

7 - Cerveau et conscience : bilan et perspectives

par J.F. Lambert

“ On ne ramènera jamais les manifestations de notre âme aux propriétés brutes des appareils nerveux pas plus qu'on ne comprendra de suaves mélodies par les seules propriétés du bois ou des cordes du violon nécessaires pour les exprimer ” (Claude Bernard)

La première, et certainement la principale difficulté que rencontre tout explorateur de la conscience est d'ordre linguistique. Que recouvrent chacun des multiples usages du terme “ conscience ”? Et qu'en est-il du “ psychisme ”, de “ l'esprit ” ou de “ l'âme ”? Une autre difficulté majeure tient à la nature même du problème à résoudre qui suppose d'être déjà résolu pour être posé. En effet, si l'on voulait éclairer les pouvoirs de la conscience de manière non circulaire, on ne devrait ni les invoquer, ni les supposer, dans l'explication des processus par lesquels ils se manifestent. Ce qu'il convient d'expliquer devrait rester un but et n'être pas utilisé comme ressource explicative, ce qui, dans le cas de la conscience, est effectivement impossible. Comme le souligne A. Damasio, “ Qu'y a-t-il de plus étourdissant que de s'apercevoir que c'est le fait même d'avoir une conscience qui rend possibles et même inévitables nos questions sur la conscience ? ”

1 – LA CONSCIENCE DANS TOUS SES ETATS

1.1. PSYCHISME, AME, ESPRIT

Le psychisme peut-être considéré comme inhérent à tout être vivant. Au sens large, ce terme désigne le fait, pour tout organisme [vivant] d'être le siège de processus (pour l'essentiel, inconscients) dont la finalité est la conservation de son identité (de sa structure). L'âme ^[1] (végétative, sensitive, intellectuelle) apparaît, quant à elle, comme détentrice et/ou génératrice de l'information (de ce qui donne la forme) à condition de ne pas réduire cette dernière (l'information) au formalisme de la théorie qui la quantifie. Cette information n'est pas non plus équivalente au contenu du programme génétique mais plutôt à la dynamique de son expression dont on sait précisément qu'elle ne dépend pas seulement de son contenu mais aussi de son contexte. Bien que non réductible à la matière-énergie qu'elle “ anime ”, l'âme n'a pas ici de connotation spirituelle : il s'agit d'une propriété “ naturelle ” des organismes vivants. Quant à la notion d'esprit ^[2], elle est, en

[1] Pour l'hylémorphisme développé par les scolastiques à partir de la philosophie d'Aristote, les êtres corporels que nous sommes sont composés de matière “ première ”, principe quantitatif d'individuation et puissance de déterminations (précédant la matière “ seconde ”, la matière physique concrète) et de forme, principe qualitatif spécifiant les opérativités de l'être corporel concret. Matière “ première ” et forme sont des composantes complémentaires n'ayant de réalité que conjointe.

[2] Selon Husserl le rôle dévolu à l'esprit doit être distingué du pôle psychique qui est celui de l'âme. Matière et âme font partie de la nature tandis que l'esprit forme “ un monde ”. Dans ce contexte, la chair est l'unité primordiale du corps et de l'âme, unité psycho-physique, unité charnelle, non pas union du corps et de l'âme. La chair désigne ici non le corps physique mais le corps vécu et c'est par cette chair que la réalité psychique se constitue. C'est bien d'unité qu'il s'agit et non pas d'unification d'entités substantielles préalablement séparées. La “ chair ” désigne précisément cette unité préalable à toute distinction. C'est pourquoi l'âme est liée, non pas au corps (réalité physique matérielle), mais à la

français, beaucoup plus ambiguë. Le même mot désigne, on le sait, à la fois le mental, la cognition (parfois réduite à la computation), c'est-à-dire le mind, et le principe même de l'être connaissant, le pneuma ou spirit. Pour compliquer les choses, l'esprit dans sa première acception (mind) peut désigner soit l'opérateur (le système de traitement), soit les opérations (la mise en œuvre de règles d'inférence), soit les contenus de ces opérations (les connaissances). Pour la philosophie classique, comme pour le sens commun, un "esprit" (spirit), c'est ce qui a de la volonté, des projets, une mémoire, une conscience [3]. Inversement, pour un matérialiste conséquent la matière se définit comme ce qui n'a pas "d'esprit", c'est-à-dire qui ne possède ni volonté, ni projets, ni mémoire, ni conscience.

1.2. PEUT-ON DEFINIR LA CONSCIENCE ?

Selon le psychiatre André Bourguignon, le problème de la conscience est mal posé, et cela pour deux raisons principales. D'abord et avant tout, parce que la conscience que nous avons de nos perceptions, pensées ou sentiments nous les fait éprouver comme indépendants de notre corps et notamment de notre cerveau, contrairement à ce que nous éprouvons lors du fonctionnement d'autres organes comme les muscles striés dont nous sentons la contraction, ou tout autre organe comme le cœur ou le tube digestif. Nous n'avons aucune connaissance directe de ce qui se passe dans notre propre cerveau lorsque nous pensons (voir § 3.5). Ce que la conscience a de désarmant, constate également G. Edelman, "*c'est qu'elle ne semble pas relever du comportement. Elle est, tout simplement, toujours là*".

Pour le philosophe J. Searle, au contraire, la définition de la conscience "n'est pas un problème sérieux". La conscience, selon lui, "fait référence à ces états de sensibilité et de connaissance immédiate qui commencent généralement lorsque nous nous éveillons [] et qui se poursuivent jusqu'à ce que [] nous devenions, d'une manière ou d'une autre, inconscients". La conscience ainsi définie est un état qualitatif interne, à la première personne. Toutefois, la conscience comme telle n'implique pas nécessairement la conscience immédiate de soi. En revanche, tous les phénomènes conscients sont des expériences subjectives qualitatives, et sont en conséquence des qualia. Searle insiste sur le fait qu'il n'y a pas deux types de phénomènes, la conscience et les qualia. Il n'y a que la conscience, qui est une série d'états qualitatifs.

Dans la perspective cognitive (voir § 3.1) les processus de traitement sont non seulement "cognitivement impénétrables" mais nécessairement non accessibles au soi conscient. Nous n'avons pas accès au traitement des symboles mais seulement au résultat. Il n'existe pas de connexion nécessaire entre l'esprit (computationnel) et la conscience. On peut même dire que l'essentiel des processus cognitifs échappe à la conscience (voir note 8).

En fait, il n'existe pas de sens premier de la conscience mais plutôt un "réseau de métaphores". On peut y voir un principe causal [explicatif] du même type que la gravité. La conscience est toujours déjà là, comme "*ce qui va sans dire*". Faut-il alors comme Wittgenstein le suggère, négliger "*ce qui va sans dire*". Une telle négligence, selon l'auteur du *Tractatus*, a seule

chair (au corps vécu) porteuse de l'unité phénoménologique première. Cette chair joue le rôle d'un agent unique de constitution du corps et de l'âme en unité psycho-physique. Le corps n'est "plein" d'âme qu'en tant qu'il est chair, unité vécue. Pour Husserl, le spirituel n'est pas dégagé du psychique mais en quelque sorte rabattu sur lui : âme et esprit sont *incarnés* (dans la "chair") et non *ajoutés* au corps.

[3] La tradition biblique et l'eschatologie chrétienne expriment fortement la complémentarité d'une diversité de point de vue sur l'unique réalité qu'est la *personne* irréductible à l'interaction de deux principes séparés. La personne unifiée et unifiante s'exprime de manière différente à travers ses divers attributs que l'on désigne par des concepts comme la chair, le cœur, l'âme ou l'esprit.

“ le pouvoir de faire ressortir en négatif ce qu’est la conscience, en suggérant que toute tentative pour la désigner pointerait vers quelque chose qu’elle n’est pas ”.

1.3. TROIS “ DIMENSIONS ” DE LA CONSCIENCE

Comme le souligne J. Searle, la conscience désigne, d’abord le fait d’être éveillé ou endormi, attentif ou distrait, c’est-à-dire plus ou moins *vigilant*. Or, il n’est pas rare que dans un même ouvrage la vigilance soit définie comme le fait d’être *conscient*. La réactivité du sujet aux stimulations extérieures est ici l’un des critères majeurs de l’état conscient. L’étude objective de la conscience en tant que niveau de vigilance a été grandement facilitée par l’évolution des techniques d’exploration fonctionnelle (électroencéphalographie et imagerie cérébrale [4]).

La conscience désigne également *ce dont un sujet a connaissance* : pensées, sentiments, perceptions, rêves, raisonnements. Elle est, dans ce cas, synonyme de *présence à soi et au monde*. Elle exprime l’intuition immédiate qu’a le sujet de ses propres actes. Il s’agit (en principe) d’une intuition claire et rationnelle qui se manifeste notamment dans la saisie immédiate de l’objet [5]. Cette conception de la conscience procède d’une théorie de la connaissance qui suppose donné ce qu’elle prétend expliquer. Une telle approche n’est pas suffisante car elle ignore la temporalité et nie le processus même d’élaboration de la conscience réfléchie. D’autre part la conscience d’objet est inséparable de la conscience de soi [6]. Enfin l’irruption de l’inconscient (cérébral avant d’être psychanalytique) [7] dans le champ de la rationalité consciente a mis un terme à l’idéal de

[4] A titre d’exemple, on peut citer les récents travaux sur l’état de repos conscient. Cet état correspond à une activité mentale effectuée sans entrée perceptive et ni sortie motrice et qui n’est pas dirigée vers un but. On a attribué à cet état mental le nom de “random episodic silent thinking” (REST) que l’on peut traduire par “pensée silencieuse épisodique aléatoire” (PSEA, prononcer psi). Pour certains, cet état cérébral peut être vu comme un état mental “brownien”, pour d’autres, il s’agit au contraire d’un état structuré. Les techniques d’imagerie fonctionnelles et la procédure de méta-analyse permettent aujourd’hui une approche “objective” de cet état de repos conscient. Il s’agit de comparer un grand nombre de tâches de références à l’état de repos conscient considéré comme tâche d’intérêt. Plus les tâches de références sont variées et nombreuses, plus faible est la probabilité qu’elles partagent un processus commun susceptible d’engendrer systématiquement une désactivation lorsqu’elles sont comparées à l’état de repos. Les résultats montrent que l’état de repos conscient met en jeu un réseau d’aires localisées principalement dans l’hémisphère gauche. Ce réseau se superpose en grande partie à celui qui est désactivé durant l’état végétatif persistant et pendant le sommeil paradoxal, ce qui renforce l’interprétation selon laquelle il s’agit bien d’un réseau correspondant à des processus actifs sous-tendant la pensée consciente. Voir également Lambert, 1987.

[5] On peut distinguer ici une conscience *perceptive* et une conscience *réflexive*. La conscience *perceptive* est associée à une expérience qui lui donne un contenu à un moment donné (une couleur, une forme, une émotion, etc.). Elle présente ainsi un caractère *subjectif* auquel est associé le concept de *qualia*. Quant à la conscience *réflexive* elle correspond à la conscience de soi, à la faculté que nous avons d’attribuer des pensées à autrui et à les interpréter, à la connaissance que nous avons de notre propre histoire.

[6] Selon W. James “La conscience humaine est un flux continu d’expériences mentales qui n’entretiennent pas de relations nécessaires avec les événements extérieurs”. Le syndrome de perte d’auto-activation psychique, décrit par D. Laplane (2001), montre que la conscience n’est pas d’abord la conscience de quelque chose mais une présence à soi-même. Des sujets souffrant de lésions des circuits passant par les noyaux gris de la base présentent une totale inertie lorsqu’ils sont abandonnés à eux-mêmes mais redeviennent normaux, ou quasi-normaux, dès qu’ils sont activés par une interaction avec leur entourage. Les patients atteints d’un tel déficit d’auto-activation affirment que dans les périodes de solitude ou d’abandon à eux-mêmes ils présentent un état de conscience “vide”, très analogue à ce que décrivent les méditants orientaux. Ce syndrome montre bien que la computation est indépendante de la conscience qui n’ajoute rien au traitement de l’information. De ce fait, comme le souligne D. Laplane, elle n’ajoute rien à la capacité de survie de l’individu ou de l’espèce et le stéréotype Darwinien est inefficace pour l’interpréter. L’existence d’un tel syndrome confirme également que la question de la conscience est irréductible à celle des modalités de traitement des contenus. La conscience ne s’identifie pas à la connaissance (voir note 8).

[7] Voir Gauchet, 1992.

“ transparence ” du sujet. Il y a comme une altérité interne en chacun, quelque chose qui pense sans moi. L'inconscient apparaît alors comme “*le chaînon manquant entre l'âme et le corps*” (P.L. Assoun).

Le terme conscience renvoie enfin au sens ultime de l'action lorsqu'il s'agit de la *conscience morale*. Le point de vue moral considère la conscience comme rapport à l'autre. La normativité est une dimension nécessaire de cette conscience morale telle qu'elle s'exprime dans le jugement éthique. Il s'agit d'une réalité axiologique qui dépasse largement le cadre des sciences de la nature.

1.4. CONSCIENCE ET DUREE

Une hypothèse générale a été proposée par le neurologue américain B. Libet selon laquelle la plupart des expériences conscientes requièrent une période minimale substantielle d'activation corticale pouvant atteindre 500ms pour une stimulation juxta-liminaire [⁸]. D'autres travaux semblent indiquer que l'élaboration du sens des mots requiert une durée du même ordre de grandeur (400-600ms). Des périodes d'activation corticales plus courtes peuvent cependant produire des opérations mentales inconscientes [⁹]. Un déterminant cérébral majeur de la différence entre événements mentaux inconscients et événements conscients pourrait bien être la durée des activités neuronales appropriées. Ceci justifierait que la plupart des opérations puissent être réalisées inconsciemment et rendrait compte du fait que le contenu de l'expérience subjective puisse être inconsciemment modifié pendant le délai substantiel nécessaire à son élaboration corticale (voir Lambert, 2000).

[⁸] Les travaux de Libet montrent qu'il existe un délai important (plusieurs centaines de millisecondes) entre le moment où une stimulation sensorielle parvient au cortex et le moment où l'activité neuronale est suffisante pour que le sujet fasse l'expérience consciente de cette stimulation. Cependant, d'autres résultats indiquent que la perception du stimulus périphérique est subjectivement *référée* au moment où le message sensoriel arrive au cortex (ce dont témoigne le potentiel évoqué primaire), soit seulement 15 à 25 msec. après la stimulation. Libet parle d'*antédatage* pour caractériser ce processus de *projection rétrograde* de la conscience perceptive. Subjectivement la sensation cutanée apparaît donc sans aucun délai significatif en dépit du fait que l'activité neuronale requise pour provoquer cette sensation ne sera atteinte que plusieurs centaines de millisecondes plus tard. Un tel décalage subjectif (*référation*) dans la dimension temporelle est analogue à celui constaté dans le domaine spatial. Ces deux modes de *référation* servent à projeter l'image subjective sur les caractéristiques spatiales et temporelles des stimulus réels, bien que la représentation neuronale distorde à la fois le pattern spatial et le déroulement temporel du processus.

[⁹] Des images dont on n'a pas conscience sont néanmoins perçues par l'organisme. Un stimulus en deçà du seuil perceptif (dont l'intensité ou la durée sont inférieures au seuil absolu) bien que déclaré non perçu par le sujet, peut être pris en compte au niveau infraconscient et influencer de manière déterminante les perceptions et conduites consécutives. On parle de perception sous-liminaire (infraliminaire ou subliminaire) ou encore de subception. Ainsi, un mot ou une image présentés de manière subliminaire peuvent faciliter ou perturber la perception d'un autre mot ou d'une autre image présentés consécutivement après un bref délai : on parle d'amorçage figural, lexical ou sémantique. Le sujet “voit sans percevoir”. Le premier stimulus furtif, n'en est pas moins traité cognitivement et affectivement comme en témoigne le déclenchement de réponses électrodermales (RED) par certains stimulus à forte valeur émotionnelle dont le sujet (conscient) déclare pourtant ne rien savoir. En outre, malgré une telle absence de discrimination consciente, les potentiels évoqués corticaux (précoces et tardifs) corréleront fidèlement les stimulus infraliminaires ce qui signifie que ces stimulus sont effectivement traités à haut niveau par le système (sub) cognitif. De nombreux résultats confirment l'effectivité d'une telle conscience implicite (masquage rétrograde, écoute dichotique, “vision aveugle”, etc.).

1.5. DETECTION ET CONSCIENCE

L'étude objective de la conscience dans sa seconde acception (conscience-présence, § 1.3) reste problématique puisqu'elle implique de prendre en compte le témoignage des sujets et donc de recourir à l'introspection. L'expression du vécu conscient suppose la médiation du langage et/ou celle d'indicateurs comportementaux (comme la posture, les mimiques faciales ou la gestuelle) ou physiologiques corrélant notamment l'activité émotionnelle (comme les variations de conductance de la peau ou celles du rythme cardiaque). Mais de tels indicateurs ne sauraient donner accès à l'expérience subjectivement vécue. Comme le fait remarquer B. Libet : *“La conscience ou l'expérience subjective ne sont effectivement accessibles qu'introspectivement par le sujet éprouvant l'expérience en question [] Même une connaissance complète de la représentation neuronale ne saurait, sans être validée par le témoignage du sujet, nous dire quelle sensation est en train d'être subjectivement vécue”*.

Une expérience du même B. Libet montre clairement qu'il ne faut pas confondre *détection* et *conscience* (la détection adéquate n'est donc pas un critère de la conscience). Cette expérience est menée sur des patients présentant des douleurs rebelles à tout traitement pharmacologique, munis d'électrodes intracrâniennes implantées, pour des raisons thérapeutiques, dans le noyau ventrobasal du thalamus. Des trains de stimulation de durée inférieure ou supérieure au seuil de perception consciente défini lors des expériences précédentes, sont délivrés unilatéralement par l'intermédiaire de ces électrodes intra-thalamiques. Après chaque stimulation on demande au patient, dans le cadre d'un protocole de choix forcé, d'indiquer de quel côté il a été stimulé, indépendamment du fait qu'il ait ou non effectivement ressenti quelque chose (même quand il dit n'avoir été conscient de rien). Les résultats montrent que le taux de détection correcte est très supérieur à 50% même avec des durées de stimulation trop brèves pour produire une expérience consciente. Ceci confirme l'existence de processus cognitifs inconscients ou préconscients conditionnant l'élaboration de la détection consciente [10].

1.6. D'IMPOSSIBLES CRITERES

A quoi peut-on reconnaître qu'une entité possède un esprit conscient ? Quel est le critère de démarcation de la pensée ou plus exactement quelle sont les conditions de possibilité d'un énoncé tel que : *“cette entité pense et/ou est consciente”* ? Comme le souligne J. Bouveresse *“la pensée ne s'exprime adéquatement dans aucun exemple d'activité pensante [] elle est une sorte de point de vue sous lequel nous acceptons ou refusons de considérer cette activité”*. Nous n'accédons à l'intériorité d'autrui que par une inférence que nous effectuons à partir de son comportement. La décision concernant l'existence ou non d'une pensée ou d'une conscience est principalement de nature éthique.

L'identité opérationnelle entre systèmes est-elle suffisante pour affirmer leur identité *“ontologique”* ? Suffit-il qu'une entité *“se comporte comme”* un humain pour lui attribuer des qualités propres à ce dernier ? Sommes-nous seulement ce que nous faisons ? Selon Hofstadter et

[10] L'existence d'une forme de détection implicite est également suggérée par le fait que de nombreux patients sont capables, dans certaines conditions (notamment sous hypnose) de rapporter des propos tenus pendant une intervention chirurgicale sous anesthésie générale ou de rapporter certains incidents peropératoires. On peut également s'interroger sur la véritable nature de *“l'inconscience”* du coma. Il a été montré, en effet, que la présentation de visages familiers à des patients inconscients, qui se trouvent dans un état végétatif persistant, active le cortex à la jonction occipito-temporale correspondant à *“la zone du visage”* (observation rapportée par A. Damasio). Divers témoignages semblent bien aller dans le sens d'une certaine permanence du psychisme inconscient comme le pensait Freud (voir Hoppenheim-Gluckman).

Dennett, il ne s'agit pas de savoir, par exemple, si une machine souffre, mais si elle se comporte "comme si" elle souffrait. Et d'ajouter que nous procédons de même entre nous. A. Turing soutient également que la question n'est pas de savoir si une machine pense ou est intelligente mais si son comportement est indiscernable de celui d'un humain dans les mêmes conditions. Le réalisme de l'imitation et la pertinence du comportement sont ici des critères suffisants. "Si un robot astucieusement conçu pouvait (avoir l'air de) nous raconter sa vie intérieure (émettre les sons appropriés dans les contextes appropriés) devrions-nous l'admettre dans notre caste?" (D. Hofstadter). Carnap considère comme légitime de conclure à la possession de sentiments, de pensées, de souvenirs, de perceptions par une entité, à partir d'un comportement extérieur déterminé. Cette inférence est tout aussi légitime, pense-t-il, que celle qui permet de conclure à la valeur de l'intensité d'un courant électrique dans un fil à partir de grandeurs mesurées comme l'élévation de la température du fil ou la déviation d'une aiguille aimantée placée à son voisinage. Il n'est plus question "de compréhension, de perceptions ou de pensées mais de faire celui qui a compris, qui voit ou qui pense".

Au demeurant, même si un homme et un automate satisfont aux mêmes épreuves ce n'est pas ce que l'on voit mais ce que l'on sait de leur comportement qui nous fait en décider. Aucun comportement ne peut être la preuve de la conscience et inversement, l'absence de comportement n'implique pas nécessairement l'absence de toute forme de "conscience".^[11].

2 - CERVEAU ET CONSCIENCE : RAPPEL DES POSITIONS CLASSIQUES

2.1. LES MONISMES

Parmi les formes de monisme matérialiste la thèse de l'identité forte préconisée par Feigl et Amstrong affirme bien l'existence d'un principe interne de la conduite (le mental), mais ce principe n'est rien d'autre que le cerveau. Il y a identité psychoneurale. Pour ce matérialisme physicaliste le mental est réductible au biologique qui est lui même réductible à la physique [12].

Selon la thèse de *l'identité faible*, défendue notamment par les fonctionnalistes, à chaque état mental correspond bien un certain état neurophysiologique mais à chaque catégorie d'état mental ne correspond pas nécessairement une catégorie d'états neurophysiologiques particuliers. L'identité est

[¹¹] Le coma est un état "d'inconscience" durable, parfois définitif. La définition du coma reste purement comportementale. L'évaluation neurologique du patient comateux se fonde sur un examen clinique (le Score de Glasgow), sur les données de l'électroencéphalographie (EEG et potentiels évoqués) et sur celles de l'imagerie structurale et fonctionnelle qui permettent notamment d'évaluer la réactivité du cortex cérébral. Pour Plum et Pozner auxquels on doit la première description du "*locked in syndrome*", les limites de la conscience sont difficiles à définir de façon satisfaisante car elle "*ne peut être appréciée chez autrui qu'à travers l'apparence de l'activité*". L'équivalence entre malade inconscient et sujet absent est d'ailleurs discutable. Un jugement de la 20^{ème} chambre correctionnelle en date du 30 mars 1989 reconnaît que "*Rien n'établit que la victime [dans le coma] n'a pas conscience de sa misère*" (voir Oppenheim-Gluckman).

[¹²] Beaucoup pensent désormais, comme J.P. Changeux, que "*l'identité entre états mentaux et états physiologiques ou physico-chimiques du cerveau s'impose en toute légitimité*" et que "*l'homme n'a dès lors plus rien à faire de l'esprit*" puisque "*il lui suffit d'être un homme neuronal*". Ainsi, pour J.D. Vincent il n'y a pas de différence entre "*ce qui serait de l'ordre de la pensée et de celui du cerveau*". Quant à Francis Crick, il estime que nous ne sommes "*rien d'autre qu'un paquet de neurone*" et que toutes nos activités mentales, nos affects, nos désirs, que "*tout cela est un truc de neurones*". Selon D. Hofstadter "*l'esprit humain est un objet physique*" ou encore pour Minsky "*la pensée est ce que font les cerveaux*" et "*La conscience s'identifie au domaine unifié de toutes les lois de la nature*".

occasionnelle, elle n'est pas catégorielle. Cette théorie autorise une certaine autonomie de l'explication psychologique à l'intérieur d'une ontologie strictement moniste (monisme non physicaliste).

Il en va de même des théories de “l'émergence” qui revendiquent un physicalisme non réductionniste. Pour *l'émergentisme* les entités mentales transcendent le physiologique, tout en émanant de lui. Le mental *émerge* de la complexification progressive des systèmes matériels auxquels il est cependant irréductible [13].

Dans cette classe de modèle, on peut inscrire également le paradigme de l'*autopoïèse* (ou de la *clôture opérationnelle*) développé par F. Varela et H. Maturana. Ce paradigme, inspiré de la phénoménologie et des traditions orientales, cherche à sortir les sciences cognitives des impasses de la computation. Il s'agit principalement, pour Varela, de comprendre comment se constitue un système autonome et comment il est couplé à son environnement. Un tel système est caractérisé par sa forte détermination interne. Les *comportements propres*, qui correspondent à des *points fixes* de sa dynamique, peuvent être couplés à des événements (des sollicitations de l'environnement) mais à la différence des modèles habituels ces événements ne constituent pas des *inputs*. Il n'existe pas de correspondance bi-univoque entre des événements externes et une quelconque *représentation* interne de ces événements. Le stimulus est ici l'agent d'une perturbation ou d'une modulation de la dynamique des comportements propres (comme le suggère l'exemple célèbre de l'embouteillage dans la métaphore du trafic automobile). Alors que l'input détermine l'état de la structure de façon univoque et prévisible (représentation par correspondance), la perturbation déclenche dans le système des adaptations dont la direction et le résultat dépendent de l'état actuel du système lui-même. Il en résulte qu'un observateur peut difficilement prédire les conséquences d'une perturbation, d'autant que les événements qu'il considère, lui, comme identiques, ne sont pas forcément perçus comme tels par le système et peuvent donc avoir - sur ce dernier - des effets forts différents. Dans ce contexte, à l'opposé de la logique de la représentation chère au cognitivisme standard (voir § 3.1) “*le système nerveux est défini essentiellement par ses divers modes de cohérence interne, lesquels découlent de son interconnectivité*”. Ainsi, “*la cognition n'est plus considérée comme une résolution de problèmes s'appuyant sur des représentations; la cognition dans son sens le plus vaste consiste plutôt en l'enaction ou le faire émerger d'un monde par le biais d'une histoire viable de couplage structurel*”. Nous sommes ici en présence d'un processus d'émergence de formes fondé sur la notion de co-détermination. L'unité fonctionnelle émerge d'un environnement qui se constitue (pour elle) en même temps qu'elle en émerge.

Le *monisme neutre* qui postule que le mental et le physique sont deux manifestations d'une unique substance [14] neutre (ni physique, ni mentale) a été particulièrement défendu par Russell et Carnap. Cette ontologie fait grand cas du principe de complémentarité en physique quantique et rejoint certaines intuitions des philosophies orientales. La physique quantique est en effet une source d'inspiration féconde pour ceux qui, comme R. Penrose cherchent une “*explication physique appropriée*” au phénomène de la conscience. Selon Penrose, la conscience résulterait de l'interaction, au niveau du cytosquelette des neurones, entre les activités de nature microscopique et macroscopique. L'esprit en tant que processus physique serait le point de rencontre des niveaux de description classique et quantique. Le niveau neuronal de description pourrait n'être que l'*ombre* d'un niveau plus profond où il convient de chercher le substrat physique de l'esprit. La description

[13] Le prix Nobel Roger Sperry considère le cerveau comme un prodigieux générateur d'émergence de phénomènes qui, à leur tour, exercent un contrôle sur les activités de niveau inférieur. Dès qu'ils sont engendrés par les événements neuronaux, les programmes mentaux acquièrent leurs propres qualités subjectives, opèrent et interagissent selon leurs propres lois irréductibles à celles de la neurophysiologie. Les entités mentales sont molaires, configurationnelles et transcendent le physiologique tout en émanant de lui. La conscience est globale, opérationnelle et fonctionnelle alors que les processus physiologiques n'ont pas une telle unité. De même, l'expérience subjective n'est pas corrélée à la mise en route des éléments excitateurs et inhibiteurs formant l'infrastructure des processus cérébraux mais à des propriétés holistiques dites mentales.

[14] Voir note 22

de ce niveau plus profond suppose une refonte de la physique actuelle et le recours au concept de non-calculabilité. Si donc la conscience est irréductible au calcul mais pas à la physique, la physique de l'esprit ne peut qu'échapper au calcul. L'explication physicaliste de Penrose suppose l'existence d'un insu non calculable qui précisément est irréductible à la physique. Comme le souligne M. Bitbol “chez Penrose [] ce sont les autolimitations de la physique, plus encore que ses inachèvements, qui ouvrent des perspectives insoupçonnées au physicalisme”.

2.2. DE LA CIRCULARITE DES ENONCES

A propos de toutes ces tentatives de réduction de l'esprit “au domaine unifié de toutes les lois de la nature” (Minsky, voir note 11) il est important de souligner que toute hypothèse prétendant que lois ordinaires de la physique suffiraient à rendre compte adéquatement de la réalité du sujet, débouche sur une évidente circularité puisque les lois en question supposent un sujet - ou pour le moins un agent cognitif - pour être pensées. *L'énoncé de la loi présuppose l'existence de ce qu'elle est censée expliquer.* Comme le souligne le physicien F. Dyson “les lois de la physique sous-atomique ne peuvent être formulées sans référence à l'observation [] Ces lois réservent une place au rôle de “l'esprit” dans la description que l'homme fait de chaque molécule”. Il n'existe pas, en effet, de lois dans la nature mais des lois de la nature qui supposent un agent cognitif pour être énoncées. L'assimilation de la pensée, de la conscience, du sujet, aux lois de la nature - ou plus précisément à ce que nous en savons - est donc fatalement tautologique. C'est avec notre cerveau que nous créons des catégories dans un monde qui n'en possède pas sauf celles déjà créées par l'homme. Il n'y a rien qui émerge s'il n'y a personne pour le signifier ou alors tout est émergence, ce qui ne signifie plus rien.

Quoiqu'il en soit, comme Georges Canguilhem (cité par Debru) le faisait observer à J.P. Changeux, lors d'un débat devant la Société Française de Philosophie en 1981, “on peut, on doit tout accorder au physiologiste en tant qu'auteur et véhicule d'un savoir relatif au cerveau comme objet d'investigation expérimentale. Mais on peut et on doit lui demander de renoncer à prétendre se promener dans la rue de son cerveau pour comprendre comment il se fait qu'il se trouve à la fenêtre”.

2.3. L'ALTERNATIVE DUALISTE

Face aux difficultés - voire à l'impossibilité - de rendre compte adéquatement de la réalité phénoménologique de la conscience subjective avec les modèles monistes, le dualisme offrirait-il de nouveau une alternative crédible ?

Le *parallélisme* cherche à rendre compte des corrélations psycho-physiologiques en postulant que le mental et le cérébral se correspondent mutuellement sans agir aucunement l'un sur l'autre. On évite ainsi le délicat problème de l'interface mais se pose alors la question de savoir comment garantir la correspondance entre états cérébraux et états mentaux. Leibniz répond avec la thèse de *l'harmonie pré-établie*. Dans sa version vulgarisée naïve, le parallélisme correspond assez bien à la croyance spontanée du psychophysiologiste mais il ne permet pas d'aller très loin dans l'exploration des corrélations entre fonctions psychiques et activité cérébrale. Considéré par Bergson comme une “*illusion philosophique*”, le parallélisme ne pourrait constituer, selon Marc Jeannerod (cité par Debru), qu'une “*philosophie d'attente*” qui ne résout rien. Ainsi, selon Claude Debru “*Toute tentative de donner au parallélisme un contenu psychophysiologique précis se heurte [] au double obstacle formé par l'imprécision des catégories psychiques et par la difficulté de donner une interprétation fonctionnelle autre que global des états ou événements neurophysiologiques correspondants*”. Au demeurant, un tel “parallélisme global” semble bien

établi. On peut d'ailleurs envisager les rapports entre l'organe psychique et l'organe cérébral en termes topologiques en soutenant qu'ils ne sont pas absolument superposables. Cette proposition n'a de sens qu'à condition de concevoir ces organes comme des *espaces de fonctions* et non comme des structures. Il s'agit d'*hyperespaces* difficilement représentables. Ainsi les corrélations psychophysiques ne reposeraient plus sur la synchronisation de deux horloges comme l'envisageait Leibniz. Toutefois, pour G. Canguilhem (cité par Debru) “ *Il faut abandonner l'idée qu'il y aurait dans le cerveau, quelque part, une sorte de surface topologique où se rejoindraient, par torsion et continuité, un endroit objectif et un envers subjectif* ”. Canguilhem récuse l'idée de représentation topologique de la corrélation psychophysique.

L'*interactionnisme* représente la thèse la plus proche du sens commun : le mental exerce bien une action causale sur le cérébral et, inversement, le cérébral exerce ses propres contraintes sur le mental. Il y a *interaction* entre les deux entités. Le prix Nobel de médecine Sir J. Eccles et l'épistémologue Sir K. Popper sont, parmi les auteurs contemporains, ceux qui ont le plus contribué à actualiser et à tenter de valider cette hypothèse [15]. Pour Eccles, l'unité de l'expérience consciente dépend de l'esprit et non de la machinerie neuronale. Les processus mentaux transcendent les phénomènes matériels. L'expérience consciente et les phénomènes mentaux s'intègrent dans une organisation hiérarchique avec un contrôle descendant du mental sur les entités neuronales. En outre, ces deux catégories de phénomènes correspondent à des régulations impliquant des lois et des “ forces ” de nature différente. Enfin, l'ensemble s'inscrit dans la logique de la théorie de l'évolution.

Les critiques qui ne manquent pas d'être formulées à l'encontre du modèle de Sir John ne sont pas seulement le fait de matérialistes. En effet, ou bien nous sommes dans le domaine de la science expérimentale et nous n'avons pas à faire référence au mystère, ou bien nous prenons la mesure d'un au-delà de la science qui, par nature n'a pas à être justifié selon les critères de cette dernière. N'y a-t-il pas contradiction entre le fait de proclamer, comme J. Eccles, que l'avènement de chacun de nous en tant qu'être auto-conscient reste “ *un miracle à tout jamais au-delà des possibilités d'explication de la science* ” et le fait de vouloir donner une explication physique (même quantique) de ce “ *mystère qui dépasse notre compréhension* ”? Le spiritualisme se nie lui-même quand il cherche à concevoir comme un principe empirique ce qui par principe ne l'est pas.

En fait, aucune de ces deux solutions n'est vraiment satisfaisante. La solution interactionniste laisse ouverte la question de la nature du lien de causalité et du point d'impact des influences mutuelles provenant de la conscience et des structures nerveuses. Quant à la solution paralléliste, elle conduit à ne voir dans le psychisme que le reflet ou l'*équivalent* subjectif de l'activité nerveuse. On peut finalement se demander à quoi sert le mental dans ce système puisque la description physicaliste suffit (du fait de l'équivalence cérébral/mental). La position paralléliste conduit donc soit à prendre effectivement au sérieux l'hypothèse d'un programme garantissant la correspondance (et ce que cela implique !) soit à ne considérer que le versant matériel de la correspondance qui se suffit à soi-même et donc à se rallier finalement au réductionnisme physicaliste.

[15] Le neurophysiologiste justifie sa position par un certain nombre d'arguments expérimentaux (comme les expériences de Libet, déjà citées) et fonde son explication sur le rapprochement de la neurophysiologie et de la physique quantique. L'esprit pourrait intervenir sous la forme d'une subtile modification de la probabilité d'émission des neurotransmetteurs au niveau synaptique. Eccles propose d'expliquer une telle modification en se fondant sur la notion de *champ de probabilité* en physique quantique qui ne véhicule *ni matière ni énergie*. La transaction pourrait ainsi n'impliquer aucune violation des lois de conservation.

2.4. D'AUTRES HYPOTHESES

2.4.1. L'isomorphisme (J. Piaget)

Piaget refuse à la fois l'interprétation interactionniste (du fait de l'hétérogénéité des éléments en présence) et l'interprétation paralléliste (du fait que les processus psychologiques n'y ont pas de fonction propre) et propose une explication fondée sur la notion d'*isomorphisme* entre *causalité physique* et *implication logique* [¹⁶]. Il s'agit ici d'une correspondance des structures logiques abstraction faite des contenus et non d'une liaison causale. Il est, selon Piaget, dénué de signification de vouloir appliquer la notion de causalité aux liaisons entre un processus physiologique et la conscience car on ne voit pas sur quel point de cette dernière (puisqu'elle n'est pas spatiale), ni de quelle manière (puisqu'elle ne mobilise ni masse, ni énergie) s'appliqueraient les forces matérielles en jeu dans la causalité physique. La notion de causalité ne s'applique donc pas à la conscience (un état de conscience n'est pas la cause d'un autre). Ainsi, ou bien la conscience n'est rien, ou bien elle relève de catégories originales et spécifiques, qu'ignorent en eux mêmes les faits matériels.

2.4.2. L'hypothèse de transformation (D. Laplane)

Le neurologue D. Laplane a proposé une interprétation originale de la relation cerveau-pensée-conscience fondée sur une hypothèse de transformation postulant le passage d'une forme de réalité (matière-énergie) à l'autre (pensée-conscience). Pour D. Laplane en effet, on ne peut affirmer simultanément l'homogénéité et l'hétérogénéité de ces deux formes de réalité qu'en supposant l'identité des termes (matière-énergie-pensée) ou en postulant leur transformation réciproque (voir Laplane, 1987). S'il y a bien identité, l'énergie répandue dans l'univers doit être aussi nécessairement pensée. “*La pensée (présence à soi) existe en dehors du cerveau et se trouve diffuse dans l'univers qui est présent à lui-même*”. La pensée est “inhérente” à la matière-énergie. C'est la thèse défendue par différents physiciens dont J.-E. Charon et avant eux par Spinoza. Mais comment le fonctionnement de notre cerveau pourrait-il interférer avec cette pensée universelle ? D. Laplane préfère donc à la thèse de l'identité matière, énergie, pensée, celle de la transformation de l'énergie en pensée et réciproquement [¹⁷]. L'axiome de transformation constitue un *a priori* qui n'est ni plus

[¹⁶] Le mode essentiel de liaison propre à la conscience logique est l'implication. La vérité $2+2=4$ n'est pas la cause de la vérité $4-2=2$ mais la première *implique* la seconde. La valeur d'un but implique celle des moyens. Cette implication des valeurs double dans la conscience la série causale des actions « de moyen à but » sans s'y réduire pour autant. La relation entre significations est, sous une forme très générale une *relation d'implication*. Pour Piaget l'implication logique ne constitue qu'un aboutissement particulier de l'implication au sens large ou implication signifiante qui commande la relation de signifiant à signifié intervenant dans les catégorisations ou désignations, dès le niveau perceptif. Or, ni le lien entre significations, ni la relation de signifiant à signifié ne relèvent de la causalité. Le mode de connexion propre aux phénomènes de conscience n'est autre que cette implication signifiante. Si le système nerveux en conditionne l'effectivité, puisqu'il engendre causalement son substrat matériel, il reste qu'il faut une conscience pour juger des vérités et des valeurs. Le parallélisme entre les états de conscience et les processus matériels concomitants revient à un isomorphisme entre les systèmes d'implications signifiantes et certains systèmes relevant de la causalité. On pourrait considérer qu'il s'agit de deux aspects d'une même réalité : un aspect extérieur lié à la causalité et un aspect intérieur lié à l'implication.

[¹⁷] A l'objection selon laquelle une telle transformation violerait le second principe de la thermodynamique, Laplane répond en rappelant que l'expérience d'Oersted violait les principes de conservation de la mécanique newtonienne, puisqu'on ne connaissait pas alors le type de relation énergétique qu'elle mettait précisément en évidence. En s'ouvrant à un nouveau type d'énergie (électromagnétique) la physique a récupéré son principe de conservation. Popper évoque la possibilité “*d'interférence de l'extérieur, à partir de quelque chose d'encore inconnu que nous aurions à ajouter au monde physique si nous désirons que la physique soit complète*”. Quant à l'objection de quantification selon laquelle, puisque la matière et l'énergie sont quantifiables, la pensée devrait l'être aussi, D. Laplane admet qu'il est vain d'imaginer un coefficient de transformation puisque la pensée n'appartient pas au monde objectif. Cependant le fait que cette pensée soit toujours associée à une certaine quantité d'information (à laquelle elle ne se réduit pas) et que la notion même d'information, notion qui est elle-même un produit de notre pensée) soit, bien que totalement abstraite,

ni moins valable que les autres interprétations monistes ou dualistes. Selon son auteur, cette hypothèse permet de dépasser les contradictions tant du matérialisme moniste que du spiritualisme dualiste en reconnaissant au cerveau son rôle de “fabriquant ” de la pensée (fabriquer au sens de mettre en forme, conditionner, donner un contenu objectif et informationnel, pas au sens de créer) tout en faisant du cerveau et de la pensée deux entités nettement distinctes. De plus, si la transformation est réciproque il n'y aurait aucune raison de se scandaliser à l'idée que la matière-énergie puisse provenir d'une pensée.

2.4.3. Le postulat de non-équivalence (S. Watanabe)

Comment construire un langage susceptible de rendre compte de la coexistence de termes mentalistes et physicalistes non-réductibles ? Dans le cadre de la logique ensembliste habituelle de l'algèbre de Boole, si la description physicaliste est complète elle n'a pas besoin d'être enrichie. La question des rapports cerveau-pensée peut être envisagée de manière non exclusive et non additive dans le cadre d'une logique non-booléenne. Dans ce cas la conjonction de deux entités ou états peut être une entité ou un état qui n'est réductible ni à l'un ou à l'autre, ni à l'un et à l'autre. Moyennant certains aménagements de la logique classique (modification de la distributivité, passage d'un treillis booléen à un treillis modulaire), on peut démontrer qu'une description complète au sens physique du terme n'est cependant pas nécessairement adéquate. Selon S. Watanabe, le langage mentaliste dirait sur le langage physicaliste ce qui ne peut être dit dans ce langage. Il a montré que dans une logique “modulaire ”, on peut accepter à la fois et sans contradiction, la prétention à la complétude du discours physicaliste et l'inadéquation de ce discours qui a toujours besoin d'être enrichi par des propositions mentalistes pour être conforme à la réalité qu'est l'homme [18]. L'intérêt d'une telle approche est de permettre une conception psychophysique de la nature humaine sans réduction ni confusion. Matière et esprit sont ici des manifestations irréductibles l'une à l'autre. Bien que la description physicaliste puisse légitimement prétendre représenter la totalité, une telle représentation a toujours besoin d'être enrichie pour rendre compte adéquatement de ce qu'elle représente.

3 – CONFIGURATIONS NEURONALES ET CONSCIENCE

3.1. LES IMPASSES DU COGNITIVISME

Comment le cerveau-esprit computationnel produit-il ou s'articule-t-il avec de l'esprit phénoménologique? Comment des opérations cognitives élémentaires inconscientes peuvent-elles être éprouvées non pas comme telles mais comme un état signifiant du soi conscient ? En fait, le cognitivisme ne propose aucune explication de ce qu'est l'expérience consciente. Selon le cognitiviste Jackendoff, les éléments de la conscience sont le résultat causal projeté des processus

parfaitement mesurable, ne rend pas complètement absurde l'idée que la pensée serait une “grandeur” non mesurable (comme le défend également R. Penrose, revoir § 2.1)

[18] Watanabe utilise la métaphore du plan et de la maison. Le plan peut (doit) fournir une description complète de la maison, mais il doit être enrichi (d'une dimension supplémentaire) pour être adéquat à la réalité qu'est finalement une maison. La démonstration de Watanabe repose sur le rejet du postulat d'équivalence (entre proposition mentaliste et proposition physicaliste). A partir du principe de conjonction restreinte et du postulat de non-équivalence il montre que la conjonction d'une proposition (m) et d'une proposition (p) est une proposition qui n'appartient ni à (m), ni à (p) et que donc une description physicaliste complète peut être enrichie. Ainsi, il n'existe pas de description physicaliste complète *et* adéquate. Une description physicaliste prétendument complète pourra toujours être enrichie par la conjonction non booléenne du physique et du mental puisque le cerveau et la pensée ne “commutent” pas.

computationnels. Mais que veut dire “projeté”? Ici le cognitivisme a besoin de l'évidence phénoménologique. Or, cela ne va pas de soi car il n'est pas du tout certain que quelque chose soit comme nous l'éprouvons.

Selon la thèse *cognitiviste standard*, la conscience est une conséquence de la manipulation réglée de symboles. Le terme de symbole utilisé dans le modèle computationnel ne doit pas faire illusion. Il est utilisé ici en un sens purement technique, désignant une chaîne de caractères assimilée à un signifiant dont il reste à préciser *qui* fournit le signifié. Quelle est précisément la nature de l'opération permettant de passer d'une chose à son symbole ? Un symbole symbolise quelque chose *pour* quelqu'un. Les règles de correspondance sont fournies, de l'extérieur, par quelque “programmeur”. L'analogie avec l'ordinateur montre, certes, comment il est possible de simuler des processus sémantiques par des processus formels, mais elle ne permet pas de comprendre *comment le sens vient aux symboles*. Comme le souligne J. Searle “*La syntaxe [] n'est pas intrinsèque à la physique du système ; elle est dans l'œil de l'observateur [] le calcul n'est pas un processus intrinsèque qui se trouverait dans la nature comme la digestion ou la photosynthèse ; il n'existe que relativement à un agent qui donne une interprétation computationnelle à la physique. Il en résulte que le calcul n'est pas intrinsèque à la nature mais relatif à l'observateur ou à l'utilisateur*”. Il s'ensuit donc que ce dernier ne saurait être lui-même réduit à un calcul. De plus, comme le souligne Wittgenstein, un tel calcul suppose d'être *effectué* : “*Le langage ne se réduit pas à un calcul, au sens d'une manipulation de symboles, sans considération de leur signification. En nous servant du langage nous devons faire quelque chose de la même manière que pour calculer il ne suffit pas de regarder les signes : c'est à nous de faire le calcul*”.

L'approche *connexionniste* ou neuro-mimétique considère au contraire que le raisonnement en termes de symboles est totalement inadapté à la problématique esprit-cerveau. Le neuro-calcul privilégie fortement la nature du substrat (réseau de neurones) et sa structure fonctionnelle (connectivité, poids synaptique...). Le recours aux modèles connexionnistes met en évidence une caractéristique essentielle du cerveau humain : la plasticité de sa structure dont l'organisation est susceptible d'être modifiée par le résultat des opérations qu'elle exécute. On peut simuler cette propriété à l'aide d'un réseau d'automates dont la connectivité (la force des liaisons entre éléments ou *poids synaptique*) se modifie pas à pas en fonction de l'écart mesuré entre le résultat attendu et le résultat effectif grâce à des propriétés de “rétropropagation”. Le traitement s'opère ici à un niveau sub-symbolique.

Dans tous les cas le système n'interprète pas lui-même les symboles ou les configurations qu'il manipule, ce sont le concepteur et, plus tard, l'utilisateur, tous deux extérieurs au système, qui projettent sur ces symboles ou configurations une interprétation. On, voit donc difficilement comment on pourrait accorder à un système informatique une *intentionnalité intrinsèque*. Son intentionnalité apparaît comme une *intentionnalité d'emprunt*, dérivée de celle du concepteur et de l'utilisateur qui lui font “crédit de sens”(voir § 3.3).

3.2. LE CERVEAU NE “VOIT” PAS

On sait, par exemple, qu'il existe une organisation hautement spécifique du traitement de l'information visuelle. Les différentes caractéristiques de l'objet sont traitées en parallèle par des canaux distincts (couleur, forme, mouvement, position, contraste...). Or le sujet qui perçoit, celui pour qui cette activation fait sens (signifie une couleur), n'est réductible ni à des aires du cortex visuel, ni à aucune autre région du cerveau susceptible de les scruter car *ce sont des couleurs* (des objets) *dans le monde* et *non des représentations dans le cerveau* que nous percevons. Le cerveau ignore tout des stimulus qu'il reçoit. C'est *moi* qui perçoit et qui pense et non mon cerveau! Ainsi la vision n'est localisée ni dans la rétine, ni dans le nerf optique, ni dans les corps genouillés, ni dans aucune aire corticale : le cerveau ne voit pas.

Toujours à propos du système visuel, S. Zeki note que *“manifestement il y a plus dans la vision que dans l'œil”*. Celui qui voit n'est *“manifestement”* pas réductible à une subdivision neuro-anatomique. Il en va de même pour la mémoire. Selon F. Crick par exemple *“notre cerveau contient une représentation inactive de la tour Eiffel et quand nous pensons à ce monument, la représentation est activée parce que les neurones appropriés déchargent”* mais ce ne sont évidemment pas les *“neurones appropriés”* qui perçoivent l'évocation de la tour Eiffel. Il existe bien d'autres exemples illustrant l'impossibilité sémantique absolue de rendre compte de quoi que ce soit en matière de perception, de mémoire, d'affects, de cognition ou d'action sans avoir recours à quelque pronom personnel désignant un sujet agissant et conscient. La question de l'interprétation (signification) des configurations ne peut recevoir de réponse non contradictoire à l'intérieur du paradigme cognitiviste. Ce que l'on appelle concept ne fait que renvoyer soit à une série d'états pris par des composants d'une machine électronique ou nerveuse et interprétés comme tel, soit à un énoncé de toute façon produit à l'origine par un humain. Dans les deux cas l'acte de signifier n'est toujours pas expliqué.

Comme l'admet Damasio, l'accès aux images sous leur forme consciente n'est possible *“qu'à partir du point de vue de la première personne”* alors qu'au contraire *“les configurations neuronales ne sont accessibles que du point de vue d'une tierce personne. Si j'avais l'occasion, grâce à une technologie de pointe de contempler mes propres configurations neuronales ce serait toujours du point de vue de la troisième personne”*. Comme nous l'avons déjà souligné, *“même une connaissance complète de la représentation neuronale ne saurait, sans être validée par le témoignage du sujet, nous dire quelle sensation est en train d'être subjectivement vécue”* (B. Libet). A. Damasio reconnaît également qu' *“Il est faux de penser qu'on puisse effectivement comprendre la nature des expériences subjectives en étudiant leurs corrélats comportementaux [] Vous ne voyez pas ce que je vois quand vous examinez mon activité cérébrale ; vous voyez une partie de mon activité cérébrale au moment où je vois ce que je vois [] Je n'ai pas besoin de savoir quoi que ce soit sur le comportement des neurones et des molécules des différentes parties de mon cerveau pour faire l'expérience de la baie de San Francisco”*

Certes ! et tout le monde est bien d'accord là-dessus. Mais alors comment passe-t-on de la configuration neuronale de la baie de San Francisco à l'expérience subjective de l'image mentale associée à cette configuration ? C'est la notion même de représentation (symbolique ou non) qui est ici en cause et qu'il convient de discuter.

3.3. COMMENT UNE “ CONFIGURATION NEURONALE ” FAIT-ELLE SENS ?

Comment une configuration neuronale fait-elle sens ? De quoi et pour qui est-elle une représentation ? Qui l'interprète ? Comment l'adéquation de la représentation au “réel” est-elle garantie ? Comme le souligne D. Dennett *“rien n'est intrinsèquement une représentation de quoi que ce soit ; quelque chose est une représentation seulement pour ou de quelqu'un ; ainsi n'importe quelle représentation ou système de représentation requiert au moins un utilisateur et un interprète externe à cette représentation [] Un tel interprète est alors une sorte d'homoncule. Il s'ensuit qu'une psychologie sans homoncule est impossible tandis qu'une psychologie avec homoncule est condamnée à une régression circulaire ou infinie, ainsi, la psychologie est impossible”*. C'est du signifié que le signe reçoit son importance. Aucun objet, dégagé de la relation sémantique triangulaire, n'est jamais par lui-même un signe, ni naturel ni artificiel.

Antonio Damasio est parmi ceux qui refusent catégoriquement la notion d'homoncule : il suffit, explique-t-il que les états successifs de l'organisme donnent lieu à des représentations neurales donnant elles-mêmes lieu à des images mentales. Les images sont produites à partir des configurations neuronales. Certes, mais le problème de la conscience est précisément de savoir *“comment le cerveau engendre les configurations mentales que nous appelons images à partir de*

configurations neuronales (c'est le problème des qualia) et comment parallèlement le cerveau engendre aussi un sentiment de soi dans l'acte de connaître ” (A. Damasio).

Mais que pourrait-il y avoir d'autre dans le cerveau que des configurations neuronales? Si des configurations neuronales sont perçues subjectivement comme des images, et que dans le cerveau on ne rencontre que des configurations neuronales (et non des images qui supposent d'être interprétées et vécues comme telles), cela implique que le passage d'une configuration neuronale à un vécu subjectif n'est pas une configuration neuronale de plus. Damasio prétend arrêter la régression au niveau des représentations dites de second ordre. “[une] *configuration d'activité peut faire l'objet d'une représentation au sein des structures de second ordre [] Le sens que nous avons de l'émotion comme objet découle de la représentation que nous formons de la configuration d'activité des sites qui induisent l'émotion ”*. Si l'on compare l'élaboration des images mentales au déroulement d'un film, il s'agit alors, selon Damasio, de comprendre comment le cerveau engendre également le sens que l'on a qu'il existe bel et bien un propriétaire et un observateur de ce film. “*Il s'agit d'engendrer l'apparence d'un propriétaire et d'un observateur du film, à l'intérieur du film lui-même [] L'histoire contenue dans les images de la conscience noyau n'est pas racontée par quelque homoncule intelligent. Et l'histoire n'est pas non plus racontée par vous en tant que vous êtes un soi parce que le vous ne naît qu'au moment où l'histoire est racontée, à l'intérieur de l'histoire elle-même ”*

Quelle qu'en soit la forme, la question de savoir comment le sens vient aux symboles (comment on passe d'une configuration neuronale objective à une représentation mentale subjective indépendamment du type de couplage supposé entre l'objet et la configuration neuronale associée), cette question ne laisse finalement le choix qu'entre trois types de réponses : celle de l'homoncule (quelqu'en soit la forme), celle du méta-niveau censé arrêter la régression et celle de “l'incomplétude ” [19].

Certains postulent, en effet, l'existence d'un “*méta-niveau inviolable ”* qui peut prendre la forme d'un “*module superviseur ”*, d'un “*système de fixation des croyances ”* (Fodor) ou encore d'un “*générateur d'hypothèses ”* (Gazzaniga), tous plus ou moins “*cognitivement impénétrables ”* (Pylyshyn). Mais alors, ou bien une telle instance est de nature différente de toutes les autres et on sort du paradigme cognitiviste, ou bien elle est de même nature (elle traite de l'information) et il ne s'agit que d'une représentation d'ordre supérieur qui ne répond pas davantage à la question de l'interprétation. On peut affirmer avec Cummins qu'un “*crédit de sens ”* est nécessaire à l'origine pour amorcer, en quelque sorte le processus d'interprétation. Certains ajoutent même que ce pseudo-cognitivism devra se démettre s'il se révèle incapable d'honorer sa dette! Il est clair pour Fodor que les processus cognitifs centraux ne sont pas modulaires et “*c'est - dit-il - une bien mauvaise nouvelle pour les sciences cognitives ”*. C'est pourquoi, poursuit-il, “*il ne faut pas se faire beaucoup d'illusions sur les perspectives d'une neuropsychologie de la pensée ”*. Quant au crédit de sens dont parle Cummings, on connaît bien l'identité du créancier dans le cas de l'ordinateur mais s'il en va de même pour l'esprit humain, qui lui a fait crédit ? et de quoi au juste lui a-t-on fait crédit ?

3.4. UN ABUS DE LANGAGE

Une manière courante de résoudre la question de savoir comment et pour qui les configurations neuronales font sens est d'attribuer à tout ou partie d'un ensemble de structures en interaction, voire à une structure unique, le pouvoir suprême d'interprétation que l'on refuse précisément à l'homoncule. Paradoxalement on retombe alors dans une forme d'animisme ou de vitalisme qui attribue à la matière cérébrale les propriétés caractéristiques de l'esprit (intention,

¹⁹ Voir Lambert, 1991

désir, volonté, conscience). Parler de “ *la manière dont le cerveau est attentif à l’image d’un objet* ” est typiquement un abus de langage. Comme le souligne le philosophe allemand E. Straus, “ *on peut difficilement parler de la compréhension qu’exerce sur lui-même le cortex [] Ce n’est pas un cerveau qui évalue mais quelqu’un [] Entre deux processus nous pouvons saisir des relations temporelles, mais un processus ne sait rien de l’autre. Un processus ne se compare pas lui-même à un autre* ”.

En fait, toutes les théories réductionnistes présupposent à leur insu ce qu’elles prétendent éliminer. Dire que le cerveau “ pense ” procède de l’amalgame de deux discours qui ne cessent d’être corrélés mais qui restent irréductibles l’un à l’autre (voir 4). La seule alternative à la régression des homoncules comme à l’attribution animiste de désirs à la matière, c’est de prendre acte du fait qu’il n’y a pas de réponse à la question de la signification dans la répétition du “ même ” (configuration de configuration) mais seulement dans la reconnaissance d’une altérité, dont l’incomplétude “ fait signe ” [²⁰]. En objectivant le monde, quelqu’un s’objective paradoxalement comme inobjectivable. Les processus neuronaux fonctionnellement nécessaires à la conscience et à la pensée impliquent l’existence d’un co-principe qui ne peut qu’échapper à l’observation empiriquement objectivante : ce qui permet de quantifier n’est pas quantifiable [²¹].

4 - CONCLUSION : UN “ TIERS ” INACCESSIBLE

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

Quelque chose (quelqu’un) résiste. “ *La totalité se dérobe nécessairement* ” (J. Ladrière). Cette conception de la totalité-qui-se-dérobe est partagée par bien des auteurs dont les convictions sont par ailleurs très diverses. C’est, en particulier, le cas de C. Castoriadis qui, pour exprimer une telle réalité a créé le concept de *magma*. “ *Un magma est ce dont on peut extraire (ou: dans quoi on peut construire) des organisations ensemblistes en nombre indéfini, mais qui ne peut jamais être reconstitué (idéalement) par composition ensembliste (finie ou infinie) de ces organisations [] Il en résulte qu’il existe des domaines auxquels des énoncés significatifs peuvent être référés, mais qui ne satisfont à aucune théorie déterministe* ”. Ainsi “ *l’activité de formalisation elle-même n’est pas formalisable* ”.

Le sujet conscient apparaît comme une absence nécessaire (une “ fiction ” agissante) pour expliquer la présence du résultat. Il est ce qui produit [co-construit] du symbolique. Le sujet n’est pas de l’ordre de la substance [²²]) mais de l’ordre de l’acte en lui-même non objectivable. Il est, en

²⁰ Voir Lambert, 1997.

[²¹] Comme ce qui permet de calculer n’est pas calculable. Si le monde est réductible à un calcul, il est ultimement inaccessible et cette inaccessibilité est la condition même de sa calculabilité (voir Lambert, 1999).

[²²] Pour Aristote la question de la substance est identique au problème de l’être. Elle constitue la première des catégories, relative à l’identité - à l’essence - de l’être. Elle exprime *le quoi c’est* contrairement à la catégorie de l’existence qui désigne *le si c’est*. La substance constitue la réponse à ce que l’on demande quand on pose la question de l’être [de l’étant]. La substance est une *abstraction* opérée à partir du concret : c’est un *principe* d’intelligibilité. Elle renvoie au *ceci-que-je-montre*, à ce qui est connaissable. Elle désigne le support de la diversité, le substrat, le *ce qu’il y a par là-dessous*. La matière est certes le support universelle des choses concrètes mais elle est *a priori* totalement

effet, impossible de “représenter” un acte : il ne peut être qu’effectué. La conscience ne peut qu’échapper à toute forme de représentation. Si la connaissance est du côté de l’être, la signification est du côté de l’agir [de l’étant]. Mais définir la conscience en termes d’expérience vécue conduit à une régression sans fin dans la mesure où, comme nous l’avons souligné en introduction, il faudrait pouvoir définir l’expérience vécue indépendamment de la conscience. C’est précisément l’arrêt de cette régression qui donne à la conscience son effectivité dans la mesure même où les modalités d’un tel arrêt lui échappent (voir Lambert, 1997, 1999).

Certes, les neurosciences se doivent de mettre entre parenthèses l’existence de la subjectivité pour ne concentrer leur effort que sur l’objectivité des mécanismes cérébraux et des comportements auxquels ils sont associés. Mais quand on prétend ensuite rendre compte de cette subjectivité que l’on a d’abord écartée du champ de son investigation - en postulant son identité aux mécanismes objectifs - on transgresse les limites de la science. Le scientisme consiste précisément à nier ce qui a été mis entre parenthèses, à occulter l’opération de clôture constitutive du savoir scientifique. Un autre type d’erreur, tout aussi fatale, consiste à oublier que les parenthèses sont nécessaires et à réintroduire subrepticement dans le discours scientifique des considérations que l’exigence critique conduit justement à mettre entre parenthèses. Prétendre montrer où et comment l’esprit interagit avec le cerveau procède d’un parallogisme de même nature que celui qui consiste à en nier l’existence parce qu’il appartient à un domaine de la réalité dont la science s’est *a priori* interdit l’accès.

C’est au prix d’une simplification abusive qu’on en vient à opposer massivement dualisme spiritualiste et monisme matérialiste. Les discours tenus d’un côté et de l’autre relèvent de deux perspectives hétérogènes, c’est-à-dire non réductibles l’une à l’autre et non dérivables l’une de l’autre. Il paraît légitime de défendre un *dualisme sémantique* exprimant une *dualité de perspective*, un *dualisme des référents* et non de substances, comme le suggère P. Ricoeur. Le mental vécu implique le corporel irréductible au corps objectif des sciences de la nature. Le corps, cet objet qui est *mien* en même temps qu’il est *moi*. Corps-objet et corps-vécu ou corps-propre. Le corps figure deux fois dans le discours, comme objet du monde et comme lieu d’où j’apprends le monde. Or, il n’y a pas de passage d’un discours à l’autre. “*Ou bien je parle des neurones ou bien je parle de pensée que je relie à mon corps avec lequel je suis dans un rapport de possession et d’appartenance réciproques : mes pieds ou mes mains ne sont miens que vécus comme tels*” (P. Ricoeur). En revanche aucun vécu ne correspond à mon cerveau. Je prends avec mes mains n’est pas identique à je pense avec mon cerveau (revoir A. Bourguignon § 1.2). Le cortex ne sera jamais dans le discours du corps propre. Comme nous l’avons déjà dit avec Damasio (§ 3.2.2), mes connaissances sur le cerveau ne change en rien l’expérience de mon corps vécu.

Nous comprenons soit un discours “psy” soit un discours “neuro” mais leur relation fait problème parce que nous n’arrivons pas à inscrire leur lien à l’intérieur de l’un ou l’autre. Nous manquons du discours tiers “ni-ni”. C’est le même corps-cerveau-esprit qui est vécu et connu, c’est le même homme qui est corporel et mental, mais *d’un point de vue que je ne sais pas*, un point de vue tiers englobant l’unité de substance. Comme Paul Ricoeur, on ne peut être que sceptique quant à la possibilité de tenir un tel discours de surplomb.

La question cruciale pour l’homme n’est pas tellement de savoir si l’on peut réduire le mental au neural ou si l’on peut, grâce à la physique quantique rendre compte de leur interaction, mais celle du fondement, de l’origine, de la sauvegarde de son humanité et cette question continuera de se poser quelle que soit l’idée que l’on se fait du rapport entre le cérébral et le mental.

indéterminée. Il s’agit d’une véritable *matière première*, support possible de toutes les déterminations donc ne portant en elle-même aucune détermination. Le dualisme correspond-il alors à une réalité ontologique ou ne traduit-il qu’une difficulté d’ordre gnoséologique exprimant seulement un problème d’intelligibilité. La dualité se résout dans l’unité de l’acte.

REPERES BIBLIOGRAPHIQUES

- AMSTRONG, D.M. *A Materialist Theory of the Mind*, Routledge & Kegan Paul, 1968.
- ANDERSON, A.R., (dir.). *Pensée et machine* (trad. franç.), Champ Vallon, 1983.
- ARSAC, J. *Les Machines à penser*, Seuil, 1987.
- BERGSON, H. “Le cerveau et la pensée: une illusion philosophique”. *L'Energie spirituelle*, P.U.F., 1972 (148^{ème} éd.).
- BITBOL, M. *Physique et philosophie de l'esprit*, Paris, Flammarion, 2000.
- BOURGUIGNON, A. “Le problème esprit-cerveau”. *Le cerveau et l'esprit*, sous la direction de H. Barreau, CNRS Editions, 1992.
- CASTORIADIS, C. “La logique des magmas et la question de l'autonomie”. *L'auto-organisation, de la physique au politique*, Seuil, 1983.
- CHANGEUX J.P. *L'Homme Neuronal*, Fayard, 1983.
- CHANGEUX J-P. et CONNES, A. *Matière à pensée*, Odile Jacob, 1989.
- CHANGEUX, J-P. et RICOEUR, P. *La nature et la règle. Ce qui nous fait penser*, Odile Jacob, 1998.
- CRICK, F. *L'hypothèse stupéfiante. A la recherche scientifique de l'âme* (trad. franç.), Plon, 1995.
- CUMMINS, R.. *The Nature of Psychological Explanation*, Bradford Books / MIT Press, 1983
- DAMASIO, A. R. *L'erreur de Descartes. La raison des émotions* (trad. franç.), Odile Jacob, 1995.
- DAMASIO, A. R. *Le sentiment même de soi. Corps, émotions, conscience* (trad. franç.), Odile Jacob, 1999.
- DEBRU, C. *Neurophilosophie du rêve*, Paris, Hermann, 1990.
- DENNETT, D.C. *Brainstorn*, Bradford Books & Harvester Press, 1978.
- DENNETT, D.C. *La conscience expliquée*, Odile Jacob, 1993.
- ECCLES, J.C., *Le Mystère Humain*, Mardaga, 1981.
- ECCLES, J.C. *Evolution du cerveau et création de la conscience. A la recherche de la vraie nature de l'homme* (trad. franç.), Fayard, 1992.
- ECCLES, J.C. *Comment la conscience contrôle le cerveau* (trad. franç.), Fayard, 1997.
- EDELMAN G.M., *Biologie de la Conscience* (trad. Franç.), Odile Jacob, 1992.
- FEIGL, H. *The Mental and the Physical*, Minneapolis, 1967.
- FODOR, J.A. *La modularité de l'esprit, essai sur la psychologie des facultés* (trad. Franç.), Editions de Minuit, 1986
- GAUCHET, M. *L'inconscient cérébral*, Seuil, 1992.
- GAZZANIGA, M. *Le cerveau social*. Robert Laffont, 1987.
- HOFSTADTER, D. et DENNETT, D. *Vues de l'Esprit*, InterEditions, 1987.
- LADRIERE, J. “Wittgenstein et la philosophie analytique”. Science et foi, Desclée International, 1982.
- LAMBERT, J.F. “Corrélat Electro-encéphalographiques des Etats de Conscience Volontairement Modifiés (ECVM)”. *Les Relaxations Thérapeutiques Aujourd'hui*. Actes du Premier Colloque International de Relaxation, l'Harmattan, 1987, tome 2, 48-58.
- LAMBERT, J-F. “L'Epreuve du Sens : Science et Incomplétude”. *Les Cahiers Jean Scot Erigène (3) : Connaissance Traditionnelle, Connaissance Rationnelle*, Trédaniel, 1991, pp 135-222.
- LAMBERT, J-F. “L'absence qui fait signe : les sciences cognitives et la naturalisation de l'esprit”. *L'Esprit et la Nature*, Cahiers du GERC (5). Archè Edidit, 1997, pp 9-35.
- LAMBERT, J-F. “Inaccessibles Commencements”. *Phrétique - Langage et création*, 88 , 1999, pp 12-24
- LAMBERT, J-F. “Durée et expérience consciente : à propos des expériences de Libet”. *Bulletin de l'Association des Scientifiques Chrétiens*, n°3, 2000, pp 25-66.

- LAPLANE, D. *La mouche dans le bocal. Essai sur la liberté de l'homme neuronal*, Plon, 1987.
- LAPLANE, D. et DUBOIS, B. "Auto-Activation Deficit : A Basal Ganglia Related Syndrome ". *Movement Disorders*, 16, 5, 2001, pp 810-814.
- LEONARD, P. *Les raisons de croire*, Fayard-Communio, 1987.
- LIBET, B et al. "Time of conscious intention to act in relation to onset of cerebral activity (readiness potential) ". *Brain*, 1983, pp 106: 623.
- LIBET, B. "The Neural Time. Factor in Perception, Volition and Free Will in Neurosciences et philosophie, le problème de la conscience ". *Revue de Métaphysique et de Morale*, N°2, 1992, pp 255-272.
- GAZZANIGA, M. *Le Cerveau social*, Robert Laffont, 1987.
- MINSKY, M. *La Société de l'Esprit*, InterEditions, Paris 1988.
- OPPENHEIM-GLUCKMAN, H. *Mémoire de l'absence. Clinique psychanalytique des réveils de coma*. Masson, 1996.
- PENROSE, R. *Les ombres de l'esprit* (trad. franç.), InterEditions, 1996.
- PENROSE, R. *Les deux infinis et l'esprit humain* (trad. franç.), Flammarion, 1999.
- PIAGET, J. "L'explication en psychologie et le parallélisme psychophysique ". *Traité de psychologie expérimentale*, tome 1, ch. III, PUF, 1966.
- POPPER, K. et ECCLES, J.C. *The Self and its Brain*, Springer-Verlag, 1977.
- PYLYSHYN, Z.W. *Computation and cognition*, Bradford Books / MIT Press, 1984.
- SEARLE, J.R., *Du cerveau au savoir*, Ed. Hermann, Paris, 1985.
- SEARLE, J.R. *Le mystère de la conscience* (trad. franç.), Odile Jacob, 1999.
- SPERRY, R. *Science and Moral Priority*, Columbia University Press, 1983.
- STRAUS, E. *Du Sens des Sens*, Ed. J. Millon, 1989 (éd. originale Berlin, 1935).
- TURING, A.M. "Les ordinateurs et l'intelligence " in Anderson, A.R. (dir.) *Pensée et machine* (trad. franç.), Champ Vallon, 1983.
- VARELA, F., THOMPSON, E., ROSCH, E.. *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, Seuil, 1993.
- VARELA, F.J. *Connaître. Les sciences cognitives, tendances et perspectives* (trad. franç.), Seuil, 1989.
- VINCENT, J.D., *La chair et le diable*, Odile Jacob, 1996.
- WATANABE, S. "A model of mind-body relation in terms of modular logic ". *Boston Studies in the Philosophy of Science*, D. Reidel Publishing Company, 1963, pp 1-41.
- WITTGENSTEIN, L. *Notes sur l'expérience privée et les sense-data* (trad. Franç.), T.E.R., 1982.
- ZEKI, S. "Les images visuelles ", *Pour la Science*, nov. 1992, pp 60-68.