

# Les rapports de l'âme et du corps dans la philosophie de l'esprit contemporaine

**Fabrice Clément**

In *Revue de théologie et de philosophie*, v.131, no.1, pp.1-24.

**Résumé:** *La philosophie de l'esprit jouit actuellement d'un développement considérable dans les pays anglo-saxons où elle s'est arrogée la position de «philosophie première». Les travaux passionnants qui s'effectuent en son sein sont cependant encore relativement peu connus dans nos contrées et c'est la raison pour laquelle une présentation de quelques-unes des facettes de cette discipline s'avère utile. Pour ce faire, l'auteur s'est attaché à la question de base de la philosophie de l'esprit: comment la matière peut-elle donner naissance aux phénomènes mentaux? Autrement dit, comment le cerveau («brain») peut-il donner naissance à l'esprit («mind»)? Les principales réponses qui ont été apportées à cette interrogation fondamentale permettront de présenter le champ des possibles philosophiques et de faire connaissance avec les penseurs qui ont marqué la discipline. La dernière partie esquisse certaines voies de solution du problème.*

Les philosophes européens, britanniques exceptés, sont encore assez peu nombreux à s'intéresser à un domaine qui jouit actuellement d'un développement considérable outre-Atlantique: la philosophie de l'esprit. Un clivage historique qui date de trois siècles est à l'origine de ces intérêts philosophiques divergents. Descartes, en effet, a tout d'abord associé son nom à la culture française et son influence, même si elle n'est souvent pas perçue en tant que telle, est encore aujourd'hui très sensible. La très nette division entre le corps et l'esprit, qui constitue une des affirmations centrales de son système, a notamment favorisé une réflexion sur l'esprit entièrement autonome et insensible aux progrès de la recherche scientifique, «condamnée» à concentrer son attention sur la *res extensa*. Il n'est donc pas si étonnant de constater que la plupart des philosophes francophones s'inspirent principalement de la phénoménologie, méthode qui insiste sur la description du réel tel qu'il apparaît à la conscience, observation à la première personne qui ne peut que rebuter les partisans d'une description qui se veut objective<sup>1</sup>.

L'héritage intellectuel des pays anglo-saxons est par contre nettement plus favorable à un rapprochement entre les réflexions philosophiques menées sur l'esprit humain et les recherches scientifiques. Une telle compatibilité puise son origine dans la vision matérialiste des empiristes, et notamment de Hobbes qui considérait que la seule entité apte à composer l'«ameublement ontologique du monde» était la matière en mouvement; du coup, l'âme se trouve soumise à une dépendance ontologique par rapport au corps. Si la philosophie anglo-saxonne ne s'est pas concentrée plus tôt sur les

---

<sup>1</sup> Pour un recueil de textes qui mettent bien en évidence les oppositions entre philosophie analytique et phénoménologie, voir JANICAUD, 1994.

problématiques complexes engendrées par cette dépendance des états mentaux par rapport aux états cérébraux, c'est probablement parce qu'un certain nombre de problèmes posés par des philosophes allemands l'ont longtemps troublée. Ainsi, Kant s'attacha à la connaissance *a priori* et à la manière dont nous *devons* penser. Frege, en s'intéressant particulièrement à des objets logiques et mathématiques, montra que les significations ne sont pas à rechercher «dans la tête» mais qu'elles sont au contraire porteuses d'une objectivité qui dépasse les esprits individuels. Pour éviter de renvoyer le sens à ce «troisième monde» frégéen quelque peu mystérieux, les philosophes analytiques se sont longuement concentrés sur le *langage* qui, en tant que réservoir public de significations, continue à inspirer de nombreux chercheurs visant à saisir les relations qui unissent les mots à ce qu'ils signifient<sup>2</sup>.

Un certain nombre d'événements ont favorisé l'éclosion et l'importance de plus en plus grande de la philosophie de l'esprit, notamment aux Etats-Unis<sup>3</sup>. Une des sources de ce changement provient de la philosophie du langage elle-même puisque celle-ci va réintroduire dans sa réflexion les états mentaux en s'intéressant aux *intentions* impliquées lors des processus de communication<sup>4</sup>. Mais l'auteur à qui la philosophie de l'esprit doit le plus son statut actuel est certainement Quine. Celui-ci, en remettant en cause la distinction entre énoncés analytiques et énoncés synthétiques, supprima l'opposition entre ce qui est vrai en vertu de la signification et ce qui est vrai en vertu de l'expérience<sup>5</sup>. Il alla même plus loin en affirmant que, si les vérités proprement conceptuelles ne sont plus séparables des vérités scientifiques, l'épistémologie ne constituait plus un domaine réservé au philosophe mais qu'elle se devait au contraire d'être *naturalisée*. Cette idée a été encouragée par l'émergence d'un courant de recherche interdisciplinaire fortement inspiré par l'intelligence artificielle et dont le but consiste à mettre au jour les mécanismes de l'esprit: les sciences cognitives<sup>6</sup>. Ce champ de recherche influence aujourd'hui considérablement les philosophes de l'esprit, comme nous le verrons dans la suite de cet article.

Le rôle de la philosophie de l'esprit est aujourd'hui crucial dans le monde anglo-saxon et elle s'est pour ainsi dire arrogée le statut privilégié de «philosophie première». Son développement a été si important depuis une quinzaine d'années qu'il serait tout à fait prétentieux de vouloir en donner un aperçu systématique et exhaustif. Dans les

---

<sup>2</sup> Sur la genèse de la philosophie analytique, voir DUMMETT, 1988.

<sup>3</sup> Ces différents facteurs sont bien décrits par ENGEL, 1994.

<sup>4</sup> L'article fondateur est GRICE, 1957.

<sup>5</sup> L'article de Quine, «Les deux dogmes de l'empirisme», a été traduit en français in JACOB, 1980. Jacob propose également dans cette ouvrage une lumineuse introduction aux problématiques philosophiques qui ont conduit à la philosophie de l'esprit contemporaine.

<sup>6</sup> Pour une histoire de l'émergence des sciences cognitives et une description de leurs différents champs d'intérêt, voir GARDNER, 1985 et DUPUY, 1994.

pages qui suivent, l'accent sera porté sur un aspect central de toute réflexion portant sur l'esprit: les rapports entre l'âme et le corps ou, pour le dire dans des termes plus modernes, entre l'esprit et le cerveau. Les différentes solutions qui ont été apportées à ce problème ont l'avantage de mettre en évidence les principales lignes de partage qui caractérisent la philosophie de l'esprit contemporaine. De plus, comme il s'agit d'une question centrale à laquelle la plupart des grands auteurs ont tenté d'apporter une réponse, cette problématique permettra de présenter quelques-uns des grands personnages qui règnent sur cette discipline. Nous terminerons en proposant une manière de voir les choses qui permette de tenir compte des avantages de différentes positions tout en évitant un certain nombre de leurs pièges.

## **LA PHILOSOPHIE DE L'ESPRIT**

Donner de but en blanc une définition du concept d'«esprit» s'avère extrêmement périlleux étant donné que, d'une part, cette notion recouvre des usages parfois divers et que, d'autre part, la plupart des définitions reflètent d'emblée une certaine conception philosophique. Rappelons tout d'abord que les anglo-saxons utilisent le mot «mind» — et non pas «spirit», par exemple — et que c'est donc la sémantique de ce terme qui doit diriger nos interrogations. Samuel Guttenplan a demandé à ses différentes volées d'étudiants ce qui, selon eux, manifeste la présence d'un esprit<sup>7</sup>. Ce «sondage d'opinion philosophique» laisse apparaître une liste de verbes qui, pour ces étudiants anglo-saxons, indiquent la présence d'un esprit. La liste mentionnait des verbes tels que croire, se représenter, rêver, imaginer, aimer, percevoir, décider, agir intentionnellement, avoir mal, etc. Pour désigner cet ensemble de facultés partagées par tout à chacun, les philosophes parlent de «psychologie populaire» (*folk psychology*), désignant par là l'ensemble des conceptions qui nous permettent de comprendre, expliquer, prédire et manipuler le comportement de nos semblables. Autrement dit, le concept d'«esprit» renvoie dans ce contexte à ce qui se passe «dans nos têtes», à la vie psychique (ou *mentale*) qui sous-tend nos comportements et nos réflexions. La philosophie de l'esprit va alors tenter de rendre compte de ces étranges entités mentales que sont les attitudes propositionnelles (comme les croyances ou les désirs), les perceptions ou les sensations (aussi appelés «qualia»), ou encore les émotions et les intentions, autant d'entités qui peinent à trouver leur place au sein de la sphère scientifique.

---

<sup>7</sup> GUTTENPLAN, 1994, «An Essay on Mind», p.6. Cet ouvrage constitue une sorte de Bible pour tous ceux qui s'intéressent à la philosophie de l'esprit. Les articles concernant les philosophes les plus importants ont généralement été écrits par ces auteurs en personne, ce qui leur permet d'explicitier leurs positions de manière synthétique. Nous nous sommes largement servi de cet outil de travail lors de la rédaction de cet article.

La question fondamentale est alors *ontologique*: de quel type de choses s'agit-il? La richesse de notre vie intime, à laquelle nous pouvons avoir partiellement accès par introspection, laisse supposer l'existence d'«objets mentaux» qui déploient pour nous quelque chose comme des paysages intérieurs. Et pourtant, nous en sommes aujourd'hui bien convaincu, tout cela se passe dans notre tête, c'est-à-dire au sein de notre cerveau. Ce dernier serait ainsi un organe — physique — susceptible de produire des images, des pensées et des sensations, autant d'entités dont on peine à discerner le caractère *matériel*. La philosophie de l'esprit tourne autour de cette difficulté paradoxale: comment un organe physique peut-il produire quelque chose d'immatériel? Comment le cerveau peut-il *causer* le mental et quel type de causalité ce dernier peut-il à son tour exercer?

Face au gouffre qui semble opposer la sphère du mental à celle du physique, le sens commun a tendance à admettre qu'il s'agit de deux choses radicalement différentes; il paraît en effet bien difficile de réduire la sensation éprouvée lors de la contemplation d'un site grandiose à une simple activation neuronale. Le cheminement méthodique de Descartes, en lui faisant douter de l'existence même de la matière tout en lui interdisant dans le même temps de douter de son esprit en train de douter, aboutit à une conclusion de ce type: le corps et l'esprit forment deux substances différentes, la seconde pouvant exister sans la première. Une telle conclusion, en plus de confirmer le sens commun, avait l'avantage d'encourager le développement de la science (dévouée à l'étude de la *res extensa*) sans pour autant réfuter l'existence de l'âme. Mais cette solution dualiste, qui ne fait d'ailleurs que repousser le problème de l'articulation du corps et de l'esprit — malgré l'hypothèse de la glande pinéale — est refusée d'emblée par la grande majorité des philosophes de l'esprit contemporains. C'est en effet à la science que notre époque s'adresse pour définir l'«ameublement ontologique» de notre monde et la philosophie de l'esprit s'inscrit dans un courant résolument moniste ou naturaliste<sup>8</sup>. Son problème consiste donc à rendre compte des entités et des processus mentaux en termes matérialistes ou physicalistes, ce qui n'est pas une mince affaire étant donnée leur hétérogénéité. Différentes solutions ont été proposées, dont les plus connues sont les suivantes: le behaviorisme logique, le matérialisme éliminativiste, le matérialisme dit «de l'identité», le monisme anomal, le fonctionnalisme computationnel, l'instrumentalisme et le fonctionnalisme téléo-sémantique.

## **ELIMINER L'ESPRIT: BEHAVIORISME LOGIQUE ET ÉLIMINATIVISME**

---

<sup>8</sup> Il existe bien sûr quelques exceptions, comme le prix Nobel Sir John Eccles qui reste résolument dualiste (ECCLES, 1989). Un auteur comme Colin McGinn avance d'intéressants arguments d'ordre conceptuel en faveur du dualisme (cf., par exemple, MCGINN, 1983).

Confrontés à l'incommensurabilité de l'univers mental et de l'univers physique, certains auteurs se sont mis à remettre en question l'existence même des «états mentaux». La première tentative de ce type remonte aux années cinquante. Gilbert Ryle s'attaqua en effet avec verve au dualisme de type cartésien dans son ouvrage fameux *The Concept of Mind*. Mais sa critique adressée à ce qu'il appelle le «dogme du fantôme dans la machine» ne s'applique pas qu'au dualisme cartésien. Croire qu'il existe des états mentaux relève en effet selon lui de «l'erreur de catégorie». Son exemple, devenu paradigmatique, permet d'illustrer ce qu'il veut dire<sup>9</sup>. Imaginons qu'un touriste visite l'Université d'Oxford. On lui montre les différents collèges, les bibliothèques et autres salles de cours. A la fin de la visite, on lui demande comment il a trouvé l'Université d'Oxford et celui-ci, déçu et étonné, répond: «Mais enfin, je ne l'ai pas vu, cette Université!». Le touriste s'attendait donc à voir un objet circonscrit appelé Université et il ne se doutait pas que ce terme se bornait à désigner l'*ensemble* des bâtiments visités. De manière similaire, pour savoir si quelqu'un est intelligent, on ne va pas se mettre à la recherche d'une substance particulière située quelque part dans son cerveau. On va plutôt se demander quels types de comportements sont susceptibles de rendre manifeste l'intelligence; ce qui est testé, ce sont des capacités et des tendances. Autrement dit, il est erroné de rechercher derrière le comportement un «double caché» qui en rendrait compte. Ryle laisse ainsi suggérer que l'esprit ne sert qu'à désigner des structures de comportement et qu'il peut en dernière instance être *réduit* au comportement. Une telle conception est d'ailleurs proche de celle de Wittgenstein, pour qui les concepts mentaux n'appartiennent pas à la catégorie des substances mais à celle des propriétés des substances<sup>10</sup>. Un auteur contemporain français, Vincent Descombes, critique lui-aussi les conceptions cognitivistes en tentant de montrer qu'elles succombent à une telle erreur de catégorie<sup>11</sup>.

Une telle solution n'est cependant pas dénuée de défauts. Ryle reste par exemple très suggestif quant à la notion de disposition. Dire qu'il est possible de réduire la croyance de Bill qu'il y a de la bière dans le frigo à un ensemble de dispositions (se lever de son fauteuil, se diriger vers la cuisine, ouvrir la porte du frigidaire, etc.) n'élimine pas pour autant toute référence à des états mentaux, bien au contraire; Bill doit avoir soif, désirer boire, se décider à se lever de son fauteuil si confortable, etc. Définir les états mentaux

---

<sup>9</sup> RYLE, 1949, p.16.

<sup>10</sup> WITTGENSTEIN, 1960, p.41. Voir également ENGEL, 1996, p.171. Engel fait remarquer ailleurs qu'une telle conception renvoie en fait à la conception aristotélicienne selon laquelle l'esprit n'est pas une substance, mais la forme ou la propriété d'une substance (cf. l'article de Pascal ENGEL, «La pensée», in *Encyclopaedia Universalis*).

<sup>11</sup> DESCOMBES, 1995 et 1996. A ce sujet, voir mes recensions critiques dans CLÉMENT, 1996 (a) et 1996 (b), ainsi que CLÉMENT [1997].

par des structures comportementales semble donc conduire à un *regressus ad infinitum*. De plus, certaines croyances peuvent ne jamais se traduire en comportements. La croyance d'Elgar, par exemple, jardinier de son état, qu'il y a eu autrefois de l'eau sur Mars a peu de chance d'engendrer des comportements qui mettraient au jour une disposition quelconque. La tentative éliminativiste de Ryle, appelée parfois «behaviorisme logique», est donc finalement bien peu satisfaisante.

Les progrès récents des neurosciences ont par contre encouragé une nouvelle tentative d'élimination des états mentaux. L'idée qui motive ces efforts est on ne peut plus physicaliste: il n'y a pas de place, au sein du monde tel qu'il est décrit par la science, pour des entités telles que les états mentaux. L'un des leaders des «matérialistes éliminativistes», Paul Churchland, qui se qualifie lui-même de «neurophilosophe», se doit donc de remettre en cause la psychologie populaire (*folk psychology*) qui présuppose l'existence d'entités telles que les croyances, les désirs ou les douleurs<sup>12</sup>. Pour ce faire, il insiste sur le fait que la psychologie populaire fonctionne comme une *théorie* scientifique: il s'agit d'un cadre spéculatif, systématique et corrigible réunissant des informations sous une forme générale et qui permet l'explication ainsi que la prévision des comportements<sup>13</sup>. Mais cette théorie souffre de défauts irrémédiables. Tout d'abord, elle ne fournit aucune explication des maladies mentales, de l'imagination créative, de la différence d'intelligence entre individus, du rêve, de la construction d'images en trois dimensions, des illusions perceptives, de la mémoire. Ensuite, elle n'a pas progressé de manière sensible depuis au moins 2500 ans; les Grecs semblaient en effet utiliser un cadre d'explication similaire<sup>14</sup>. Enfin, elle peine à s'inscrire dans le développement exponentiel des neurosciences. Bref, tout porte à croire que cette théorie n'est qu'un programme scientifique dégénéré dont il faut se débarrasser au plus vite. Du coup, les croyances, désirs ou autres douleurs ne sont que des configurations d'événements neuronaux dont les neurosciences arriveront bientôt à rendre compte au sein d'une théorie «scientifiquement correcte», ramenant ainsi ces notions au rang de mythes utiles au même titre que le phlogistique des chimistes du XVIII<sup>e</sup> siècle. Selon Patricia Churchland-Smith et Terence Sejnowski, la nouvelle théorie de l'esprit, radicalement réductionniste, devrait beaucoup au connexionnisme, modèle qui permet de simuler les interconnexions des neurones<sup>15</sup>. Or ce sont précisément les connexions

---

<sup>12</sup> Churchland a notamment écrit *A Neurocomputational Perspective*. Il collabore régulièrement avec son épouse, Patricia Smith Churchland, neurophysiologiste, auteure de *Neurophilosophy: Toward a Unified Science of the Mind/Brain*.

<sup>13</sup> CHURCHLAND, «Folk psychology» in GUTTENPLAN [1994], p.308.

<sup>14</sup> CHURCHLAND, «Eliminative Materialism and the Propositional Attitudes», in LYCAN, 1990, p.210.

<sup>15</sup> CHURCHLAND & SEJNOWSKI, «Neural Representation and Neural Computation», in LYCAN, 1990.

synaptiques entre les neurones qui permettent à un organisme d'incarner des connaissances générales et des habiletés (*skills*). Selon les éliminativistes, c'est à ce niveau que les capacités mises en oeuvre dans l'interprétation des comportements d'autrui sont à rechercher, et non dans une pseudo-théorie psychologique dont les «lois» seraient connues de tous.

Mais est-il si simple d'éliminer la psychologie populaire et les états mentaux qui lui sont associés? Certains arguments prouvent qu'une telle élimination peut être remise en cause. Tout d'abord, la thèse de Churchland n'est valide que dans la mesure où la psychologie populaire est bel et bien une théorie, au sens scientifique du terme. Or, un tel qualificatif ne semble pas lui convenir, car il est bien difficile d'admettre que la psychologie populaire soit véritablement une théorie empirique du mental. Plutôt que des lois, elle nous fournit des recettes, des bricolages psychologiques. Si, comme le montre Pascal Engel, il s'agit bien d'un savoir, celui-ci est essentiellement *pratique* et il pourrait être décrit au moyen de règles normatives de rationalité à la lumière desquelles nous interprétons le comportement de nos semblables<sup>16</sup>. De plus, la possession d'une théorie neurobiologique précise ne nous empêcherait pas nécessairement de recourir à des notions comme celles de croyance ou de désir pour rendre compte du comportement de nos semblables; après tout, le terme «chaise» continue à être utilisé même si on peut en rendre compte dans les termes de la physique<sup>17</sup>... Enfin, l'éliminativisme ne rend pas compte de la force de nos intuitions concernant les états mentaux et de leur succès prédictif. Là encore, l'option éliminativiste n'est finalement guère satisfaisante.

Éliminer les états mentaux sous prétexte qu'ils ne s'intègrent pas à la vision du monde délimitée par les sciences physiques nous apparaît ainsi être une option radicale dont les défauts ne compensent pas la bienséance ontologique. Le pari consiste alors à proposer une interprétation des états mentaux qui cadre avec la vision moniste préconisée par les philosophes de l'esprit. Certains auteurs s'y sont essayés, avec plus ou moins de succès.

## **LA THÉORIE DE L'IDENTITÉ ET LE MONISME ANOMAL**

Pour sauver les états mentaux tout en respectant une conception matérialiste, une solution se présente assez spontanément à l'esprit, et c'est d'ailleurs une des premières à avoir été développées chronologiquement: il suffit de dire que les états mentaux sont identiques aux états physiques ou neurologiques<sup>18</sup>. Cette hypothèse, appelée

---

<sup>16</sup> ENGEL, 1994 (a), p.62.

<sup>17</sup> ENGEL, 1994 (b), p.541.

<sup>18</sup> LEWIS, 1966 et LEWIS, «Reduction of Mind», in GUTTENPLAN, 1994, p.412.



«matérialisme de l'identité», est attractive car, en identifiant chaque type d'état mental avec un type d'état physique, elle semble s'inscrire dans la logique même du progrès scientifique. Prenons en effet l'exemple de la lumière<sup>19</sup>. Avant que ne se développent les théories de l'électricité, la lumière devait constituer un phénomène véritablement étrange. Une fois l'électricité découverte, la lumière a pu être ramenée à un rayonnement électromagnétique. Attention: il ne s'agit pas de dire que la lumière est *causée* par un tel rayonnement, mais qu'elle n'est *rien d'autre* que ce rayonnement électromagnétique. Dans le même ordre d'idée, le but n'est pas de dire que les états mentaux sont causés par des décharges neuronales, mais qu'ils ne sont *rien d'autre* que des configurations de décharges neuronales. Pour les défenseurs de ce type de physicalisme, il est possible d'établir des «énoncés-passerelles» permettant de subsumer les régularités de la psychologie sous celle de la neurophysiologie. Si les propriétés mentales de l'esprit sont identiques à des propriétés sous-jacentes du cerveau, le problème des rapports entre l'âme et le corps n'est plus inextricable.

Les défauts de cette théorie ont cependant rapidement été mis en évidence. Tout d'abord, s'il est vrai que les propriétés mentales sont *identiques* à des propriétés physiques ou neurophysiologiques, alors, selon la «loi de Leibniz», toutes les propriétés qui sont vraies dans le domaine du mental doivent être vraies dans le domaine physique. Or la nature du mental a quelque chose de fondamentalement subjectif et qualitatif; il est en effet difficile de soutenir qu'une sensation de douleur ou de honte n'est, en fin de compte, rien d'autre qu'une décharge neuronale. Une autre difficulté est issue du progrès simultané de l'intelligence artificielle et des neurosciences; certaines démarches que l'on croyait strictement réservées à l'esprit humain, comme le calcul, le raisonnement ou la prise de décision, ont ainsi pu être «implémentées» dans des machines. De telles réussites ont encouragé les philosophes à envisager la possibilité de réaliser certains états mentaux indépendamment du substrat cérébral humain<sup>20</sup>. En cas de succès, il deviendrait impossible de dire que tout état mental n'est rien d'autre qu'un état physique donné puisque différents systèmes physiques pourraient donner naissance au même état mental. Une telle conclusion, qui renvoie à ce que les philosophes appellent la «réalisabilité multiple» des états mentaux, est d'ailleurs soutenue par les travaux des neuropsychologues et des neurophysiologistes qui ont pu montrer que les structures cervicales sous-tendant la dynamique psychologique pouvait varier d'un individu à l'autre<sup>21</sup>. On voit dès lors mal comment continuer à soutenir qu'à un certain type d'état mental correspond systématiquement un type d'état physique donné.

---

<sup>19</sup> Cet exemple est proposé par GUTTENPLAN, 1994, p.91.

<sup>20</sup> D'autres philosophes sont beaucoup plus dubitatifs à ce sujet, notamment John Searle [1992].

<sup>21</sup> Les neurophysiologistes ont ainsi pu mettre en évidence la «plasticité» du cerveau, certaines capacités fonctionnelles pouvant dans certains cas être prises en charge par d'autres tissus cérébraux

Une autre solution se profile alors à l'horizon. Elle consiste à reconnaître le caractère irréductible des états mentaux tout en acceptant que ceux-ci ne doivent leur existence qu'à la présence d'états physiques — on dit alors que les états mentaux «surviennent» sur les états physiques. Pour ce faire, il suffit de refuser l'identité des types (à un type d'état mental correspond un type donné d'état physique) tout en acceptant qu'à chaque occurrence (*token*) d'un événement mental corresponde une occurrence physique, sans que pour autant une réduction systématique du mental au physique soit envisageable<sup>22</sup>. Donald Davidson est le philosophe qui a le mieux défendu cette option, généralement appelée «monisme anomal». Son argument comporte trois parties<sup>23</sup>: (1) tous les événements mentaux sont reliés causalement à des événements physiques, (2) si deux événements sont couplés par une relation de cause à effet alors il existe une loi stricte sous laquelle ils peuvent être subsumés, et (3) il n'existe pas de lois psychophysiques strictes qui permettraient de relier les événements mentaux avec des événements physiques. Aucune réduction du monde mental au monde physique n'est ainsi envisageable, même si tous deux appartiennent au même univers ontologique. Pour justifier une telle impossibilité, Davidson s'appuie sur le caractère holistique du mental. Les états mentaux, que l'on décrit sous la forme d'attitudes propositionnelles (comme «croire que...», «désirer que...», etc.), n'existent pas isolés les uns des autres: leur identité est au contraire en partie due à leur position au sein d'un large réseau d'autres attitudes. Ces associations peuvent varier d'individu à individu (la croyance en une forme de vie extra-terrestre peut prendre des formes fort différentes) et chez un même individu au cours du temps. Il est par conséquent absurde de penser que des lois psychophysiques valables universellement puissent être découvertes. La sphère du mental ne peut donner lieu qu'à des interprétations et celles-ci sont à chaque fois étroitement dépendantes d'un contexte.

La solution de Davidson, si subtile soit-elle, comporte néanmoins de sérieux inconvénients. Elle encourt tout d'abord le risque d'«épiphénoménalisme». On désigne par ce terme la thèse selon laquelle les phénomènes mentaux ne peuvent exercer d'effets causaux; dans cette optique, les phénomènes mentaux sont causés par des phénomènes physiques mais ils ne peuvent quant à eux rien causer du tout. Une telle perspective, si elle est ontologiquement correcte, laisse notre bon sens en rade. En effet, l'enchaînement des mouvements physiques qui président à l'ouverture de la porte du frigidaire ne pourraient en aucun cas être les effets de mon désir de boire une bière. De même, on ne voit pas comment la rougeur qui envahit mon visage pourrait résulter de

---

lorsque les zones prévues génétiquement pour cette tâche sont rendues inopérantes. Voir par exemple CYRULNIK, 1995.

<sup>22</sup> En langage technique, on parle dans ce cas d'identité «token-token».

<sup>23</sup> DAVIDSON, «Davidson, Donald», in GUTTENPLAN, 1994, p.231.

ma honte de m'être laissé emporter, sous le coup de l'alcool, à insulter un ami proche. Le risque d'épiphiénoménalisme menace d'ailleurs bon nombre des théories qui tentent de résoudre le «mind-body problem», comme nous le verrons un peu plus loin avec Dennett. Mais la proposition de Davidson souffre d'un autre désavantage. En refusant de reconnaître des lois psychophysiques et en insistant sur le caractère holistique du mental, il condamne toute approche scientifique de la sphère du mental. A la limite, Davidson condamne les efforts des psychologues en vue de construire une science de l'esprit et les encourage à interpréter au coup par coup les différents réseaux d'attitudes propositionnelles qui sous-tendent, de façon à chaque fois originale, les comportements des individus. Abandonner l'ambition de donner une description du mental en termes scientifiques représente cependant un bien grand prix à payer. Cela d'autant plus que, comme le remarque Pascal Engel, il existe dans un grand nombre de cas des relations suffisamment systématiques entre la sphère de mental et la sphère du physique pour que l'on puisse envisager des réductions partielles<sup>24</sup>. Les régularités qui sous-tendent la vie mentale sont donc susceptibles de donner lieu à des lois psychologiques de nature statistique plus précises que de simples lois *ceteris paribus*<sup>25</sup>, même si leur précision est loin d'être équivalente à celle des lois physiques.

## LE FONCTIONNALISME

Un courant philosophique très important en philosophie de l'esprit s'est fait le farouche défenseur de l'existence d'un ordre de phénomènes mentaux pouvant donner lieu à une étude scientifique: le *fonctionnalisme*. Ce mouvement s'est opposé à la fois au matérialisme de l'identité et au behaviorisme. On a vu que pour les partisans de l'identité du physique et du mental, les états et événements mentaux étaient identiques aux états et événements physiques ayant lieu dans le système nerveux de leurs «propriétaires». Ressentir de la douleur, par exemple, ne serait rien d'autre que d'avoir des fibres C en activité<sup>26</sup>. Putnam fit alors remarquer<sup>27</sup> qu'une telle théorie entraîne une conséquence démesurée: seuls les organismes munis de fibres C seraient en effet susceptibles de ressentir de la douleur. Mais n'est-ce pas faire preuve d'un «chauvinisme de l'espèce» que de dire que seuls les organismes doués des mêmes constituants chimiques que les êtres humains pourraient ressentir de la douleur (et toutes sortes d'autres états mentaux)? Selon Putnam, il n'est pas important en soi que ce soient

---

<sup>24</sup> ENGEL, 1996, p.203.

<sup>25</sup> «Toutes choses étant égales par ailleurs». Cette expression est utilisée pour décrire des relations causales qui ne sont valables que dans des conditions bien précises, susceptibles de varier.

<sup>26</sup> Cf. LYCAN, «Functionalism», in GUTTENPLAN, 1994, p.317-318.

<sup>27</sup> Dans un article intitulé «Minds and Machine», in HOOK, 1960.

des fibres C qui entrent en activité; ce qui importe, c'est la contribution de ces fibres à l'organisme pris comme un tout. En fait, le rôle des fibres C aurait pu être tenu par un autre composant physique qui se serait chargé de la même *fonction*. Du coup, les différents types d'états mentaux n'ont plus à être identifiés à certains types d'états physiques; il faut plutôt les rapporter à des rôles fonctionnels plus abstraits définis par leurs relations avec des entrées, des sorties et d'autres états psychologiques de l'organisme.

La proposition de Putnam, en attribuant une certaine dignité ontologique à des états de l'organisme situés entre le stimulus et la réponse comportementale, avait également l'avantage de s'opposer aux préceptes behavioristes, qui dominaient encore la psychologie expérimentale. Le behaviorisme, pour satisfaire à tout prix les critères scientifiques, refusait d'accorder aux états mentaux, inobservables par définition, une place dans ses explications, et se bornait à travailler sur des régularités stimulus-réponse. Avec le fonctionnalisme, les psychologues découvraient une nouvelle possibilité de sonder la «boîte noire» que constituait pour eux l'esprit: la voie était ouverte au *cognitivism*<sup>28</sup>. Ce terme désigne les travaux de très nombreux chercheurs qui partagent tous une vision commune: (1) les psychologues doivent s'attacher à la description des états et des épisodes mentaux pour décrire le comportement et (2) les êtres humains et les autres organismes psychologiques doivent être considérés comme des systèmes de traitement de l'information<sup>29</sup>. Avec de telles prémisses, l'étude de l'esprit trouvait alors une alliée de poids: l'intelligence artificielle.

Jerry Fodor est un des philosophes qui a le plus travaillé au rapprochement de disciplines aussi diverses que la philosophie, la psychologie, l'intelligence artificielle et la linguistique<sup>30</sup>. Il admet tout d'abord que les états mentaux sont doués de pouvoirs causaux. Mais il précise immédiatement qu'ils doivent également être sémantiquement évaluables: une croyance, par exemple, renvoie à quelque chose et elle peut donc être vraie ou fausse selon ce qu'elle affirme de cette chose. Les états mentaux sont ainsi des entités paradoxales: d'un côté, ils se comportent comme des propositions en étant sémantiquement évaluables — mais les propositions sont des objets abstraits dénués de pouvoir causal, et d'un autre côté ils sont semblables à des objets physiques en étant munis d'une certaine force causale. Pour résoudre ce problème, Fodor s'inspire du fonctionnement des ordinateurs et met les *symboles* au centre de son explication. Les symboles ont en effet à la fois des propriétés sémantiques et des propriétés matérielles. Si j'écris par exemple sur une pancarte «le chat est sur le paillason» en grosses lettres

---

<sup>28</sup> Pour une introduction fouillée aux différents domaines regroupés sous ce terme, voir le recueil édité par ANDLER, 1992.

<sup>29</sup> LYCAN, in GUTTENPLAN, 1994, p.319.

<sup>30</sup> Voir, par exemple, FODOR, 1987.

noires, les symboles utilisés ont une existence matérielle et disposent par là-même de certaines propriétés (ils reflètent la lumière d'une certaine manière, ils exercent une influence gravitationnelle minuscule mais en principe mesurable, etc.). Mais ce que j'ai écrit est aussi sémantiquement évaluable; cette phrase est vraie si et seulement s'il y a effectivement un chat sur le paillason<sup>31</sup>. Pour se sortir d'affaire, il faudrait donc pouvoir trouver un moyen de «traduire» le fonctionnement de l'esprit au moyen de symboles. Or les développements de l'informatique ont montré qu'il était possible de prendre au sérieux l'analogie entre l'esprit et l'ordinateur. A la suite de Turing, on a construit des machines capables de manipuler des informations grâce aux propriétés matérielles (électromagnétiques) des symboles impliqués dans les opérations. De plus, une machine correctement programmée est capable de garantir la cohérence sémantique (les valeurs de vérité) des symboles qu'elle manipule. Dans une telle perspective, les processus mentaux peuvent alors être considérés comme des interactions causales entre des symboles mentaux ou des «représentations», interactions réglées par une syntaxe qui préserve la cohérence sémantique des contenus mentaux. Autrement dit, la succession de nos états mentaux est un processus causal gouverné par les règles d'une grammaire mentale, un «Langage de la Pensée» — ou «mentalais». L'analogie entre les états mentaux et les propositions est donc, c'est le cas de le dire, prise à la lettre. Les pensées sont identifiées à des «phrases dans la tête» qui peuvent se combiner, selon les règles syntaxique du langage interne de l'esprit pour articuler les processus mentaux. Dans notre exemple favori, la croyance qu'il y a de la bière dans le frigo donne lieu à une phrase en mentalais de la forme: «il y a de la bière dans le frigo» et celle-ci est placée dans ce que Fodor appelle, à la suite de Schiffer, une «boîte à croyance». De même, le désir de boire une bière est placé dans une autre «boîte», la «boîte à désir». Il s'ensuit une série d'opérations d'ordre syntaxique (combinaisons de symboles dont le parcours est gouverné par des règles) qui débouche sur la mise en oeuvre d'un comportement — mon déplacement vers la cuisine. Avec une telle manière de décrire les phénomènes de l'esprit, l'objet d'étude de la «science du mental» est tout trouvé: il s'agit de mettre à jour la grammaire de ce «mentalais» qui règle la succession de nos états mentaux, eux-mêmes à la source de nos actes intentionnels.

La conception de Fodor a l'avantage de constituer une sorte de paradigme qui jouit aujourd'hui encore d'une importance considérable et qui est à la source de bon nombre des recherches menées dans les sciences cognitives. Deux grands types de critiques peuvent cependant lui être adressées: les premières insistent sur la difficulté que ce programme rencontre dans l'explication de phénomènes pourtant cruciaux pour la philosophie de l'esprit; les secondes mettent en doute le réalisme des entités mentales

---

<sup>31</sup> FODOR, «Fodor», in GUTTENPLAN, 1994, p.292-300.

postulées en s'appuyant sur les recherches récentes menées sur le cerveau. Les critiques du premier type s'adressent principalement au caractère «internaliste» du point de vue fodorien — appelé aussi «computationnel». En se focalisant sur les règles du langage interne de l'esprit, Fodor prêche en effet pour une forme de «solipsisme méthodologique». Mais une telle option pose problème lorsqu'on s'intéresse à des entités qui sont définies par leur intentionnalité, c'est-à-dire par le fait qu'elles sont *dirigées vers* ou *à propos* d'objets ou d'états de choses situés dans le monde. Peut-on dès lors se borner à définir les états mentaux par leur rôle fonctionnel *à l'intérieur* d'un organisme? Putnam, dans une expérience de pensée fameuse, montre qu'il n'en est rien. Il imagine en effet qu'il existe une terre jumelle semblable en tout point à notre planète excepté une chose: l'eau n'a pas pour composition chimique H<sub>2</sub>O mais XYZ. S'il existe sur chacune de ces planètes un homme, appelons-les respectivement Oscar 1 et Oscar 2, alors ces jumeaux sont en tout point similaires, à la molécule près. Et pourtant, même si aucune différence fonctionnelle ne permet de les différencier, leurs états mentaux ne sont pas strictement les mêmes lorsque, face à un lac, ils disent: «ceci est de l'eau», puisque la signification du terme «eau» n'est pas la même dans les deux cas: le référent est H<sub>2</sub>O dans un cas et XYZ dans l'autre. Par cet exemple, Putnam montre que l'on ne peut faire abstraction des relations avec l'environnement par lesquelles se définissent les états mentaux. Un autre problème se pose avec les *qualia* — les sensations subjectives associées au fait d'éprouver un état mental, difficulté qui a elle aussi été soulevée par une expérience de pensée, dite du «spectre inversé». Il se pourrait en effet qu'une chose dite «rouge» apparaisse en fait verte à certains individus. Et pourtant, comme leur «erreur» est automatiquement corrigée par leur inversion de vocabulaire, elle passe inaperçue aussi bien pour les uns que pour les autres. Bref, toutes les personnes impliquées dans cette expérience peuvent être considérées comme fonctionnellement identiques tout en éprouvant des *qualia* fort différents. Il faut donc bien admettre que l'hypothèse fonctionnaliste peine à rendre compte de la nature des états mentaux, que ce soit sous l'angle de l'intentionnalité ou de l'expérience sensible.

Une autre ligne d'arguments, à notre avis beaucoup plus difficile à récuser, remet en question le fondement même des thèses fonctionnalistes-computationnalistes, et plus particulièrement l'hypothèse du langage de la pensée, en doutant de leur réalisme. Autrement dit, le fonctionnalisme est critiquable parce qu'il est difficilement compatible avec les connaissances actuelles sur le cerveau humain. Il existe ainsi aujourd'hui une querelle au sein des sciences cognitives entre les partisans du cognitivisme classique et les défenseurs du connexionnisme<sup>32</sup>. Pour ne pas nous perdre dans les dédales de cette

---

<sup>32</sup> Pour une présentation simple proposant un panorama des théories proposées et l'esquisse d'une nouvelle solution, dite de «l'énaction», voir VARELA, 1996.

dispute, nous allons nous appuyer sur un auteur, Daniel Dennett, pour montrer comment le fonctionnalisme de Fodor peut être attaqué. Cela nous permettra par la même occasion de présenter la solution de Dennett au problème des rapports entre le corps et l'esprit.

## **LA STRATÉGIE INSTRUMENTALISTE DE DANIEL DENNETT**

Dennett se considère lui-même comme le fruit d'un croisement entre Ryle et Quine, mâtiné d'une curiosité d'ingénieur le conduisant à s'intéresser sans relâche à la manière dont les choses fonctionnent<sup>33</sup>. Une telle généalogie permet de mieux suivre le cours de sa pensée.

Dennett hérita tout d'abord de Ryle une forme de scepticisme vis-à-vis des états mentaux; relever des régularités comportementales ne suffit pas à prouver qu'il existe un état mental sous-jacent bien délimité qui en serait responsable. De la même manière, ce n'est pas parce que certains termes ont une existence dans le langage qu'ils renvoient à un état de chose existant «dans le monde» — le spécialiste des licornes ne s'attend (généralement) pas à en croiser au coin de la rue. Fort de cet a priori méthodologique, Dennett s'est alors efforcé de comprendre comment l'esprit-cerveau fonctionne. Les découvertes que lui ont réservé les travaux des scientifiques ont confirmé ses intuitions: les états mentaux sont des entités dont on serait bien en peine de donner une description physique, tant les entités cérébrales censées leur correspondre leur sont peu ressemblantes. Sur quoi d'ailleurs se reposer pour décrire ces états mentaux? En tout cas pas sur l'introspection, tant celle-ci paraît peu fiable. Dennett décrit par exemple une expérience ahurissante menée par le neurochirurgien Walter Grey. Celui-ci avait installé des électrodes directement dans le cortex moteur de ses sujets, ce qui permettait à leur cerveau de déclencher une action dans un laps de temps très court. Il leur proposait ensuite de se passer, au rythme désiré, des diapositives en se servant d'un bouton. Mais ce bouton était en fait factice et il n'exerçait aucun effet sur le projecteur. Le résultat fut surprenant: les sujets avaient tous l'impression que le projecteur anticipait leur décision, alors qu'il ne se déclenchait en fait *que* lorsque le potentiel d'action était lancé et enregistré par les électrodes<sup>34</sup>. Une telle expérience, corroborée par d'autres du même type, semble montrer que la conscience est «à la traîne» par rapport aux processus cérébraux qui contrôlent effectivement notre corps. Il est par conséquent hautement critiquable de baser une théorie de l'esprit sur nos introspections qui, par définition, n'ont accès qu'aux phénomènes conscients. En fait, notre esprit se décompose en une

---

<sup>33</sup> DENNETT in GUTTENPLAN, 1994, p.236

<sup>34</sup> DENNETT, 1991, p.212.

myriade de micro-mécanismes «stupides» mais calibrés par l'évolution de manière à produire des comportements intelligents. Ces mécanismes opèrent en nous sans que nous nous en apercevions et, lorsque nous prétendons utiliser nos pouvoirs d'observation intime, nous sommes en fait toujours en train de nous engager dans une forme de *théorisation* improvisée<sup>35</sup>. En fait, lorsque nous nous interrogeons sur ce qui a présidé à notre comportement, nous appliquons à nous-mêmes un type de raisonnement comparable à celui que nous appliquerions pour expliquer le comportement de quelqu'un d'autre.

C'est à ce point que nous rencontrons la seconde inspiration de Dennett: Quine. Si les états mentaux ne renvoient en fait pas à des entités discrètes et isolables, comment se fait-il que nous utilisions des concepts comme «croyance», «désir» ou «intention» dans nos explications quotidiennes, stratégie qui entraîne d'ailleurs un certain succès prédictif? En fait, la vie en société exige que nous soyons capables d'interpréter le comportement de nos semblables. Or, si ce comportement résulte effectivement des activations simultanées de millions de neurones reliés entre eux par des connexions qui ont été calibrées par la sélection naturelle et l'apprentissage individuel, alors le système en cause est d'une telle complexité que tout comportement resterait définitivement imprévisible. L'évolution aurait alors sélectionné dans le règne humain une *stratégie interprétative* facilitant la vie en commun: la psychologie populaire<sup>36</sup>. Cette dernière permet à chacun de nous de mettre spontanément en oeuvre des concepts mentalistes qui produisent une approximation tout à fait satisfaisante de la plupart des comportements. Il s'agit donc d'une stratégie qui consiste à attribuer à un système quel qu'il soit — il s'agit le plus souvent d'êtres humains mais également d'animaux, et même de certains artefacts (un ordinateur) ou événements naturels (une tornade) — les croyances et les désirs qu'il est censé avoir étant donné sa place et ses objectifs. Mais attention, pour Dennett le fait qu'une telle stratégie fonctionne dans de nombreux cas ne signifie pas que le système possède effectivement des états comme des croyances ou des désirs — cela est particulièrement évident dans le cas où l'on dit qu'un ordinateur «ne veut pas exécuter une opération parce qu'il n'a pas assez de mémoire vive», par exemple. Au contraire, les entités mentales sont à son avis des *abstracta*, des concepts abstraits comparables à ceux de «centre de gravité» ou de «ligne de l'équateur». Les croyances ou les désirs ne font ainsi pas partie de l'ameublement du monde: ils sont des *instruments* qui nous permettent d'approximer le comportement de nos semblables et qui ont été sélectionnés pour cette raison par l'évolution. Et comme le langage joue un rôle crucial dans la maîtrise d'une telle «stratégie intentionnelle», il peut nous conduire,

---

<sup>35</sup> DENNETT, 1991, p.93.

<sup>36</sup> DENNETT, 1987 est consacré à cette stratégie de l'interprète.



en nous forçant à donner des formes verbales aux états mentaux, à supposer que les croyances et les désirs sont des phrases stockées quelque part dans notre tête<sup>37</sup>. Selon Dennett, c'est précisément l'erreur commise par Fodor qui se serait laissé berner par les caractéristiques «de surface» de la psychologie populaire en attribuant une dignité ontologique à toutes les entités supposées par ce qui n'est finalement rien d'autre qu'une «technologie sociale». Pour reprendre ses termes, la psychologie populaire n'est qu'«une méthode d'interprétation idéalisante, abstraite et instrumentale, qui est un produit de l'évolution parce qu'elle marche et qui marche parce que nous sommes des produits de l'évolution»<sup>38</sup>.

La pensée de Dennett, qui utilise les recherches scientifiques récentes pour tordre le coup à nos intuitions, est l'une des plus stimulantes qui existent sur le «marché des idées» en philosophie de l'esprit. Elle permet par ailleurs de profiter des avantages de plusieurs des conceptions présentées jusqu'ici. Tout d'abord, en s'attachant à la manière dont «se fabriquent» concrètement nos pensées, Dennett s'inscrit pleinement dans le courant naturaliste et profite des avancées récentes des neurosciences. Mais, d'autre part, il ne rejette pas à proprement parler le fonctionnalisme et l'ambition corrélatrice d'étudier en tant que telle la dynamique des états mentaux; c'est surtout au fonctionnalisme de Fodor et à son hypothèse du langage de la pensée qu'il en veut. Si Dennett peut cumuler autant d'avantages, c'est parce qu'il pense, tout comme David Marr, que les choses de l'esprit peuvent se décrire selon trois niveaux différents, et donner ainsi lieu à trois «stratégies» descriptives<sup>39</sup>. On peut tout d'abord adopter le «point de vue du physicien» en décrivant le substrat physique de nos activités physiques: ce point de vue est le seul auquel s'attachent les neurophilosophes. Mais il est également possible d'adopter le «point de vue du concepteur» (*design stance*) qui prend en compte le «ce pour quoi» une chose a été conçue: c'est la stratégie du fonctionnaliste «téléologique» — sur lequel nous revenons plus bas — qui vise à comprendre la nature d'un mécanisme en se demandant quel rôle il joue dans le fonctionnement d'un système plus vaste. Enfin, il est possible d'appréhender les choses du «point de vue de l'interprète», en leur attribuant des croyances et des désirs qui permettent de rendre compte de leur comportement: il s'agit là plus spécifiquement de l'option instrumentaliste de Dennett. Le problème principal de la stratégie globale de Dennett réside dans la difficulté à relier ces différents niveaux d'analyse; chacune des stratégies semble exister de manière autonome, le choix de l'une d'entre elles reposant sur le bon vouloir de l'interprète et sur les contraintes de la situation. Mais comment les capacités mises en oeuvre au sein de la stratégie intentionnelle reposent-elles sur des

---

<sup>37</sup> DENNETT, 1991, p.155.

<sup>38</sup> Ibid., p.68.

<sup>39</sup> MARR, 1982.

fonctionnalités de l'organisme, elles-mêmes rendues possibles par un arrangement physique cérébral? Là, Dennett n'apporte pas de réponses. Pire, la stratégie intentionnelle ne consiste qu'en une projection d'états mentaux «théoriques» qui ne jouent donc en tant que tels aucun rôle causal. Bref, la théorie de Dennett, aussi séduisante soit-elle, n'échappe pas au reproche d'épiphénoménalisme.

### **Vers une solution «téléosémantique»**

Dans la dernière partie de ce travail, nous allons nous demander s'il est possible de «sauver» les états mentaux, de leur attribuer une nature qui ne les réduirait pas à n'être que des épiphénomènes dépourvus de toute efficacité causale. Pour ce faire, il faut arriver à les inscrire dans l'ordre naturel et causal. Fred Dretske et Ruth Millikan se sont efforcés d'insuffler à la philosophie de l'esprit une inspiration biologique. Ainsi, pour Millikan, si nous sommes capables de croire ou de désirer, c'est parce que de telles capacités (et les organes qui les portent) ont rendu historiquement un service qui nous a aidés à proliférer<sup>40</sup>. Pour comprendre le rôle des états mentaux dans l'adaptation de notre espèce au milieu, il faut alors s'attacher à décrire la manière dont l'organisme est relié à son environnement. Dretske remet ainsi nettement en cause le fonctionnalisme «classique», qui ne s'intéressait qu'aux interactions des états mentaux au sein d'un esprit individuel (solipsisme méthodologique), pour se concentrer sur les relations entre le cerveau et le monde extérieur. Son pari consiste à rendre le contenu de l'état mental indissociable de sa fonction: de cette manière, les états mentaux possèderaient en tant que tels un pouvoir causal et le danger d'épiphénoménalisme serait repoussé.

En fait, l'organisme doit être capable, pour survivre dans des conditions environnementales changeantes, de se tenir informé des modifications de son milieu. Selon Dretske, il existe dans la nature des informations objectives, c'est-à-dire des relations régulières entre des phénomènes dont l'existence ne repose pas sur la présence d'un observateur. Ainsi, par exemple, les cercles concentriques d'un arbre indiquent ou signifient son âge, que quelqu'un y prenne garde ou non<sup>41</sup>. Certains traits propres au monde physique peuvent ainsi jouer le rôle de signaux ou d'«indicateurs» par rapport à un certain type d'information. Dans certains cas, il est même possible que certains systèmes physiques aient été sélectionnés par l'évolution précisément parce qu'ils avaient la possibilité de détecter un type d'informations dont dépendait la survie d'un organisme. Tel est précisément le cas de l'appareil sensoriel des animaux qui a été «aménagé» par la nature pour que ce qui se passe à l'intérieur de l'organisme dépende

---

<sup>40</sup> MILLIKAN 1984, p.7.

<sup>41</sup> DRETSKE 1988, p.65.

de ce qui se passe à l'extérieur. Les cellules neuronales du cerveau auraient alors été sélectionnées du fait qu'elles étaient capables de *représenter* certains traits de l'environnement. De ce fait, le système sensoriel est composé d'organes grâce auxquels des dépendances continues sont maintenues entre l'extérieur (événements publics) et l'intérieur (processus neuronaux)<sup>42</sup>. Ainsi, un état du cerveau acquiert des propriétés mentales lorsque sa fonction consiste à transporter de l'information (dans le cas de la perception) ou à l'utiliser (dans le cas de la croyance). Les états mentaux peuvent ainsi être considérés comme des états internes d'un système ayant pour fonction de fournir et de mobiliser de l'information afin de participer au contrôle du comportement. Comme le dit joliment Dretske, l'esprit, terme permettant de désigner l'ensemble des fonctions permettant à l'organisme de rester en contact avec le milieu afin de s'y adapter, serait en quelque sorte «la face externe du cerveau»<sup>43</sup>.

Dans une telle conception, les états mentaux, tout en n'étant «rien d'autre» qu'une certaine organisation de traits cérébraux en relation avec des traits physiques externes, exercent bel et bien une causalité qui leur est propre. S'ils ont été «recrutés» parmi d'autres états physiques (réseaux de neurones), c'est parce qu'ils ont la propriété de pouvoir représenter un trait de l'environnement dont la «mise en mémoire» constitue un avantage pour l'organisme. Ces états complexes du système nerveux central jouent donc un rôle causal en vertu du contenu représentationnel qui est le leur. Autrement dit, les caractéristiques mentales qui leur sont propres sont fondamentales et exercent en tant que telles un pouvoir causal; bref, les états mentaux sont bel et bien «réels». Si ma soif me conduit en direction du frigo, c'est parce que je dispose d'une structure cérébrale représentationnelle qui joue pour moi le rôle d'une «carte» cognitive; et c'est l'activation de celle-ci qui me permet d'adopter un comportement efficace (me diriger vers le frigo). Les états mentaux ne sont alors pas des *irrealia* mais bien des structures dotées d'efficacité causale.

La «téléosémantique» de Dretske et Millikan comporte de nombreux avantages. Elle s'inscrit tout d'abord dans une perspective biologique qui permet de considérer les états mentaux comme le fruit d'une série d'adaptations physiques sélectionnées au cours de l'histoire de l'espèce. Une telle approche permet de simplifier quelque peu les problèmes posés par la complexité de l'esprit humain en remontant aux facteurs qui ont favorisé l'émergence de telle ou telle capacité; le cerveau peut ainsi être vu comme un système qui a dû s'équiper de différents mécanismes fonctionnels pour résoudre certains types de problèmes. Comme les nombreuses difficultés auxquelles sont confrontées les espèces animales sont souvent homologues, l'étude de l'esprit dans une perspective

---

<sup>42</sup> Ibid., p.62.

<sup>43</sup> DRETSKE, in GUTTENPLAN 1994, p.260.

naturaliste peut ainsi s'appuyer sur le fonctionnement d'organismes plus simples afin de saisir les mécanismes fondamentaux à l'oeuvre dans l'esprit humain; cette approche évolutionniste nous paraît plus que prometteuse<sup>44</sup>. De plus, un tel cadre théorique accorde aux états mentaux une véritable dignité ontologique et permet d'envisager leur étude dans une perspective qui soit à la fois naturaliste et fonctionnaliste tout en évitant, tant que faire se peut, la dérive épiphénoménaliste<sup>45</sup>. Ce programme de recherche n'est néanmoins pas dépourvu de défauts — de jeunesse? Tout d'abord, des notions fondamentales comme celles d'information, de fonction ou de contenu représentatif sont relativement peu développées. Dretske donne l'impression de considérer les états mentaux comme des «enregistrements» qui, une fois stockés, semblent se caractériser par une certaine inertie. Comment dès lors rendre compte du dynamisme des états mentaux, et notamment du caractère plus ou moins flexible des croyances? De plus, l'approche téléologique s'est focalisée sur la description d'états mentaux élémentaires, élaborés en réponse à des régularités environnementales. Il est nettement plus difficile de voir comment aborder dans une telle perspective des états mentaux plus complexes qui mettent en jeu des raisonnements, des échanges interpersonnels ou encore des idéologies.

## **LA NATURE DES REPRÉSENTATIONS MENTALES**

Mis à part l'effort d'organisation inhérent à l'élaboration d'un état des lieux en philosophie de l'esprit, l'ordre de présentation des différents courants de pensée que nous avons suivi reflète bien évidemment des préférences et des options personnelles. A vrai dire, nous avons vu que chacune des tentatives visant à résoudre le problème des relations entre le corps et l'esprit apportait certains éléments satisfaisants sans combler pour autant notre curiosité.

Le «dualisme» flattait notre sens commun mais heurtait notre «Weltanschauung» scientifique du monde; le «behaviorisme logique» mettait fort justement l'accent sur la nature dispositionnelle de bon nombre d'états mentaux mais ne parvenait pas véritablement à les éliminer; le «matérialisme éliminativiste» s'appuyait sur une description du cerveau pertinente mais sans nous convaincre que nous pourrions nous passer des états mentaux; le «matérialisme de l'identité» montra rapidement ses limites; le «monisme anomal» acceptait de prendre en compte les états mentaux mais se refusait

---

<sup>44</sup> Il suffit par exemple pour s'en convaincre de lire les travaux que Cosmides & Tooby ont mené dans une optique évolutionniste (TOOBY & COSMIDES, 1990, COSMIDES & TOOBY, 1994).

<sup>45</sup> Pour prendre conscience des possibilités d'une approche de ce type, voir PROUST, 1997.

à en donner une description scientifique; le «fonctionnalisme» ouvrait la voie à une science du mental mais donnait, dans sa version «classique» ou «computationnelle», trop d'importance au modèle langagier; enfin, l'«instrumentalisme» offrait un aperçu saisissant des apports que les neurosciences peuvent avoir pour la réflexion philosophique sans éviter pour autant l'épiphénoménalisme. L'option «téléosémantique» nous paraît finalement constituer un formidable pari explicatif de type naturaliste. Le recours à la théorie de l'évolution permet de concevoir l'esprit comme un ensemble de «bricolages» évolutifs qui ont servi, à un moment de l'histoire d'une espèce, à résoudre un problème de survie, et qui continuent à (co)opérer au sein d'un système cognitif devenu fort complexe. De plus, ce programme de recherche repose sur l'idée que les états mentaux sont bien réels et il permet de reprendre à son compte de vieilles ambitions: comprendre les «lois» de la pensée, les mécanismes qui président à la succession des idées, à l'acceptation des croyances.

Comment participer à l'accomplissement d'un tel programme? Tout d'abord, il nous paraît essentiel de s'attacher à la notion de «représentation». Actuellement, c'est en s'appuyant principalement sur ce concept que les critiques d'une conception réaliste des états mentaux exercent leur travail de sape. Pour Vincent Descombes, par exemple, il est absurde de penser que le rapport au monde est médiatisé par des représentations conçues d'après le modèle des objets physiques<sup>46</sup>. En effet, si l'on ne pouvait entrer en contact avec le monde que par l'intermédiaire de représentations conçues sous la forme de «présences à l'esprit», alors elles devraient se manifester à nous d'une manière ou d'une autre. Mais où pourraient-elles bien faire «acte de présence»? Y aurait-il représentation sur la scène du «théâtre cartésien de l'esprit»? Si tel était le cas, il faudrait imaginer un public, quelqu'un qui puisse contempler les idées représentatives, autrement dit des «homoncules» capables d'observer de l'intérieur ces entités représentatives. Une telle conséquence, absurde d'un point de vue scientifique, conduit de plus à une régression à l'infini car on peut ensuite se demander ce qui se passe dans l'esprit de ces homoncules. Bref, pour sauver les états mentaux, il faut davantage s'interroger sur la nature des représentations. Ce travail peut bénéficier des apports des neurosciences, qui utilisent abondamment la notion de représentation. Mais celle-ci n'est pas tout à fait un objet comme un autre, qui en tant que tel pourrait entrer dans des relations causales du type «boules de billard». Sans entrer dans les détails, les neurophysiologistes considèrent que les représentations renvoient à des assemblées de neurones reliées par des connexions dont le nombre et la force sont pour la plupart déterminés dans les interactions avec le milieu. Ce sont ces assemblées de neurones qui jouent le rôle d'indicateur dont parlait Dretske. Lorsqu'une assemblée entre en activité

---

<sup>46</sup> DESCOMBES, 1996.

alors que quelque chose d'important pour la survie de l'organisme est capté par le système perceptuel, elle peut être sélectionnée et devenir «porteuse» de l'information en question<sup>47</sup>. Mais un tel enregistrement n'est pas à considérer comme une fiche sur laquelle serait définitivement inscrite l'information ou comme une image collée dans un «album mental». Le réseau neuronal en question est plutôt à considérer comme une sorte de «mode d'emploi» permettant de reconstituer, un peu sous la forme d'un hologramme, l'information qui a laissé une trace dans le système nerveux au moyen de connexions synaptiques. Ainsi, ce ne sont pas à proprement parler des objets mentaux qui sont stockés, mais la «marche à suivre» pour reconstituer, selon les besoins de la situation, les informations qui ont donné lieu à une inscription<sup>48</sup>. C'est la raison pour laquelle les états mentaux, qui peuvent dès lors être vus comme des ensembles complexes d'activations reconstituant au coup par coup des pans entiers d'informations, ne peuvent être vus comme des objets exerçant une causalité en vertu de leur forme ou de leur poids — comme peuvent le faire des boules de billard. Leur pouvoir causal, les représentations le doivent aux multiples trajets électriques qui se propagent au sein des assemblées de neurones qui les constituent. Mais ces dernières ne peuvent mettre en oeuvre leur *force* que dans la mesure où elles sont rassemblées pour donner lieu à un état mental. C'est en ce sens que l'on peut dire que les états mentaux *surviennent* sur les états physiques. Vues sous cet angle, les représentations ne sont plus alors à considérer comme un *quelque chose* que l'on pourrait compter, ou dénombrer. En tant que supports d'information, les assemblées de neurones, dont les interconnexions sont quasi infinies, peuvent donner naissance à des représentations sensiblement différentes selon le contexte qui leur donne vie. Ainsi, une représentation qui s'est constituée principalement par le biais des activations visuelles ne renvoie pas forcément à une image définie une fois pour toute de façon immuable; elle peut prendre plusieurs formes et être activée de manière différentielle selon le contexte. (Imaginez une bière — est-elle en bouteille, dans une chope, blonde, brune, froide, tiède?) Cette malléabilité des représentations correspond d'ailleurs bien à notre expérience et laisse envisager une description plus fine du dynamisme de nos processus mentaux.

Ce caractère dynamique de la vie mentale a d'ailleurs été peu traité jusqu'à présent en philosophie de l'esprit. Ce désintérêt est dû pour une part à l'importance donnée à la notion d'*état* mental, notion qui encourage inévitablement une perspective assez statique. Mais surtout, la philosophie de l'esprit, conformément à l'esprit des sciences cognitives, s'est jusqu'ici peu attachée à la dimension émotionnelle, qui joue pourtant sans aucun doute un rôle fondamental dans la dynamique du mental. Si les émotions ont

---

<sup>47</sup> Ce mécanisme est bien décrit par CHANGEUX, 1983.

<sup>48</sup> Damasio parle à ce sujet de «représentations potentielles», DAMASIO, 1994, p.139.

peu retenu l'attention des philosophes de l'esprit contemporains, c'est en partie en raison de la difficulté à aborder ce sujet sous un angle qui corresponde aux ambitions naturalistes. Cette angoisse légitime peut cependant être tempérée, ici aussi grâce aux travaux des neuroscientifiques et des biologistes. Damasio a ainsi démontré que les émotions jouent un rôle tout à fait fondamental au coeur même des processus cognitifs de traitement de l'information, remettant ainsi en cause la division traditionnelle entre la raison et les passions<sup>49</sup>: dépourvu de passions, l'homme est un animal qui raisonne fort mal. L'hypothèse de Damasio est que les systèmes impliqués dans le raisonnement et la prise de décision sont étroitement reliés avec ceux qui sous-tendent la régulation biologique<sup>50</sup>. Les circuits bio-régulateurs doivent en effet être tenus au courant de ce qui se passe dans toutes les parties de l'organisme. Le caractère "bon" ou "mauvais" des situations rencontrées leur est ainsi constamment signalé, ce qui leur permet de réagir de façon "préprogrammée" sur les processus en cours. La capacité à percevoir les émotions constitue ainsi un mécanisme permettant de détecter la bonne ou la mauvaise adéquation entre les adaptations de l'organisme et les circonstances extérieures<sup>51</sup>. Les émotions jouent alors un rôle «directeur» fondamental dans la conduite de nos activités mentales en les orientant vers l'endroit où la bonne décision, en terme de survie, doit être prise grâce à la mise en oeuvre de capacités cognitives plus récentes en termes d'évolution. On le voit, le rôle des émotions paraît dès lors si fondamental qu'il semble impossible de s'en passer pour rendre compte des phénomènes mentaux<sup>52</sup>.

Enfin, la volonté de comprendre comment l'esprit peut émerger du cerveau, l'âme du corps, nous a forcé à nous focaliser sur des phénomènes mentaux élémentaires. Il ne faudrait néanmoins pas oublier qu'une part importante de notre activité cérébrale met en jeu des processus de haut niveau qui, il faut l'avouer, sont encore très difficilement saisissables du point de vue neurophysiologique. Dans une perspective naturaliste, pour laquelle il ne saurait être question de causalité qu'en termes physiques, il est ainsi fort difficile de comprendre comment certains raisonnements peuvent être doués d'une force quasiment causale. La manière dont les «lois» de la logique exercent leur pouvoir pourra éventuellement être décrite un jour en termes neurologiques<sup>53</sup>, mais il restera difficile de dire s'il s'agit de nécessités propres au cerveau ou de nécessités propres au langage et à la communication qui ont ensuite été implémentées neuronalement. Dans le même ordre d'idées, il nous arrive de temps à autres d'éprouver des prises de

---

<sup>49</sup> C'est à cette tâche qu'il s'attelle avec succès dans DAMASIO, 1994.

<sup>50</sup> DAMASIO, 1994, p.117.

<sup>51</sup> Ibid., p.12.

<sup>52</sup> Un philosophe de l'esprit, Ronald De Sousa, avait pressenti cette importance des émotions dans la conduite du comportement (DE SOUSA, 1987).

<sup>53</sup> Comme tente actuellement de le faire Changeux (CHANGEUX, 1992 et CHANGEUX & DEHAENE, 1989).

conscience qui nous conduisent à réviser, de manière réfléchie et parfois très rapide, des croyances qui bénéficiaient jusque-là de toute notre confiance<sup>54</sup>. C'est alors comme si des représentations de niveau supérieur (des «métareprésentations») bénéficiaient d'une force causale supérieure à celle des milliers de connexions qui sous-tendaient jusqu'ici nos représentations<sup>55</sup>. Le pouvoir des représentations consciemment entretenues et des raisonnements auxquels elles donnent forme peut donc être immense et il nous faut avouer que la complexité abyssale de l'architecture de l'esprit sur laquelle repose sa puissance laisse exsangues nos propres capacités représentationnelles.

Fabrice Clément

---

<sup>54</sup> Le problème de la conscience intéresse de plus en plus le champ des sciences cognitives. Pour de bonnes précisions conceptuelles, voir BLOCK, 1995. Varela propose une approche originale de la conscience en se basant aussi bien sur les neurosciences que sur le bouddhisme et la phénoménologie (VARELA, THOMPSON & ROSCH, 1993).

<sup>55</sup> La question des métareprésentations a donné lieu à une littérature considérable et passionnante ces dernières années. Voir par exemple PERNER, 1991, LESLIE, 1987, SPERBER, 1990, ENGEL, in KAMBOUCHNER, 1995. Annette Karmiloff-Smith et Andy Clark ont tenté de donner une description naturaliste des mécanismes en jeu (CLARK & KARMILOFF-SMITH, 1993).