprendre l'esprit humain de manière adéquate s'il ne connait d'abord de manière adéquate la nature du corps. En effet, *« L'Ame humaine ne connait le Corps humain lui-même et ne sait qu'il existe que par les idées des affections dont le Corps est affecté »* [45]. Or, cette connaissance n'est pas claire et distincte mais mutilée et confuse, parce que l'esprit n'a pas une connaissance adéquate immédiate des parties qui composent son corps, ni des corps extérieurs qui affectent son corps. *« L'idée d'une affection quelconque du corps humain n'enveloppe pas la connaissance adéquate du Corps humain lui-même »*[46].

Enfin, soulignons que si l'esprit humain est l'idée du corps et des affections du corps, il perçoit encore l'idée de ses idées. On pourrait se demander s'il n'y a pas là l'esquisse d'un modèle spinozien de rapports esprit-corps, de type

---

[43] *Cours de philosophie biologique et cognitiviste*, Paris, Odile Jacob, 2018 p. 121.

[44] *Ethique*, II° partie, 16ème proposition, corollaire II, p. 93.

[45] *Ibid.* II° partie, 19ème proposition, p. 97.

[46] *Ibid.* II° partie, 27ème proposition, p. 103.

« idée – objet » avec *« l'aspect « double » de la pensée, car une seule et même idée est dite avoir deux rapports différents, quoique symétrique, entre deux modes appartenant à des attributs différents : l'idée d'une affection d'un corps enveloppe ... la nature de ce corps et, parce que l'esprit est l'idée de ce corps, cette idée ... convient avec la nature de l'esprit. Et c'est seulement à cause de ce rapport que l'idée de cette idée ou encore, l'idée qui porte sur cette idée de l'affection, enveloppe la connaissance de l'esprit ... »*[47].

4° Ainsi, si le corps et l'esprit constituent une seule et même réalité modale, une seule et même chose singulière, chacun relève d'un attribut différent, l'étendue pour l'un, la pensée pour l'autre. Autrement dit, ni les attributs, ni les modes ne peuvent se confondre. *« Les modes de chaque attribut ont pour cause Dieu en tant seulement qu'il est considéré sous l'attribut dont ils sont des modes et non en tant qu'il est considéré sous un autre attribut »*[48]. On notera qu'il existe une certaine correspondance entre les deux attributs, selon les termes de la 7ème Proposition, de la 2ème Partie de l'*Ethique* : *« L'ordre et la connexion des idées sont les mêmes que l'ordre et la connexion des choses »*[49] précisés dans le scolie comme suit : *« ... aussi longtemps que les choses sont considérées comme les modes du penser nous devons expliquer l'ordre de la Nature entière ... par le seul attribut de la Pensée ; et en tant qu'elles sont considérées comme des modes de l'Etendue, l'ordre de la Nature entière doit être expliqué aussi par le seul attribut de l'Etendue ... »*[50]. Il s'ensuit que

---

[47] Lia Levy, « Les rapports entre l'esprit et le corps dans la proposition 23 II° partie de l'Ethique » in *La théorie spinoziste des rapports corps/esprit* sous la dir. de Chantal Jacquet, Pascal Severac et Ariel Suhamy, Paris, Hermann éd, coll « Philo », 2009, p. 43-44.

[48] *Ibid.* II° partie, 6ème proposition, p. 74.

[49] *Ibid.* II° partie, 7ème proposition, p. 75.

[50] *Ibid.* II° Partie, 7ème proposition, scolie p. 76.

le corps, mode de l'étendue, ne peut être déterminé par l'esprit, mode de la pensée, pas plus que l'esprit ne peut être déterminé par le corps. Corps et esprit ne peuvent agir l'un sur l'autre, l'un contre l'autre. Il n'y a donc pas d'interactions, de rapports de causalité entre les deux, l'âme n'agit pas sur le corps, ni le corps sur l'âme. C'est ce qu'énonce cette double affirmation à la 2ème Proposition de la 3ème Partie de l'*Ethique* : *« Ni le corps ne peut déterminer l'Ame à penser, ni l'Ame le Corps au mouvement ou au repos où à quelque autre manière d'être que ce soit (s'il en est une autre) »*[51]. Cette élimination, cette négation des rapports de causalité psychophysique entre les deux modes, constitue le point central de la philosophie de Spinoza.

Mais, l'esprit et le corps, chacun dans son ordre propre, en tant que mode de la pensée ou de l'étendue, agit par lui-même sur lui-même, sans l'aide de l'autre. Il en résulte deux séries de conséquences. La première est relative aux modalités de la connaissance que l'on peut avoir des deux modes : celle-ci n'est accessible, a) qu'au sein de chaque ordre, sous chaque attribut, b) qu'en fonction du principe de causalité, d'enchainement causal, selon les lois de son ordre propre, de sa propre nature, qui sont les lois de la pensée et les lois physiques. Dès lors, s'il n'y a pas de causalité d'états cérébraux sur des états mentaux (et réciproquement), il y a une causalité intracérébrale, d'états cérébraux sur d'autres états cérébraux, et une causalité intracognitive, des activités cognitives les unes sur les autres. La seconde conséquence concerne la conception spinozienne de l'attribut pensée, de l'activité de l'esprit, de l'activité mentale ou cogitative. Les modes de l'attribut pensée, les idées, ne peuvent être causés que par d'autres modes de ce même attribut, d'où le fait que les idées ne sont pas de nature « représentative » des modes de l'attribut étendue. Cette conception implique la critique

[51] *Ibid.* III° partie, 2ème proposition, p. 136.

voire le rejet de celle cartésienne de l'image, de la représentation des choses, du modèle représentationnel de l'esprit, assimilé à une sorte de « théâtre intérieur ». L'idée, pour Spinoza, n'est pas de nature représentative, au sens cartésien d'une image interne à l'esprit qui représenterait son objet. Cette réfutation spinoziste de l'idée-image se comprend dans le cadre de la critique de l'hypothèse cartésienne du libre arbitre et de la volonté libre. En effet, pour Spinoza, le libre arbitre n'est qu'une illusion, l'âme étant soumise aux déterminations et aux nécessités naturelles. *« Il n'y a dans l'âme aucune volonté absolue ou libre ; mais l'âme est déterminée à vouloir ceci ou cela par une cause qui est aussi déterminée par une autre, et cette autre l'est à son tour par une autre, et ainsi à l'infini »*[52]. Ainsi, l'esprit relèverait d'un ordre abstrait, causal, mathématique car il agirait sur, enchainerait les idées selon des lois précises qui sont les lois de la pensée, de « l'entendement », de l'activité rationnelle, d'où l'idée d'un déterminisme cogitatif dans la mesure où la causalité mentale, idéelle est complète, distincte de tout autre type notamment de la causalité cérébrale, physique. On peut voir là, l'explication de l'affirmation de Spinoza, au § 46 du *Traité de la Réforme de l'Entendement*, ouvrage antérieur à l'*Ethique* selon laquelle il aurait été le premier à concevoir *« l'âme agissant selon des lois déterminées et telle un automate spirituel »* (« *automa spirituale* »)[53]. A l'encontre de cette explication, trois objections peuvent être opposées : d'abord, l'esprit, agissant dans l'ignorance des causes qui le déterminent, peut être aussi activé par le hasard ; ensuite, la 3ème partie de l'*Ethique* est consacrée à l'étude des affects (sentiments, passions) qui sont produits par l'esprit ; enfin on ne retrouve aucune occurrence de l'expression « automate spirituel » dans l'*Ethique*. En

[52] *Ibid.* II° partie, 48ème proposition, p. 123.

[53] *Traité de la Réforme de l'Entendement*, § 46, trad. par Charles Appuhn, Paris, Flammarion, coll « GF », 1964, p. 210.

revanche l'idée sera reprise, modifiée et développée par Leibniz. Quoi qu'il en soit, Spinoza rejette bien la théorie représentationnelle de l'esprit et engage la critique de la connaissance par le sujet cartésien de ses propres actes, produits de sa volonté absolue, de son libre arbitre.

Il importe de souligner que la pensée de Spinoza présente un fort caractère « contre-intuitif », en décalage avec la position du « sens commun », de l'opinion générale - ce qui peut la rendre difficilement accessible -, car elle fait tenir ensemble deux hypothèses radicalement différentes : l'unicité de la substance et l'infinité de ses attributs, dont deux seulement, la pensée et l'étendue, sont accessibles à la connaissance humaine.

Aussi, comme pour Descartes, la philosophie de Spinoza, touchant au problème du corps et de l'esprit, ouvre droit à deux lectures selon que l'on insiste, soit sur son monisme ontologique, sa théorie de l'identité au sein d'une substance unique, soit sur la distinction entre les deux attributs, pensée et étendue. Selon la première approche, il n'y aurait aucun sens à distinguer entre les déterminations idéelles ou psychiques et les déterminations matérielles ou physiques puisqu'elles agiraient simultanément au sein de la seule et même substance, ce qui rendrait impossible toute interaction entre le corps et l'esprit. Dans la seconde lecture, l'existence primordiale de la substance unique et infinie, ne pourrait interdire la distinction entre les deux attributs, qui sont aussi deux ordres distincts d'enchainement causal fonctionnant simultanément, mais en parallèle, ce que d'aucuns appellent le « parallélisme » de Spinoza. Il importe d'observer que cette approche, qui conduit à « cartésianiser » Spinoza, est ontologiquement erronée, puisque ce parallélisme n'implique aucune forme de dualisme ontologique mais qu'il est plutôt de nature épistémologique. En tout cas, la philosophie de Spinoza reste, elle aussi, d'actualité, voire, occupe une place de plus en plus importante dans la mesure où elle sert de point d'appui métaphysique aux neurosciences.

Chapitre 3 : Descartes, Spinoza et les neurosciences

Après avoir rappelé les différences radicales qui les opposent l'un à l'autre, on ne manquera pas de souligner que leurs philosophies présentent nombre de points communs, avant d'évoquer leurs places respectives dans les neurosciences.

§ 1 : De Descartes à Spinoza, des divergences aux convergences

1° Leurs divergences sont bien réelles et manifestes.

a) Sans doute la plus importante tient dans la « source », le point de départ, le fondement de leurs philosophies qui s'avère, au moins en apparence, en totale contradiction : dualité de substances d'un côté, unicité de l'autre, dualisme contre monisme.

Il faudrait nuancer cet antagonisme car leurs conceptions sont « contenues », limitées : le dualisme cartésien, par l'union substantielle du corps et de l'esprit et le monisme spinozien, par la considération des seuls attributs, étendue et pensée.

b) Mais de cette divergence fondamentale résultent d'autres différences significatives en ce qui concerne la conception des rapports entre l'âme, l'esprit et le corps. Si, pour Descartes, l'esprit humain est plus aisé à connaitre que le corps, pour Spinoza, à l'inverse, seule la découverte du corps permet de comprendre ce qu'est l'esprit. De même, si Descartes admet la réalité des interactions, des rapports de causalité entre l'esprit et le corps, Spinoza rejette cette hypothèse.

c) Surtout, Descartes postule l'existence du libre arbitre du sujet tandis que Spinoza soutient que l'individu humain est soumis à la nécessité en raison de son ignorance des causes des choses, ignorance de laquelle il peut toutefois se libérer en accédant à la connaissance de celles-ci.

Enfin Descartes énonce la théorie représentationnelle de l'esprit que Spinoza réfute.

En marge du problème corps-esprit, on ajoutera que leurs positions sont également contraires, sur la question des rapports de l'homme à la nature : pour Descartes, l'homme peut prétendre se *« rendre comme maitre et possesseur de la nature »*[54] ; alors que pour Spinoza, c'est une erreur de concevoir *« l'homme dans la Nature comme un empire dans un empire »*[55].

2° Mais leurs convergences. sont également fortes et nombreuses :

a) Pour l'un et l'autre, dans la continuité de la pensée du Moyen-Âge, la philosophie trouve sa source dans la métaphysique, l'ontologie notamment, la définition de l'Être. A cet égard, le concept de base de leurs réflexions respectives est bien la Substance.

b) Mais, l'essentiel tient à ce que ces deux philosophes s'inscrivent, clairement et pleinement, dans le courant intellectuel dominant à partir du XVII° siècle : le rationalisme. A ce titre, en rupture avec l'école scolastique, Descartes d'abord (après Francis Bacon) et Spinoza ensuite, rejettent la croyance dans les causes finales, le finalisme, c'est-à-dire ce en vue de quoi les choses sont faites. Le premier affirme : *« ... nous rejetterons entièrement de notre philosophie la recherche des causes finales ... »*[56] ; tandis que le second précise : *« La Nature n'agit pas pour une fin »*[57]. C'est cette position philosophique qui va ouvrir la porte au développement de la connaissance objective et scientifique de la Nature.

c) On peut aussi relever une certaine proximité

---

[54] *Discours de la Méthode*, 6ème partie, 2ème §, (AT VI, 5) Paris, Editions Garnier Frères, 1960, p. 102.

[55] *Ethique*, III° partie, préface, Paris, Flammarion, coll « GF », 1965, p. 133.

[56] *Les principes de la philosophie*, 1ère partie, § 28, Paris, Libraire philosophique J. Vrin, 1989, p. 72.

[57] *Ethique*, V° partie, préface, Paris, Flammarion, coll « GF », 1965, p. 218.

entre la conception cartésienne de l'union substantielle de l'âme et du corps et celle spinozienne de l'identité de l'homme avec ses deux faces. Ces deux approches ont le même « arrière-plan » : une vue mécaniste, physique notamment du corps humain qui s'inscrit, là encore, dans l'émergence des sciences de la Nature à partir de l'Âge classique.

Il résulte de ces convergences que, si Spinoza se pose effectivement en critique de Descartes, la question de savoir s'il ne serait pas dans le prolongement de ce dernier, plutôt qu'en rupture avec lui, est bien posée. Ceci a conduit certains chercheurs, tel Denis Kambouchner, à relever qu'il y a une *« solidarité historique des* (deux) *philosophies … et … les solidarités sont … importantes … il* (faut) *plutôt parler d'une transformation spinoziste des vues de Descartes »*[58].

§ 2 : Les places respectives de Descartes et Spinoza dans les neurosciences

1° A première vue, les philosophies de Descartes et de Spinoza sont inopérantes dans la solution du problème corps-esprit, et, en particulier dans ce qui en est une question essentielle, à savoir la réalité et la nature (causalité psychophysique ou non) des interactions entre le cérébral et le mental. Descartes n'apporte pas d'explication suffisante et fiable, Spinoza ignore plutôt la question en postulant l'absence d'interactions.

Dès lors, pourquoi ces deux conceptions restent-elles, comme on l'a dit, co-présentes dans les neurosciences ?

On peut se demander si leur véritable intérêt philosophique n'est pas, au-delà de leurs divergences métaphysiques et ontologiques, dans leurs modes d'approche du problème corps-esprit sur le terrain de l'épistémique et de la

[58] Denis Kambouchner, « L'erreur de Damasio », in *La théorie spinoziste des rapports corps/esprit et ses usages actuels*, sous la dir. de Jacquet Chantal, Séverac Pascal et Suhamy Ariel, Paris, Hermann éd., coll « Philo », 2009, p. 202.

méthodologie.
Autrement dit, le cœur dudit problème ne serait pas la question de la substance mais plutôt celle de la voie à mettre en œuvre pour tenter de connaître la « chose ».
Force est de constater que le fait que le cerveau (ses structures, son activité) soit devenu un objet de science, l'objet des neurosciences, a considérablement « banalisé », « normalisé », « objectivé » son approche. Ce point est encore renforcé par le fait que les neurosciences nous apprennent que la majeure partie de notre activité mentale est inconsciente. *« Même des opérations complexes qui relient la perception à l'action, se déroulent sans conscience : nous fonctionnons souvent en « pilotage automatique ». ... nous ne cessons de surestimer notre pouvoir de décision consciente – alors qu'en vérité notre degré de contrôle est sévèrement limité »*[59]. Ceci pose la question, du point de vue philosophique, de l'existence du libre arbitre, de la place de la liberté et de la volonté, laquelle pourrait être reformulée dans ces termes : le libre-arbitre ne serait-il pas finalement un « obstacle épistémologique » au développement des neurosciences ?
Face à cette donnée liée à l'évolution du contexte du problème corps-esprit, les neuroscientifiques ont développé deux types de positions sur le plan philosophique : la première tendant à la critique sinon au rejet du dualisme cartésien, la seconde considérant que le monisme spinozien peut servir de référentiel aux neurosciences.
2° Au sujet de la critique du dualisme cartésien et du libre-arbitre, on peut évoquer deux lectures :

a) La première est celle du neurologue américain Antonio Damasio (né en 1944) qui, dans son livre célèbre *L'erreur de Descartes*, affirme que celui-ci : *« ... a instauré une séparation catégorique entre le corps ... et l'esprit ... Et spécifiquement, il a posé que les opérations de l'esprit*

[59] Stanislas Dehaene, *Le Code de la conscience*, Paris, Odile Jacob, coll « Sciences », 2014, p. 75.

les plus délicates n'avaient rien à voir avec l'organisation et le fonctionnement d'un organisme biologique »[60]. Cette interprétation, qui a connu un fort retentissement éditorial et médiatique est manifestement erronée, si l'on considère que le dualisme cartésien est limité par la théorie de l'union substantielle du corps et de l'esprit.

b) La seconde est celle du neuropsychologue français Stanislas Dehaene (né en 1965) qui dans son récent ouvrage *« Le Code de la conscience »* soutient que Descartes doit être considéré, si l'on se réfère à son texte intitulé *L'Homme*, écrit autour de 1632-1633 et publié, après sa mort, en 1664, comme *« ... un scientifique visionnaire, et ... un réductionniste dont les analyses mécanicistes de l'esprit humain, très en avance sur leur temps, peuvent être lues comme le texte fondateur de la biologie synthétique et de la modélisation des réseaux neuronaux »*[61]. En effet, son *« insistance ... sur l'immatérialité de l'âme était, peut-être au moins en partie, une façade destinée à protéger sa vie »* à le mettre à l'abri d'un conflit avec l'Eglise[62] ; (thèse également évoquée par Damasio).

Dans les deux cas, c'est bien l'esprit, la conscience qui est mis en cause et avec lui le libre-arbitre, tandis que le corps est porté au premier plan.

Il s'ensuit une conséquence : si l'on admet qu'entre le corps et l'esprit il existe des relations causales, le développement des connaissances du corps, du cerveau finira par expliquer la conscience.

En attendant, celle-ci est réduite soit à sa dimension d'activité cérébrale, c'est la position des tenants du matérialisme éliminativiste ou de la neurophilosophie, tels

---

[60] trad. de l'anglais (USA) par Marcel Blanc, Paris, Odile Jacob, coll « Sciences », 1995, p. 312.On remarquera que 5 pages seulement (sur 349) sont consacrées à l'exposé de « l'erreur de Descartes) alors que celle-ci constitue l'objet du livre …

[61] Stanislas Dehaene, *Le Code de la conscience*, Paris, Odile Jacob, 2014, p. 18.

[62] Paris, Odile Jacob, 2014, p. 363, note 5.

Paul Churchland et Patricia Smith Churchland, soit à l'état d'illusion ou d'épiphénomène, c'est la position de Daniel C. Dennett.

3° Le monisme spinozien, comme référentiel des neurosciences

Le médecin, biologiste et philosophe français Henri Atlan (né en 1931) dans son livre *Cours de philosophie biologique et cognitive* opère un retour à la philosophie de Spinoza, en particulier à deux propositions de l'*Ethique* qu'il appelle « fondatrices », évoquées plus haut : la 7ème proposition de la 2ème Partie et surtout la 2ème proposition de la 3ème partie avec son scolie. Il considère en effet que l'identité de l'homme supprime toute relation de causalité entre le corps et l'esprit, entre le cérébral et le mental, dès lors qu'il s'agit de deux expressions ou deux modes d'attributs différents mais relevant d'une seule et même réalité. Pour Atlan, le monisme ontologique spinozien, ni dualiste, ni matérialiste, ni idéaliste, ou plus exactement à la fois matérialiste et idéaliste dans sa méthode, fournit un cadre interprétatif, qui permet de résoudre le problème corps-esprit en considérant qu'il s'agit d'un faux problème ou d'un problème mal posé à partir de préjugés métaphysiques inadéquats. Selon lui, seule la philosophie de Spinoza offre un cadre opérant aux travaux actuels sur les rapports entre le corps et l'esprit qui, grâce aux techniques d'exploration du cerveau, permettent de mettre au jour et d'analyser des corrélats cérébraux de la conscience. La thèse d'Atlan *« est que le programme de naturalisation de l'esprit* (comme expression) *du monisme spinozien* (n'impose pas) *le matérialisme éliminativiste, ni au plan ontologique, ni au plan anthropologique, pas plus que les solutions idéaliste et dualiste »*[63].

[63] *Cours de philosophie biologique et cognitive*, Paris, Odile Jacob, 2018, p 468.

Néanmoins, la solution spinozienne est contre-intuitive, on l'a dit, et aussi radicale en ce qu'elle conduit à admettre que le *« libre arbitre est réduit à la portion congrue »*[64].
Malgré cela, elle offre l'avantage de lever l'obstacle épistémologique. En sens inverse, elle se heurte inévitablement à la volonté de ceux qui entendent préserver voire réhabiliter le libre arbitre.
En outre, le fait que le corps et l'esprit constituent deux modes de deux attributs relevant d'une seule et unique substance, permet, à tort, à d'aucuns, de réintroduire le dualisme, comme l'atteste la théorie du double aspect de Thomas Nagel ou la thèse du monisme anomal de Donald Davidson.

[64] *Ibid.* p 247.

IIème partie
Le renouvellement de la problématique de la nature de la conscience comme objet de science au XX° siècle

Introduction

1° On peut considérer que les courants philosophiques ou plus largement culturels qui ont accompagné le développement de la science du cerveau et de la conscience peuvent être distingués selon qu'ils relèvent de l'une ou de l'autre de deux approches radicalement différentes. La première s'inscrit dans le sillage d'un large mouvement philosophique qui a conduit à la naissance de la philosophie dite « analytique » (largement de source anglo-saxonne) ; elle emprunte donc une voie plus philosophique que scientifique. La seconde se rattache au développement des sciences du vivant, de la biologie et de la neurologie ; elle est clairement d'orientation (neuro)scientifique. Il s'ensuit que le couple philosophie – science va se référer à deux modèles de l'esprit, l'un plutôt « cognitiviste » où l'esprit est défini comme une machine abstraite et l'autre plutôt « neurologiste » dans lequel la conscience est caractérisée par les termes d'« états internes, psychiques ou mentaux ».

a) La première approche, suivie par la philosophie des sciences cognitives, se forme, on l'a dit, dans le prolongement du mouvement philosophique qui, à partir de la fin du XIX° siècle, à la suite de l'empirisme et du positivisme, va tendre à rompre avec la métaphysique et l'ontologie pour élaborer la philosophie analytique. Tout d'abord, les progrès des sciences de la nature ont relancé les recherches en mathématiques lesquelles entraînent un renouvellement de la logique. Cette évolution a eu lieu, au cours de la première moitié du XX° siècle, notamment avec les travaux du logicien et mathématicien Kurt Gödel (publiés en 1931). Parallèlement, la philosophie analytique se diversifie et donne naissance à la philosophie du langage. La convergence de ces diverses disciplines va

conduire à la formulation des théories de l'information dont celle dite « théorie mathématique de la communication » de Claude Elwood Shannon. C'est dans ce contexte qu'est née l'idée que l'esprit, la conscience se caractérise non comme telle ou telle substance, mais comme un système de traitement de l'information. Il s'ensuit que la philosophie quitte le terrain de l'ontologie, pour aborder celui du fonctionnement du système. La dimension « substantialiste » ou « essentialiste » ne disparaît pas complètement, puisqu'elle se réfère à ce qui sert de substrat, de support au système qui peut être soit naturel, soit artificiel. Mais ce qui importe le plus c'est la fonction, c'est pourquoi cette première approche a été appelée : le « fonctionnalisme ».

b) La seconde approche résulte de ce que le progrès des connaissances du cerveau, de ses différents niveaux d'organisation (neurone, réseau de neurones, assemblée de réseaux de neurones) a eu un fort impact en philosophie, dès lors qu'il est de plus en plus démontré scientifiquement que la conscience et ses manifestations (sensations, pensées, émotions) sont en lien avec le cerveau, existent en raison de leur support physique, de leur substrat neurobiologique. Autrement dit, toute recherche sur le cerveau conduit à aborder des questions touchant à la conscience, et inversement toute interrogation relative à l'esprit ne peut (pour le moins) ignorer les connaissances du cerveau.

Cette dynamique a conduit à ce que l'on appelle la « naturalisation de la conscience », et à son « programme », c'est-à-dire l'intégration des états et des opérations de la vie mentale, de notre subjectivité, appréhendés comme des phénomènes neurologiques, dans l'objet même des neurosciences. Cette approche va conduire à l'émergence au sein de la philosophie des neurosciences, d'une option en faveur d'un monisme ontologique, adossé à l'idée que la conscience doit être considérée comme un fait, un objet de la nature, d'où le concept de « naturalisme ». Au sujet de celui-ci, on peut

faire trois remarques : la première étant que d'autres types de monismes ont été théorisés, tels le matérialisme (la substance unique est la matière) voire le physicalisme (la substance unique est constituée d'éléments physiques) ; la deuxième tient à préciser que le naturalisme englobe tous les monismes, la notion d'objets naturels s'étendant au-delà des seuls objets matériels ou physiques, en considérant que *« la nature embrasse tout ce qui existe »*, selon l'expression de Daniel Andler[65]. La troisième remarque conduit à souligner que la conscience naturalisée ne peut être dissociée du monde vivant.

Cette option pour le naturalisme qui relève dans une large mesure de l'ontologie, est celle de la majeure partie de la communauté neuroscientifique. Elle a été affirmée, plus ou moins explicitement, par nombre d'acteurs de renom en cette matière. Ainsi, Gerald Edelman : *« … Toutes les théories de la conscience s'accordent aujourd'hui à dire qu'elles doivent se fonder sur les mécanismes physiques et chimiques à l'œuvre dans le cerveau. … Tout le monde est à présent d'accord pour dire que la sélection des groupes neuronaux est un mécanisme essentiel de la conscience »*[66] Aussi, Jean-Pierre Changeux : *« Les spécialistes de neurosciences considèrent la conscience comme une propriété ou une fonction du cerveau …* (qui) *devrait se comprendre en termes d'activité des neurones, des synapses et de leur régulation par des signaux chimiques »*[67].

c) A ce stade, la question est de savoir qu'est-ce qui caractérise, spécifie le plus la conscience. Celle-ci est-elle « naturelle » en totalité ou seulement en partie ? Autrement dit, la conscience peut-elle être réduite, en totalité ou seulement en partie, à l'activité du cerveau ? C'est la

---

[65] Daniel Andler, *La silhouette de l'humain*, Paris, Gallimard, coll « *nrf* essais », 2015, p 11.

[66] Gerald Edelman M., Entretien, in Les Dossiers de la Recherche, n° 30, 02/2008, p 10.

[67] Jean-Pierre Changeux, *L'Homme de vérité*, Paris, Odile Jacob, 2002, p 114.

question dite du réductionnisme, à propos duquel une observation doit être faite. Il s'agit ici d'un réductionnisme plutôt ontologique ou « fort » ou « de fond », où ce qui est en jeu c'est la question de l'essence de la conscience : est-elle plutôt une fonction, ou une substance (matérielle, physique, biologique) ? Ce réductionnisme ne doit pas être confondu avec le réductionnisme épistémologique, « faible » ou « de méthode » auquel recourt la science pour diviser son objet et réduire ses champs de recherche. De ce point de vue, on peut considérer que le fonctionnalisme et le naturalisme conduisent, l'un et l'autre, à une identification conceptuelle du cerveau et de la conscience. En effet, par la réduction de la conscience, le premier à un système de traitement de l'information dont le cerveau est le support, le second à un fait de la nature indissociable du cerveau, on peut concevoir un monisme épistémologique, le cerveau et la conscience constituant un seul et même objet de science.

2° Dans les années 1990, la confrontation entre le naturalisme et le fonctionnalisme a paru insuffisante, sinon inefficace, pour apporter des éléments de réponse pertinents au problème corps-esprit. Cet état de fait a conduit certains philosophes et scientifiques à concevoir une nouvelle approche du problème, appelée la « neurophénoménologie », tendant à la conjonction de la phénoménologie et, des neurosciences et des sciences cognitives.

Dans un premier temps, nous exposerons les thèses du fonctionnalisme et de ses dérivés, et du naturalisme, (Chap. 1). Dans un deuxième temps, nous passerons en revue deux théories principales du naturalisme ontologique : le naturalisme biologique et le physicalisme (Chap. 1). Enfin, dans un troisième temps, nous évoquerons la neurophénoménologie (Chap. 3).

Chap 1 : le fonctionnalisme et ses dérivés

En nous situant dans une perspective historique, nous allons tour à tour évoquer, d'abord la philosophie d'Hilary Putnam puis deux courants de la philosophie de l'esprit qui s'y réfèrent : le computationnalisme et le connexionnisme et enfin, deux champs scientifiques qui y sont liés : le cognitivisme et l'intelligence artificielle.

§ 1 : Hilary Putnam et le fonctionnalisme

Ce n'est plus la substance ou l'essence, la structure ou le contenu qui est au fondement du fonctionnalisme, mais le rôle, la fonction des éléments du système. Avec cette théorie, le débat philosophique sort de l'ontologie.

1° Le fonctionnalisme, concernant le problème corps-esprit, a été formulé par le philosophe américain Hilary Putnam (1926 - 2016) notamment dans son article « la nature des états mentaux » (« The Nature of mental states ») publié en 1967.

2° Deux idées en sont à la base :

Une même fonction (le traitement d'informations ou de données) peut être effectuée, selon un mode d'opération préprogrammé, quel que soit le dispositif, le substrat qui exécute, réalise ladite opération. Il y a déconnection, indépendance totale entre le « programme » et son « exécutant ». Le support est donc totalement indifférent : organisme vivant ou machine peu importe dès lors que le système se charge de la même fonction. Un tel schéma appliqué au problème corps-esprit, implique que l'esprit ne consiste pas (ou plus) en un certain substrat, en une certaine constitution, matérielle (physique, mécanique, électronique ou biologique) ou immatérielle. Dès lors, la question de la substance ne se pose plus.

3° La seconde idée se réfère à l'article « Théorie des nombres calculables » (« On Computable Numbers ») publié en 1937, dans lequel le mathématicien Alan Turing décrit son modèle de « machine à calculer » ou « machine computationnelle » (qu'on appellera la « machine de

Turing ») permettant de calculer toute fonction mathématique, c'est-à-dire l'ordinateur. Il s'agit d'un système qui correspond à une succession d'états reliés entre eux par une relation de causalité, selon un ordre programmé, un ensemble de règles préétablies, constituant un algorithme. Dans cette séquence causale les états sont tour à tour des états déterminés, « entrants » (« inputs ») puis des états déterminants, « sorties » (« outputs »). C'est ce schéma fonctionnaliste qui est à la base de l'informatique. Ce même schéma peut être étendu, par analogie, au problème corps-esprit. C'est la métaphore de l'esprit conçu comme un simple ordinateur biologique. Ainsi Putnam considère que l'esprit exécute des tâches computationnelles et est fonctionnellement identifiable à une « machine à calculer de Turing » et à la logique de calculabilité qu'elle met en œuvre. Les états mentaux (croyances, désirs, sensations etc.) sont définis selon leur rôle, fonction et leurs relations de cause à effet, par rapport aux autres états mentaux, les perceptions et les comportements. L'esprit consiste en une disposition à suivre des règles, à exécuter des tâches impliquant des procédures prédéterminées. On retrouve là l'idée d'un esprit-automate logique (version contemporaine de l'automate spirituel formulée par Spinoza et développée par Leibniz) qui admet des états entrants (les *« inputs sensoriels »* et des états sortants (les *« outputs moteurs »*). Chaque état mental est défini par son origine causée et son rôle causal. Ce schéma repose sur le postulat d'un déterminisme ou plutôt d'un « probabilisme » logique qui apparaît comme le corrélat de l'indépendance de l'esprit-automate, au moins « probabiliste », à l'égard de toute constitution particulière. L'esprit *« est défini de la même manière que* (la) *machine de Turing, sauf que les « états » peuvent avoir différentes probabilités au lieu d'être « déterministes »*[68]. Le système *« spécifie pour chaque*

[68] Hilary Putnam « The Nature of mental States », publié en 1967, « La nature des états mentaux », traduit de l'anglais (USA) par

combinaison possible d'un « état ... une « instruction » qui détermine la probabilité de l'« état » suivant »[69].

4° On peut observer que le fonctionnalisme implique la réduction des états mentaux, non à des états « substantiels » mais à des états « fonctionnels ». *« ... la douleur n'est pas un état cérébral au sens d'état physico-chimique du cerveau (ou même de l'ensemble du système nerveux) mais un état d'une toute autre espèce. Je propose l'hypothèse que la douleur, ou l'état d'éprouver de la douleur, est un état fonctionnel d'un organisme tout entier »*[70]

5° Avec le fonctionnalisme, on le voit, la question du réductionnisme ontologique - la réduction ou non des états mentaux à des états physiques ou biologiques - est moins prégnante. Certains fonctionnalistes ne sont pas réductionnistes, tel Putnam, d'autres le sont, comme le philosophe australien David Lewis (1941 – 2001).

Notons enfin que Putnam finira par reconnaitre *« le caractère utopique du fonctionnalisme »* dans un article « Aller et retour au pays du fonctionnalisme » paru en 2001[71].

Pour autant, le fonctionnalisme a connu un développement important dans le cadre de deux courants successifs : le premier, dit « computationnalisme », se référant directement au modèle de l'ordinateur, base des théories de l'information et de la cybernétique ; le second appelé « connexionnisme » s'inspirant du modèle des réseaux de neurones « formels ».

§ 2 : Du computationnalisme au connexionnisme

Jean-Michel Roy, revue et corrigée par Dominique Boucher, in *Philosophie de l'esprit*, T1, Paris, Librairie philosophique J. Vrin, coll. « textes clés », 2002, p 276.

[69] *Ibid.* p 277.

[70] *Ibid.* p 276.

[71] traduit de l'américain par Mireille Duchastelle-Cabanes, in *Philosophie*, n° 71 2001/4 p. 80-90.

1° Le computationnalisme a été initié par le philosophe américain Jerry Fodor (1935 - 2017) qui a élaboré une théorie « symbolique ou computo-représentationnelle de l'esprit », sur laquelle se sont développées, à partir des années 1970, les sciences cognitives et le mouvement intellectuel y correspondant, appelé le « cognitivisme ».
La théorie de Fodor repose sur deux piliers, combinant le langage et le modèle de l'ordinateur.

a) Dans son livre *Le Langage de la pensée* (*« The Language of Thougt »*) publié en 1975, Fodor affirme que les productions mentales (pensées, intentions, désirs) sont traitées sous forme de propositions (représentations ou symboles) à caractère sémantique, qui possèdent une structure interne essentiellement linguistique notamment une structure syntaxique prédicative, plus ou moins innée, selon l'analyse du linguiste américain Noam Chomsky (né en 1928). Ces représentations sont reliées entre elles par des relations qui jouent un rôle causal en vertu de leur forme logique (si P alors Q).

b) Comme Putnam, Fodor admet que l'esprit humain fonctionne comme une « machine à calculer » de Turing. Il ajoute que l'architecture de son système est modulaire. Dans son ouvrage *La modularité de l'esprit* (*« The Modularity of Mind »*) paru en 1982, il explicite son hypothèse « modulariste », en la restreignant aux processus « inférieurs » (perceptifs et moteurs) et au langage. Le cerveau est composé de modules spécialisés, adossés à des aires cérébrales et dédiés à une fonction spécifique qui s'active de manière autonome et inconsciente. Le cerveau coordonne les modules entre eux par un « système central » qui, lui, est non spécialisé et conscient.
A noter qu'en 2001, Fodor publiera un livre intitulé *L'Esprit, ça ne marche pas comme ça !* (*« The Mind doesn't work that way »*) dans lequel il remet en cause sa propre théorie.

2° Le connexionnisme est une autre approche du fonctionnalisme qui, après la découverte des réseaux de neurones cérébraux, va se développer dans le sillage des avancées de la science du vivant et de la biologie et de leurs concepts spécifiques, notamment les théories de l'auto-organisation et de l'émergence.

Comme dans le computationnalisme où l'esprit était issu d'un système fonctionnant selon le modèle de l'ordinateur, dans le connexionnisme, l'esprit, la pensée émerge d'un système conçu sur le modèle des réseaux de neurones du cerveau. En effet, les neurones naturels sont massivement connectés entre eux, fonctionnent en parallèle et génèrent un comportement global qui leur permet de repérer et de coder des « formes », c'est-à-dire des objets ou des catégories d'objets (lettres, visages ...) et leurs modifications.

C'est dans les années 1940-1950 que les fondements théoriques du connexionnisme sont établis. Une première étape est franchie par le neuropsychiatre Warren S. Mac Culloch (1898-1969) et le logicien Walter Pitts (1923-1969), tous deux américains, qui exposent le fonctionnement du système nerveux à partir du comportement des neurones, dont ils font une description simplifiée dans leur article « Un calcul logique des idées immanentes dans l'activité nerveuse » (*A Logical Calculus of Ideas Immanent in Nervous Activity*) publié en 1943.

Ces travaux seront développés notamment par le neuropsychologue canadien Donald Hebb (1904-1985) qui découvre les assemblées de réseaux de neurones en 1949.

Enfin, en 1982, le physicien américain John Joseph Hopfield (né en 1933) établit le modèle (qui porte son nom) des réseaux de neurones dits « formels » ou « récurrents ». Il s'agit de réseaux de neurones artificiels très simplifiés, qui réalisent des systèmes dynamiques plus ou moins complexes, évoluant dans et avec le temps.

C'est sur ces bases que s'élabore le connexionnisme, dans le prolongement de la biologie et de la théorie des réseaux

de neurones artificiels et en parallèle à la théorie computationnelle de l'esprit ou le computationnalisme.
A noter, d'une part, que cette approche connexionniste est sensiblement différente du computationnalisme car elle fait une grande place aux notions de flux et de dynamique dans les systèmes ; d'autre part, qu'elle a fortement contribué à relancer le secteur de l'intelligence artificielle à partir des années 1990.

§ 3 : Le cognitivisme et l'intelligence artificielle (IA)

1° La source de l'IA réside dans une composante fondamentale du modèle classique de la science dès ses débuts aux XVI° et XVII° siècles : le modèle de la mathématisation de la connaissance qui intègre l'abstraction ou la formalisation, c'est-à-dire la représentation des choses au moyen de symboles, et qui finira par s'imposer car sa « matière », les mathématiques, constituent le seul langage commun. L'astronome Galilée (1564-1642) l'applique à la physique : « *L'univers est écrit en langue mathématique ...* » dit-il dans *L'essayeur* (*Il saggiatiore*) en 1623. Le philosophe anglais Thomas Hobbes (1588-1679) - pour qui, contrairement à Descartes, le corps et l'esprit ne sont pas deux substances séparées - reprend l'idée lorsqu'il écrit *« Penser c'est calculer »*. Gottfried Wilhelm Leibniz entreprend à la fin du XVII° siècle, de rapprocher les mathématiques et la logique, un projet qui sera développé au XIX° siècle puis au début du XX° avec le Positivisme et le Cercle de Vienne.
2° Dans ce contexte, le logicien et mathématicien Kurt Gödel (1906-1978) dans la 1ère moitié du XX° siècle, (ses travaux sont publiés en 1931) montre que la logique peut être arithmétisée. De son côté, dans son fameux article « Théorie des nombres calculables » (« On Computable Numbers »), évoqué plus haut, Alan Turing décrit son modèle de l'ordinateur. En 1950, il publiera un autre

célèbre article « Les ordinateurs et l'intelligence » (« Computing Machinery and Intelligence ») dans lequel il propose un test permettant d'établir si une machine possède ou non une mémoire (le fameux « test de Turing »). Les neuroscientifiques américains Warren S. Mac Culloch et Walter Pitts, dans leur article publié en 1943, affirment que le cerveau peut être assimilé à un mécanisme de type « machine de Turing ». A la même époque, de 1946 à 1953, se tiennent dix conférences (à New York puis à New Jersey pour la dernière) organisées par la fondation privée Josiah Macy Jr., entrées dans l'histoire sous le nom de conférences Macy qui réunissent des scientifiques d'horizons divers (mathématiciens, logiciens, psychologues, physiologistes, anthropologues, etc.…). Leur projet est d'étudier les mécanismes des systèmes complexes en vue d'élaborer une *« théorie entière de la commande et de la communication, aussi bien chez l'animal que dans la machine »*[72]. Les résultats de ces travaux seront exposés par le mathématicien américain Norbert Wiener (1894-1964) dans son livre intitulé *La cybernétique : Information et régulation dans le vivant et la machine (Cybernetics, or Control and Communication in the Animal and the Machine)* publié en 1948.

En 1956, les mathématiciens américains John Mac Carthy, Claude Shannon et Marvin Minsky, organisent un séminaire au Darmouth Collège (New Hampshire, Etats Unis) au cours duquel est employé pour la 1ère fois l'expression d'« intelligence artificielle ».

Le lancement de l'IA se fait dans l'enthousiasme. Mais au cours des décennies qui suivent, les résultats vont se révéler décevants ou tout au moins difficiles à obtenir si

[72] Norbert Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, Cambridge (Mass.), MIT Press, 1° éd. 1948, *La cybernétique : Information et régulation dans le vivant et la machine*, traduit de l'anglais par Ronan Le Roux, Robert Vallée et Nicole Vallée-Lévy, « Introduction », Paris, Seuil, coll. « Sources du savoir », 2014, p 70.

bien que l'IA va connaître une période entre stagnation et déclin, une sorte de traversée du désert, jusqu'à la fin des années 1990.

C'est alors que sont mises au point les théories et techniques de l'apprentissage machine ou apprentissage automatique (machine learning) *« Branche de l'intelligence artificielle, fondée sur des méthodes d'apprentissage et d'acquisition automatique de nouvelles connaissances par les ordinateurs, qui permet de les faire agir sans qu'ils aient à être explicitement programmés »*[73]. La mise en œuvre de ces nouvelles techniques va permettre le développement exponentiel d'applications dans de nombreux secteurs, d'où la relance de l'intérêt pour l'IA. En 1996, le français Yann Le Cun (né en 1960) met au point un logiciel de lecture de chèques fondé sur la technique dite du réseau de neurones « convolutif », logiciel qui sera déployé en France et dans le monde.

3° Ces avancées propres à l'IA ont lieu en même temps que celles intervenant dans d'autres domaines, ce qui est théorisé sous le nom de « la convergence des NBIC » : N pour les nanotechnologies, B pour les biotechnologies, I pour l'informatique (avec les algorithmes) et le numérique (avec les big-datas) et C pour les technologies cognitives (dont l'IA elle-même et les neurotechnologies). Au-delà de ce qui peut être un état de fait, la convergence des NBIC, c'est d'abord et surtout une théorie inventée aux USA en juin 2002 dans un rapport de la National Science Fondation et du Department Of Commerce intitulé *Nanotechnology, Biotechnology, Information technology and Cognitive science : Converging Technologies for improving Human Performance* (qui contient cette formule *« si on annonce telle chose, elle se réalisera »*).

---

[73] CNIL (Commission Nationale Informatique et Libertés), Rapport *Comment permettre à l'homme de garder la main ?*, Glossaire, 12/2017, p 75.

A noter que cette théorie constitue la rampe de lancement du post- ou trans-humanisme, idéologie elle-même liée au modèle de la croissance illimitée, comme nous le verrons dans la dernière partie.

Chap 2 : Le monisme en force sous le label du naturalisme

Le naturalisme recouvre des positions plus ou moins radicales, comme on va le voir à l'examen de deux pensées marquantes de la philosophie de l'esprit : celle de John Searle d'abord, de Jaegwon Kim, ensuite.

§ 1 : John R. Searle et le naturalisme biologique

1° Le philosophe américain John R. Searle (né en 1932) se prononce, de son côté, pour un *« naturalisme biologique »*. Pour cela, il se fonde sur la conception contemporaine, globale du monde qui s'est mise en place à partir du XVII° siècle, sur la base de deux théories scientifiques : la théorie atomique de la matière et celle de l'évolution en biologie. Il souligne que cette conception « scientifique » du monde *« que cela plaise ou non ... n'est pas une option »*[74]. Searle vise ainsi « *à réintroduire la conscience dans le domaine scientifique ... »*[75]. Selon lui, *« A la base de* (cette) *conception du monde se trouve l'idée que les êtres humains ... font partie de l'ordre biologique ... sont en continuité avec le reste de la nature. ... la conscience, est un trait biologique du cerveau humain ... Elle est causée par des processus neurobiologiques et fait autant partie de l'ordre biologique naturel que n'importe quels autres traits biologiques tels que la photosynthèse, la digestion, la mitose »*[76]. Sa conception est donc à la fois naturaliste en ce sens que la conscience, l'esprit fait partie de la nature

---

[74] *La redécouverte de l'esprit*, traduit de l'anglais (USA) par Claudine Tiercelin, Paris, Gallimard, coll « *nrf* Essais », 1995 p 134.

[75] *Ibid.* p 128.

[76] *Ibid.* p 133.

et biologique en ce sens que le mode d'explication des phénomènes mentaux relève plus du champ des sciences du vivant et notamment de la biologie, que des sciences physiques. Ce faisant, cette position le conduit à critiquer « *ce qui nous empêche de voir le caractère biologique naturel de la conscience ... notre tradition philosophique, qui fait du « mental » et du « physique », deux catégories mutuellement exclusives. Pour en sortir, il faut rejeter et le dualisme et le matérialisme ... »*[77].

2° Pour résoudre le problème des relations corps-esprit, Searle se fonde sur deux principes : a) *« Tout d'abord, les opérations mentales sont causées dans le cerveau par des processus neuronaux. Nous supposons que c'est un processus neuronal, mais c'est aux spécialistes d'en décider. ... Le premier principe est donc simple : le cerveau est la cause de l'esprit. »* b) *« Le second principe ... est que la réalité mentale causée par les phénomènes neurobiologiques n'est pas une substance séparée. La vie de l'esprit ... est plutôt l'état ou la configuration générale d'un système qui est affecté par l'ensemble des micro-évènements se produisant dans le cerveau. »* Autrement dit : *« ... la relation entre le cerveau et l'esprit est causale, mais en même temps, l'esprit n'est qu'une qualité particulière du cerveau »*[78].

3° Le premier principe de Searle, que l'on peut rapprocher des conceptions de Diderot et de Changeux, le conduit à reprendre la thèse de l'émergence de la conscience, qui a été développée à partir des années 1970 en lien avec le concept d'auto-organisation issu de travaux en biologie. Une thèse que l'on pourrait formuler ainsi : soit un système S, constitué d'éléments a, b, c. Certaines caractéristiques de S peuvent être déduites de a, b ou c. Mais d'autres caractéristiques du système ne peuvent se concevoir à partir de la seule composition des éléments ; il faut les

---

[77] *Le Mystère de la conscience*, traduit de l'américain par Claudine Tiercelin, Paris, Odile Jacob, coll « Philosophie », 1999, p 12.

[78] Entretien, in Philosophie magazine, n° 24, 11/2008 p 61-62

expliquer en termes d'interactions causales entre ceux-ci ; ce sont des *« caractéristiques du système causalement émergentes »*[79]. Or, le cerveau est un système biologique. Dès lors, la capacité d'auto-organisation de ce système et de ses éléments, constitués eux-mêmes en systèmes de plus en plus complexes (neurones, puis réseaux de neurones puis encore assemblées de réseaux de neurones) aboutit à l'émergence d'une caractéristique, d'une propriété, d'une qualité : la conscience. *« Etant donné que c'est une caractéristique qui émerge à partir de certaines activités neuronales, nous pouvons penser qu'il s'agit d'une « propriété émergente » du cerveau »*[80].

Il faut noter que la théorie de l'émergence, a également été instrumentalisée dans le cadre du connexionnisme, comme nous l'avons vu plus haut, pour tenter de démontrer que des activités conscientes peuvent être produites par des entités informatiques. Il s'agit là d'un point de contact entre les neurosciences et les développeurs de ce que l'on appelle dans le langage courant l'« intelligence artificielle » dite « forte » (c'est-à-dire égale à celle de l'être humain). A son égard, Searle s'est montré particulièrement critique sinon hostile, notamment dans son fameux article « esprits, cerveaux et programmes » (« Minds, brains and programs ») publié en 1980. Pour lui, en effet, *« … l'ordinateur ne fait que combiner et manipuler des symboles …* (il) *exécute des programmes …* (sans qu'il) *y comprenne qlq chose. … L'esprit n'est pas un ordinateur dans le cerveau, ni un programme, ni son exécuteur. L'esprit, lui va plus loin, il comprend les symboles qu'il manie »*[81].

4° Pour autant, Searle admet que *« Les états et processus mentaux conscients ont une caractéristique particulière … la subjectivité »*[82], caractéristique qu'il considère,

---

[79] *La redécouverte de l'esprit* p 160

[80] *Le Mystère de la conscience* p 30

[81] Entretien précité in Philosophie magazine p 62

[82] *La redécouverte de l'esprit* p 138

prioritairement, sous un angle ontologique. Ainsi, l'énoncé « J'ai une douleur au dos » correspond à un fait réel (la douleur) lequel a *« un mode subjectif d'existence ... une existence à la première personne »*[83]. Cette approche ontologique conduit Searle a intégré dans sa théorie de la conscience, une part de non-réductionnisme. En effet, en affirmant que l'esprit est causé par le cerveau, il admet l'existence de la réduction du mental au cérébral, le mental n'étant finalement *« rien que »* du cérébral. Pour autant, quand bien même « *dans l'histoire de la science, les réductions causales* (conduisent) *à des réductions ontologiques »*[84], en ce qui concerne la conscience, il est impossible de réduire celle-ci aux phénomènes neurologiques. Ceci revient à dire que la conscience est irréductible, c'est-à-dire « inéliminable », et que le naturalisme biologique n'est donc pas pleinement réductionniste. Ceci n'empêche pas Searle d'affirmer, d'une part, qu'*« on peut ... accepter l'existence et l'irréductibilité de la conscience comme phénomène biologique, sans accepter ... le dualisme »*[85] ; et, d'autre part, que *« l'irréductibilité de la conscience ... n'a aucune conséquence métaphysique quant à ... notre vision scientifique globale du monde »* [86].

D'aucuns voient dans cette affirmation une réhabilitation du dualisme.

5° Mais, cette position a une conséquence importante sur le plan épistémologique. En effet, notre conception scientifique du monde nous conduit à appréhender celui-ci du point de vue de sa réalité objective, c'est-à-dire de l'extérieur ; notre accès au monde s'effectuant par l'observation, l'acte d'observer qui est lui-même un acte conscient. Or, *« là où il est question de subjectivité, il n'y a aucune distinction entre l'observation et la chose*

---

[83] *Ibid.* p 139
[84] *Ibid.* p 164
[85] *Le Mystère de la conscience* p 201
[86] *La redécouverte de l'esprit* p 173

observée, entre la perception et l'objet perçu ... »[87]; il est impossible de séparer la chose vue et le fait de la voir. C'est la raison pour laquelle nous avons du mal à accepter la subjectivité car l'acte d'observer, est bien produit par un processus biologique naturel mais il constitue aussi un phénomène mental subjectif donc inobservable. « *Notre modèle ... de la relation entre réalité et observation ... est un modèle d'observateurs objectifs (au sens épistémique) observant une réalité objectivement existante (au sens ontologique)* (qui) *ne permet* (pas) *d'observer l'acte d'observer lui-même »*[88].

6° Enfin, Searle est également intervenu sur la question relative aux impacts des progrès des neurosciences en matière de libre-arbitre, notamment dans une de ses conférences publiées en 2004 dans un recueil intitulé *Liberté et neurobiologie*.

Pour Searle *« ... il n'y a que deux obstacles qui nous empêchent d'avoir une pleine compréhension des rapports entre le corps et l'esprit : le préjugé philosophique* (selon lequel) *le mental et le physique sont deux domaines distincts* (c'est-à-dire le dualisme) *et notre ignorance des opérations du cerveau »*[89]. Pour autant, ceci n'empêche pas Searle d'affirmer que *« Le mystère de la conscience disparaîtra lorsque nous* (aurons résolu) *le problème biologique de la conscience »*[90].

§ 2 : Jaegwon Kim et le physicalisme

Le « physicalisme » est un autre naturalisme ontologique dont l'un des représentants les plus imminents - bien que méconnu - est le philosophe américain d'origine sud-coréenne : Jaegwon Kim (1934 – 2019).

1° Selon cet auteur, le physicalisme repose sur un double postulat : a) Tout évènement est un évènement physique

---

[87] *Ibid.* p 142
[88] *Ibid.* p 144
[89] *La redécouverte de l'esprit* p 146
[90] *Le Mystère de la conscience* p 207

qui a une cause physique laquelle s'inscrit dans une chaîne causale qui sépare le physique du non physique ; le monde physique est donc causalement clos ; c'est ce qu'énonce le principe de la *« clôture causale du domaine physique »*, seul susceptible de fonder la possibilité d'une théorie complète et englobante de tout le monde physique ; b) Pour autant, la réalité des états mentaux ne peut être niée.

Se pose alors la question : comment admettre, la réalité du mental, de *L'esprit dans un monde physique* (« *Mind in a physical world* »)[91] pour reprendre le titre d'un ouvrage de Kim, paru en 1998 ? Selon son auteur, cette problématique recouvre deux questions essentielles : celle des relations entre le corps et l'esprit d'abord, celle de la conscience et de la subjectivité, ensuite.

2° La première question est certainement la plus importante. Kim y répond en faisant jouer la préséance du physique : les faits physiques déterminent tous les autres faits, les propriétés physiques d'une chose déterminent toutes ses autres propriétés. C'est l'*« argument de la survenance »* qui est une notion fondamentale de la philosophie kimienne, introduite dans le débat sur le problème corps-esprit, par Donald Davidson, dans son article les *« Evènements mentaux »*, publié en 1970 et qui va connaître son essor à la fin de cette décennie[92]. Kim définit la survenance appliquée aux relations corps-esprit, de manière simplifiée, comme suit : *« ... une propriété physique P, base de la propriété mentale M, garantit comme nécessaire l'occurrence de M : nécessairement, si une chose instancie P, elle instancie aussi M »*. Il précise qu'on peut associer *« la survenance avec l'idée de <u>dépendance</u> ou de <u>détermination</u> »*[93]. L'argument de la survenance implique que parmi toutes les propriétés que peut posséder un être humain - qui est une chose

[91] Cambridge, (Mass.) MIT Press, 1998, Traduit de l'américain par François Athané et Edouard Guinet, Paris, Ithaque Editions, 2014,

[92] *Ibid.* p 6 et 7

[93] *Ibid.* p 11

entièrement physique -, les propriétés mentales, psychologiques (sensations, désirs, croyances) bien qu'elles ne semblent pas être des propriétés physiques, dérivent, procèdent des propriétés physiques elles-mêmes, « surviennent » sur celles-ci. Autrement dit, les propriétés physiques sont des propriétés de base, de premier ordre, « subvenantes » tandis que les propriétés mentales sont de second ordre, « survenantes ». Ajoutons à cela, que la notion de survenance conduit à l'idée d'un *« modèle feuilleté »*[94], c'est-à-dire organisé en une structure hiérarchique, comme stratifié en différents « niveaux », de bas en haut, du plus simple au plus complexe, à partir d'un niveau de base, constitué d'entités sans parties propres, de « particules élémentaires » en quelque sorte. C'est donc à partir des propriétés de niveau inférieur que s'opère la « réalisation » (terme introduit par Hilary Putnam dans son article « Minds and Machins » publié en 1960[95] - on dit aussi exécution ou implémentation, voire instanciation) les propriétés du niveau supérieur. A noter que, le physicalisme considère les propriétés physiques comme les « réaliseurs », du fait de leur causalité, de leurs rôles spécifiés par leurs relations causales-nomologiques (dépendant des lois du domaine physique), aux autres propriétés et non pas du fait de leur caractère intrinsèque.
Néanmoins, selon Kim, cet argument n'est pas une théorie explicative de la relation corps-esprit, mais seulement un modèle de relation de dépendance entre les deux.
3° En effet, se pose la question de la causalité mentale. *« Comment l'esprit peut-il exercer ses pouvoirs causaux dans un monde ... fondamentalement physique ? »*[96] ; dans un tel monde physique, une cause mentale est-elle possible ? Au terme d'un raisonnement construit, en dix étapes, sous forme de dilemme, Kim conclut que la causalité mentale est inintelligible[97].

---

[94] *Ibid.* p 16
[95] *Ibid.* p 8 Note 1
[96] *Ibid.* p 33
[97] *Ibid.* pp 42 à 48

4° Toutefois, pour que l'argument de la survenance devienne explicatif, pour affirmer le principe d'une dépendance causale du mental à l'égard du physique, de la causation du mental par le physique, il faut procéder à la réduction du mental, considérer que le mental est réductible au physique, c'est-à-dire est susceptible d'être « physicalisé », ce qui revient à s'opposer au postulat de l'irréductibilité de la conscience et au dualisme cartésien. Cette nécessité de la « réduction » constitue, pour Kim, un réquisit incontournable. Le réductionnisme permet d'établir le réalisationnisme physique, base de l'efficacité explicative de la survenance corps-esprit, du fait de la force de la nécessité nomologique et non d'allégations métaphysiques ou de constructions logico-conceptuelles.
5° A cet effet, Kim va concevoir un modèle de réduction du mental au physique, qui fait intervenir la « fonctionnalisation » des propriétés mentales, c'est-à-dire l'attribution à celles-ci d'un pouvoir, d'un rôle causal en lien avec celui des entités physiques, qui les « réalisent ». Autrement dit, le physicalisme implique la conjonction de deux thèses distinctes, la thèse réductionniste et la thèse fonctionnaliste. *« Dans la conception fonctionnaliste, les propriétés mentales sont spécifiées par leur rôle causal, c'est-à-dire en termes de relations causales exécutées par des propriétés physiques de premier ordre. En ce sens, les propriétés mentales paraissent être des propriétés extrinsèques ou relationnelles, des* (entités) *qui les possèdent »*[98]. Autrement dit, *« Si nous voulons réduire une propriété M à un domaine de propriétés de base, nous devons « préparer » M pour la réduction, en l'interprétant ou en la réinterprétant relationnellement ou extrinsèquement. ... M* (est) *une propriété de second ordre définie par son rôle causal ... M est la propriété d'avoir une propriété* (P) *qui correspond à* (une) *spécification causale. C'est ce qui fonde l'identification de M à P »*[99].

[98] *Ibid.* p 21 et 22
[99] *Ibid.* p 101

6° A ce stade, la question est : le mental est-il susceptible de fonctionnalisation en vue d'une réduction ? Kim admet que l'on peut donner une réponse positive si on retient une conception large du domaine physique, autrement dit l'idée de la généralisation de l'argument de la survenance. Le point de départ est de savoir s'il existe, en dehors des propriétés physiques fondamentales, de base, d'autres propriétés susceptibles d'être admises dans le monde physique. A cet égard, on peut considérer, sans difficulté majeure, que les propriétés des sciences spéciales telles que les propriétés chimiques, géologiques, biologiques voire neurologiques surviennent sur les propriétés physiques et que leur causalité – c'est-à-dire les relations causales impliquant des propriétés chimiques, géologiques, biologiques ne semble pas poser de problèmes particuliers. Selon Kim, il y a *« trois conditions de délimitation du domaine physique : 1° est physique toute entité … physique, 2° est physique toute propriété … microbasée dans des entités ... physiques, 3° est physique toute propriété ... de second ordre par rapport à des propriétés physiques »*[100]. Or, les propriétés mentales, on l'a vu, peuvent être réduites à des propriétés fonctionnelles de second ordre comme le sont les propriétés chimiques, géologiques, biologiques.

Toutefois, Kim voit une limite dans le domaine neuropsychologique. En principe, *« … les propriétés fonctionnelles … propriétés de second ordre … n'ont pas de pouvoirs causaux* (supplémentaires) *qui aillent au-delà des pouvoirs causaux de leurs réaliseurs de premier ordre »*[101]. Or, s'agissant *« des propriétés psychologiques survenantes,* (leurs) *propriétés physiques de base - neurobiologiques - se situent au même niveau qu'*(elles) *: les unes et les autres appartiennent aux êtres humains … C'est ce qui fait le problème de la causalité mentale : le rôle causal d'une propriété mentale … est menacé de*

[100] *Ibid.* p 118
[101] *Ibid.* p 119

préemption par une ... propriété neurale … Il s'ensuit que, pour les niveaux supérieurs, nous devons accorder de nouveaux pouvoirs causaux aux propriétés microbasées – nouveaux (car non) *détenus par aucune des propriétés de niveau inférieur ... »*[102].

6° Reste la question de l'énigme de la « conscience » (« consciousness ») et de la subjectivité. Kim considère qu'il existe *« deux grandes catégories dans le domaine du mental : l'intentionnel et le phénoménal* (le quale) *»*. S'agissant de la première, il affirme sans autre précision : *« Je ne vois pas d'objections de principe à une explication fonctionnaliste de l'intentionnalité »*. Au sujet de la seconde, il écrit : « (je pense) *que la difficulté principale vient des qualia »*[103], estimant peu probable la possibilité de leur fonctionnalisation. Kim reconnait ainsi la réalité, du *« quale »*, de l'expérience subjective et intersubjective et de son caractère irréductible.

Il s'agit là d'un jugement qui vient nuancer son physicalisme.

Chap 3 : La recherche d'une conjonction entre la phénoménologie, les neurosciences et les sciences cognitives

La phénoménologie se définit comme l'approche descriptive, la « science » des phénomènes (du grec *phainomenon* « ce qui apparaît »). Au rang des phénomènes figure la conscience, l'expérience subjective. On peut donc établir un lien entre la phénoménologie et la philosophie de l'esprit. Mais alors, quelle place la phénoménologie occupe-t-elle en philosophie de l'esprit ? C'est ce que l'on va tenter de voir en invoquant trois figures majeures de cette discipline philosophique : Edmund Husserl, son fondateur ; Maurice Merleau-Ponty, qui a joué le rôle d'introducteur de la phénoménologie

[102] *Ibid.* p 121
[103] *Ibid.* p 104 et 105

dans la philosophie de l'esprit et enfin Francisco Varela, l'initiateur de la neurophénoménologie (ou de sa tentative) au travers de la théorie de la cognition incarnée.

§ 1 : Edmund Husserl

Né dans l'empire autrichien puis naturalisé allemand, Husserl (1859 – 1938) peut être considéré comme un des philosophes les plus influents du XX° siècle. Il n'a eu de cesse de faire de *« la philosophie « comme science rigoureuse » »*.

Avant d'exposer ses conceptions relatives à l'esprit, la conscience et le corps, il importe de présenter les grandes lignes de sa philosophie, particulièrement au regard de l'ontologie.

1° Son point de départ – à l'instar de son contemporain Henri Bergson -, consiste en la critique du positivisme et du scientisme. Surtout, Husserl opère un retour à Platon et à son idéalisme. La réalité première, est celle des idées ou des *essences*, sans lesquelles les choses ne seraient pas ce qu'elles sont. La philosophie est dès lors une *science eidétique* (du grec *eido*, idée, essence) c'est-à-dire qu'elle porte sur les idées. Elle exige d'adopter une attitude, une démarche pour passer du phénomène à l'essence, ce que Husserl appelle la « réduction eidétique ». Il s'agit, selon son célèbre mot d'ordre, de *« faire retour aux choses mêmes ! »*[104].

Pour autant, la philosophie de Husserl n'entre pas dans le sillage du dualisme ontologique cartésien. Il s'agit plutôt d'un monisme idéaliste qui rejette notamment toute relation de causalité entre l'âme, l'esprit et le corps au profit du concept de la motivation. Ces éléments permettent de comprendre son antinaturalisme qui réduit

---

[104] *Recherches logiques*, t1, *Prolégomènes à la logique pure*, Introduction, traduit de l'allemand par Hubert Elie, Arion L. Kelkel, René Schérer, Paris, Puf, coll « Epiméthée », Paris, PUF, 1990, p 171.

« à un fait de nature, la conscience (et) *les idées »*[105] et qui l'amène à combattre *« la philosophie ... de la naturalisation de la conscience »* c'est-à-dire *la psychologie* (psychophysique) *et plus particulièrement la psychologie expérimentale »*[106].

2° Selon Husserl, la conscience est la clé pour accéder à l'essence des choses. Mais, pour cela, il faut abandonner l'attitude, naturelle, « naïve », consistant à confondre spontanément l'essence des objets et sa représentation.

Surtout, la conscience est liée à la perception, celle-ci se produisant dans celle-là, l'élément perçu n'apparaissant que progressivement. Cette prise en compte de la perception établit, non pas une filiation mais tout au moins, un lien avec les empiristes anglais des XVII° et XVIII° siècles dont Locke et Hume, qui font de la perception et de l'immanence, le vecteur de l'expérience. La phénoménologie va développer cette approche de la conscience au travers de la vie, mieux du vécu. En effet, selon Husserl, la conscience est constituée par un flux continu de vécus.

En outre, contrairement au phénomène physique, le phénomène psychologique, l'acte de conscience (sensation, sentiment, souvenir ...) vise une chose précise. *« ... tout vécu de conscience en général est, en lui-même, conscience de ceci ou de cela* (nb : nous soulignons) *... tout cogito ; tout vécu de conscience ... vise quelque chose, et porte en soi, sur le mode de ce qui est visé, ce qui est dans chaque cas son cogitatum ; chaque cogito le fait à sa manière »*[107]. L'acte – la noèse - et l'objet visé – le noème - sont liés de manière très particulière. L'objet visé, n'est pas l'objet tel qu'il existe objectivement, dans son

---

[105] *La philosophie comme science rigoureuse*, traduit de l'allemand par Marc de Launay, Paris, Puf, coll « Epiméthée », 4° édition, 2003, p 20

[106] *Ibid.* p 23 et 24

[107] *Méditations cartésiennes*, 2ème, § 14, traduit de l'allemand par Marc de Launay, Paris, Puf, coll « Epiméthée », 1994 p 77

essence ; il n'est que le contenu que produit l'acte de conscience, ou noèse. A noter encore que l'acte de conscience se définit comme un acte de relation par lequel le sujet se rapporte au monde. Cette propriété de la conscience d'exister toujours comme visée, d'être conscience de quelque chose, mais d'autre chose qu'elle-même, est conceptualisée par Husserl sous le concept d'intentionnalité. *« ... Le terme d'intentionnalité ne signifie rien d'autre que cette propriété fondamentale et générale de la conscience qui est d'être conscience de quelque chose, de porter en soi, en tant que cogito, son cogitatum »*[108].

3° La perception, on l'a dit, joue un rôle central dans la réduction eidétique, rôle qui tient à la prise en compte de l'expérience, du « vécu ». Ceci renvoie à la place qu'occupe le corps dans la phénoménologie. Importante chez Husserl, cette thématique sera approfondie et développée par Merleau-Ponty, comme nous le verrons plus loin. A ce stade, il importe de souligner la distinction opérée par Husserl entre : le « Körper », le corps extérieur, objectif, physique, localisable, composé de ses organes et le « Leib », le « corps propre », « vivant », subjectif, ou « chair », corps uni à la conscience.

4° Notons enfin que la phénoménologie distingue trois « régions » de l'être : La nature, l'animalité, l'humanité. Il s'ensuit que l'expérience, le « vécu » n'est pas celui d'un être isolé, mais d'un être en rapport avec les autres, avec la nature-monde. L'intersubjectivité est un concept fondamental de la phénoménologie. Comment alors percevoir autrui comme un autre moi ? C'est parce que je sais que mon corps extérieur, objectif recèle aussi mon corps propre, vivant, ma chair, que je peux saisir intuitivement que le corps physique d'autrui recèle aussi une vie charnelle.

---

[108] *Ibid.* p 78

§ 2 : Maurice Merleau-Ponty

Bien qu'il soit généralement rattaché au courant existentialiste, le philosophe français Maurice Merleau-Ponty (1908-1960) occupe une place importante parmi ceux qui introduisirent la phénoménologie en France.
De surcroit, c'est à partir de la phénoménologie qu'il s'est interrogé sur le problème corps-esprit (aux côtés de Jean-Paul Sartre, mais dans une moindre mesure en ce qui concerne ce dernier.)
1° La perception est le concept fondamental de la pensée de Merleau-Ponty : c'est au moyen de la perception que nous prenons conscience de notre appartenance au monde qui nous entoure. Pour lui, la conscience est principiellement perceptive. Il complète l'affirmation de Husserl "toute conscience est conscience de quelque chose", en ajoutant "toute conscience est conscience perceptive". Merleau-Ponty explique que l'on ne peut séparer la conscience qui perçoit, de l'objet perçu. C'est pourquoi il critique à la fois la tradition idéaliste de Descartes et la tradition empirique de Locke. Il oppose aux idéalistes que la conscience ne peut être isolée, séparée du monde dans lequel elle est immergée, et, aux empiristes, que les phénomènes n'existent pas par eux-mêmes. La perception est donc ce par quoi je ne fais qu'un avec le monde dans lequel je me situe. C'est pourquoi lorsque je perçois un objet, il manque nécessairement une partie de celui-ci, puisque je suis aussi au-dedans de cet objet.
Cette conception de la perception simultanée de soi et du monde, de soi « percevant » et du monde « perçu », constitutive de la démarche phénoménologique, Merleau-Ponty l'illustre dans sa description de la vision et du regard : *« ... puisque la vision est palpation par le regard, il faut qu'elle s'inscrive dans l'ordre d'être qu'elle nous dévoile, il faut que celui qui regarde ne soit pas lui-même étranger au monde qu'il regarde. Dès que je vois, il faut que la vision soit doublée d'une vision complémentaire ou*

d'une autre vision : moi-même vu du dehors, tel qu'un autre me verrait, installé au milieu du visible, en train de le considérer d'un certain lieu »[109].

2° Le primat de la perception implique la place importante faite au corps dans la philosophie de Merleau-Ponty. Il lui confère un nouveau sens, un nouveau statut, celui de la corporéité qui n'est pas une superposition de réalités, mais une dialectique entre le corps, l'esprit et le monde. En effet, la perception n'est rendue possible que par le corps, puisqu'il y est d'emblée et en premier lieu impliqué. En outre, le corps est à la fois « objet du monde », en tant qu'il est lui-même, une chose ou un phénomène et un « point de vue sur le monde », puisque son rapport au monde est spontané. Le corps ici correspond à ce que Husserl nommait le « corps propre », défini non pas comme ensemble d'organes et de fonctions, mais par le seul fait qu'il rend possible la perception ; comme ce que ma conscience "habite", qui habite le monde et qui partage le monde avec d'autres corps. Le corps propre n'est pas réductible à un simple objet extérieur, c'est une présence au monde qu'il habite en partage avec d'autres. On retrouve là le thème husserlien de l'intersubjectivité qui s'étend avec Merleau-Ponty à l'intercorporéité car autrui ne se réduit jamais à un corps matériel. Il est saisi d'emblée comme autre sujet dans l'intersubjectivité.

3° Avec le concept de « chair », Merleau-Ponty élabore une définition incarnée de l'être : *« La chair n'est pas matière, n'est pas esprit, n'est pas substance. Il faudrait pour la désigner, le vieux terme d'« élément », au sens où on l'employait pour parler de l'eau, de l'air, de la terre et du feu, c'est-à-dire au sens d'une chose générale, à mi-chemin de l'individu spatio-temporel et de l'idée, sorte de principe incarné qui importe un style d'être partout où il s'en trouve une parcelle. La chair est en ce sens un*

[109] *Le visible et l'invisible*, Paris, Gallimard, coll « Bibliothèque des idées », 1964, p 177

« élément » de l'Être »[110]. Aussi : *« Il faut penser la chair, non pas à partir des substances, corps et esprit, car alors elle serait l'union de contradictoires, mais ... comme élément ... »*[111]. Ainsi, Merleau-Ponty extrait la chair du dualisme cartésien qui, en la substantialisant, en fait « l'union de contradictoires ». Avec le concept de chair, sa philosophie dépasse le dualisme substantiel corps-esprit, en qualifiant dans un même mouvement, le sujet et le corps. Le sujet est incarné et habite un corps ; le sujet est corporel et le corps est subjectif. La chair qui unit le visible et l'invisible régit l'ensemble de la pensée ontologique de Merleau-Ponty.

4° Finalement, on peut considérer que celle-ci établit un « pont » entre la phénoménologie et Spinoza. Dans ce sens, on relèvera que, dès ses premiers travaux, Merleau-Ponty soulignait que *« ... matière, vie, esprit ne* (peuvent) *être définis comme trois ordres de réalité ou trois sortes d'êtres, mais comme trois plans de significations ou trois formes d'unité »*[112]. Plus loin, il ajoutait *« Le corps et l'âme sont des significations et n'ont donc de sens qu'au regard d'une conscience »*[113].

Avec Merleau-Ponty, le "retour aux choses mêmes" de Husserl s'opère donc par la perception, grâce à l'intermédiaire du corps.

Ceci ouvre des perspectives sur la « possibilité » de « naturaliser la phénoménologie ».

§ 3 : Francisco Varela

1° Le neurobiologiste et philosophe chilien Francisco Varela (1946 – 2001) est à l'origine de la théorie de l'« énaction » (de l'anglais to enact « susciter », « faire « advenir ») ou de la « cognition incarnée » ou plutôt

---

[110] *Ibid.* p 184

[111] *Ibid.* p 193

[112] *La structure du comportement*, Paris, PUF, coll « Bibliothèque de philosophie contemporaine », 7° édition, 1972, p 217.

[113] Ibid p 233.

« énactive »[114] qui vise à intégrer à la fois nos capacités « intérieures » physico-psychiques et nos liens avec notre environnement « extérieur » biologique et culturel. Elle est une voie moyenne qui consiste à « *contourner cette … logique « intérieur / extérieur » en étudiant la cognition non comme reconstitution ou projection, mais comme action incarnée … mot* (qui souligne) *deux points :* a) *la cognition dépend des types d'expérience qui découlent du fait d'avoir un corps doté de diverses capacités sensori-motrices ;* b) (lesquelles) *s'inscrivent dans un contexte biologique, psychologique et culturel … le terme action souligne les processus sensoriels et moteurs … perception et action sont inséparables dans la cognition vécue … elles ont évolué ensemble* »[115]. Par la suite, Varela indiquera que cette théorie comporte deux idées : a) « *L'action est guidée par la perception* ». Il ajoute : « *comme le note Merleau-Ponty, l'organisme donne forme à son environnement en même temps qu'il est façonné par lui. … nous devons appréhender l'organisme et l'environnement dans la sélection et la détermination réciproques qui les lient l'un à l'autre* »[116] ; b) « *les structures cognitives se dégagent de types récursifs de motifs sensori-moteurs qui permettent à l'action d'être guidée par la perception* ». Varela déclare ici se référer à Jean Piaget et à « *son programme d'épistémologie génétique* (en vue d') *expliquer comment le développement de l'enfant transforme un organisme biologique immature en un être doué de raison abstraite à l'âge adulte* »[117]. Plus loin, il admet aussi la nécessité de prendre en compte la théorie de l'évolution. L'énaction

---

[114] Dans un entretien avec Réda Benkirane publié dans *La Complexité, vertiges et promesses* Paris Editions le Pommier 2002, Varela indiquait préférer le terme « énactive », car « *le mot incarné … a une résonance théologique* »

[115] *L'inscription corporelle de l'esprit*, traduit de l'anglais par Véronique Havelange, Paris, Seuil coll » La couleur des idées », 1993, p 234.

[116] *Ibid.* p 235 et 236.

[117] *Ibid.* p 239.

serait *« le produit de l'évolution biologique et de son mécanisme de sélection naturelle »* qui permettrait l'*« ajustement plus ou moins optimal au monde »*[118].

2° Cette thèse de l'énaction, de la cognition énactive est au fondement de la neurophénoménologie, c'est-à-dire de la convergence de la phénoménologie et des neurosciences. Dans ce sens, elle se situe dans le prolongement de la phénoménologie de Merleau-Ponty et de sa théorie de la corporéité, le corps – *« ... à la fois structure vécue et contexte ou lieu des mécanismes cognitifs »*[119] - étant en lien avec le monde. Mais elle intègre aussi le développement des sciences cognitives – d'où l'emploi du terme « cognition » -, notamment les concepts et le langage de la neurobiologie (auto-organisation ou autopoïèse, émergence). Varela explique que la cognition énactive s'inscrit dans le courant des théories de l'information et de sa traduction, après le computationnalisme, dans le connexionnisme[120].

3° Ainsi, se fait jour, avec Varela, l'idée d'une alliance entre la phénoménologie et les neurosciences / sciences cognitives. Mais cela implique de *« couper la phénoménologie husserlienne de ses racines antinaturalistes ou la naturaliser »*[121]. Ce projet de naturalisation de la phénoménologie a été engagé notamment par la tenue du colloque « Actualité cognitive de la Phénoménologie : les défis de la naturalisation » en octobre 1995 à Bordeaux à l'initiative de Francisco Varela, du psychiatre Bernard Pachoud et des philosophes Jean Petitot et Jean-Michel Roy. Il s'agissait de constater que

---

[118] *Ibid.* p 244.

[119] *Ibid.* p 18.

[120] Entretien avec Réda Benkirane précité

[121] « Essai introductif - Combler le déficit : Introduction à la naturalisation de la phénoménologie » in *Naturaliser la phénoménologie*, sous la dir. de Bernard Pachoud, Jean Petitot, Jean-Michel Roy et Francisco Varela Paris CNRS Editions 2002 p 54

l'approche strictement neuroscientifique, voire neurophysiologique, empruntée aux sciences « dures » de la nature, s'avère incomplète ou limitée du fait de son incapacité à appréhender la subjectivité et qu'elle ne parvient pas à combler le déficit explicatif, « explanatory gap » (selon l'expression de Lévine) que constitue la conscience phénoménale – la dimension subjective de la vie mentale – représentée notamment par les qualia.

Il s'agit là d'une question non plus du ressort de l'ontologie mais de celui de l'épistémique.

Pour y parvenir, des phénoménologues et des neuroscientifiques, tels les initiateurs du colloque susvisé, en sont venus à l'idée qu'il est nécessaire d'intégrer au sein des sciences cognitives, dont la nature s'avère de plus en plus pluridisciplinaire, une voie explicative supplémentaire, celle ouverte par la phénoménologie.

En effet, celle-ci apporte aux sciences cognitives, deux éléments nouveaux. En premier lieu, sa démarche descriptive complète celle axiomatico-déductive, objective des neurosciences. En second lieu, elle permet de saisir la subjectivité et l'intersubjectivité, puisqu'en appréhendant le sujet pensant dans son inscription à la fois corporelle et mondaine (dans le monde) elle prend en compte la réalité des vécus individuels de conscience et l'étend à la réalité du monde extérieur. Ainsi, la « naturalisation de la phénoménologie » doit s'entendre, non comme l'annexion de cette dernière par les sciences cognitives, mais comme une liaison entre l'étude neuroscientifique du cerveau et la mise en lumière des vécus de conscience.

Aujourd'hui, ce projet est toujours en cours, comme l'a montré la tenue d'un nouveau colloque en avril 2014 à Strasbourg organisé par le philosophe Jean-Luc Petit.

Sans doute, serait-il opportun de faire le point sur l'état d'avancement de ces recherches.

IIIème PARTIE
La problématique des limites à apporter aux applications des connaissances en neurosciences et sciences cognitives

« Science sans conscience n'est que ruine de l'âme » Rabelais

Introduction

1° Les neurosciences s'insèrent dans le champ plus vaste des sciences cognitives (Voir leur « Hexagone » en appendice de l'introduction) qui intègre une discipline appelée, dans le langage courant « intelligence artificielle », dont on a déjà dit qu'il serait plus juste de la nommer « computation artificielle », mais que nous continuerons à désigner par son expression habituelle, voire par son acronyme « IA ».
2° Que faut-il entendre par intelligence artificielle ? Nous avons déjà évoqué, la distinction souvent opérée entre IA « faible » et « IA forte ». Le rapport de la Commission Nationale Informatique et Libertés *Comment permettre à l'homme de garder la main ?* de décembre 2017 distingue bien : *« ... IA faible (IA capable de simuler l'intelligence humaine pour une tâche bien déterminée) et IA forte (IA générique et autonome qui pourrait appliquer ses capacités à n'importe quel problème, répliquant en cela une caractéristique forte de l'intelligence humaine, soit une forme de « conscience » de la machine »*[122]. Sur un plan général, on retiendra la définition actualisée des « systèmes d'IA », retenue par le Groupe d'experts indépendants de haut niveau (ci-après, le GEIHtN)) mis en place par l'Union Européenne : ce *« sont des systèmes logiciels (et éventuellement matériels) conçus par des êtres humains et qui, ayant reçu un objectif complexe, agissent dans le monde réel ou numérique en percevant leur environnement par l'acquisition de données, en interprétant les données structurées ou non structurées*

[122] Glossaire, Intelligence artificielle (IA), p 75

collectées, en appliquant un raisonnement aux connaissances, ou en traitant les informations, dérivées de ces données et en décidant de la/des meilleure(s) action(s) à prendre pour atteindre l'objectif donné. Les systèmes d'IA peuvent soit utiliser des règles symboliques, soit apprendre un modèle numérique. Ils peuvent également adapter leur comportement en analysant la manière dont l'environnement est affecté par leurs actions antérieures »[123].

3° Au-delà du caractère scientifique de l'IA, il s'agit là d'une discipline qui connait, depuis quelques années seulement, un fort retentissement dans les médias, eu égard à la puissance d'action qui pourrait en découler pour modifier, en profondeur, nos systèmes de pensée, nos modes de vie, tant individuels que collectifs. De surcroit, ce retentissement est d'autant plus important en raison de ses liens de proximité avec a) le « mythe de la singularité technologique », c'est-à-dire la croyance selon laquelle l'IA va non seulement égaler l'intelligence humaine mais la dépasser, être à l'origine d'une superintelligence extérieure à la nature humaine ; b) le post- ou trans-humanisme, un courant d'idées dont l'objectif est de préparer la venue de l'« *homme humain, demain* » - qui sera un « homme augmenté » - d'où la nécessité de prôner *« l'usage des sciences et des techniques afin d'améliorer les caractéristiques physiques et mentales humaines »*[124]. Dès lors, cet objectif conduit à mettre en œuvre une action, un projet et une stratégie, tendant à faire évoluer la société dans le sens préconisé : aussi, bien plus qu'une croyance ou une philosophie, le post- ou trans-humanisme constitue une idéologie politique. On soulignera que celle-ci fait l'objet de nombreuses et sévères critiques, de divers horizons, philosophiques ou scientifiques, qui dénoncent

---

[123] Document « Définition de l'IA : principales capacités et disciplines. », avril 2019, p 6.

[124] Page d'accueil du site de l'Association française transhumaniste Technoprog https://transhumanistes.com

son caractère manifestement antihumaniste. A cet endroit, il importe de rappeler que le transhumanisme trouve ces racines dans ce que l'on a appelé plus haut la « convergence des NBIC », un phénomène directement et étroitement lié au modèle économique dominant de la croissance illimitée, dont le triomphalisme est inquiété (est-il besoin de le souligner ?) par l'entrée de plus en plus marquée dans l'ère de l'anthropocène : dérèglement climatique, épuisement des sources d'énergies fossiles, effondrement de la biodiversité - cause d'une « épidémie des pandémies »[125]-. Enfin, il importe de souligner qu'il existe entre IA, singularité technologique et transhumanisme plus de rapports de complémentarité que de concurrence voire d'exclusion. Ainsi, *« Les instituts qui travaillent sur ces sujets* (d'IA) *comme l'Institut du Futur de l'Humanité ou l'Université de la singularité aux Etats-Unis sont financés par des compagnies comme Google ou Cisco ou … Tesla. Des groupes qui … investissent massivement dans ces technologies. … Dans le même temps, ces firmes promettent de nous aider à* (les) *maitriser, car elles en ont la meilleure connaissance. C'est une manière de garder la main »*[126].

Alors n'est-il pas temps de réhabiliter la conscience humaine, voire de la tirer de l'oubli, au sens où l'expression « conscience humaine » est aussi l'objet d'un programme de recherches tendant à augmenter le niveau de conscience de l'être humain, à rendre l'humanité de plus en plus éclairée, à *« Ce qui conduit les hommes vers la Société commune c'est-à-dire fait qu'ils vivent dans la concorde »*, pour reprendre les termes de Spinoza[127]

Ainsi, s'ouvre le champ du questionnement en matière éthique qui nous conduit à nous interroger sur trois points. En premier lieu, il s'avère nécessaire de clarifier les enjeux

---

[125] Cf : Marie-Monique Robin *La fabrique des pandémies* La Découverte coll « Cahiers libres » 2021.

[126] Jean-Gabriel Ganascia Entretien in Le Monde du 11 mars 2017

[127] Baruch Spinoza *Ethique* IV° partie, 40ème proposition trad. par Charles Appuhn, Paris, Flammarion, coll « GF », 1965, p. 258.

actuels ou futurs, notamment les conséquences socioculturelles réelles et potentielles, du développement des connaissances et des techniques des sciences cognitives et particulièrement de l'IA (Chap. 1). En second lieu, il faudra examiner ce que l'éthique appliquée à la science en général, aux neurosciences en particulier, et enfin à l'IA elle-même peut ou pourrait apporter, sur le plan principiel et normatif, pour prévenir les risques susceptibles de découler de ses usages (Chap. 2). En troisième lieu, il faudra envisager, les mécanismes politiques et sociaux, qui, sur la base des principes et normes éthiques prédéfinis, devraient être mis en place (Chap. 3).

Chap 1 : l'intelligence artificielle, ses enjeux actuels et futurs

La question ici posée est de savoir quelles conséquences réelles et potentielles l'IA est susceptible d'entraîner pour l'être humain, notamment sur le terrain socioculturel. Quel peut-être le « rapport bénéfices / risques » de l'IA ?
Pour répondre à cette question, il convient, au préalable, d'analyser les facteurs qui influencent la capacité de l'IA à « agir » sur l'individu et la société. Certains sont intrinsèques, internes aux systèmes d'IA eux-mêmes ou systémiques ; d'autres sont extrinsèques, externes, ce sont les facteurs qui concernent les individus et la société en général.
Mais ce qui est le plus important, c'est d'analyser prioritairement les facteurs qui régissent le développement exponentiel de l'IA dans le monde contemporain, voire dans notre vie quotidienne.

§ 1 : Les facteurs du développement actuel de l'IA

Il faut distinguer entre les facteurs qui jouent en faveur de ce développement et ceux qui, au contraire, sont susceptibles de le freiner.

I. Les facteurs favorables

Ils sont de deux ordres :
1° Sur le plan interne, trois facteurs sont déterminants : a) l'augmentation de la puissance de calcul des ordinateurs. Aujourd'hui, le supercalculateur le plus puissant au monde, n° 1 du classement publié le 22 juin 2020, est installé à Kobé, au Japon. Sa puissance de calcul atteint 415,5 pétaflops (1 pétaflops = 1 million de milliards d'opérations par seconde)[128]. b) L'explosion du Big data, c'est-à-dire, le volume des données historiques disponibles, du fait de leur captation au moyen d'internet. c) La complexification des algorithmes (définis comme une suite d'instructions à exécuter) avec le déploiement des techniques dites d'apprentissage machine profond (« deep machin learning »). Il s'agit là *« de méthodes pour entraîner un système* (d'IA) - *à accomplir une tâche - au lieu de le programmer explicitement »*. On soulignera qu'il en existe trois niveaux, selon que l'apprentissage est *« supervisé : le système est entrainé … à partir d'exemples d'entrée et de sortie correspondante »* ; *« par renforcement : le système, s'entraîne … par essais et erreurs »*, ou *« non supervisé : Le système découvre les interdépendances entre les variables d'entrée sans être entrainé pour une tâche particulière. … »*[129]. En clair, ce dernier niveau conduit à produire des systèmes d'IA en partie autoorganisés c'est-à-dire autonomes. On comprendra aisément pourquoi le GEIHtN pose la question de la *« Boîte noire de l'IA … : Certaines techniques d'apprentissage automatique, tout en donnant de très bons résultats du point de vue de la précision, sont très opaques pour ce qui est de comprendre la manière dont elles prennent leurs décisions. La notion de boîte*

[128] Ridha Loukil, « le supercalculateur le plus puissant au monde est japonais », in *L'Usine nouvelle*, du 23 juin 2020
[129] Yann Le Cun, directeur du pôle de de recherche fondamentale sur l'IA de Facebook, *Quand la machine apprend* Paris Odile Jacob 2019

noire de l'IA fait référence à de tels scénarios, dans lesquels il n'est pas possible de déterminer la raison de certaines décisions … »[130].

2° Extrinsèquement, le développement de l'IA tient au contexte dans lequel il intervient, un contexte dominé par l'idéologie néo-libérale et le modèle de la croissance illimitée, déjà critiqué plus haut, dont il résulte que toute innovation technologique ne peut être que positive, en ce sens qu'elle entraine une amélioration des conditions de la vie humaine. Ainsi, l'IA devrait *« faire baisser la pénibilité en* (effectuant) *les tâches répétitives* (ingrates et sans intérêt) *diminuer le stress en augmentant la qualité de la production …* (d'où une) *libération* (qui) *favorisera l'évolution des capacités d'abstraction et … permettra à l'homme de faire progresser son intelligence … L'IA … n'est autre qu'une nouvelle façon d'amener de la valeur »*[131]. On pense naturellement aux robots collaborateurs, (« collaborative robots » ou cobots) qui vont décharger l'être humain des missions, des tâches les plus ingrates, éprouvantes et répétitives. On retrouve là, la croyance selon laquelle, seule la technologie peut apporter une solution à tous les problèmes auxquels se trouvent confrontées nos sociétés. Pourtant l'expérience montre que toute innovation entraîne aussi des conséquences, des « externalités » négatives.

II. Les facteurs défavorables

Ils sont de deux catégories :

1° Certains sont internes aux systèmes d'IA. Comme tout artefact, toute machine, ces systèmes ne sont jamais parfaits. Si l'on utilise le langage informatique, on dira que

---

[130] Document « Définition de l'IA : principales capacités et disciplines. », avril 2019, p 6.

[131] Stéphane Roder, DG exécutif d'AI Builders, cabinet de conseil en IA, « N'ayons pas peur de l'IA au travail », in Le Monde du 1er février, 2019.

le « hardware » peut connaitre des bugs, des dysfonctionnements et des pannes. Mais c'est le « software » et particulièrement les algorithmes d'apprentissage machine profond (deep machin learning), on l'a vu, qui posent des questions majeures relatives à notre compréhension réelle du « choix » ou de la « décision » automatique du système. C'est le problème de la « boite noire ». Une autre question concerne notamment les algorithmes de sélection dont certains résultats sont quasi inévitablement de « faux positifs » ou « négatifs ».

2° Il existe aussi des facteurs externes de deux types : a) Du fait des utilisateurs : d'abord, le public a souvent tendance à « surévaluer » ce que peut faire un système d'IA. Ensuite, un tel système peut être détourné de ses fonctionnalités d'origine, à des fins de cybercriminalité voire de cyberterrorisme. Ainsi, *« … il faut s'attendre à des détournements de systèmes … mal sécurisés par des groupes de hackers avec des objectifs de cybercriminalité ou de cyberterrorisme »*[132] ; b) Du fait des concepteurs. Ceci tient à ce que les algorithmes reproduisent les biais cognitifs et les préjugés socioculturels. Ainsi les algorithmes sont toujours vecteurs du sexisme : *« l'IA reproduit les préjugés genrés de notre société, les diffuse et les amplifie »*[133]. Ils reflètent aussi les préjugés et stéréotypes racistes, homophobes, religieux, physiques avec toutes les discriminations et les inégalités qui en découlent. En définitive, les algorithmes et l'IA reproduisent les structures mentales traditionnelles et de domination. C'est ce que montre l'exemple de Tay, « chatbot » (robot conversationnel) de Microsoft devenu antisémite en quelques heures, après avoir été « trollé »

---

[132] Thierry Berthier, Maître de conférences en mathématiques, chercheur en cybersécurité, Entretien in *Voyage au cœur de l'IA* Libé – France Inter 12/2017-02/2018 p 105.

[133] Aude Bernhein et Flora Vincent, *L'intelligence artificielle, pas sans elles*, Paris, Belin / Humensis, coll « Egale à Egal », 2019 p 23.

(entrainé) par des internautes[134]. Ces biais ont une double origine : - Les algorithmes sont constitués à partir d'une « matière première », les bases massives de données (big data) générées par les humains qui contiennent les stéréotypes « populaires », « de sens commun » ; - les développeurs eux-mêmes, consciemment (voire volontairement) ou non intègrent leurs propres préjugés dans les annotations apportées aux données de base ou dans les instructions, règles données à l'algorithme (dans les critères de sélection, par exemple). Du fait que les algorithmes et leurs règles, sont conçus par des humains, les biais sont aggravés. *« … c'est vrai que … si les données d'apprentissage sont biaisées, les systèmes d'apprentissage entrainés sur ces données vont refléter le biais. C'est inévitable. …* (Or) *Beaucoup de données sont biaisées, car elles sont collectées à partir d'actions produites par des gens qui eux-mêmes sont biaisés. Ça crée des problèmes … »*[135].

La question se pose alors de savoir comment les résorber. Trois types d'opérations peuvent être menées : - en premier lieu, il importe de les détecter ; - en deuxième lieu, il peut être envisagé non pas tant de les corriger que de les réduire par « forçage », en intégrant des limites, en fixant des contraintes dans les bases de données et les algorithmes d'apprentissage machine. Mais, il s'ensuit un problème de fond : le résultat de l'algorithme ne reflète plus la réalité du monde. Se pose alors le problème de la légitimité des réductions : sur quel objet doivent-elles porter ? Qui peut en être l'auteur, le décisionnaire ? Qui en assure le contrôle et l'évaluation de leur application ? ; - plus généralement, il faut effectuer un vrai travail d'éducation pour aboutir à une nécessaire régulation des comportements et des mentalités …

---

[134] Morgane Tual « A peine lancée, une IA de Microsoft dérape sur Twitter », in Le Monde du 24 mars 2016.

[135] Yann Le Cun, Entretien in *Voyage au cœur de l'IA*, co-édition Libération - France Inter 12/2017-02/2018, p 23.

§ 2 : La nature des effets et impacts de l'IA, leurs dangers et risques réels et potentiels

Trois séries de conséquences au moins, souvent négatives, sont mises en avant.

I. Les mutations affectant la place et l'autonomie de l'individu dans la société

A cet égard, deux catégories de conséquences sont particulièrement évoquées :
1° Le développement de la robotique et, corollairement, de l'automatisation et de la robotisation des tâches en matière d'organisation du travail, va entrainer de profondes transformations structurelles du tissu économique (déjà en cours, au demeurant) et, par ricochet, du niveau et de l'organisation de l'emploi. A noter que ces mutations (qui vont affecter le secteur des services après avoir touché l'industrie) seront amplifiées sous l'effet de deux autres facteurs (déjà à l'œuvre) : la mondialisation et le vieillissement de la population. Nombre d'analyses font ressortir cette évolution attendue dans les années à venir. Ainsi, l'étude de la Oxford Martin School a fait grand bruit en 2013 en montrant que 47% des emplois aux Etats-Unis seraient menacés. Plus récemment, le rapport de l'OCDE *L'Avenir du travail* publié le 25 avril 2019 indique qu'au cours des vingt prochaines années, l'automatisation des tâches devrait, dans les Etats membres de ladite organisation, engendrer la disparition de 14% des emplois et la transformation profonde de 31,6% des postes. La France serait un peu plus exposée avec respectivement 16,4% des emplois et 32,8% des postes affectés. Certes il a été constaté que l'évolution technologique crée, en parallèle, de nouveaux emplois, - voire même que le taux d'emploi (nombre de personnes en poste parmi celles en âge de travailler) augmente -, mais aussi que la nature des emplois est transformée, ceux-ci étant moins stables sinon beaucoup plus précaires.

2° De même, le développement des systèmes d'aide, d'assistance et de délégation en matière de prise de décision pourrait, à l'extrême limite, conduire à une « déresponsabilisation » des acteurs sociaux, voire une perte d'autonomie de ceux-ci par rapport à la « machine ». Naturellement, l'IA a déjà permis et permet encore le développement des statistiques, dont les applications sont courantes dès lors qu'il s'agit d'effectuer des opérations de tri, de classification, de sélection etc. … Si l'usage des statistiques s'inscrit seulement dans une perspective prévisionnelle, l'IA joue alors le rôle de système d'aide, d'assistance à la prise de décision. Mais qu'en est-il si les statistiques sont utilisées dans un but prédictif ? N'y aura-t-il pas délégation de la prise de décision à l'IA et à ses algorithmes (biaisés) ? Cette question se pose avec d'autant plus d'acuité que cette délégation peut très bien concerner des secteurs tels que l'Education (l'orientation des élèves) la Justice (l'application des peines) la Médecine (le diagnostic et la prescription). Et nous n'en sommes qu'au début : *« … la quantité de données engendrée par notre société numérique croît de manière exponentielle à la même vitesse que les réseaux de communication et les systèmes de stockage. … La capacité des gens à analyser ces données est, elle, quasi constante. Inévitablement, … la quantité de données … dépasse* (cette capacité). *Aujourd'hui, la plupart du savoir issu des données réside dans les ordinateurs, pas dans les crânes … C'est une réalité … »*[136].

II. Les atteintes aux Droits de l'Homme, à la vie privée et aux libertés publiques

Deux séries de conséquences notamment, font l'objet de débats :

[136] *Ibid.* p 22-23.

1° La captation des données personnelles (sans consentement préalable) constitue une atteinte grave à la vie privée. Le problème est « sur-aggravé » par le fait que les données captées peuvent être vendues et réutilisées à des fins (commerciales, politiques) qui sont inconnues des personnes. L'affaire désignée sous le nom de la « société Analytica » en est un parfait exemple : la captation de données personnelles sur Facebook a permis à ladite société de créer des algorithmes pour établir des profils psychologiques et politiques d'électeurs permettant ensuite l'élaboration de stratégies de marketing électoral.
2° Le développement du contrôle et de la surveillance des personnes grâce à des techniques de plus en plus sophistiquées, efficaces et indétectables, telle la reconnaissance faciale, remet sérieusement en cause l'existence même des libertés individuelles.
On peut imaginer à l'avenir, la mise au point de systèmes permettant la prédictivité des comportements voire la « saisie » des intentions.

III. La mise en question de la nature onto-anthropologique de l'Homme et de son évolution

La question posée ici est de savoir dans quelle mesure le développement de l'IA peut être une menace ou, tout au moins, représenter un risque de mutation voire de rupture concernant la nature intrinsèque de l'être humain.
A cet égard, trois points en particulier sont à mentionner :
1° Les menaces sur l'intégrité psycho-affective de l'être humain. C'est la question de l'interface homme-machine.
Si, depuis longtemps déjà, la machine a remplacé l'homme (ou l'animal) pour effectuer certains travaux ; si, plus récemment l'ordinateur a été substitué à l'être humain pour procéder à des calculs de plus en plus complexes, aujourd'hui la question posée est celle de la possibilité d'établir des liens affectifs entre des agents conversationnels, des robots dits « émotionnels » et des personnes. Ces machines, dotées de la capacité

d'interaction sociale pourraient (peuvent déjà) avoir des fonctions d'accueil dans les boutiques, les gares, les hôtels, les établissements publics, de collaboration mais aussi d'assistance auprès de personnes dépendantes, handicapées ou malades. Quelle serait la nature du lien entre la personne et la machine dès lors que les robots ne font que simuler des émotions et ne seront jamais des êtres vivants dotés de sensibilité ?

2° Les menaces sur l'intégrité physique de l'être humain.

Elles sont liées aux possibilités d'amélioration, d'augmentation des capacités, des performances de la personne humaine, de ses fonctions intellectuelles (mémoire, attention) ou végétatives (sommeil, sexualité) mais aussi de ses comportements sociaux (humeurs). Il en est ainsi déjà, non pas tant avec des machines qu'avec des produits relevant de la neuropsychopharmacologie, à base de molécules psychotropes : le Modafinil® pour lutter contre le sommeil, le Ritaline® pour soutenir l'attention, le Provigil® pour combattre la fatigue et d'autres encore (Prozac® et autres Deroxat®). D'autres techniques sont aussi à indiquer : les cellules souches régénératives, les nanoprothèses. Dans le prolongement de ce qui vient d'être dit on peut penser que des « robots décupleurs de performances » pourront voir le jour.

On notera la tendance constante selon laquelle de telles pratiques sont d'abord mises au point sur le plan médical, pour « réparer » le corps, avant d'être « dérivées », détournées sur le plan du confort, pour augmenter les performances. D'où la question : jusqu'où aller dans la dénaturation du corps, dans la modification de la nature biophysique de l'être humain, dans la mécanisation, la réification, l'artificialisation de ce dernier ? Ceci nous ramène à l'interrogation fondamentale d'ordre onto-anthropologique : peut-on réduire la nature de l'homme à sa dimension biologique, neurologique ? N'est-ce pas une façon d'ouvrir la porte au transhumanisme qui prône,

justement, le passage de l'homme réparé, régénéré à l'homme amélioré, « augmenté » ?
3° La question des rapports Homme / Nature, nature « naturelle » et nature « artificielle ». Un des principaux arguments des transhumanistes consiste à soutenir que l'homme est toujours en « évolution » et que le transhumanisme n'est qu'une étape « naturelle » de celle-ci. Cette allégation doit être critiquée en soulignant que l'évolution de l'homme a répondu à la nécessité de s'adapter à son environnement naturel c'est-à-dire au monde extérieur à sa personne, et non pas aux contraintes qu'il s'impose à lui-même.

Chap. 2 : L'Ethique

Introduction

La prise en compte des conséquences et impacts, déjà en cours ou seulement potentiels, résultant du développement des sciences cognitives et notamment de l'IA en tant que science, conduit à s'interroger sur le plan éthique.
La question est de définir les principes et normes qui, compte-tenu desdits conséquences et impacts de l'IA, sont nécessaires pour permettre d'atteindre le but, la fin de toute action humaine, individuelle ou collective : avoir, vivre une vie bonne et heureuse.
A cet égard, de multiples questions se posent sur deux niveaux différents. Le premier, qui concerne l'éthique fondamentale, générale ou théorique, touche aux rapports que celle-ci entretient ou doit entretenir avec la connaissance, la science en général, et la recherche c'est-à-dire la science « en train de se faire ». A noter qu'il existe, à l'inverse, une autre question, en marge de notre problématique : celle de savoir si le développement des neurosciences et de la science du cerveau peut être utile à l'éthique elle-même. Le second niveau relève de l'éthique appliquée et concerne les principes et normes qui doivent

guider les neurosciences d'une part et l'IA, considérée comme science, d'autre part.

§ 1 : L'éthique fondamentale

I. L'éthique de la connaissance

« *... le plus redoutable de tous les problèmes : les rapports entre l'éthique et la connaissance* » écrit Jacques Monod en réponse aux critiques de son livre *Le hasard et la nécessité*[137]. Or, à la base de la connaissance se trouve le postulat d'objectivité. Selon Monod : « *Du moment où l'on pose le postulat d'objectivité comme condition nécessaire de toute vérité dans la connaissance, une distinction radicale, indispensable à la recherche de la vérité elle-même, est établie entre le domaine de l'éthique et celui de la connaissance. La connaissance en elle-même est exclusive de tout jugement de valeur (autre que « de valeur épistémologique ») tandis que l'éthique, par essence* <u>*non objective*</u>*, est à jamais exclue du champ de la connaissance* ». Ainsi, il considère que connaissance et éthique constituent deux domaines distincts, séparés ; mais il se propose de les relier, de les rendre compatibles en ajoutant : « *... poser le postulat d'objectivité comme condition de la connaissance vraie* <u>*constitue un choix éthique*</u> *...* ». On peut comprendre en effet que fonder la connaissance, notamment scientifique c'est-à-dire objective, sur ce postulat, qui est un <u>axiome indémontrable</u>, correspond effectivement à un choix, d'abord épistémologique mais aussi et surtout, éthique, car il implique une façon d'être et d'agir, une exigence qui doit

[137] Court texte de 10 pages paru d'abord sous forme de tapuscrit intitulé « Autocritique » avant d'être publié, en 1976, sous le titre « Notes de bas de pages » dans la revue Prospectives et Santé 3. 1976 ; repris par Lionel Simonneau, « L'« Autocritique » de Jacques Monod : 40 ans après ... », in Bulletin d'histoire et d'épistémologie des sciences de la vie, Volume 17, 2010/2, p. 125 à 145.

être observée tant sur le plan individuel que collectif. Et Monod conclut ainsi : *« Accepter le postulat d'objectivité, c'est donc énoncer la proposition de base d'une éthique : l'éthique de la connaissance »*[138]. Dans ce sens, l'éthique participe au fondement de la connaissance objective. En revanche, celle-ci ne peut être, à la base de l'éthique. En effet, comme Monod le montre lui-même la réalité n'est pas seulement le produit de la « nécessité », de la détermination, mais aussi du « hasard », de l'aléa. Néanmoins, la connaissance peut être utile à l'éthique. Ainsi, en matière de neurosciences : si la connaissance du cerveau ne peut être le fondement de l'éthique, elle peut permettre de comprendre, d'expliquer et finalement de juger moralement, certains comportements. A titre d'exemple : une étude publiée le 16 novembre 2020 dans la revue American Journal of Biothics-Neuroscience a montré que l'admission de la violence pour défendre des convictions était confortée par l'activation de processus cérébraux liés au circuit de la récompense et du plaisir[139].
Le questionnement relatif à l'éthique de la connaissance a des prolongements concernant la conception de la science en général et aussi de son activité « opérationnelle », la recherche.

II. L'éthique de la science

Deux conceptions sont possibles :

a) La première est classique : la science « neutre »
Selon Henri Poincaré : *« La morale et la science ont leurs domaines propres qui se touchent mais ne se pénètrent pas. L'une nous montre à quel but nous devons viser, l'autre, le but étant donné, nous fait connaître les moyens de l'atteindre. Elles ne peuvent donc jamais se contrarier puisqu'elles ne peuvent se rencontrer. Il ne peut pas y avoir*

[138] *Le hasard et la nécessité*, Paris, Seuil, 1970, p 189 – 191.
[139] Florence Rosier, « la face sombre de la moralité », in Le Monde supplément Science et médecine du 25 novembre 2020

de science immorale, pas plus qu'il ne peut y avoir de morale scientifique »[140].

Selon cette approche, la science est autonome (par rapport aux champs de l'activité humaine) impartiale (elle ne prend pas partie sur les différentes valeurs en concurrence) et neutre (elle n'entraine elle-même directement aucune conséquence sur la vie des humains).

b) La seconde est « nouvelle » : la « science impliquée ». En effet, il ressort de l'observation des faits que l'activité scientifique nécessite, souvent, la mise en œuvre de moyens nombreux et couteux, que le développement des connaissances scientifiques permet d'accroître les capacités d'action de l'Homme sur le monde et sur lui-même et accélère les mutations de l'un et de l'autre. Parallèlement, il semble que les sociétés humaines subissent l'influence de deux facteurs d'évolution radicalement opposés : l'un dans le sens de l'uniformisation du fait de la mondialisation économique, l'autre, dans le sens inverse, de la diversification voire de la différenciation les unes par rapport aux autres, du fait de la prégnance des identifiants ou marqueurs socio-culturels. Ainsi, le monde aujourd'hui est caractérisé par sa complexité et la pluralité. Dès lors, on peut considérer que la science est *de facto* « impliquée » dans l'évolution des sociétés. Ceci conduit à s'interroger sur ce que pourrait être le statut de la « science impliquée ». Léo Coutelec, maître de conférences en éthique et épistémologie à l'université Paris-Saclay, en donne la définition suivante : *« Ce que* (j'appelle) *une science impliquée est une science qui tient sa pertinence à sa tolérance au pluralisme et aux valeurs et qui abandonne son idéal d'autonomie et de neutralité sans abandonner son exigence d'impartialité »*[141]. On peut objecter que la prise en compte de la pluralité ne doit pas

[140] *La valeur de la science*, 1905, Paris Flammarion Coll « Champs » 1970 p 20

[141] *La science au pluriel*, Paris, Ed. Quae, coll « Sciences en question », 2015, p 39.

être opérée au détriment de la perspective universelle de la science.

III. L'éthique de la recherche : l'intégrité scientifique

1° Celle-ci est la contrepartie de la position d'expert que la société confère au savant, sinon au chercheur. Si la fonction de chercheur entraine la nécessité de bénéficier d'un « statut », comme nous le verrons au Chap. 3, la position d'« expert » comporte des exigences éthiques (de comportement : faire preuve d'humilité et de modestie, répondre aux questions du « public » en se mettant à sa portée) et déontologiques. Ce qui est en jeu ici, c'est la question de la fraude en matière de recherche scientifique. L'exigence d'une recherche intègre et responsable a une dimension collective et sociale.

2° Sur ce terrain, il existe des textes de référence, notamment :

- La Charte française de déontologie des métiers de la recherche signée en janvier 2015 par six établissements de recherche (Cirad, CNRS, Inrae, Inria, Inserm, IRD), la fondation de la recherche de l'Institut Curie et la Conférence des Présidents d'Université (CPU)
- Le Guide *Pratiquer une recherche intègre et responsable* établi en 2017 par le CNRS et la CPU.

3° L'application et le développement notamment de la Charte, entrent dans la mission de l'Office français de l'intégrité scientifique (OFIS), créé spécifiquement à cet effet en mars 2017, avec à sa tête, le Conseil français de l'intégrité scientifique (Cofis), composé de 12 membres nommés intuitu personae et à parité de genres qui oriente et supervise l'OFIS.

Il existe également des instances de réflexion sur ce thème, créés à l'initiative de certains établissements de recherche et d'universités, notamment le Conseil pour l'éthique de la

recherche et l'intégrité scientifique - PolETHIS - de l'Université de Paris Saclay.

§ 2 : L'éthique appliquée aux neurosciences, la neuroéthique

I. Histoire du concept et avènement de la neuroéthique

Il semblerait que le terme « neuroéthique » ait été utilisé, pour la première fois, dans le titre « Neuroéthique. Cartographier le domaine » (« Neuroethics. Mapping the Field ») des actes du congrès de la Fondation Dana, tenu en mai 2002 à San Francisco, rassemblant cent cinquante chercheurs (en neurosciences, psychiatrie, bioéthique, droit, philosophie) sur les questions éthiques, juridiques et sociales posées par les neurosciences[142].

L'éthique des neurosciences, c'est-à-dire la branche de l'éthique qui concerne les neurosciences comporte deux domaines : a) l'éthique de la pratique des neurosciences, de la science du cerveau ; b) l'étude des implications éthiques des neurosciences. La première tend à établir des principes et des règles pour guider la recherche, la seconde vise à étudier les effets des avancées de la connaissance du cerveau et de la conscience.

La neuroéthique, définie comme éthique en neurosciences doit répondre aux questions qui concernent : - le sujet lui-même c'est-à-dire l'être humain, qui est soit un sujet sain, « normal » ou un sujet-patient, « anormal », ce qui ouvre les questions de la distinction entre le normal et le pathologique et de la validité des interprétations ; - les matériels et techniques d'expérience : jusqu'où aller dans

[142] Bernard Baertschi, « Évolutions en neurosciences, les enjeux éthiques », in *Traité de bioéthique*, T IV, sous la direction d'Emmanuel et François Hirsch, Toulouse, Editions Eres, 2018, p 378

l'insertion d'implants (électrodes, nanoprothèses) dans le cerveau activés électriquement ? ; - les conséquences des opérations de recherche en neurosciences : jusqu'où aller dans leurs incidences sur la capacité de raisonnement, le comportement, la personnalité du sujet, sur leur caractère réversible ou non ?

Dans ce sens, la neuroéthique se situe dans le prolongement de la bioéthique.

II. Les principes et les règles en matière de neuroéthique

Nous n'évoquerons pas l'aspect institutionnel de celle-ci (les espaces de discussion et de proposition) afin de nous limiter au point de vue normatif, c'est-à-dire aux principes généraux et aux règles juridiques en vigueur, en France.

1° Les principes généraux

La loi bioéthique du 7 juillet 2011 a confié au Comité Consultatif National d'Ethique (CCNE) créé en 1983, une mission de veille éthique sur les progrès des neurosciences. Dans ce cadre, le CCNE a émis quelques avis en lien avec la neuroéthique parmi lesquels : - l'avis du 23 février 2012 n° 116 sur les enjeux éthiques de l'IRM fonctionnelle ; - l'avis du 12 décembre 2013 n° 122 sur (les enjeux éthiques du) recours aux techniques biomédicales en vue de « neuro-amélioration » chez la personne non-malade.

A noter que le Comité d'éthique de la défense, installé auprès du Ministère des Armées le 10 janvier 2020, a présenté, le 18 septembre de la même année, un avis « sur le soldat augmenté », rendu public le 4 décembre.

2° Les règles juridiques en vigueur

Pour ce qui concerne la France, celles-ci sont énoncées dans le cadre général des règles applicables en matière de bioéthique. Ainsi, la neuroéthique a fait l'objet d'une décision législative dans le cadre de la loi n° 2011-814 du 7 juillet 2011 relative à la bioéthique (révisant la loi du 6 août 2004) au Titre VIII unique « Neuroscience et

imagerie cérébrale ». Son article 45 unique crée deux dispositions :

- L'article 16-14 du Code civil (au Livre 1er, Titre 1er, Chapitre IV : De l'utilisation des techniques d'imagerie cérébrale) : *« Les techniques d'imagerie cérébrale ne peuvent être employées qu'à des fins médicales ou de recherche scientifique, ou dans le cadre d'expertises judiciaires. Le consentement exprès de la personne doit être recueilli par écrit préalablement à la réalisation de l'examen, après qu'elle a été dûment informée de sa nature et de sa finalité. Le consentement mentionne la finalité de l'examen. Il est révocable sans forme et à tout moment »* ;
- L'article L 1134-1 du Code de la Santé publique (Partie législative, 1ère Partie, Livre 1er, Titre III bis, Chapitre unique) : *« Un arrêté du ministre chargé de la santé définit les règles de bonnes pratiques applicables à la prescription et à la réalisation des examens d'imagerie cérébrale à des fins médicales. Ces règles tiennent compte des recommandations de la Haute Autorité de Santé »*.

En outre, le projet de loi relatif à la bioéthique actuellement en cours de discussion, comporte un Titre III : « Appuyer la diffusion des progrès scientifiques et technologiques dans le respect des principes éthiques ». Précisons qu'après son adoption, en deuxième lecture, par l'Assemblée Nationale (AN) le 31 juillet 2020, puis par le Sénat le 4 février 2021, la Commission mixte paritaire ayant constaté ne pas pouvoir parvenir à élaborer un texte commun, ledit projet doit faire l'objet d'une troisième et dernière lecture par l'AN. En formulant l'hypothèse que celle-ci reconduise le dernier texte qu'elle a voté, on a tout lieu de penser que les dispositions nouvelles en matière de neuroéthique seraient les suivantes :

- L'article 12 modifierait l'article 16-14 du Code Civil en disposant que l'utilisation des techniques d'imagerie médicale dans le cadre d'expertises judiciaires

serait possible : *« … à l'exclusion, dans ce cadre, de l'imagerie cérébrale fonctionnelle »*.

- L'article 13 prévoit l'ajout d'un article L 1151-4 au Code de la Santé Publique (Partie législative, 1ère Partie Livre 1er, Titre V, Chapitre 1er Mesures de protection) : *« Les actes, procédés, techniques, méthodes, équipements ayant pour objet de modifier l'activité cérébrale et présentant un danger grave ou une suspicion de danger grave pour la santé humaine peuvent être interdits par décret, après avis de la Haute Autorité de Santé. Toute décision de levée de l'interdiction est prise dans la même forme »*.

§ 3 : L'éthique de l'IA
Les recommandations et les normes en matière d'IA

C'est à partir de la fin des années 1990 et plus encore du milieu de la décennie 2010, à la suite de quelques évènements mettant en scène la « victoire » de systèmes d'IA face à l'intelligence humaine, que s'ouvre le débat sur les enjeux éthiques de l'IA.
Là aussi, nous nous limiterons à présenter les recommandations formulées et les (quelques) normes retenues jusqu'ici ou en cours d'élaboration, au niveau français, européen et international en vue d'une régulation de l'IA.

I. Les recommandations

On peut mentionner :
1° En France

- Deux rapports, l'un *Éthique de la recherche en robotique* de novembre 2014, l'autre *Éthique de la recherche en apprentissage machine* de juin 2017 de la Commission d'éthique de la recherche en sciences et technologies du numérique d'Allistene (CERNA), créée en 2012 par la Conférence des Directeurs des Ecoles Françaises d'Ingénieurs (CDEFI), le CEA, le CNRS, la

Conférence des Présidents d'Université (CPU), l'Inria et l'Institut Mines-Télécom ;

- le rapport, déjà évoqué, de la Commission Nationale Informatique et Libertés présenté le 15 décembre 2017 intitulé *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle* ;
- le rapport de la Mission *Donner un sens à l'IA* confiée au mathématicien et député Cédric Villani publié en mars 2018 qui comporte une Partie V titrée « Quelle éthique pour l'IA ? » (A noter également sa Partie VI « Pour une IA inclusive et diverse » sous-titrée « Mixité et diversité : agir pour l'égalité ») ;
- Deux avis du CCNE : l'un du 19 novembre 2018 sur *Numérique et Santé : quels enjeux éthiques, pour quelles régulations ?*, l'autre du 29 mai 2019 n° 130 sur les *Données massives et santé : Etat des lieux, prospective et nouvelles questions éthiques* ;
- Le premier avis du Comité national pilote d'éthique du numérique (CNPEN) - créé en décembre 2019 au sein du CCNE - du 14 mai 2020 relatif aux *enjeux d'éthique concernant des outils numériques pour le déconfinement*

2° En Europe : On a déjà mentionné le rapport *Lignes directrices en matière d'éthique – Pour une IA digne de confiance*, du Groupe des Experts Indépendants de Haut Niveau de l'Union Européenne actualisé le 08 avril 2019 qui préconise sept « exigences » et place en premier objectif le contrôle par l'humain du système IA en recourant à des mécanismes de gouvernance tels que « l'humain supervise le processus » (« human-on-the-loop ») et « l'humain aux commandes » (« human-in-command »). Ces propositions ont été reprises par la Commission Européenne dans son Livre blanc *IA Une approche européenne axée sur l'excellence et la confiance* publié le 19 février 2020.

3° A l'international : parmi les textes élaborés par de nombreuses instances de réflexion mises en place depuis

le début des années 2000 et particulièrement à compter de 2015, on peut citer :

- La Déclaration dite des « vingt-trois principes d'Asilomar » établie en janvier 2017 par le Future of Life Institute (Institut pour l'Avenir de la Vie) fondé en mars 2014 notamment par le cosmologiste Max Tegmark avec le soutien de Stephen Hawking et Elon Musk ;
- La Déclaration pour un développement responsable de l'intelligence artificielle, en dix principes, dite Déclaration de Montréal, adoptée en 2018 à l'issue d'un processus de co-construction citoyenne, lancé par l'Université de Montréal, le 3 novembre 2017 ;
- A noter, surtout, que l'Unesco dans le cadre du mandat que lui ont confié ses 193 États membres lors de la 40e session de la Conférence générale de novembre 2019, procède actuellement à l'élaboration d'une série de recommandations sur l'éthique de l'intelligence artificielle, dont le projet final sera soumis à la Conférence générale prévue fin 2021. Il s'agira du premier instrument normatif mondial sur cette question.

II. Les normes

1° En France, certaines dispositions légales, relatives notamment à la gestion des données personnelles de santé, ont été adoptées ou sont en passe de l'être :

. La Loi du 26 janvier 2016 de modernisation de notre système de santé a prévu les modalités de mise à disposition des données de santé[143] ;

. La Loi du 7 octobre 2016 pour une République numérique instaure l'obligation d'ouverture du code source des algorithmes ;

. La Loi du 20 juin 2018 relative à la protection des données personnelles introduit de nouvelles obligations

[143] Chapitre V, article 193 codifié au Titre VI, Livre IV, 1ière Partie du Code de la Santé publique

concernant les traitements algorithmiques automatisés (sans intervention humaine).
. Enfin, l'Article 11 du projet de loi relatif à la bioéthique adopté par l'Assemblée Nationale en 2ème lecture le 31 juillet 2020, prévoit d'insérer dans le Code de la Santé Publique, un article L 4001-3 relatif aux données massives de santé.
2° En Europe
L'Union européenne a adopté le Règlement n° 2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel, entré en vigueur à compter du 25 mai 2018. A noter que son article 4 dispose que *« Le traitement des données à caractère personnel devrait être conçu pour servir l'humanité. Le droit à la protection des données à caractère personnel n'est pas un droit absolu ; il doit être considéré par rapport à sa fonction dans la société et être mis en balance avec d'autres droits fondamentaux, conformément au principe de proportionnalité »*.

Chap. 3 : Politique

Introduction

On vient de constater que le champ de l'éthique appliquée aussi bien à la connaissance, à la science et à la recherche en général qu'aux neurosciences et à l'IA en particulier est encore largement à explorer.
La question de savoir si la science est ou non sous la domination de la pensée sinon de l'idéologie économique dominante, à savoir, le néo-libéralisme, doit en outre être effectivement posée. La réflexion menée ici n'a pas pour objet d'y répondre. Néanmoins, il est clair que certains éléments de fait montrent notamment que les acteurs les plus en pointe de l'économie productiviste et capitaliste mondialisée exercent une forte pression, par le biais de leurs investissements sur la science et la recherche, dans le sens de leur but premier et unique : la rentabilité du capital investi.

C’est en considération de cet élément de contexte déterminant qui n’est certes pas exclusif, puisqu’il existe dans le monde d’autres « forces » en rapport avec la science et la recherche, qu’il importe de s’interroger sur les voies et moyens permettant de garantir la mise en œuvre effective des principes et normes préalablement définis en vue de permettre le partage des connaissances, l’accès de tous aux améliorations qui en découlent et finalement, l’existence d’un monde en équilibre et en paix.
Au bout du compte, l’objectif est d’assurer la régulation des relations entre toutes les parties prenantes en soulignant que cette expression vise aussi bien la communauté humaine, que l’ensemble des êtres vivants et la Terre elle-même, le premier de tous les « biens communs ».
Il s’agit donc d’analyser les mécanismes politiques et sociaux, les structures et les modes de fonctionnement qui permettraient cette régulation.
Cette analyse doit être menée à deux niveaux : collectif, celui des citoyens, de la société civile et de l’Etat, individuel, celui du chercheur.

§ 1 : La place des sciences dans la société et le rôle des citoyens

1° Il importe de rappeler que la fonction sociale, le but premier de la recherche scientifique est la production de savoir, plus largement, que la science a une utilité sociale, dès lors qu’elle développe une vision objective, globale, systémique du monde. Ce qui fait la validité d’une découverte scientifique ce n’est pas son utilité pratique, voire économique, mais le fait qu’elle augmente le niveau des connaissances, même si elle n’engendre pas, immédiatement, une application technique. Autrement dit, le but premier de la science n’est pas l’innovation mais l’intelligence collective commune des êtres humains.
Il s’ensuit que sa production, le savoir donc, constitue un bien commun universel, qui doit pouvoir faire l’objet

d'une appropriation publique, c'est à dire qui doit pouvoir être accessible à tous les membres de la société humaine. Pour cela, il faut faire en sorte que le public ait un certain niveau de culture générale et scientifique.

2° La conception des rapports de la science et de la société du XVII° siècle tend à s'effacer devant une conception nouvelle :

a) La conception classique tient dans l'idéal de la « science pure », d'une science élitiste, « coupée » de la société. Selon Gaston Bachelard : *« La science, dans son besoin d'achèvement comme dans son principe, s'oppose absolument à l'opinion ... L'opinion pense mal ; elle ne pense pas : elle traduit des besoins en connaissances. En désignant les objets par leur utilité, elle s'interdit de les connaître ; on ne peut rien fonder sur l'opinion : il faut d'abord la détruire. Elle est le premier obstacle à surmonter ... »*[144]. Du XVII° au XIX° siècles, le citoyen a été à l'écart de la science, considéré comme simple usager de celle-ci.

b) La conception nouvelle tient dans deux idées-forces :

- La science n'est pas neutre, elle s'insère dans un contexte historique, socio-culturel avec lequel elle est en interaction. C'est ce que démontre l'anthropologue Giulia Anichini qui ayant passé deux ans dans les laboratoires de neuroimagerie à observer : *« ... le quotidien de neuroscientifiques, et en déconstruisant leurs travaux à chaque étape,* (a) *constaté que leur démarche était souvent guidée par des choix techniques ou théoriques, tantôt délibérés tantôt imposés, par des contraintes économiques ou sociales ... »*[145].
- Il faut analyser en effet le lien entre la recherche et

---

[144] *La formation scientifique* Chap I : « La notion d'obstacle épistémologique », Librairie philosophique J. Vrin, 1999, p 13.

[145] Giulia Anichini, Entretien *Sciences Humaines* n° 310 01/2019 p 36.

le(s) monde(s) non académique(s) notamment celui des entreprises et aussi celui de la société civile. La science et la recherche ont une interface avec la société et ses différents composants. La question posée est notamment celle du choix des priorités : qui les définit, qui finance les opérations de recherche, le chercheur doit-il rendre compte de ces travaux, si oui, auprès de qui, selon quelles modalités, sous quelles éventuelles sanctions ? Plus largement, quel type de lien établir entre la science et son environnement ?

Dans les sociétés démocratiques, seules prises en compte ici, ce lien ne peut être conçu sans référence aux citoyens, pris dans leur dimension collective. En premier lieu, comme on l'a dit plus haut, il est essentiel que le citoyen, l'« honnête homme d'aujourd'hui », conscient et responsable puisse accéder aux résultats de la recherche ; le chercheur, le savant a, à cet égard, le devoir de rendre compréhensibles ses découvertes auprès du public. C'est ce qu'énonce le concept de « science ouverte » dont les données et les résultats sont libres d'accès. Le Plan national pour la Science ouverte lancé par le Ministère de l'Enseignement Supérieur et de la Recherche le 4 juillet 2018 va dans ce sens. En second lieu, le lien entre la recherche et le public peut être renforcé en permettant à ce dernier de contribuer, de participer à la recherche. On parle alors de « sciences citoyennes » ou de « science participative ». Dans cette perspective, deux initiatives sont à signaler. La première est celle de l'Association Sciences Citoyennes[146], fondée par le biologiste Jacques Testart, qui a publié en octobre 2015 un « Manifeste pour une recherche scientifique responsable » dans le cadre du Forum Mondial Sciences et Démocratie. La seconde se rapporte au programme d'expériences intitulé « Les nouveaux commanditaires - science »[147] (NC-S) mis en œuvre depuis 2013, sous l'égide du philosophe Bruno

[146] https://sciencescitoyennes.org
[147] https://www.joursavenir.org

Latour, par la société coopérative et à but non lucratif « L'Atelier des jours à venir », animée notamment par les biologistes Claire Ribrault et Livio Riboli-Sasco, rédacteurs du Protocole des NC-S d'avril 2017. Ledit programme *« permet à des citoyens d'assumer la responsabilité d'une communauté de recherche »*. Il s'agit d'un *« modèle de commande en sciences* (qui) *s'inscrit dans une dynamique de développement des formes de recherche participative »*, avec l'intervention d'un médiateur dont le *« rôle est de garantir à la fois que la connaissance produite … respecte les exigences de qualité scientifique, et que la recherche réponde aux besoins des citoyens »*[148].

3° Il importe de souligner que ce nouveau type de lien entre chercheur et citoyen n'est possible que si deux conditions préalables sont réalisées. D'une part, la société considérée doit répondre aux critères de la démocratie libérale. La Démocratie, les libertés publiques et individuelles, la Justice sont des conditions d'existence de la Science. Cette exigence implique l'instauration de modalités particulières d'intervention des citoyens, de mécanismes de démocratie participative a priori (avant la prise de décision) permettant dans le cadre de débats publics, l'organisation de conférences citoyennes ou de citoyens en vue de discuter des politiques d'investissement et d'assurer le contrôle a posteriori ou le suivi de leur application. A cet égard, la recherche concernant le vivant et la conscience doit être reconnue comme entrant dans le domaine de compétences et de décision des citoyens dans les sociétés démocratiques. D'autre part, il doit exister une profonde relation de confiance mutuelle entre la communauté scientifique et la société. Cette relation suppose, là aussi, pour cette dernière, l'application de certains mécanismes institutionnels et pour la première, le respect de certaines

[148] Vinciane Despret, Livio Riboli-Sasco et autres, « La place donnée aux citoyens dans la recherche scientifique doit être repensée », in Le Monde du 25 juin 2020.

normes éthiques. Ainsi, les chercheurs doivent individuellement observer les règles de l'intégrité scientifique (notamment d'indépendance) et collectivement veiller au bon fonctionnement de leur communauté. De son côté, la société « organisée » doit garantir la liberté et l'indépendance des chercheurs. De ce point de vue, il est primordial de reconnaître le rôle des lanceurs d'alerte et d'instituer un régime juridique de protection de ceux-ci. C'est à cela qu'a répondu l'instauration de la Commission nationale Déontologie Alerte Santé publique et Environnement (DASPE) et des règles qui s'y rattachent : saisine directe par les lanceurs d'alerte (LA) et prise en charge de leur défense par le Défenseur des Droits. Par ailleurs, les citoyens doivent veiller à dénoncer et combattre les opérations de discrédit telles les « procès baillons », les licenciements-sanctions, dont peuvent être victimes certains scientifiques, ainsi que les opérations de dénigrement de certaines recherches, notamment à travers les entreprises de « fabrication de doutes ». Enfin, les citoyens, à titre individuel doivent exercer leur esprit critique et faire preuve de patience, car la recherche a besoin de temps long. Evidemment ce lien de confiance est encore plus nécessaire en temps de crise sanitaire, telle celle de la pandémie de Covid 19 que l'humanité connait depuis fin 2019. On signalera, la déclaration sur le Covid-19 de l'Unesco du 6 avril 2020, qui comporte onze principes parmi lesquels : 1° face aux incertitudes, la nécessité du *« dialogue pluridisciplinaire »*, 2° l'exigence d'une politique sanitaire fondée sur *« des connaissances et des pratiques scientifiques solides »*, 8° le fait qu'une crise sanitaire *« ne doit pas empêcher les pratiques de recherche responsables* (et) *les chercheurs* (de) *se conformer aux principes éthiques de la recherche »*[149].

[149] UNESCO, *Déclaration sur le Covid-19 : Considérations éthiques selon une perspective mondiale*, du 6 avril 2020.

§ 2 : La place et le rôle du chercheur, la nécessité d'un statut de la recherche

A titre liminaire, il importe d'indiquer qu'il aurait été utile de donner quelques précisions, concernant d'une part, l'évolution considérable de la position occupée par le chercheur qui est passée de celle de savant universaliste au XVII° siècle à celle d'enseignant-chercheur au XX°, voire de chercheur-entrepreneur au XXI°[150] ; d'autre part, la distinction importante au sein de la recherche scientifique entre la recherche fondamentale et la recherche appliquée. Sur ce dernier point, on se limitera à dire que si, dans les deux cas, les travaux entrepris visent à acquérir de nouvelles connaissances, celles-ci, en recherche fondamentale sont élaborées dans une perspective purement théorique, sans envisager leur application ou leur utilisation particulière, alors qu'en recherche appliquée leur acquisition vise à discerner les applications possibles des résultats de la recherche fondamentale.

I. Le principe de liberté (et de responsabilité) du chercheur et le principe d'indépendance et d'autonomie de la recherche

Ce qui est en jeu ici, c'est la question de savoir qui définit ce sur quoi va porter la recherche autrement dit, sur un plan général, qui va décider de son objet. Dans ce sens, il y a lieu de distinguer :

1° Le principe de liberté (et de responsabilité) du chercheur, communément désigné sous le terme de « liberté académique ». Ce principe vaut surtout en recherche fondamentale. Il signifie que l'objet de celle-ci est issu de choix qui sont effectués, de problématiques qui

[150] Yves Gingras Yves, Peter Keating, Camille Limoges « Du savant au chercheur entrepreneur » in Histoire et philosophie des sciences éd SH 2013 p 129 – 138

sont énoncées, au sein même de la communauté scientifique, par les chercheurs eux-mêmes. C'est ce principe qui constitue le seul vrai moteur de la recherche, de la créativité scientifique.

2° Le second principe, celui d'indépendance et d'autonomie de la recherche concerne particulièrement la recherche appliquée. Il résulte logiquement de l'idée que le savoir constitue un bien commun universel, qu'il doit profiter à tous les êtres humains. Dans cette perspective, le choix des buts assignés aux opérations de recherche appliquée, c'est-à-dire la mise au point des applications attendues, doivent relever, dans les sociétés démocratiques, du pouvoir politique, autrement dit de la compétence des citoyens. Ceci pose le problème, examiné ci-après, de :

II. L'organisation des structures de la recherche

Il faut veiller à ce que la recherche notamment la recherche appliquée ne soit pas sous la domination des principaux acteurs de l'économie néo-libérale, sous l'emprise des lois du marché et du profit des investisseurs privés ; domination et emprise qui se réalisent au travers du financement des laboratoires de recherche par les organismes privés. Seul le modèle classique de l'organisation publique de la recherche semble susceptible de répondre à cette exigence, dès lors qu'il remplit trois séries de conditions pratiques :

1° Sur le plan institutionnel :

On l'a vu au point 1., le mode d'organisation de la recherche doit garantir le choix des thèmes, par les chercheurs eux-mêmes, dans le domaine de la recherche fondamentale et par le pouvoir politique en matière de recherche appliquée.

En outre, le chercheur doit bénéficier d'un statut public, lui garantissant la stabilité dans l'emploi.

Enfin, le système de gestion et de pilotage des structures doit être basé sur l'égalité entre pairs et une gouvernance collective.

2° sur le plan budgétaire et financier

Le financement de la recherche doit être majoritairement un financement public, pérenne, indépendant des résultats obtenus.

3° Le mode d'organisation de la recherche doit intégrer la contrainte du temps long lequel constitue une ressource préalable à la réussite de cette activité.

Conclusion

Au terme de cette réflexion, il semble possible d'apporter quelques éléments de réponse aux questions posées à son ouverture, dans le cadre de la problématique générale des implications et des enjeux philosophiques du développement des connaissances scientifiques concernant le cerveau et la conscience.

1° Il est clair à présent que la conscience, est pleinement devenue un objet de la science. Il s'ensuit que son étude ressort plus de celle-ci que de la philosophie. De ce dernier point de vue, on peut considérer que la « prise en charge » de la conscience par la science a mis fin à la question qui relevait de l'ontologie. La conscience est désormais « naturalisée », il est difficile de soutenir voire de contester qu'elle ne soit pas autre chose qu'un élément de la Nature. Il faut en conclure que le dualisme cartésien a perdu la partie au profit du monisme naturaliste notamment d'un monisme de type spinozien qui permet de concevoir une pluralité hiérarchisée de « niveaux » de la réalité de la conscience : le niveau de base neurobiologique, des niveaux intermédiaires : psychologique, fonctionnel, évolutionniste et le niveau phénoménologique, le plus « abstrait ». Ce dernier permet d'intégrer dans le domaine scientifique des questions qui constituent encore aujourd'hui des sujets de recherche, tels l'intentionnalité (est-elle de nature mentale ou non ?) le libre arbitre (qu'en reste-t-il si l'essentiel de notre vie mentale est inconsciente ?) la conscience phénoménale et les qualia, (expériences subjectives) qui constituent un vrai problème car toujours inexpliqués.

2° C'est surtout sur le plan pratique et prospectif que le développement des connaissances scientifiques du cerveau et de la conscience et de leurs applications techniques actuelles ou potentielles, pose un problème car il est susceptible d'avoir des effets négatifs sinon dangereux pour l'être humain lui-même et le monde, quand bien

même le mythe de la Singularité technologique et la dérive de l'idéologie post- ou transhumaniste semblent pouvoir être efficacement combattus. La prise en compte de ces menaces et périls confirme la nécessité et l'urgence de prendre « à bras-le-corps » » la question de l'éthique. Il s'ensuit que la science, particulièrement la recherche (la science « en train de se faire ») et la technique ne peuvent se passer, pour répondre aux exigences de l'éthique, de faire appel à la philosophie. Il est temps de comprendre que les neurosciences et les sciences cognitives comportent un enjeu civilisationnel qui doit être pris en charge par la société et, en première instance par les citoyens, tant sur le plan théorique que pratique. Or, force est de constater qu'à ce jour, peu de choses ont été engagées.

3° L'étude de l'intersection de la connaissance et de la conscience, dans la perspective d'élaborer une théorie générale et scientifique de la conscience, ne peut être menée sans une convergence entre la science et la philosophie. Ceci conduit à évoquer à nouveau la question de l'épistémologie et particulièrement des conditions à remplir, des voies et moyens à mettre en œuvre pour traiter les problèmes spécifiques aux neurosciences et aux sciences cognitives. A cet égard, on a déjà mentionné la question de savoir comment l'être humain peut être acteur-sujet de recherche lorsque celle-ci porte sur la conscience, c'est-à-dire lorsqu'il est lui-même objet de sa recherche. Sur ce point, l'approche neurophénoménologique pourrait être explorée pour apporter des éléments de réponse. On a vu également que le problème de la nature des rapports corps – esprit est toujours sans solution. Ici, l'approche spinozienne telle qu'elle a été analysée par Henri Atlan semble la plus opérante, utile pour y faire face. Reste la question - peut-être la plus importante – liée au fait que l'extension des connaissances scientifiques du cerveau et de la conscience s'accompagne d'un recul de la notion de libre-arbitre qui vient remettre en cause l'image, la représentation de l'Homme, du Sujet comme être libre et

souverain. On peut se demander s'il n'y a pas là une sorte de barrière au progrès de la connaissance scientifique, voire de « refus de connaître », d'obstacle épistémologique comme l'entendait Gaston Bachelard lorsqu'il écrivait : « *… c'est dans l'acte même de connaître … qu'apparaissent … les causes de stagnation et même de régression … d'inertie que nous* (appelons) *obstacles épistémologiques* »[151].

4° Sur un plan plus général, ceci conduit à s'interroger sur la nécessité de repenser les rapports de l'Homme avec la Nature en considérant que celui-ci n'est finalement ni séparé, ni supérieur à celle-là, mais qu'il n'en est qu'une partie, singulière certes, mais qu'une partie seulement. Ce qui est en question ici, ce sont les relations que l'être humain doit établir avec les autres éléments constitutifs de son « milieu », appréhendé comme un Tout, en y incluant le vivant, le non-vivant et même l'artefact. Au bout du compte, ce questionnement constitue « l'arrière-plan » de la problématique de notre réflexion et de celle qui nous est posée aujourd'hui par l'émergence de l'anthropocène. Il s'ensuit que la tâche qui incombe au philosophe, au scientifique et au citoyen est d'assumer - et d'assurer - le passage à une autre conception de l'homme et du monde pour construire une humanité meilleure et plus éclairée.

[151] Gaston Bachelard, *La formation de l'esprit scientifique*, Paris, Librairie philosophique Vrin, coll « Biblio Textes Philo », 1986, p 14.

Bibliographie

Auteurs classiques

DESCARTES René

Œuvres, édition de référence, Paris, Adam et Tannery, 11 tomes, T I à XI, 1897-1909, abrégée par la suite, (entre parenthèses), en AT suivie du tome en chiffres romains et de la page en chiffres arabes

Discours de la méthode, 1637, (AT VI), Paris Editions Garnier Frères, 1960.

Meditationes de prima philosophia, 1641, *Méditations Métaphysiques*, trad. du latin par le duc de Luynes, 1647, (AT IX), Paris Flammarion coll « GF », édition revue et corrigée, 2011.

Principia Philosophiae, 1644, *Principes de la philosophie*, trad. du latin par l'abbé Claude Picot, 1647, (AT IX) ; 1ère partie, Paris, Librairie philosophique J. Vrin, coll. « Bibliothèque des Textes Philosophiques » 1989.

Les passions de l'âme, 1649, (AT XI) ; Paris Flammarion coll « GF » 1996.

Correspondance, avril 1622 – février 1650, (AT I à V) ; *Correspondance avec Elisabeth et autres lettres* Paris Flammarion coll « GF » édition mise à jour 2018.

SPINOZA Baruch

Ethique 1663 – 1674 publié en 1677 ; Œuvres complètes éd. J. van Vloten et J.P.N. Land La Haye 1882-1884 ; trad. par Charles Appuhn Paris Flammarion coll « GF », 1965.

Tractatus de Intellectus Emendatione 1661 publié en 1677 ; Œuvres complètes éd. J. van Vloten et J.P.N. Land La Haye 1882-1884 ; *Traité de la réforme de l'entendement*; trad. par Charles Appuhn, Paris, Flammarion, coll « GF », 1964.

Auteurs contemporains

ANDLER Daniel, *La silhouette de l'humain*, Paris, Gallimard, coll « *nrf* essais », 2015.

ANICHINI Giulia, Entretien in *Sciences Humaines*, n° 310, 01/2019 p 36.

ATLAN Henri,

Entre le cristal et la fumée, Paris, Seuil, 1979.

Cours de philosophie biologique et cognitive, Paris, Odile

Jacob, 2018.

BACHELARD Gaston, *La formation de l'esprit scientifique*, Paris, Librairie philosophique J. Vrin, coll « Bibliothèque des Textes Philosophiques », 1986 et 1999.

BAERTSCHI Bernard, « Évolutions en neurosciences, les enjeux éthiques », in *Traité de bioéthique*, T IV, sous la direction d'Emmanuel et François Hirsch, Toulouse, Editions Eres, 2018.

BERNHEIN Aude et VINCENT Flora, *L'intelligence artificielle, pas sans elles*, Paris, Belin / Humensis, coll « Egale à Egal », 2019.

BERTHIER Thierry, Entretien in *Voyage au cœur de l'IA* Libé – France Inter 12/2017-02/2018 p 105.

CANGUILHEM Georges, *La connaissance de la vie*, Paris, Librairie philosophique J. Vrin, coll « Biblio des Textes Philosophiques », 1993.

CHANGEUX Jean-Pierre, *L'Homme de vérité*, Paris, Odile Jacob, 2002.

CHNEIWESS Hervé, Entretien in *La Recherche* n° 557, 03/2020.

CNIL (Commission Nationale Informatique et Liberté) Rapport *Comment permettre à l'homme de garder la main ?* Glossaire, 12/2017, p 75.

COUTELEC Léo, *La science au pluriel*, Paris, Ed. Quae, coll « Sciences en question », 2015.

DAMASIO Antonio *L'Erreur de Descartes*, trad. de l'anglais (USA) par Marcel Blanc, Paris, Odile Jacob, coll « Sciences », 1995.

DEHAENE Stanislas, *Le Code de la conscience*, Paris, Odile Jacob, coll « Sciences », 2014.

DESPRET Vinciane, RIBOLI-SASCO et autres, « La place donnée aux citoyens dans la recherche scientifique doit être repensée », in Le Monde du 25 juin 2020.

EDELMAN Gerald M., Entretien, in *Les Dossiers de la Recherche* n° 30 02/2008

FODOR Jerry Alan, *The Mind Doesn't Work that Way*, Cambridge MIT Press, 2000 ; *L'esprit, ça ne marche pas comme ça*, trad. de l'anglais (USA) par Claudine Tiercelin, Paris, Odile Jacob, 2003.

GALLESE Vittorio, Entretien in *Philomagazine*, n° 135, 12/2019 – 01/2020.

GANASCIA Jean-Gabriel, Entretien, in Le Monde du 11 mars 2017.

GEIHtN (Groupe d'experts indépendants de haut niveau) de l'UE, Document « Définition de l'IA : principales capacités et disciplines. », avril 2019.

HUSSERL Edmund,

Recherches logiques, t1, *Prolégomènes à la logique pure*, Introduction, trad. de l'allemand par Hubert Elie, Arion L. Kelkel, René Schérer, Paris, Puf, coll « Epiméthée », 1990.

La philosophie comme science rigoureuse, trad. de l'allemand par Marc de Launay, Paris, Puf, coll « Epiméthée », 2003.

Méditations cartésiennes, 2ème, trad. de l'allemand par Marc de Launay, Paris, Puf, coll « Epiméthée », 1994.

KAMBOUCHNER Denis, « L'erreur de Damasio », in *La théorie spinoziste des rapports corps/esprit et ses usages actuels*, sous la dir. de JACQUET Chantal, SEVERAC Pascal et SUHAMY Ariel, Paris, Hermann éd, coll « Philo », 2009.

KIM Jaegwon, *Mind in a physical world*, Cambridge, (Mass.), MIT Press, 1998 ; *L'esprit dans un monde physique*, traduit de l'américain par François Athané et Edouard Guinet, Paris, Ithaque Editions, 2014.

LE CUN Yann,

Entretien, in *Voyage au cœur de l'IA*, co-édition Libération - France Inter 12/2017-02/2018.

Quand la machine apprend, Paris, Odile Jacob, 2019

LE RU Véronique, *Oser penser avec E. du Châtelet, D'Alembert, Poincaré .... La distinction entre croire et savoir*, Paris, Editions Matériologiques, coll « Essais », 2020.

LEVY Lia, « Les rapports entre l'esprit et le corps dans la proposition 23 II° partie de l'Ethique », in *La théorie spinoziste des rapports corps/esprit*, sous la dir. de Chantal Jacquet, Pascal Severac et Ariel Suhamy, Paris, Hermann éd, coll « Philo », 2009.

LOUKIL Ridha, « le supercalculateur le plus puissant au monde est japonais », in *L'Usine nouvelle*, du 23 juin 2020.

MERLEAU-PONTY Maurice,

La structure du comportement, Paris, PUF, coll « Bibliothèque de philosophie contemporaine », 7° édition, 1972.

Le visible et l'invisible, Paris, Gallimard, coll « Bibliothèque des idées », 1964.

MONOD Jacques,

Le hasard et la nécessité, Paris, Seuil, 1970.

« Autocritique » retitré « Notes de bas de pages » in *Prospectives et Santé 3*, 1976 ; repris par Lionel Simonneau, « L' « Autocritique » de Jacques Monod : 40 ans après ... », in Bulletin d'histoire et d'épistémologie des sciences de la vie Volume 17, 2010/2.

NACCACHE Lionel, « Observer la conscience », in *Pour la Science*, n° 500, 06/2019.

POINCARE Henri, *La valeur de la science*, 1905, Paris, Flammarion, coll « Champs », 1970.

PUTNAM Hilary

« The Nature of mental States », publié en 1967, « La nature des états mentaux », traduit de l'anglais (USA) par Jean-Michel Roy, revue et corrigée par Dominique Boucher, in *Philosophie de l'esprit*, T1, Paris, Librairie philosophique J. Vrin, coll. « textes clés », 2002.

« Aller et retour au pays du fonctionnalisme », traduit de l'américain par Mireille Duchastelle-Cabanes, in *Philosophie*, n° 71 2001/4.

ROBIN Mari-Monique, *La fabrique des pandémies* La Découverte coll « Cahiers libres » 2021.

RODER Stéphane, « N'ayons pas peur de l'IA au travail », in Le Monde du 1er février, 2019.

ROSIER Florence, « la face sombre de la moralité », in Le Monde supplément Science et médecine, du 25 novembre 2020.

SEARLE John R.,

La redécouverte de l'esprit, traduit de l'anglais (USA) par Claudine Tiercelin, Paris, Gallimard, coll « *nrf* Essais », 1995.

Le Mystère de la conscience, traduit de l'américain par Claudine Tiercelin, Paris, Odile Jacob, coll « Philosophie », 1999.

Entretien, in Philosophie magazine, n° 24, 11/2008.

TUAL Morgane, « A peine lancée, une IA de Microsoft dérape sur Twitter », in Le Monde du 24 mars 2016.

UNESCO, *Déclaration sur le Covid-19 : Considérations éthiques selon une perspective mondiale*, du 6 avril 2020.

VARELA Francisco,

L'inscription corporelle de l'esprit, traduit de l'anglais par Véronique Havelange, Paris, Seuil coll » La couleur des idées », 1993.

« Essai introductif - Combler le déficit : Introduction à la naturalisation de la phénoménologie » in *Naturaliser la phénoménologie*, sous la dir. de Bernard Pachoud, Jean Petitot, Jean-Michel Roy et Francisco Varela Paris CNRS Editions 2002.

Entretien avec Réda Benkirane, in *La Complexité, vertiges et promesses*, Paris, Editions le Pommier, 2002.

WEINBERG Achille, « L'évolution conjointe du cerveau et de la culture », in *Sciences Humaines*, n° 309, 12/2018.

WIENER Norbert, *Cybernetics, or Control and Communication in the Animal and the Machine*, Cambridge (Mass.), MIT Press, 1° éd. 1948, *La cybernétique : Information et régulation dans le vivant et la machine*, traduit de l'anglais par Ronan Le Roux, Robert Vallée et Nicole Vallée-Lévy, « Introduction », Paris, Seuil, coll. « Sources du savoir », 2014.

Table des matières

Remerciements (p 2)

Introduction (p 3)

Ière PARTIE : La mise en place de la problématique de la nature de la conscience comme objet de science au XVII° siècle

Introduction (p 13)

Chap 1 : René Descartes

Introduction (p 14)

§ 1 (p 15) : Le dualisme ontologique
Théorie de la dualité des substances de l'âme et du corps

§ 2 (p 18) : L'union substantielle de l'âme et du corps et leurs interactions

Chap 2 : Baruch Spinoza

Introduction (p 23)

§ 1 (p 24) : Le monisme ontologique
Théorie de l'unité de la substance

§ 2 (p 26) : L'identité individuelle de l'homme

Chap 3 : Descartes, Spinoza et les neurosciences

§ 1 (p 34) : De Descartes à Spinoza, des divergences aux convergences

§ 2 (p 36) : Les places respectives de Descartes et Spinoza dans les neurosciences

IIème PARTIE : Le renouvellement de la problématique de la nature de la conscience comme objet de science au XX° siècle

Introduction (p 41)

Chap 1 : le fonctionnalisme et ses dérivés

§ 1 (p 45) : Hilary Putnam et le fonctionnalisme

§ 2 (p 47) : Du computationnalisme au connexionnisme

§ 3 (p 50) : Le cognitivisme et l'intelligence artificielle

Chap 2 : Le monisme en force sous le label du naturalisme

§ 1 (p 53) : John R. Searle et le naturalisme biologique

§ 2 (p 57) : Jaegwon Kim et le physicalisme

Chap 3 : La recherche d'une conjonction entre la phénoménologie, les neurosciences et les sciences cognitives

§ 1 (p 63) : Edmund Husserl

§2 (p 65) : Maurice Merleau-Ponty

§ 3 (p 68) : Francisco Varela

IIIème PARTIE : La problématique des limites à apporter aux applications des connaissances en neurosciences et sciences cognitives

Introduction (p 72)

Chap 1 : l'intelligence artificielle, ses enjeux actuels et futurs

Introduction (p 73)

§ 1 (p 75) : Les facteurs du développement actuel de l'IA

I. (p 76) Les facteurs favorables
II. (p 77) Les facteurs défavorables

§ 2 (p 80) : La nature des effets et impacts de l'IA, leurs dangers et risques réels et potentiels

I. (p 80) Les mutations affectant la place et l'autonomie de l'individu dans la société
II. (p 81) Les atteintes aux Droits de l'Homme, à la vie privée et aux libertés publiques
III. (p 82) La mise en en question de la nature onto-anthropologique de l'Homme et de son évolution

Chap 2 : L'Ethique

Introduction (p 84)

§ 1 (p 85) : L'éthique fondamentale

I. (p 85) L'éthique de la connaissance
II. (p 86) L'éthique de la science

III. (p 88) L'éthique de la recherche : l'intégrité scientifique

§ 2 (p 89) : L'éthique appliquée aux neurosciences, la neuroéthique

I. (p 89) Histoire du concept et avènement de la neuroéthique
II. (p 90) Les principes et les règles en matière de neuroéthique

§ 3 (p 92) : L'éthique de l'IA

I. (p 92) Les recommandations
II. (p 94) Les normes

Chap 3 : Politique

Introduction (p 95)

§ 1 (p 96) : La place des sciences dans la société et le rôle du citoyen

§ 2 (p 101) : La place et le rôle du chercheur, la nécessité d'un statut de la recherche

I. (p 101) Le principe de liberté (et de responsabilité) du chercheur et le principe d'indépendance et d'autonomie de la recherche
II. (p 102) L'organisation des structures de la recherche

Conclusion (p 104)

Bibliographie (p 107)

Table des matières (p 112)