

**Modèles du rapport neuro-fonctionnel, ou comment la cognition
serait relative au cerveau :
du dualisme à l'intrication quantique.**

Bernard Claverie¹

Résumé :

À partir de l'expérience clinique neuropsychologique de l'auteur, on présente différents modèles explicatifs des relations potentielles entre le cerveau et les activités cognitives au sens large du terme. Après avoir discuté des positions qui animent les sciences cognitives et l'intelligence artificielle, on propose un inventaire critique des principales positions du champ des « mind/brain theories ».

Mots-clefs :

cerveau/pensée, cognition, neuropsychologie, pensée/cerveau, sciences cognitives,

Abstract :

Based on the author's clinical neuropsychological experience, one presents various explanatory models of the potential relationships between the brain and the cognitive activities in the broad sense of the term. After discussing the positions that drive cognitive science and artificial intelligence, one offers a critical inventory of the main positions in the field of "mind / brain theories".

Keywords :

brain/mind, cognition, cognitives sciences, mind/brain, neuropsychology.

1 - ENSC - HEAL et IMS-UMR5218 (CNRS/U.Bordeaux/Bordeaux INP)

Charlotte est une jolie jeune femme : c'est au moins notre avis. Nous avons présenté quelques photographies d'elle lors d'une conférence et ce point de vue était partagé, bien qu'un certain malaise semblât planer sur l'auditoire.

Brune aux yeux noisette, elle n'en voit pourtant plus qu'un seul, le droit, celui qu'elle s'applique le matin à maquiller, devant son mi-miroir, tout ceci avec beaucoup d'attention, cette attention qui lui fait pourtant défaut et lui fait négliger toute la partie gauche de son corps. Debout, alors qu'elle avance avec difficulté, elle se heurte aux objets situés à sa droite, devenus autant d'obstacles supplémentaires à sa mobilité. Dès le réveil, la journée est une suite d'épreuves : se lever du seul côté du lit qui existe, l'autre n'étant même pas du néant : il n'est même pas rien, il n'est pas. Son jeune garçon lui beurre ses tartines, elle ne les voit que lorsqu'il les pose à sa vue, sur le bon côté, le seul côté de l'assiette. Elle remercie sa fille qui l'aide en serrant sa main autour de l'anse de son bol et lui permet de porter le café à sa bouche. Encore faut-il que tout soit du bon côté pour qu'il existe.

Toute sa vie est réduite à un demi-monde, sa droite à elle, un héli-espace qui varie en fonction du regard et de l'orientation de la tête, avec une persistance mémorielle très faible qui atteint la conscience du monde, de la télé qu'elle ne regarde plus, de son repas dont elle ne mange que des moitiés d'assiette. Il faut l'aider pour s'habiller, pour la toilette, pour la vie entière...

Les notes qu'elle tente d'écrire sont illisibles, en lignes torves entassées en marge de la feuille. Elle a perdu le goût de lire ; à quoi bon lire la moitié des histoires, c'est épuisant et surtout sans grand intérêt. Le plus étrange est que Charlotte ne s'en aperçoit pas. Inconsciente de son trouble et de la disparition mobile de la moitié de l'espace, elle reste chez elle faute de plaisir à sortir dans un hémimonde, malgré les sollicitations de ses enfants. Elle se renferme sur elle-même, et sans sollicitations dépérirait dans le seul hémicorps qui lui reste.

Cerveau et cognition.

Chaque année, vingt-cinq à trente mille personnes sont en France frappées d'une pathologie neuropsychologique. Le plus souvent, ce sont des victimes d'un accident vasculaire cérébral : un défaut d'irrigation au sein du cerveau. Dans d'autres cas, le trouble fait suite à un traumatisme, le développement d'une tumeur ou une intervention chirurgicale dans le cas du traitement de maladies neurologiques telles que les épilepsies pharmaco-résistantes.

Certaines fois, il s'agit de l'atteinte d'un seul hémisphère, entraînant des hémiplégies mais également des déficits de perception et d'attention de ce qui est latéralisé. Dans certains cas, comme Charlotte, le malade devient hémignégligent : il perd la conscience de la moitié de l'espace situé face à lui, de

ce qui s'y trouve ou s'y passe et, plus paradoxal encore, de la moitié de son corps. Près de deux tiers de ces patients recouvrent sous trois à quatre mois une perception correcte ou pour le moins suffisante à vivre dans le monde de tous les jours. Pour les autres, la négligence peut se manifester plus longtemps et rendre l'existence quasiment impossible à vivre.

D'autres pathologies concernent la mémoire : le patient ne peut exploiter certaines informations, à long terme ou très rapidement, celles acquises il y a bien longtemps et sur lesquelles repose une bonne part de la personnalité du sujet, ou immédiatement dans une sorte d'incapacité à retenir de nouvelles choses. L'étude des amnésies ouvre un immense champ d'observation de cas d'atteinte, voire de dissolution de la pensée : ne traite-t-on pas une information parce qu'on ne la perçoit pas, parce qu'on n'y porte pas attention, parce qu'on n'en a pas conscience, ou parce qu'on ne s'en souvient plus faute de l'avoir enregistrée, ou parce que sa trace est perdue ?

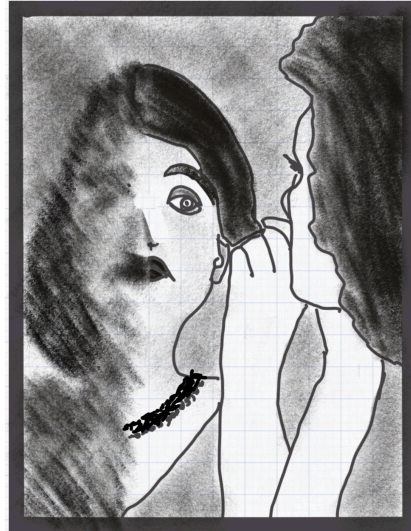


Figure n°1 : L'héminégligence et la conscience de la moitié du Monde.

Ces pathologies que l'on qualifie de neuropsychologiques illustrent deux principes fondamentaux de la vie du cerveau : les atteintes de cet organe provoquent toujours un trouble cognitif, et cela même si le patient ne s'en aperçoit pas ; le cerveau peut parfois récupérer et le sujet progresse alors souvent partiellement dans la restauration de ce qu'il a perdu.

Ce domaine des relations étranges entretenues entre le cerveau et la pensée est l'une des questions centrales de la philosophie de l'esprit connue sous le titre de « mind/brain paradox ». Néanmoins, et sans entrer dans le domaine des philosophes, c'est aussi l'un des problèmes majeurs auxquels sont confrontés les cliniciens (voir par exemple Sacks, 1992). Quelle conséquence cognitive entraîne l'altération du cerveau, qu'elle soit acquise comme dans le cas d'un accident vasculaire cérébral, d'un choc, d'une intoxication..., ou développementale par organisation différente de certaines parties du cerveau lors de son évolution, de la vie fœtale à l'involution démentielle du grand âge ? Quelles sont les perspectives de rééducation ou de restauration de la fonction perdue ou altérée, en agissant sur le comportement ou directement sur le cerveau ? Tels sont les problèmes de l'éducation

adaptée, de la rééducation et de la prise en charge des patients cérébrolésés, ou sous l'effet de substances neuroactives ou de la neurochirurgie qui, pour traiter ceci du cerveau vont perturber cela du comportement et de la pensée.

L'approche objective des fonctionnalités neurocognitives.

Une des avancées majeures de la neuropsychologie est due au progrès impressionnant des techniques d'imagerie médicale fonctionnelle. On ne minimise évidemment pas l'intérêt de la clinique neurologique, avec l'étude des comportements altérés par les atteintes cérébrales repérées autrefois à l'autopsie et aujourd'hui au scanner, à l'IRM ou par artériographie, ni celui de la psychophysologie qui met en corrélation le déroulement des processus cognitifs avec celui de leurs concomitants électro- ou magnéto-encéphalographiques. Mais l'avènement des techniques d'imagerie neurofonctionnelle a permis de repérer précisément des zones cérébrales impliquées dans telles ou telles activités mentales. Le but n'est pas de réduire une fonction à une zone cérébrale, bien que certains s'égarent encore dans cette relation simpliste, mais de montrer la nécessaire implication ou l'habitude de mobiliser certains ensembles cérébraux dans l'expression de certaines cognitions.

L'accumulation des études donne aujourd'hui une idée raisonnable de certaines correspondances entre activité cérébrale et fonctions mentales. On

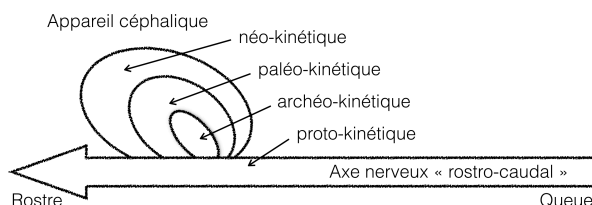


Figure n°2 : Les niveaux architectoniques des vertébrés.

peut citer les études qui explorent à la fois des latéralisations et dominances hémisphériques, des localisations tant corticales que sous-corticales, et des niveaux d'implication avec la mise en relation successive de certains

étages de l'encéphale. On peut également décrire certaines relations entre des aires fonctionnelles et leur composition cyto-architectonique de ce que l'on définit alors comme territoires. Cette dimension non invasive de l'étude cérébro-fonctionnelle vient de prendre une nouvelle dimension grâce aux récentes techniques d'étude des liaisons nerveuses utilisées entre les territoires, les connectomes, et qui s'intéresse moins à la circonscription des localisations qu'à la mise en évidence des relations qu'elles entretiennent entre elles : une nouvelle systémique du cerveau impliqué dans les actes de pensée est née.

Chez le sujet atteint d'une blessure du cerveau, l'étude imagée des pathologies et des erreurs cognitives permet une approche objectiviste de la correspondance neurofonctionnelle. On peut alors en tester la validité des modèles qui sont soit réfutés soit proposés comme explications potentielles. Malgré leur grande diversité, elles ne sont pas exclusives les unes des autres. Elles visent à une compréhension du trouble spécifique, et celle des morceaux de pensée qui restent préservés. Le but en est d'abord clinique, la rééducation et le traitement sont alors contraints par les modèles explicatifs, sans que l'on ne puisse encore approcher objectivement la nature de la cognition, de la pensée ou de la conscience considérées comme entités globales, aussi mystérieuses que mal définies. Pourtant, il faut comprendre, expliquer, et pour cela nous disposons de nombreux modèles et de bien peu de certitudes.

Les modèles neuro-cognitifs.

Le premier constat d'expérience est, pour le clinicien du cerveau lésé, celle des évidentes complémentarités structuro-fonctionnelles : les fonctions sont liées aux structures neuronales comme celles-là permettent d'exprimer les premières. C'est à force de sélection naturelle que les êtres vivants sont devenus pluricellulaires, et que certains d'entre eux, les animaux, ont développé à force de perfectionnement un axe de symétrie et peu à peu un système nerveux central. En même temps, ceux qui ne se déplaçaient pas disparaissaient peu à peu des milieux écologiques favorables à l'émergence de la mobilité.

Située près de la bouche, à force de perfectionnement, l'ampoule cérébrale s'est mise à commander le corps, du rostre à la queue, dans une logique propice à la meilleure adaptation : la fuite face à la menace, l'agression en présence de la proie. Elle a permis de développer la hiérarchie au sein des groupes qui se constituaient. Parallèlement, la douleur permettait l'évitement et naissaient, avec l'apparition de zones spécifiques, les capacités d'apprentissage, de mémoire et d'anticipation.

La motivation colonisait de nouveaux territoires cérébraux et l'homéostasie s'harmonisait d'allostasie pour la survie des individus et celle de l'espèce. Le cerveau s'est d'abord organisé pour la survie et le perfectionnement des



Figure n°3 : Homonculus moteur : représentation des proportions de territoire cortical dévolu à la commande des certaines parties du corps humain.

espèces. Et la vie de l'espèce nécessite bien entendu celle des individus qui la composent.

Alimentation et équilibre dipsique, territorialité, sexualité et reproduction, attachement et émotion, peur et colère, parmi d'autres fonctions nécessaires, ont spécialisé des réseaux et des structures dites archéo-kinétiques. Elles se sont liées entre elles, autour de systèmes de plaisir/déplaisir et de motivation supportés par les étages cérébraux paléo-kinétiques. L'amour, le contact, le bien-être sont apparus grâce à un système limbique dont on connaît l'importance dans l'organisation de la mémoire et la régulation des émotions. Puis sont apparues des structures néo-kinétiques, dites supérieures, et dont la vocation a été de maîtriser l'expression des premières, de s'y substituer partiellement dans des équilibres qu'il nous reste à comprendre. Une pensée de haut niveau qu'on désigne aujourd'hui par le terme de cognition a pu bénéficier de ces niveaux les plus récents et sophistiqués tout en s'appuyant sur les premiers. C'est la perfection progressive de ces structures en même temps que l'affinement des comportements qu'elles permettent d'exprimer qui ont amené lentement les prosimiens à devenir singes, ceux-là à devenir grands singes, puis hominidés il y a quelques milliards d'années, enfin hommes et tout récemment, depuis quelques centaines de milliers d'années, homo sapiens sapiens, l'homme qui sait qu'il sait : un être métacognitif.

Les modèles développementaux du cerveau prennent en compte différents étages dont la mise en relation successive a structuré la cognition et les comportements, comme les groupes et les prototypes sociaux. Le modèle le plus connu est celui du « cerveau tri-unique » que théorisa McLean en 1967 (Claverie, 2018). Bien que contesté pour son simplisme radical, il a donné à l'époque un cadre général de compréhension de l'indépendance des fonctions neuropsychologiques et a fourni d'autres modèles qui s'accordaient avec la clinique humaine. On citera ceux de la diaschisie² de Von Monakov en neuropsychologie et de la substitution hiérarchique³ de Jackson en épiléptologie (Séron, 1979), ceux de la dualité hémisphère droit et hémisphère gauche après les travaux de Sperry (Root-Bernstein, 2005), et les modèles localisationnistes liés aux pathologies lobaires avec l'édition de véritables cartographies cognitives, nouvelle forme d'une phrénologie instrumentée.

L'homonculus de Penfield (1959) est une représentation de la quantité de territoire cortical, des aires frontales et pariétales jouxtant la scissure de Rolando, consacrées à la commande ou à la sensation de l'hémicorps contralatéral. Il montre surtout combien l'équilibre est donné entre quatre grands

2 - La diaschisie postule qu'une atteinte d'un territoire cérébral a des conséquences sur les territoires connectés à ce premier, jetant les bases d'une clinique connectique cérébrale, notamment en termes de récupération complexe et de réadaptation.

3 - Le modèle postule l'organisation progressive de fonctions de bases non substituables qui sont associées à plusieurs niveaux pour contribuer à une fonction globale distribuée.

domaines de la communication et de l'exploration du monde : la voix et le regard pour la communication proche, la face et la tête pour le faciès, la main pour le geste, puis le reste du corps pour la posture. Ici encore on observe une correspondance entre des quantités de territoires et des données comportementales utiles à la vie et à la survie.

L'évolution du cortex cérébral⁴, cette fine couche de neurones qui enveloppe le cerveau pour lui remettre la représentation du monde, le langage et les pensées les plus sophistiquées, n'a laissé que peu de traces à travers les âges.

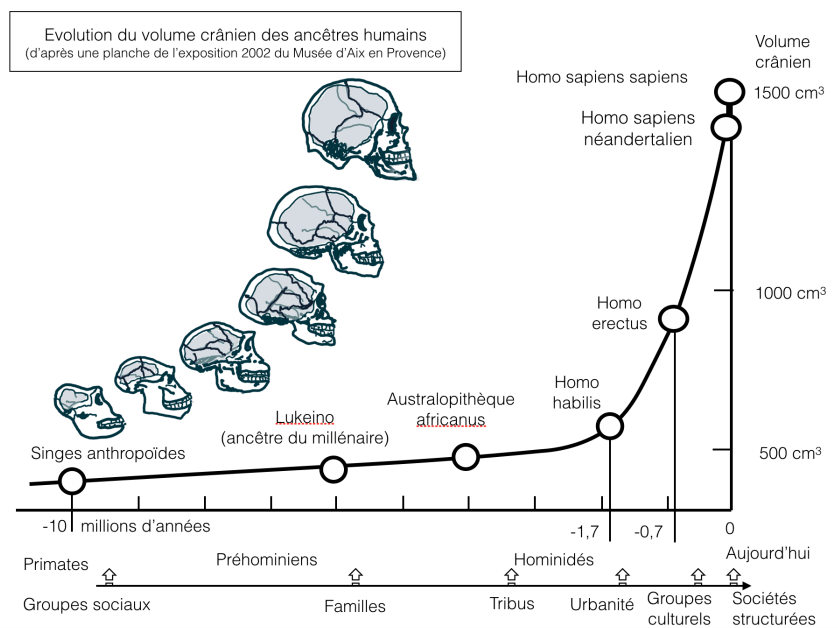


Figure n°4 : histoire de l'évolution du volume du cerveau.

On ne connaît que celles du cortex imprimées sur les fossiles que le genre homo a laissé depuis les temps récents à l'échelle géologique. On se fixe donc au volume du crâne qui abritait le cerveau et que le cortex semble avoir poussé tout en augmentant la surface de ses plissements. Notamment, le cortex préfrontal dont on connaît l'importance dans l'idéation, les fonctions exécutives, le projet et l'imagination, semble avoir pris de plus en plus d'im-

⁴ - Cortex signifie en Latin enveloppe, écorce, et par extension couverture ou manteau. En anatomie, il s'agit de la partie externe de certains organes. Pour le cerveau, ce manteau est constitué de substance grise, de systèmes neuronaux complexes (colonnes, couches, etc.) qui supportent la mise en oeuvre des grandes fonctions motrices, perceptives et cognitives du plus haut niveau évolutif.

portance à mesure que les comportements s'affinaient et que la culture permettait l'évolution sociale et la complexification des groupes et leurs échanges. C'est principalement ce cortex qui est l'objet des études neuro-cognitives, et les techniques nouvelles d'imagerie fonctionnelle s'attachent à en étudier les détails du fonctionnement en relation avec des éléments de performance comportementale ou cognitive.

Behaviorisme et cybernétique.

Le structuralisme a été un paradigme majeur de la pensée de la fin du XIXe et du XXe siècles dans sa première moitié. Cette époque féconde a permis de mettre en évidence les structures organisationnelles sous-jacentes des systèmes et notamment des productions humaines, et la psychologie s'est attachée à développer la méthode expérimentale et les statistiques. C'est principalement grâce à la mathématisation de l'observation de relations entre stimulation et réponse que des chercheurs ont supposé l'existence de composantes élémentaires de la pensée et leur structuration en systèmes complexes alors seulement accessibles par l'introspection.

Cette riche démarche a été balayée par le pragmatisme anglo-saxon qui s'affranchit de l'impossibilité de vérifier expérimentalement l'introspection, en proposant une approche purement comportementaliste : le behaviorisme.

Le cerveau est, dans cette théorie, considéré comme une machine fermée, inaccessible, en boîte noire, et la seule connaissance qu'on peut en avoir semble être de le considérer comme une sorte de processeur modifiant des informations vectorisées dont on s'attache à fournir une description matricielle pour en donner des modifications

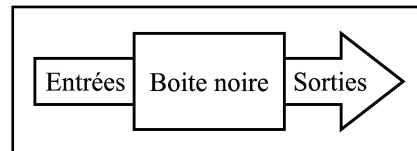


Figure n°5 : Le modèle en « boîte noire » du behaviorisme..

purement algorithmiques. La pensée est alors considérée comme un simple calcul et le psychologue ou le neurologue cherchent à en découvrir les lois de relation entre les entrées et les sorties. La preuve est faite qu'il faut désormais maîtriser la connaissance de l'environnement et celle du comportement en réponse aux modifications du premier.

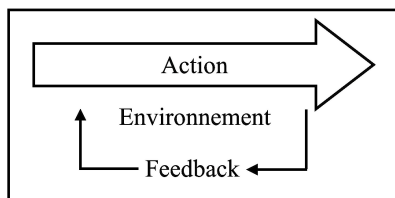


Figure n°6 : La régulation cybernétique par bouclage externe.

De remarquables succès, influencés par la théorie du signal, ont été obtenus dans le domaine de l'apprentissage associatif, chez l'animal, en clinique rééducative, en éducation chez l'enfant, etc. autour du concept central de

conditionnement opérant. Pour les behavioristes, la pensée est gérée par le cerveau autour des récompenses et punitions naturelles ou éducatives qu'impose au sujet l'environnement. Les états mentaux étant expérimentalement inaccessibles, on va cependant essayer de repousser de chaque côté les limites de la boîte noire en s'attachant notamment aux concomitants des processus attentionnels, de vigilance et de préparation motrice grâce à l'outil électroencéphalographique qui permettra de postuler quelques organisations sous-jacentes dans un traitement complexe de l'information utilisant en parallèle plusieurs canaux organisés de manière séquentielle.

Ces travaux de la psychophysologie cérébrale ont rencontré ceux de la cybernétique, avec les fameuses conférences Macy⁵. Intervient alors le nécessaire recours à la notion de feedback qui provoque l'adaptation continue entre stimulus et réponse, et qui apparaît également active dans des étapes du traitement interne.

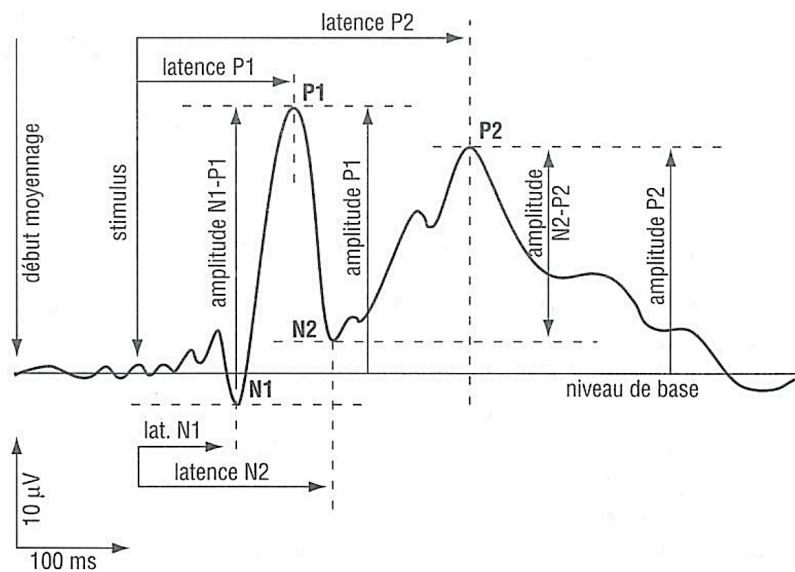


Figure n°7 : Chronométrage des opérations mentales à l'aide de potentiels électroencéphalographiques évoqués (in Claverie, 1999).

La chronométrage des opérations mentales (Claverie, 1999) montre que les processus de la pensée prennent du temps, ne se mobilisent pas n'importe

5 - Les conférences organisées à New York par la fondation Macy réunirent de 1942 à 1953 des mathématiciens, logiciens, anthropologues, psychologues et économistes fondateurs du courant cybernétique, et précurseurs des sciences de la complexité, des sciences cognitives et des sciences de l'information.

comment, et que l'esprit n'est pas un simple réservoir mais dispose de compartiments et mobilise des dispositifs emboîtés dynamiques. La cybernétique se distingue du béhaviorisme en prenant en compte des caractéristiques sur le cerveau, et s'affranchit de l'arc réflexe en postulant l'existence d'étages cognitifs plus propices à la description de la complexité de certains comportements humains. Elle a donné de nombreux modèles comparatifs des phénomènes attentionnels, avec des théories de filtres multiples et de la mémoire conçue comme état dynamique de structures internes acquises par apprentissage. La motivation est alors imaginée comme activation d'ensembles de programmes prédisposés à un déclenchement de comportements orientés qui s'adaptent en s'ajustant à l'environnement (Claverie & Le Blanc, 2016).

La révolution cognitive : cognitivisme versus connexionnisme.

Ce retour vers l'hypothèse de structures fonctionnelles a rencontré les travaux de la linguistique chomskienne⁶ et ceux portant sur la catégorisation. Ils postulaient l'existence de répertoires cognitifs internes, notamment en ce qui concerne le langage, et de stratégies mentales différentes en fonction de caractéristiques de la tâche et de stratégies mentales dans la résolution de problèmes. L'étude de la mémoire et de ses limites (7 items plus ou moins deux ; Miller 1956) a montré que le cerveau a tendance à associer des informations en grappes (chunks), dans une forme de traitement logique guidé à la fois par la motivation à la performance et par les caractéristiques de la chose à traiter. Le cerveau devient donc un organe de « traitement de l'information » composé de multiples sous-structures ayant chacune une spécificité organisée selon des règles d'une grammaire universelle.

Ce « langage de la pensée » va renouer avec le dualisme cartésien et la conviction turingienne⁷ d'une cognition calculatoire. La métaphore fodorienne⁸ des années 1970, à la base d'une démarche qui se déclare « science du mental », va ainsi s'ancrer sur l'analogie avec l'ordinateur : la pensée est

6 - La linguistique générative, développée par Noam Chomsky, est une théorie qui s'oppose à la fois à la linguistique structurale de Saussure et au béhaviorisme. Les deux idées principales résident dans la séparation de la compétence (la capacité de langage) et la performance (l'acte de parole) selon laquelle le langage est un processus cognitif résultant de la faculté de langage, et dans l'interaction de modules linguistiques qui permet la production et à compréhension globales.

7 - La calculabilité au sens de Alan Turing serait une des conditions de la pensée. Toute pensée est le résultat d'un calcul opérable sur une machine universelle de Turing (voir Claverie, 2005).

8 - Dans la suite des idées de Chomsky et de Turing, la cognition est, selon Jerry Fodor, fondée sur l'articulation de modules cognitifs encapsulés et hautement spécialisés qui produisent une computation sur l'information mentale. Fodor est le fondateur de la théorie computationnelle de l'esprit, à l'origine du cognitivisme. Pour Fodor, les individus fonctionnent avec des croyances (beliefs) fondées sur des représentations, et elles ont un pouvoir causal sur le comportement.

au cerveau ce que le programme est à la machine. Celui-là est alors un énoncé, une suite d'instructions et de symboles organisés entre eux par des règles logiques, et son implémentation produit un calcul : « thinking is computing ». Cette théorie est alors nommée symbolisme en informatique et cognitivisme en psychologie.

Le cognitivisme est une approche fonctionnaliste de l'esprit qui postule l'indépendance du calcul et du support (Claverie, 2010). Dans cette perspective, la psychologie s'affranchit des neurosciences et l'esprit considéré comme un ensemble désincarné de modules spécialisés indépendants les uns des autres. La perception, le langage comme la mémoire sont considérés comme des facultés dépourvues d'essence matérielle, chacune accessible à une éducation, une rééducation ou même une imitation en intelligence artificielle complètement indépendante du support physique, qu'il soit neuronal ou électronique. Chaque module est une entité spécifique à une opération mentale définie, autonome et dont le fonctionnement est inconscient (Fodor, 1983).

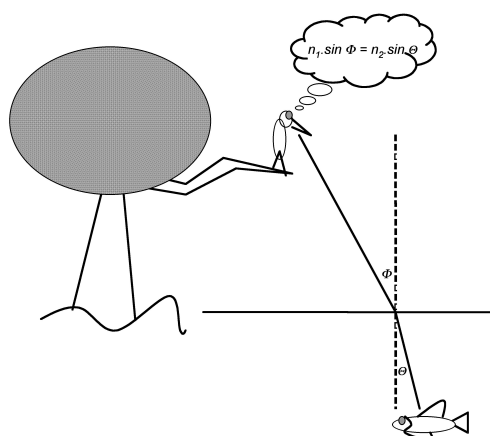


Figure n°8 : L'hypothèse cognitiviste (ou symboliste) : l'oiseau possède une implémentation de la Loi de Descartes/Snell au sein de son cerveau (incarnation) qui compte un comportement adapté grâce à elle (d'après Varela, 1989, in Claverie, 2010).

Ces sous-systèmes distincts ne peuvent d'ailleurs être « pénétrés » par la conscience, les rendant similaires à des boîtes noires théoriques, inaccessibles. Ils possèdent néanmoins une implémentation précise sur un substrat neuronal qui n'est là que pour servir de support, au même titre qu'un circuit électronique n'a rien à voir avec le calcul qu'il peut supporter.

Un administrateur central serait alors chargé d'en contrôler l'action afin de coordonner le traitement des informations pour une cognition globale. Le cerveau n'est alors qu'une super machine cybernétique colonisée par les activités cognitives qui sont organisées entre elles pour une pensée complexe.

tives qui sont organisées entre elles pour une pensée complexe.

La comparaison de l'ensemble esprit-cerveau avec celui représenté par le couple programme-ordinateur a fait l'objet de beaucoup de débats, modèles et fantasmes. Le cerveau est-il similaire à un ordinateur et la pensée l'est-elle

au programme (ou selon les auteurs, l'esprit, la cognition, etc.) ? Et de là, un ordinateur peut-il penser⁹ ?

Telle est la question qui sépare radicalement Searle (1980) de Turing (1959). Dans son expérience de pensée de la « chambre chinoise », Searle juge que

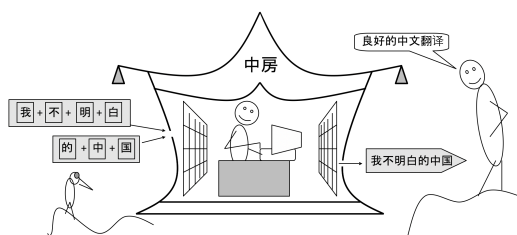


Figure n° 9 – Expérience de la chambre chinoise. Un opérateur isolé (disposant de toutes les documentations, bibliothèques et sources informatiques nécessaires) reçoit par une boîte à lettres des symboles chinois et les organise pour les distribuer en phrases par une autre boîte aux lettres, selon des procédures définies, sans pour autant comprendre le chinois. Pour Searle, ce n'est pas l'opérateur, mais bien l'ensemble chambre-opérateur qui effectue la tâche sémantique. Ni la chambre, ses composants ou l'opérateur n'ont accès à une "conscience globale". Seul un observateur (chinois) extérieur peut avoir accès à cette conscience sans pour autant savoir ce qu'il y a dans la chambre. Searle utilise cette métaphore pour contester le fait qu'un ordinateur puisse avoir accès à la compréhension des résultats du programme, et ainsi dénoncer l'IA forte. (d'après Searle, in Claverie, 1990).

les ordinateurs tels que nous les connaissons ne peuvent pas produire une pensée, car le sens (niveau symbolique) leur échappe et qu'ils ne sont capables que d'une simple application de procédures (niveau purement syntaxique).

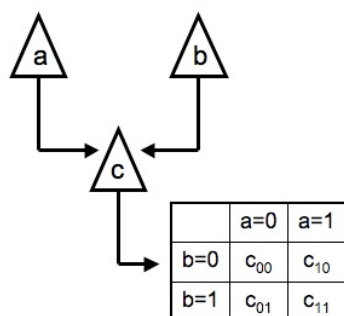
Le connexionnisme, parallèlement, se différencie de la pure métaphore cerveau-ordinateur pour revenir à des contraintes que la neurobiologie cybernétique avait mises en évidence : plasticité cérébrale, loi de Hebb¹⁰, inhibition parallèle, etc. C'est une perspective qui s'attache d'abord au substrat et à ses caractéristiques, propriétés et limites, et qui essaie de comprendre la cognition à partir de modules concrets, les réseaux de neurones.

L'idée connexionniste est née avec les premiers travaux de la cybernétique et du modèle de « neurone formel » de McCulloch et Pitts, mais ce sont les premiers travaux sur le per-

9 - Question que tiennent pour résolue des nombreux penseurs et industriels de la côte Ouest américaine, proposant de relier le cerveau/pensée à des structures numériques tels que supercalculateurs ou même Internet (société Neuralink de Elon Musk) pour un « extended mind » (Clark & Chalmers, 1998).

10 - La loi de Hebb correspond à l'énoncé du principe sur lequel repose l'apprentissage : la plasticité des circuits neuronaux permet la modification de façon durable de l'efficacité de leur transmission synaptique. Plus une synapse est sollicitée, plus elle devient performante à long terme.

ceptron¹¹ de Roseblatt qui lui ont donné ses fondements. Cette machine neuro-inspirée est capable d'apprendre à reconnaître des images et des caractères de lecture par essais-erreurs (voir Claverie, 2010). Elle est structurellement basée sur un principe de connectivité analogue à celui d'unités logiques, les neurones naturels, interconnectées entre elles, et travaillant en parallèle sur des éléments d'une information qui émerge comme résultats du traitement du réseau qui s'auto-organise après de nombreux allers-retours entre des entrées et des sorties imposées. L'information n'est plus stockée à des « adresses » précises mais est distribuée massivement en parallèle (PDP pour « parallel distributed processes », Rumelhart & McClelland, 1986).



ET : c₀₀=0 ; c₁₀=0 ; c₀₁=0 ; c₁₁=1
 OU : c₀₀=0 ; c₁₀=1 ; c₀₁=1 ; c₁₁=1

Figure n°10 : Réseau de neurone permettant d'implémenter les opérateurs logiques «et» et «ou» (in Claverie, 2010).

Cette auto-organisation échappe à une logique de manipulation de symboles, et seul le résultat donné par le réseau peut être, a posteriori, analysé de l'extérieur comme ayant une valeur symbolique. La signification réside alors dans l'interaction de plusieurs ensembles de niveau subsymbolique.

Ce paradigme connexionniste s'affranchit et abandonne plusieurs contraintes du cognitivisme. La notion de modularité n'existe plus et la localisation perd son sens au bénéfice de celui de configuration globale de réseaux. De même, on n'a plus besoin d'un hypothétique « centre de contrôle » cognitif ou « unité centrale » informatique, et le traitement de l'information comme cette information elle-même sont inséparables de l'adaptation structurelle du réseau par le simple jeu des poids synaptiques. Le cerveau est alors considéré comme un super réseau de haut niveau, composé de plusieurs sous-réseaux articulés entre eux en fonction de spécialisations qui émergent de la nécessité d'évolution de l'ensemble. Le spécialiste d'Intelligence artificielle Marvin Minsky (1997) s'est inspiré de ce paradigme pour considérer le système cognitif comme une macrosociété de micro-agents, telle une fourmi-

11 - Le perceptron est un appareil théorique développé à partir des neurones formels de McCulloch et Pitts auxquels ont été imposés des modes de fonctionnement inspirés des lois neurologiques : Loi de Hebb et rétropropagation, inhibition parallèle, etc. Les travaux sur le perceptron ont donné naissance au mouvement PDP qui sont la base des progrès actuels du deep-learning et de l'intelligence artificielle connexionniste.

lière, susceptible de résoudre des problèmes spécifiques de manière auto-organisée.

L'auto-organisation est un paradigme qui, à partir du connexionnisme, s'est différencié en s'appuyant sur l'un des principes de structuration des systèmes complexes : du produit d'organisation d'éléments coordonnés se dégage un autre qui est irréductible aux premiers. En clair, les propriétés d'un ensemble de parties sont, tant aux plans structurel que fonctionnel, supérieures à la « somme » de celles-là. La différence est une caractéristique émergente de cette somme. La cognition est alors considérée comme étant l'émergence d'états globaux

dans un réseau de composants simples, gérés par des règles locales de liens et de changements entre ces éléments. La fonction cognitive est dès lors un ensemble de propriétés émergentes adéquates à la résolution d'une tâche donnée.

Le néoconnexionnisme intègre à la fois cette fonction d'émergence et l'existence du sens, au-delà des simples symboles. Le sens n'est pas contenu dans le symbole, mais il émerge de l'activité d'un ensemble d'unités inférieures, constituant un niveau cognitif subsymbolique. Le sens n'existe pas dans chaque unité subsymbolique prise isolément, mais émerge de la collaboration de ces unités lors de leur activation en parallèle. L'organisation (connexions) de ce réseau s'organiserait progressivement, dans des étapes de maturation, puis d'éducation et/ou de rééducation, au fur et à mesure des expériences de confrontation à l'environnement, physique ou culturel. Le cerveau est donc une machine à émergence et l'intelligence est l'aptitude à faire émerger du sens, cela en dehors de toute règle programmée de logique extérieure qui s'imposerait à l'humain, comme le conçoivent les cognitivistes.

Sans remettre en question l'existence de l'émergence du sens à partir du fonctionnement de l'appareil cognitif, qu'il soit ou non identifié au cerveau,

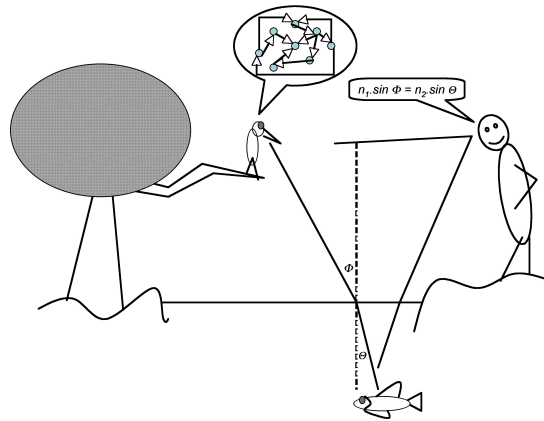


Figure n°11 : L'hypothèse connexionniste : le cerveau de l'oiseau s'auto-organise sans intervention de quelque niveau symbolique pour une performance adaptée. La Loi de Descartes/Snell n'existe que par la théorisation de l'observateur (d'après Varela, 1989, in Claverie, 2010).

de nombreux chercheurs considèrent que les modèles cognitivistes et connexionniste ont chacun des vocations et des limites. Une solution est alors de considérer qu'il existe des étages de la cognition en retenant que celle-là, sur le plan global, émerge du fonctionnement parallèle de parties symbolistes et d'autres connexionnistes. De la même manière interviendraient des processus symboliques et subsymboliques qui s'auto-enrichiraient pour atteindre une performance adéquate, ajustée aux conditions et aux problèmes posés par l'environnement.

Certains chercheurs vont alors proposer des hypothèses sur l'articulation possible entre cognitivisme et connexionnisme selon des principes de spécialisation, selon lequel chacun s'occupant spontanément de ce en quoi il est performant, et d'emboîtement, selon lequel les réseaux de neurones restent de niveau inférieur et le symbolisme organisé grâce à un langage de haut niveau, avec un administrateur central hypothétique qui coordonne le tout.

Modèles hybrides et énaction.

Un des constats, dans ce débat, est l'opposition conceptuelle majeure entre connexionnisme et symbolisme (cognitivisme), l'un basé sur l'organisation de fonctions supérieures à partir de la seule organisation en réseau du tissu neuronal (ou de son imitation numérique dans la machine), l'autre sur l'intervention de ce support dans la computation de symboles et des fonctions supérieures qui en résultent. Ces conceptions sont donc diamétralement opposées et l'argument d'exclusion réside dans celui-ci : une quelconque intervention d'un niveau symbolique, à quelque niveau que ce soit, disqualifie tout connexionnisme. Pourtant, comme dans tous débats, on trouve des adeptes de la voie moyenne, celle qui prend le meilleur de deux mouvements antagonistes, prônant plus leur complémentarité que leur opposition.

Dans cette perspective, l'appareil cognitif est toujours considéré comme un système de traitement de l'information couplé de manière cybernétique à l'environnement. Les entrées (niveau esthétique) correspondent aux ascendances sensorielles pour alimenter les processus cognitifs afférents tels que perception, reconnaissance, représentation... Les sorties (niveau dynamique) regroupent des processus de programmation et de préparation motrices, ainsi que sont les différents types d'action ou de communication. Entre les deux (niveau zygonique ou associatif) s'élaborent les processus de la transformation et de la gestion de l'information, tels que le tri et les filtres attentionnels, la généralisation et la catégorisation, l'association et inférence et la coordination de l'action vers les buts à atteindre.

La position de certains spécialistes de l'intelligence artificielle, dont il faut rappeler le projet princeps de compréhension du cerveau et de la pensée, s'inspire de la « *scienza nuova* » de Vico, grand-père du constructivisme

moderne et en son temps critique du cartésianisme. L'une des convictions viciennes est celle de connaissance de l'inconnu à partir du connu et du pouvoir explicatif de l'analogie. En quelque sorte, lorsque quelque chose de complexe dont on sait tout se comporte ou produit les mêmes effets qu'une autre chose complexe dont on ne sait rien ou peu, la connaissance de la première est éclairante de la connaissance de la seconde. Il s'agit alors de savoir comment on construit des machines à penser (Claverie, 2005) pour comprendre la machine cognitive naturelle.

Les travaux de l'intelligence artificielle se répartissent exactement en deux domaines : les systèmes experts et les algorithmes programmés sur des machines séquentielles ou partiellement parallèles, les réseaux massivement parallèles et distribués. Le premier utilise des structures de calcul de type Von Neumann¹² : ce sont les ordinateurs classiques disposant d'un ou de plusieurs processeurs, sorte d'administrateur central du traitement de l'information stockée et gérée grâce à des interfaces d'entrées-sorties et des registres

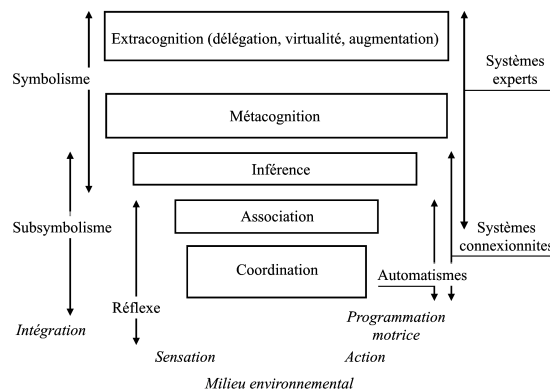


Figure n°12 : Articulation des deux modèles cognitive et connexionniste dans une complémentarité de traitement de l'information par niveaux.

mémoires dont, pour ce qui est des machines modernes, l'un au moins est à conservation à long terme (mémoire de masse) et un autre à court terme (mémoire vive). La simulation actuelle des performances cognitives humaines par ces systèmes est excellente pour les niveaux symboliques. Le second domaine est celui des machines connexionnistes fonctionnant par « deep-learning », dans les-

quelles l'information est relative à la structure évolutive des réseaux de neurones formels de la machine. La performance de ce type de systèmes pour

12 - Les architectures computationnelles dues à Von Neumann sont à la base des ordinateurs conventionnels actuels. La rupture opérée en 1946 repose sur deux innovations. La première est la séparation d'une unité de commande qui traite le séquencement des instructions d'un programme, et une unité arithmétique qui exécute ces instructions. La seconde innovation est celui du stockage de l'information et du programme dans des registres mémoires internes et repérables selon des adresses conventionnellement codées.

les niveaux perceptifs, de catégorisation, subsymboliques et associatifs est indiscutable, avec des vitesses de traitement parfaitement performantes. Ces spécificités fonctionnelles par niveaux permettent de proposer un modèle d'articulation des deux hypothèses, avec des spécificités pour les bas et hauts niveaux et une forme de complémentarité et de recouvrement de performances pour les niveaux intermédiaires. On obtient alors une machine et donc un organe disposant de processus de ségrégation, où chaque niveau s'occupe de ce pour quoi il est performant, de hiérarchie, qui délègue aux réseaux de neurones un rôle de base de traitement au bénéfice des processus symboliques alors considérés en surplomb, et d'inclusion, pour lesquels les symboles sont à concevoir comme une description de niveau supérieur de propriétés supportées par le système distribué sous-jacent.

Bien entendu, ces modèles hybrides, sans revenir sur l'efficacité calculatoire attestée, ne sont ni plus ni moins que des conceptions cognitivistes « colonialistes » renvoyant les niveaux connexionnistes au statut de systèmes esclaves.

Pour le connexionnisme, l'activité du cerveau est de traiter l'information en se modifiant sans cesse lui-même. On entre alors dans une conception constructiviste de l'activité cognitive ; les objets mentaux sont des coconstructions continues du cerveau et de l'information qui s'exprime par des émergences corrélées à ce mouvement comme représentations. Le cerveau est alors à la fois la structure en modification permanente, le processus neuropsychologique et l'information qui s'en exprime, inscrit dans une perspective matérialiste affirmée.

Une perspective généralisée considère à la fois le cerveau comme organe de plasticité permanente, mais son inscription bouclée à l'intégralité de son enveloppe vivante, le corps, lui-même en relation permanente avec l'environnement qu'il modifie de ses actions. Cette conception varélienne de l'« éraction » postule que la pensée ne peut s'extraire ni de l'organisme ni de la situation dans laquelle il se trouve. Un tel bouclage conçoit à la fois une « cognition incarnée » et une « cognition située » (Varela, 1989).

Les « mind/brain theories ».

Au-delà des théories des informaticiens de l'intelligence artificielle, des psychologues de l'intelligence naturelle ou des biologistes de la neuropsychologie, un ensemble de conceptions s'inscrit en concurrence en fonction des convictions irréfutables qui animent les chercheurs et cliniciens. Le lecteur pourra trouver un excellent recueil critique, à propos des neurosciences et de la neuropsychologie, dans Missa (1993).

Ces théories portant sur le rapport de la pensée au cerveau concernent plus généralement celles de la nature de la cognition et des opérations cognitives : ont-elles des caractéristiques physiques, partiellement ou exclusivement matérielles, sont-elles au contraire relatives à une forme d'immatérialité indépendante et inaccessible aux lois physiques, etc. ? Ce type d'interrogations remonte probablement aux origines de l'humanité, et de nombreuses croyances et mouvements religieux en ont fait leurs choux gras. Autant d'auteurs, autant de variétés théoriques, et il ne faut pas s'étonner que la religion ou l'idéologie ne soient pas, pour la plupart, bien loin du discours affiché. Le but n'est évidemment pas ici d'en donner un inventaire exhaustif ; un traité n'y suffirait pas. Il n'est pas question de prendre position pour l'une plutôt que l'autre, il convient au contraire d'ouvrir les perspectives et d'éviter les dogmatismes. Mais on peut indiquer quelques grands courants qui se réfèrent à certains chefs de file des concepteurs de la pensée et de ses rapports au cerveau, et vice versa.

Le dualisme interactionniste et les trois mondes de Popper.

Selon Eccles et Popper (1977), on peut concevoir, dans une tradition cartésienne, un espace à deux dimensions. L'homme est alors double : a) l'esprit ou son action, la pensée, sont de nature spirituelle et b) le corps, et notamment le cerveau, sont de nature matérielle. Les nécessaires échanges entre les deux niveaux nécessitent un dispositif intermédiaire, un locus particulier doté de propriétés spécifiques, une sorte de « cerveau de liaison » tel que Descartes concevait la glande pinéale (l'épiphyse) ou plus récemment Baddeley et son mystérieux « administrateur central ». Dans une telle conception, L'esprit, de niveau psychologique, dispose d'un statut distinct de celui du cerveau, du niveau neurobiologique.

Néanmoins, cette dualité ne suffit pas pour les auteurs, et il convient de rattacher au débat une forme de conception platonicienne de la connaissance elle-même. Popper et

Eccles proposent donc une articulation en trois espaces distincts, trois mondes complémentaires en interaction deux à deux. Le monde 1, physique, est celui de la réalité matérielle (y compris les cerveaux humains). Le monde

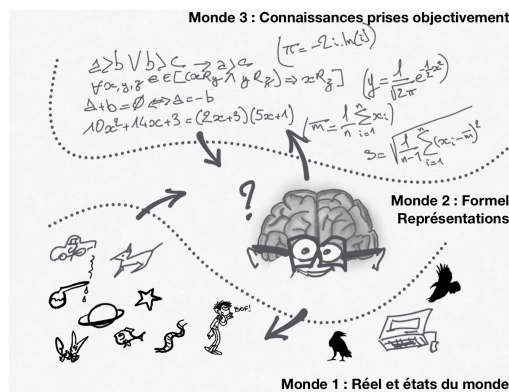


Figure n°13 : Les trois mondes de Popper (in Claverie, 2019).

2, psychologique, concerne la réalité psychique, les contenus de pensée et les états de conscience ; il est en relation directe avec le monde 1 par le moyen des représentations. Le monde 3 est spirituel ; c'est celui des productions de l'esprit humain et des connaissances prises objectivement. Seuls les mondes 1 et 2 d'une part, et les mondes 2 et 3 d'autre part, sont en relation. Il y a une impossibilité de relation directe, et donc une inaccessibilité du monde 1 au monde 3 sans passer par les états de conscience.

Cette conception du rapport entre le corps et l'esprit est la plus fréquemment rencontrée dans une psychologie du sens commun et la plus cohérente avec de nombreux systèmes de pensée, postulant une forme de permanence immanente du monde 3 avec ou non une présence désincarnée du monde 2 qui peut potentiellement être opéré en l'absence du support inscrit dans le monde 1.

Le parallélisme psychophysique de Bergson.

Bergson (1959) propose un dualisme fonctionnel, qui repose à la fois sur le temps de la pensée et sur la mémoire. Contrairement à une position cartésienne, on dispose deux mondes fonctionnels d'indépendance relative.

Avoir conscience de quelque chose, c'est voir quelque chose de passé. La simple réaction implique le corps, sans conscience. La prise de conscience implique un temps, donc un affranchissement du présent, c'est-à-dire en étant dans le passé et à la lumière du mémorisé. Cette articulation du temps permet une réaction appropriée dans le futur. Passé, présent et futur sont liés grâce au cerveau.

Bergson affirme que la mémoire est d'essence spirituelle et que le cerveau permet de l'orienter vers l'action et le présent. Le cerveau a une fonction pratique et tout état cérébral produit un état psychologique. La conscience ne traduit que ce que fait le cerveau, mais selon une autre forme. Le corps est le centre matériel de l'action mais pas celui du souvenir. En effet, les lésions cérébrales n'abîment pas le souvenir, mais elles perturbent leur incarnation.

Il y a donc deux mondes, mais articulés dans une indépendance relative du psychologique par rapport au biologique (Jaffard et al, 1998). Le cerveau est l'organe de soutien sur lequel s'accroche de la pensée, comme le manteau accroché au mur est soutenu par le portemanteau. Sans lui, elle s'écroule.

Le matérialisme émergentiste et le dualisme de point de vue.

En distinction des dualismes précédents, Varela (1989) propose un monisme ontologique qui s'appuie sur un pluralisme fonctionnel. Les états mentaux sont d'ordre matériel puisqu'ils sont des processus du cerveau, mais ils en émergent sous un statut qu'il reste à définir. Cette émergence apparaît à tous les niveaux d'organisation cérébrale : moléculaire et donc infracellulaire,

neuronal, en réseaux, au sein des structures, au niveau du cerveau. À chaque niveau émergent des processus nécessaires, voire constitutifs du fonctionnement du niveau supérieur. Ces niveaux en interaction s'organisent entre eux. C'est de ces interactions qu'émergeraient les représentations, traduisant une correspondance entre l'état global des systèmes et des propriétés du monde.

Le passage de règles régissant un fonctionnement local à un état cohérent global repose sur une « auto-organisation » des niveaux. Leur modélisation repose sur la théorisation des systèmes non linéaires ou des systèmes complexes. On a ainsi recours à la notion d'« attracteurs » élaborée dans la théorie des systèmes dynamiques, la représentation pouvant être assimilée à une incarnation de tels êtres mathématiques. Le phénomène est le même et seule sa lecture diffère selon les outils utilisés.

Le statut de l'émergence fait débat. Si le matérialisme émergentiste considère que les produits émergents sont d'ordre physique, une position proche est un épiphénoménalisme dans lequel l'émergence n'a rien à voir avec cela. Elle acquiert par un processus fonctionnel qu'il reste à formaliser un statut mental ou psychique qui sépare ainsi d'une psychologie s'attachant aux objets mentaux des objets d'étude concrets de la neurologie matérialiste.

Ces positions sont voisines d'une « théorie du double aspect » (Nagel, 1987) qui propose un monisme ontologique qui s'appuie sur un dualisme de point de vue : l'esprit constitue la face subjective et le cerveau la face objective d'une même entité « esprit-cerveau ». Esprit et cerveau constituent un double aspect d'une seule unité. Pour Nagel, la conscience est un phénomène incarné, une expérience sensible personnelle qui est à l'origine du sentiment d'un « effet que cela fait » d'être soi. Cette dualité réside dans le caractère subjectif de l'expérience phénoménale et oppose deux points de vue de la même chose, l'un extérieur et notamment celui de la science qui permet l'étude des activités du cerveau, l'autre correspondant à l'expérience vécue.

On peut rapprocher ce dualisme de point de vue de certains phénomènes physiques tels que la dualité onde-corpuscule qui est relative à l'instrument de mesure plus qu'à la nature de la lumière. Cette dualité est d'ailleurs, de manière surprenante, à l'œuvre au sein même de l'appareil neuro-ophtalmique. Des cellules réceptrices sont spécialisées dans l'électroconversion des photons, les bâtonnets en charge de la vision des niveaux de gris et de la détection des contrastes, permettant une vision fondée sur la détection de corpuscules. D'autres sont spécialisées dans le codage des fréquences des bleu, vert et rouge pour le codage des couleurs de base à partir desquelles s'organise la perception du spectre lumineux, les cônes donnant une vision basée sur l'analyse des longueurs d'onde. Cette dualité est pourtant intégrée dans une conscience unitaire de la scène visuelle comme pourraient l'être deux points de vue d'une même réalité neuropsychologique.

Le matérialisme éliminativiste.

Le matérialisme postule que, même en l'absence de théorie explicative actuelle suffisante, la réalité des états mentaux est de l'ordre de la physique. Pour Feyerabend (1963), chef de file de la lignée des matérialistes de l'esprit, il convient en premier lieu de redéfinir le vocabulaire et d'évacuer les notions approximatives de la psychologie spontanée (psychologie du sens commun ou « folk psychology »). Une telle position naturaliste était déjà chez les Grecs, mais surtout émerge au début du siècle passé chez des Anglais tels que Maudsley, Laycock, Bailey..., des Français comme Flourens, Ribot, Delmas-Marsalet..., l'américain James ou les Allemands promoteurs d'une psychologie als Naturwissenschaft (in Missa, 1993).

La thèse néomatérialiste dite « éliminativiste » dispose que les notions d'états mentaux, de conscience, de pensée, ainsi que leurs propriétés et conséquences sont posées dans des termes préscientifiques. La science du cerveau doit s'enraciner dans la physique ; elle permettra alors d'évacuer des notions dénuées de pertinence telles que désir, croyance, conscience réflexive, etc. Dans cette lignée d'auteurs qui promeuvent l'idée de traiter la psychologie comme une science naturelle (voir Cooper, 1990), Smith-Churchland (1986) et Churchland (1999) sont deux des principaux tenants d'un programme naturaliste de la pensée dit « éliminatif ». Pour ce matérialisme éliminativiste, la conception dualiste des phénomènes psychologiques est fautive à un tel point que les neurosciences les élimineront à mesure qu'elles évolueront vers une pleine maturité. L'esprit est un simple élément de croyance séparant de manière animiste la nature des âmes, et pour les animaux cette notion sera éliminée par la science comme le panpsychisme et l'esprit des plantes l'ont été radicalement par la physique ou l'agronomie.

Cette critique éliminativiste repose justement sur l'analogie avec la physique. Si la mécanique quantique pour le monde microscopique et la physique relativiste pour les grandes dimensions n'ont pas réussi à converger dans une théorie physique unifiée, la physique newtonienne, bien que fautive, permet néanmoins de produire des constructions efficaces. Il en est de même de la psychologie que cette future physique unifiée permettra d'expliquer.

Selon le courant éliminativiste, tout état psychologique est un état physique du cerveau, toute pensée est un processus physique porté par le cerveau. On ne peut donc trouver un état mental désincarné parallèle ou même identique à un état neurobiologique, et vice-versa, puisqu'il n'y a pas d'état mental sauf l'illusion que l'observateur colle à l'état du cerveau qui seul existe. (argument éliminativiste). Si l'argumentation est convaincante, il convient néanmoins de la situer dans l'action. Bien que l'explication physique naïve soit erronée, elle est pourtant suffisante pour une compréhension du monde et une action raisonnable sur lui (Claverie, 2019). Il en est alors de même

d'une psychologie ordinaire qui permet de comprendre la pathologie et d'agir concrètement sur le comportement dans une perspective clinique adaptative ou réadaptative. Aussi imparfaites qu'en soient les notions, elles ont un pouvoir modélisateur utile à la prise en charge raisonnée de patients qui n'ont concrètement pas le temps d'attendre l'évolution espérée de la science. Ainsi le neuropsychologue clinicien se contentera, en toute connaissance de cause d'une erreur et d'une incertitude probable, de la physique naïve et d'une psychologie du sens commun pour peu qu'elles soient efficaces dans sa démarche clinique. Tout en restant persuadé qu'elles n'approximent que de loin une future vérité scientifique plus cohérente avec la pureté des raisons éliminativistes, elle permet une démarche pragmatique temporaire.

Le physicalisme quantique.

Inspiré par le paradoxe de la dualité onde-corpuscule, et dans la suite de l'argument éliminativiste d'une physique explicative en marche, Penrose (1995) s'interroge sur le lien étrange qui semble exister entre l'esprit humain et la mécanique quantique. Les physiciens considèrent ainsi que la conscience influence le monde physique à une échelle quantique. Elle pourrait à la fois en résulter comme la réalité en résulterait elle-même. De cet argument, les physiciens pensent que la théorie quantique est nécessaire à la compréhension du fonctionnement du cerveau.

Revenant à l'argument de Turing, « thinking is computing », selon lequel la pensée relève de la calculabilité, on sait que certains raisonnements ne sont réductibles à aucun algorithme (incomplétude de Gödel). Une algorithmique cognitive doit alors relever d'une physique qui permet l'affranchissement de cette limite.

Si la mécanique quantique est la seule théorie que nous ayons pour décrire le monde au niveau des atomes et des particules subatomiques, elle n'en reste pas moins mystérieuse. Elle repose ainsi sur deux notions princeps qui sont la superposition d'états et l'intrication. Un autre inexplicable est celui selon lequel un résultat d'une même expérience un niveau quantique change selon que l'on mesure certaines propriétés des particules concernées. Cette réalité pose la question de la réalité même d'un extérieur objectif, indépendant de l'observateur et donc de son cerveau et/ou de sa pensée. Selon la théorie, la façon dont le monde se comporte dépend de la façon dont il est mesuré ouvrant la voie d'un matérialisme parapsychique et de la potentielle superposition d'un ou des plusieurs mondes matériels et mentaux.

À un autre plan, de manière analogique à ce que les éléments quantiques peuvent se trouver à plusieurs endroits ou se déplacer à plusieurs vitesses (superposition), un cerveau quantique pourrait adopter une logique dans la-

quelle les éléments peuvent s'exclure en même temps dans une algorithmique classique. Une telle idée est pour l'heure le meilleur moyen d'approcher une logique naturelle qui semble à l'œuvre dans le fonctionnement cérébral, loin, très loin de la logique booléenne.

Penrose propose une théorie de la conscience qui relève de cette future physique nécessaire à unifier les mondes relativiste et quantique. Pour lui, la pensée s'organiserait au niveau des protéines constituées en microtubules intraneuroniques (neurotubules). Ces éléments seraient à l'origine d'influx cohérents à grande échelle qui permettraient l'émergence du fonctionnement global au niveau du cerveau, peut-être dans une forme d'orage de décohérence généralisée. Penrose nomme cette théorie « Orch-OR »¹³.

Si des critiques se sont élevées, elles sont toutes fondées sur l'impossibilité du maintien d'une cohérence quantique dans des échelles macroscopiques telles que celles du cerveau ou à des températures éloignées du zéro absolu.

Or Fisher (2015), coordinateur du Quantum Brain Project de l'Université de Californie à Santa Barbara, a proposé une réponse à ces critiques. Il développe un modèle selon lequel le cerveau, et plus généralement les ensembles vivants, pourraient être des lieux « actifs » de décohérence à long terme et à température biologique. Ainsi, les atomes de phosphore qui sont largement distribués dans les cellules sous forme d'ions phosphate pourraient jouer le rôle de qubit¹⁴ et fournir le support nécessaire à un calcul quantique : une sorte de « qubit neural ». Cela veut dire que le cerveau serait capable de fonctionner selon des lois quantiques.

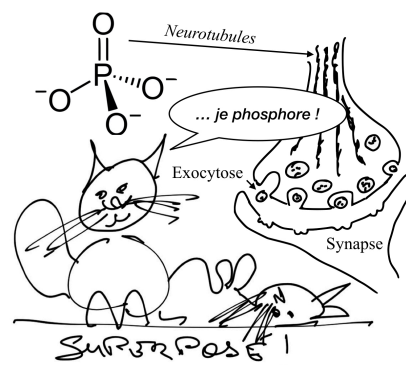


Figure n°14 : Modèle de superposition d'état, grâce au phosphore, au sein des neurotubules et des synapses.

13 - Orch-OR ou « réduction objective orchestrée ». L'idée est que les interférences et la superposition quantiques qui s'expriment au sein des structures intraneurales s'effondrent (accrétion du paquet d'ondes) en produisant des éléments de conscience. Un même mécanisme pourrait être à l'œuvre dans d'autres entités biologiques, par exemple au niveau de la synthèse biochimique dans les plantes pour la photosynthèse, ou au niveau comportemental chez les animaux migrateurs pour l'orientation spatiale (oiseaux, poissons, etc.).

14 - Analogie du bit en informatique binaire pour l'informatique quantique. Un qubit, qu-bit ou qbitt est la plus petite unité de stockage d'information quantique. Il porte la superposition de plusieurs états de base et permet aux calculateurs quantique d'obtenir une puissance de calcul qui varie de manière exponentielle en fonction du nombre de qubits.

Dans le modèle de Fisher, chaque phosphate porte un spin quantique, deux ions peuvent s'intriquer mutuellement. Chaque ion phosphate serait associé à quatre atomes d'oxygène ou être incorporé dans de plus gros objets appelés « molécules de Posner ». Ces grappes de six ions phosphate, combinées avec neuf ions calcium sont des composants des cellules vivantes, et seraient capables de protéger des états de superposition quantique. Si les ions d'une paire de phosphates intriqués sont séparés en deux molécules de Posner, ils resteront intriqués quelle que soit la distance de ces molécules, pouvant ainsi constituer de multiples systèmes de cohérence qui s'affranchissent du temps. Le modèle suggère donc des ions pouvant être superposés ou intriqués, tout en restant protégés de la décohérence¹⁵ sur une durée compatible avec les constantes de la biologie cellulaire.

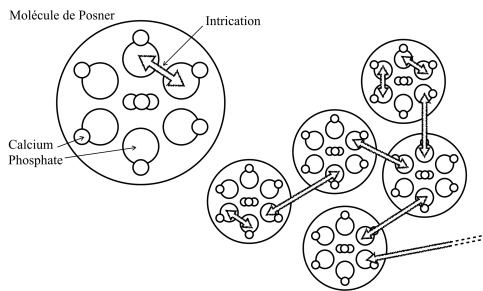


Figure n°15 : Modèle Orch-OR d'intrication multiple au sein de molécules de Posner favorable à une décohérence globale à distance en un « orage » de décohérence (Inspiré de Lucy Reading in Quanta magazine, 2016)

Des mécanismes de transport des molécules de Posner dans les terminaisons neurales présynaptiques et d'exocytose¹⁶ pendant la libération de neurotransmetteur contribueraient au déclenchement neuronal post-synaptique. Les multiples molécules de Posner enchevêtrées fourniraient alors le mécanisme clé du traitement quantique neuronal, et les influx résultants pourraient à leur tour s'entremêler, dans une sorte de superposition d'une « pensée » quantique. L'intrication et la supposition pouvant être ainsi maintenues bien au-delà de la distance et du

temps imposés à un ordinateur classique, le cerveau serait alors à considérer comme un organe de suprématie quantique¹⁷.

Certains autres théoriciens pensent que, même en l'absence des conditions physiques d'un calcul quantique, le cerveau serait capable d'imiter les pro-

15 - La décohérence est le phénomène qui permet de passer d'un état quantique superposé à un seul état cohérent avec la physique classique.

16 - L'exocytose est le processus de sécrétion de substances neuroactives habituellement stockées dans des vésicules des boutons terminaux de l'axone d'un neurone, au sein de l'espace des synapses entretenues avec les dendrites et corps cellulaires des neurones voisins.

17 - Notion désignant la supériorité d'un ordinateur quantique capable de calculer en quelques fractions de secondes ce qui prendrait des dizaines, milliers, voire millions d'années aux ordinateurs classiques, y compris les supercalculateurs les plus puissants et avancés.

cessus quantiques, un peu comme certains calculateurs numériques actuels imitent le calcul quantique en fournissant des modèles computationnels de portes quantiques. Le cerveau pourrait alors, a minima et à un niveau biologique, mettre en œuvre des états et processus analogues aux mécanismes quantiques, et servant ainsi de mémoire quantique fonctionnelle.

Un débat actuel permet de poser les hypothèses matérialistes d'un tel fonctionnement et d'aborder une logique quantique pour la computation cérébrale, avec l'économie des principes aristotéliens de non-contradiction de tiers exclus, ou de disjonction d'états binaires. Un argument, et non des moindres, est qu'en l'absence d'une physique unifiée et à cause de la superposition d'états, rien ne peut être raisonnablement « computé » et donc simulé par un algorithme. Penrose et Fisher signent là l'inutilité de toute tentative de computation binaire et de l'intelligence artificielle traditionnelle, qui ne peuvent donc servir de modèle à la compréhension de la pensée.

Conclusion.

Ce parcours des modèles de la relation entre le cerveau est l'esprit a traversé plusieurs champs d'hypothèses sans pour autant apporter quelque certitude sur la nature de ce rapport. On peut pourtant poser une description factorielle des positions (convictions) qui s'attachent à décrire ce problème. Un premier facteur oppose à la fois connexionnisme du cognitivisme, avec la place particulière du symbole et celle de la localisation cérébrale. Le second facteur est plus complexe, opposant les spiritualistes aux matérialistes éliminativistes pour finir dans les mystères de la mécanique quantique.

Pour parodier Penrose, si personne ne comprend ce qu'est la mécanique quantique ou comment cela fonctionne, personne ne comprend la conscience et la pensée non plus. Et bien malin serait alors celui qui comprendrait la neuropsychologie et des entités aussi mal définies que l'esprit, la pensée ou la conscience et leurs troubles.

Une première difficulté réside dans le fait que nous ignorons ce que sont ces phénomènes, et par voie de conséquence la cognition. La seconde est que nous ne possédons ni théorie, ni langage commun pour les décrire et les étudier objectivement. Une troisième concerne le fait que nos conceptions du monde sont prisonnières de nos croyances et que les théorisations en sont contingentes. On voit cependant un premier mouvement qui part du moléculaire pour comprendre l'organisation émergente de ces phénomènes mal définis, alors qu'un second s'attache à localiser et décomposer des entités de niveau psychologiques pour les rattacher ou non à des niveaux matériels. Rien n'est clair, le débat est rude et les enjeux dépassent les simples positions de principe ou la raison pragmatique.

Adapté de Eccles (1992), on peut tenter de résumer ces multiples positions de la manière suivante. On peut en premier lieu admettre un Monde 1 (M1) correspondant à l'ensemble d'objets concrets, d'existence matérielle (p), dans lequel on trouve les éléments physiques vivants et non vivants, y compris donc les corps et les cerveaux, et un Monde 2 (M2) correspondant à l'ensemble potentiel des expériences subjectives, des pensées ou des états mentaux (m) dont on ne possède pas de définition physique. On peut distinguer deux sous éléments : un premier sous-monde strictement physique (M1p) réunissant tous les éléments matériels dépourvus d'état mental, de conscience ou de pensée, et un sous-monde hybride (M1m) correspondant à une partie potentielle de M1 qui est associée, sous pour les produire, soit pour les supporter ou pour en être dépendant, à ces états mentaux.

On peut alors donner un formalisme au dualisme strict :

$$M1 = M1p, M2 = M1m ;$$

et au dualisme interactionniste :

$$M1 = M1p + M1m ; M1m \longleftrightarrow M2 \text{ par interaction d'un cerveau de liaison CL, d'où si } CL = M1m \text{ alors } M1 = M1p + M1CL \text{ et } M1CL \longleftrightarrow M2.$$

Le matérialisme strict s'exprime ainsi :

$$M1 = M1p, M1m = M2 = \emptyset ;$$

et le matérialisme identitaire (théorie de l'identité) :

$$M1 = M1p + M1m ; M1m = M2 ; M2 = \emptyset.$$

Le matérialisme émergentiste s'inscrit ainsi :

$$M1 = M1p ; M1p \rightarrow M1m ; M2 = M1m ;$$

et l'épiphénoménisme :

$$M1 = M1p + M1m ; M1m \rightarrow M2.$$

Enfin, le panpsychisme matérialiste dispose que :

$$M = M1 + M2 ; M1 \neq M2 ; M1 \longleftrightarrow M2, \text{ aucun n'existant séparément ;}$$

et la cognition quantique :

$$M = M1 + M2 \text{ (superposition de M1 et M2) ; } |\psi M\rangle = 1/\sqrt{a} |M1\rangle + 1/\sqrt{b} |M2\rangle ; |\psi M1\rangle = 1/\sqrt{\alpha} |M1p\rangle + 1/\sqrt{\beta} |M1m\rangle ; a \text{ et } b, \text{ et } \alpha \text{ et } \beta \text{ sont les probabilités des états superposés (a+b = } \alpha+\beta = 1). \text{ La décohérence est une émergence de l'état psychologique M2 (haut niveau) ou M1m (bas niveau), ou reste purement physiologique, sans accès au registre mental.}$$

On peut aisément supposer que le dualisme interactionniste et la cognition quantique sont les moins simples à concevoir. On peut alors considérer que la quantité d'instructions d'un programme d'une machine de Turing (ou de son équivalence en termes d'architectures Von Neumann implémentant une programmation binaire) les imitant est d'autant importante, ce qui nous éloigne d'un argument de simplicité. L'analogie computationnelle en est d'autant plus compliquée. Cela nous semble en défaveur d'une part des réa-

lisations cognitivistes et d'autre part des imitations par calcul binaire d'une cognition quantique. Notre opinion est que la pensée n'est raisonnablement envisageable qu'en deux termes de niveau différent : une activité quantique intracellulaire et/ou une organisation connexionniste en réseaux de neurones, et qui s'affranchit de la calculation symbolique. Le modèle séduisant d'une émergence par niveaux peut alors contribuer à admettre une articulation quantique moléculaire et cellulaire, et connexionniste en réseaux et zones qui se spécialisent dans l'expression de fonctions selon des modalités probablement (inévitablement ?) dissipatives dans une interaction dynamique au monde (Vitiello, 2001), et qu'il reste à comprendre à haut niveau.

En tous les cas, on remarquera que le champ de la théorisation de la relation du corps à l'esprit, et plus particulièrement du cerveau à la pensée, est loin d'être résolu, et que les convictions profondes de chacun l'emportent souvent sur la raison. Pour revenir à la neuropsychologie, peu importe ces convictions si elles permettent qu'émerge la compassion pour le patient, lui qui vit le trouble de sa pensée et sait, souvent de manière aussi révoltante que désespérante, pour longtemps, qu'il est prisonnier de son cerveau blessé.

Références bibliographiques

- Bergson H., 1896, Matière et mémoire : Essai sur la relation du corps à l'esprit, Paris : Félix Alcan (réédition 1959, Paris : Presses Universitaires de France).
- Churchland PM., 1999, Matter and consciousness : A contemporary introduction to the philosophy of mind, Cambridge : The MIT Press.
- Clark, A., Chalmers DJ., 1998, The extended mind, *Analysis*, 58, p. 10 - 23.
- Claverie B., 1999, Les méthodes en psychophysiologie, in *Les méthodes de recherche en psychologie*, s/d Rossi JP., Paris : Dunod Editions.
- Claverie B., 2005, *Cognitique : science et technique des relations à la machine à penser*, Paris : L'Harmattan.
- Claverie B., 2010, *L'homme augmenté : Néotechnologies pour un dépassement du corps et de la pensée*, Paris : L'Harmattan.
- Claverie B., 2018, Pour une Histoire Naturelle de l'Intelligence Artificielle, in *Le devenir cyborg du Monde*, s/d Lafargue B., Pau : Presses universitaires de l'Université de Pau et des pays de l'Adour.
- Claverie B., 2019, Introduction à l'épistémologie et à la méthode de recherche à l'usage des ingénieurs et autres scientifiques de l'industrie, Paris : L'Harmattan.
- Claverie B., Le Blanc B., 2016, Entre sensation et cognition : l'illusion explicative; *Hermès*, 74, p. 66 - 70.
- Cooper WE., 1990, William James's theory of mind, *Journal of the History of Philosophy*, 28, p. 571-593.
- Eccles JC., 1992, *Evolution du cerveau et création de la conscience*, Paris : Fayard.
- Feyerabend PK., 1963, Mental events and the brain, *Journal of Philosophy*, 60, p. 295 - 296.

Fisher MPA., 2015, Quantum cognition : The possibility of processing with nuclear spins in the brains, *Annals of Physics*, 362, p. 593 - 602.

Fodor JA., 1983, *The modularity of mind*, Cambridge : MIT Press.

Jaffard R., Claverie B., Andrieu B. (& col.), 1998, *Cerveau et mémoires : Bergson, Ribot et la neuropsychologie*, Bordeaux : Editions Osiris.

Miller GA., 1956, The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, 63, p. 81 - 97.

Minsky M., 1997, *La société de l'esprit*, Paris : Interéditions.

Missa JN., 1993, *L'esprit-cerveau : La philosophie de l'esprit à la lumière des neurosciences*, Paris : Vrin.

Nagel T., 1987, *What does it all mean ? A very short introduction to philosophy*, Oxford : Oxford University Press.

Penfield W., 1959, The interpretive cortex, *Science*, 129, 3365, p. 1719 - 1725

Penrose R., 1995, *Shadows of the mind : A search for the missing science of consciousness*, Oxford : Oxford University Press Inc.

Popper K., Eccles JC., 1977, *The Self and Its Brain*, Berlin : Springer-Verlag.

Prieur B., 2019, *Informatique quantique : De la physique quantique à la programmation quantique en Q#*, Saint-Herblain : Éditions ENI.

Root-Bernstein R., 2005, Roger Sperry: Ambicerebral Man, *Leonardo*, 38, 3, p. 225-224

Rumelhart DE, McClelland JL, 1986, *Parallel distributed Processing, Explorations in the microstructure of cognition: foundations*, New-York : Bradford Books.

Sacks O., 1992, *L'homme qui prenait sa femme pour un chapeau*, Paris : Seuil

Sallaberry JC, Claverie B., 2018, *Introduction aux sciences humaines et sociétales*. Paris : L'Harmattan.

Searle JR., 1980, Minds, brains, and programs, *Behavioral and Brain Sciences*, 3, p. 417 - 457.

Seron X., 1979, *Aphasie et neuropsychologie*, Bruxelles : Pierre Mardaga Éditeur.

Smith-Churchland P., 1986, *Neurophilosophy : Toward a unified science of the mind-brain*, Cambridge: The MIT Press.

Turing, A., 1959, Can machines think ?, *Mind* 59, p. 433-460.

Varela F., 1989, *Invitation aux sciences cognitives*, Paris : Seuil

Vitiello G., 2001, My double unveiled : The dissipative quantum model of brain. *Advances in Consciousness Research*, 32, 16, Amsterdam : John Benjamins Publishing Company.