

## LE CERVEAU EST-IL LE PILOTE DU NAVIRE DE L'ESPRIT?

Pascal Engel

Université de Genève

L'intérêt des philosophes pour le cerveau et le système nerveux n'est pas récent : de Descartes à Bergson en passant par Diderot, La Mettrie, Maine de Biran et James, ce qu'ont dit médecins et physiologistes sur cet organe a toujours retenu l'attention des philosophes. Pourtant ce n'est que durant la seconde moitié du dix-neuvième siècle avec les travaux de Flourens, Broca, Wernicke, Helmholtz, et surtout Ramon y Cajal, qui fut le premier à décrire l'organisation cérébrale et les connexions neuronales, que le cerveau est devenu un objet philosophique, et le lieu par excellence de l'étude des relations du corps et de l'esprit<sup>1</sup>. Plus personne ne doute aujourd'hui du fait que s'il y a quelque chose à comprendre à l'esprit, c'est là que cela se passe.

Les philosophes ont pourtant mis du temps à mettre le cerveau au centre de leurs réflexions sur les relations entre l'esprit et le corps. Bergson, malgré son intérêt pour la neurologie, considérait le cerveau simplement comme un central téléphonique et dévaluait sa fonction cognitive pour accentuer son rôle de planification de l'action. Le cerveau n'avait pas plus de place dans la conception béhavioriste de l'esprit, puisque le mental s'y réduit à sa face strictement externe et aux seules relations entre des *stimuli* et des réponses comportementales, laissant de côté la « boîte noire » interne. La conception philosophique qui incarne le mieux, au vingtième siècle, cette mise à distance du pouvoir causal du

---

<sup>1</sup> Voir Jeannerod, M. 1983 *Le cerveau machine*, Paris, Fayard

cerveau est le « béhaviorisme logique » défendu par des positivistes logiques comme C. Hempel<sup>2</sup>, qui entend réduire la signification des énoncés psychologiques à celle d'énoncés portant sur le comportement observable. Par exemple, « Il a mal aux dents » ne signifie pas autre chose que : « Il grimace en se tenant la mâchoire et fait tels ou tels gestes ». Le problème est que c'est circulaire : on ne peut définir les états mentaux par des comportements que moyennant la supposition que les comportements en question sont causés par des états mentaux. Dans *The Concept of Mind*<sup>3</sup>, Gilbert Ryle soutient une version subtile de cette thèse, contre le mythe « cartésien » de l'esprit comme « fantôme dans la machine », et en défendant l'idée selon laquelle l'esprit n'est pas quelque chose d'interne, physique ou mental, mais un ensemble complexe de dispositions. Bien qu'elle soit plus complexe, la conception de Wittgenstein<sup>4</sup>, selon laquelle un processus interne a besoin de critères externes, participe de la même méfiance vis-à-vis de l'idée que l'on pourrait loger l'esprit *au dedans*. Comme le dit W.V.O. Quine, reprenant le slogan des peintures Sherwin-Williams, « sauvez la surface et vous sauvez le reste »<sup>5</sup>.

La réaction est venue de philosophes comme H. Feigl, U. T. Place et J.J.C Smart<sup>6</sup>, qui ont défendu la thèse de l'identité de l'esprit et du cerveau, que l'on a appelée aussi « matérialisme de l'état central ». Ils soutiennent

---

<sup>2</sup> C.Hempel, "L'analyse logique de la psychologie", *Journal de Synthèse*, 10, 1935 repris in Fiset, D. et Poirier, P. eds. *Philosophie de l'esprit*, textes clés, Paris Vrin 2002 vol. 1

<sup>3</sup> Ryle, G. 1949 *The concept of Mind*, London, Hutchinson, tr.fr. *Le concept d'esprit*, Paris, Payot 1976

<sup>4</sup> Wittgenstein, L. 1953 *Philosophische Untersuchungen*, Blackwell, Oxford, tr.fr. Paris, Gallimard 2005

<sup>5</sup> W.V.O. 1990 *Pursuit of Truth*, Cambridge Mass, Harvard University Press, tr. fr. *Poursuite de la vérité*, Paris, Seuil 1993

<sup>6</sup> Place, U.T. 1956 « Is consciousness a Brain-Process ? », *British Journal of Psychology*, 71, Smart, J.J. 1963 *Philosophy and Scientific Realism*, London, Routledge

que les états mentaux, et en particulier les états de conscience, sont identiques à des états physiques du cerveau. Quand on parle d'identité en ce sens, on veut dire que ce sont des types de propriétés mentales (la douleur, les états conscients, les croyances, les désirs, etc.) qui sont identiques à des types d'états cérébraux et que l'on doit pouvoir réduire scientifiquement les premiers aux seconds, au même titre que l'on réduit l'eau à de l'H<sub>2</sub>O ou la chaleur à de l'agitation moléculaire. Mais cette thèse se heurte à des objections sérieuses. La première est que la réduction proposée n'est que de principe, car elle n'est pas empiriquement vérifiée. La seconde est que la théorie de l'identité présuppose un modèle très fort de la réduction scientifique, qui requiert que les propriétés mentales et les propriétés physiques soient co-extensives. Selon la conception de la réduction des théories d'E. Nagel<sup>7</sup>, cela requiert l'existence de « lois-ponts » permettant de réduire les lois d'un domaine aux lois d'un autre domaine. Mais l'existence de tels principes « ponts » pour le mental et le physique est précisément ce qui pose problème. S'il existait des lois-ponts de ce genre, il faudrait que toute propriété mentale « réalisée » dans un organisme ait un même substrat physique dans les autres organismes. Cela supposerait, par exemple, que la douleur soit réalisée par le même substrat neuronal chez un humain, un chat, un poulpe ou une grenouille. Mais même s'il a bien quelque chose de commun quand ces différents types d'organisme éprouvent de la douleur, il est clair que les événements qui se déroulent dans le système nerveux de chacun ne sont pas identiques. C'est vrai aussi d'un individu à l'autre au sein d'une même espèce. C'est ce que l'on a appelé la « réalisabilité multiple » des états mentaux, et c'est un obstacle sérieux à l'idée d'une réduction esprit-cerveau. La seconde

---

<sup>7</sup> Nagel, E. 1960 *The Structure of Science*, London, Routledge

objection est que l'identité a sa logique propre : si on entend dire que tout état mental conscient (la propriété d'éprouver une douleur ou la perception visuelle, par exemple) est identique à un état neuronal ou à une configuration d'états neuronaux, il faut admettre que selon le principe d'indiscernabilité des identiques toute propriété de propriété mentale est aussi une propriété d'un état neuronal. Or c'est absurde : si vous voyez du rouge, votre expérience a la propriété de porter sur du rouge, mais cela n'implique pas que votre état neuronal soit rouge. Ici le matérialiste peut répondre qu'il faut distinguer le processus en quoi consiste la propriété d'avoir une douleur de la sensation qualitative éprouvée quand on a une douleur, tout comme il faut distinguer la référence des expressions « l'étoile du soir » et « l'étoile du matin » (qui est la même) du sens de ces expressions (qui est différent). Mais cette distinction entre ce que serait « en soi » la douleur (un état cérébral) et la manière dont elle serait ressentie est justement ce que nie le dualisme: selon un dualiste au contraire, l'essence de la douleur, et d'autres états qualitatifs ou phénoménaux *est* son mode d'apparaître et il y a un gouffre irréductible entre le cerveau et les *qualia* conscients<sup>8</sup>.

Face à ces objections le théoricien de l'identité peut adopter deux sortes de stratégies. La première consiste à nier que la manière dont nous apparaissent les événements mentaux et le contenu de nos expériences ait quelque importance que ce soit quant à leur nature et leurs pouvoirs causaux. Selon cette conception, une fois que les bases neuronales de la conscience auront été établies, il n'y aura tout simplement plus lieu de continuer à parler d'«expériences conscientes», de «perceptions», de

---

<sup>8</sup> Kripke, S. 1980 *Naming and Necessity*, Oxford, Blackwell, tr.fr. *La logique des noms propres*, Paris, Minuit

«souvenirs», de «croyances», de «désirs», etc. La version la plus radicale de cette thèse « éliminativiste » est défendue par P. Churchland<sup>9</sup>, pour qui notre « psychologie populaire » ou naïve n'est qu'une mythologie ou une théorie fautive, au même titre que la chimie du phlogistique, qui sera remplacée, quand notre science du cerveau sera plus avancée, par des taxinomies appropriées en termes de connexions neuronales. Éliminer le mental et le réduire ne reviennent pas à la même chose. Réduire une théorie à une autre, par exemple la thermodynamique phénoménologique à la thermodynamique statistique, c'est admettre que les phénomènes décrits par la première sont réels, alors qu'éliminer une théorie c'est soutenir qu'elle est fautive et que ses termes ne désignent rien. Mais l'élimination de la psychologie populaire par les neurosciences ne s'est pas produite, parce que la première n'est tout simplement pas une théorie scientifique ou proto-scientifique de l'esprit, mais un schème général de compréhension de nous-mêmes, une « posture intentionnelle » inéliminable<sup>10</sup>. Même à supposer que des Martiens omniscients en neurosciences et en physique viennent sur Terre, et prévoient par des calculs très compliqués des comportements humains ils auraient bien du mal à pouvoir adopter la perspective que les humains ont sur leurs actions et leurs sentiments.

La seconde stratégie que peut adopter le théoricien de l'identité consiste à affaiblir sa thèse, en renonçant à l'idée qu'on puisse identifier des types d'états mentaux à des types d'états physiques, pour adopter seulement la thèse selon laquelle tout événement mental particulier est identique à un événement physique particulier *ou un autre*. C'est une conséquence directe

---

<sup>9</sup> Paul Churchland, *Matter and consciousness*, Oxford University Press 1984, tr. Fr. *Matière et conscience*, Champ Vallon, 1996

<sup>10</sup> Dennett, D. 1991 *Consciousness Explained*, Littlebrown, NY, tr.fr. *La conscience expliquée*, Paris, O.Jacob

de la réalisabilité multiple. On admet ainsi que la douleur peut être réalisée par l'activation de certaines fibres neuronales chez les humains, mais qu'elle se réalise par d'autres événements physiques chez des organismes différents. Au lieu de soutenir que des propriétés comme «douleur» ou «croyance» sont co-extensives avec des propriétés physiques, on soutient que les propriétés mentales sont coextensives avec une disjonction d'occurrences particulières de propriétés mentales. C'est ce que l'on appelle la théorie de l'identité des occurrences. Mais il en existe, là aussi, deux versions très différentes.

La première version n'est que le développement de la vieille idée aristotélicienne selon laquelle l'âme est la forme du corps: la nature d'un état mental n'est pas constituée par son substrat matériel, mais par sa fonction. A la suite de Turing<sup>11</sup>, les philosophes fonctionnalistes contemporains, comme H. Putnam, D. Lewis et J. Fodor<sup>12</sup> ont exprimé cette idée en comparant l'esprit au logiciel d'un ordinateur qui peut s'adapter sur de multiples bases matérielles. La métaphore a fait florès, mais la thèse fonctionnaliste a une portée plus large. Elle dit qu'un état mental est constitué par sa fonction, laquelle est le rôle causal joué par cet état, constitué par ses causes caractéristiques et ses effets caractéristiques. Ainsi, la douleur est l'état mental ayant pour rôle de recevoir des entrées sensorielles détectant les dommages subis par l'organisme et qui a comme effets caractéristiques l'évitement de ces dommages, ou la croyance est l'état dont le rôle est, avec les désirs, de produire des actions. La thèse peut

---

<sup>11</sup> Turing, A. "Computing Machinery and Intelligence", *Mind* 1950, tr. fr. in AR. Anderson, ed. *Pensée et Machine*, Dijon, champ Vallon

<sup>12</sup> Putnam H. 1967 « The nature of mental States » in Fissette et Poirrier, op.cit, 1989 *Representation and Reality*, Cambridge Mass, MIT Press, tr.fr. *Représentation et réalité*, Paris, Gallimard ; Lewis D. 1966 "How to define theoretical terms", *Philosophical Papers*, 1992, Oxford, Oxford University Press; Fodor, J. 1975 *The language of thought*, Cambridge, Mass, MIT Press

être étendue à l'idée de fonction biologique, ce qui la détache de la métaphore de l'esprit comme ordinateur. Pendant près de trente ans, le fonctionnalisme a été la thèse officielle des sciences cognitives. Il a permis l'essor d'un type de recherche qui se trouvait au croisement des neurosciences, de la psychologie cognitive, de l'intelligence artificielle et de la robotique. Le vieux projet positiviste d'une unité de la science, qui allierait physique, biologie, neurosciences et sciences sociales a vécu alors une seconde jeunesse. On s'est cependant rendu compte assez vite que l'unification n'était pas aussi aisée.

Le fonctionnalisme est un physicalisme faible. Il admet que tout événement mental est un événement physique, mais il n'admet pas que tout type d'événement mental soit un type d'événement physique. Il n'implique donc pas le réductionnisme. En ce sens, ni la psychologie, ni la biologie, ni les sciences sociales et ce que l'on appelle les « sciences spéciales » ne sont réductibles à la physique, parce qu'elles reposent toutes sur des généralisations structurales. La psychologie cognitive a tiré parti de cette thèse en admettant une méthodologie fonctionnaliste qui fait largement abstraction du niveau neuronal de description.

Jusqu'à quel point le fonctionnalisme peut-il faire abstraction non seulement des détails physiques, mais aussi de certaines caractéristiques mentales? Il semble s'adapter à des états mentaux qui jouent un rôle explicite dans l'action, tels que les croyances et les désirs, mais il est bien plus problématique quand il s'agit d'états mentaux dont la nature n'est pas relationnelle, mais intrinsèque, tels que les expériences vécues et les propriétés phénoménales. Même si on définit la douleur par son rôle fonctionnel, il n'en demeure pas moins qu'elle est toujours associée à une

certaine qualité ressentie, un « quel effet cela fait »<sup>13</sup>. Mais un « zombie » qui serait dépourvu d'expériences conscientes bien qu'il eût toutes les fonctions humaines habituelles aurait-il vraiment des états mentaux<sup>14</sup>? Un individu placé dans une chambre où il recevrait des instructions en chinois qu'il ne comprendrait pas mais les couplerait à des réponses correctes au moyen d'un ensemble de règles associant des phrases chinoises à d'autres phrases chinoises pourrait-il être crédité d'une compréhension du chinois<sup>15</sup>? Chacune de ces expériences de pensée laisse entendre que les mailles du filet fonctionnaliste sont trop larges.

La seconde version de la thèse de l'identité des occurrences refuse l'idée même qu'il puisse y avoir des lois, fussent-elles fonctionnelles, portant sur les relations entre états mentaux. Le fonctionnalisme admet des généralisations du type: « Toutes choses égales par ailleurs, si X a soif et si X croit qu'il a un verre d'eau devant lui, il boira ce verre ». Mais jusqu'à quel point les choses sont-elles « toutes choses égales par ailleurs » ? Si X, croit que l'eau du verre est non potable, il n'accomplira pas l'action spécifiée. Il faut présupposer toutes sortes de conditions de normalité et de rationalité. A cela s'ajoute le fait que les états mentaux sont holistiques, et ne peuvent s'attribuer que sur l'arrière-plan d'autres attributions : on ne peut croire que les vaches sont des ruminants que si l'on a le concept de ruminer, celui de digestion, celui de mammifère, etc. Ce qui semble interdire la possibilité de lois psychologiques strictes, isolables à la manière dont peuvent l'être les lois physiques. Le déni de l'existence de lois psychologiques va de pair avec le déni de la théorie de l'identité des types,

---

<sup>13</sup> Nagel, T. 1979 *Mortal Questions*, Cambridge, Cambridge University Press, tr. fr. Questions mortelles, Paris, PUF 1983

<sup>14</sup> Chalmers, D. 1996, *The conscious mind*, Oxford university press

<sup>15</sup> Searle, J. 1992 *The Rediscovery of the Mind*, Cambridge, Mass, MIT Press, tr.fr. La redécouverte de l'esprit, Paris, Gallimard

car s'il n'y a pas de lois mentales, elles ne peuvent pas non plus être réduites à des états physiques. Mais il est cependant possible, comme l'a montré D. Davidson<sup>16</sup> de défendre la conjonction de thèses suivantes: (i) il n'y a pas de lois psychologiques ni psychophysiques strictes (ii) les événements mentaux et physiques entretiennent des relations causales, (iii) les événements mentaux occurrents sont identiques à des événements physiques. Ce « monisme anomal » est un matérialisme. Il accepte l'idée que le mental dépend du physique, et le principe dit de *survenance* (selon lequel le physique est suffisant pour le mental, ou selon lequel toute différence mentale implique une différence physique), mais il nie que cela implique la réduction. Le problème du matérialisme faible est qu'on ne voit plus en quoi les propriétés mentales peuvent avoir la moindre efficacité causale : elles deviennent purement épiphénoménales, et l'esprit se réduit au schème d'interprétation par lequel on attribue des états mentaux. A quoi peut bien alors servir l'idée que les événements mentaux sont des événements cérébraux ? Si l'on veut garder à la fois le matérialisme et la thèse selon laquelle les propriétés mentales ont un pouvoir causal et explicatif, le cerveau doit pouvoir être la cause de l'esprit et non pas seulement sa condition suffisante.

On semble arrivé à une impasse : ou bien l'identité du cerveau et de l'esprit est trop forte quand elle implique un réductionnisme ou un éliminativisme, ou bien elle est trop faible quand elle ne l'implique pas. Quelque chose ne va pas dans l'argument de la réalisabilité multiple du fonctionnaliste. Pourquoi devrions-nous supposer que la relation entre un esprit et un cerveau est aussi lâche que celle qui existe entre un logiciel et

---

<sup>16</sup> Davidson 1980, *Essays on actions and events*, Oxford, Oxford University Press, tr. fr. *Actions et événements*, Paris, PUF

sa réalisation matérielle dans un ordinateur. Et n'y a-t-il pas des limites dans le degré auquel on peut s'abstraire de la base matérielle ? Et jusqu'à quel point la variation dans le substrat change-t-elle la nature du type d'objet ? Ce qui pose problème, dans l'image fonctionnaliste est l'idée que les événements neuronaux qui instancient des lois fonctionnelles seraient des occurrences *disjointes*. Mais si les propriétés fonctionnelles (comme la douleur) surviennent sur des disjonctions d'événements neuronaux, comment est-il possible que ces événements n'aient aucun rapport entre eux ? Peut-on admettre qu'un gène ait des propriétés causales à un niveau supérieur ou « macro » sans que les configurations moléculaires qui les sous-tendent au niveau inférieur ou « micro » soient fortement unifiées<sup>17</sup> ? J. Kim<sup>18</sup> a montré qu'on devait adopter une conception de la réduction des propriétés mentales aux propriétés physiques plus forte que celle du fonctionnalisme. Il propose que l'on assigne une caractérisation fonctionnelle à toutes les propriétés psychologiques, qui spécifie le rôle causal de celles-ci, et qu'on admette que ces propriétés sont réalisées dans des configurations causales neuronales qui tombent elles-mêmes sous des types. Cela autorise au moins des réductions de propriétés au sein de chaque espèce : si la douleur n'est pas la même chose chez les poulpes et les mammifères elle doit être fortement unifiée au sein de chaque espèce respective. Le physicalisme doit être plus robuste qu'un physicalisme fonctionnaliste ou « anomal ».

La résistance philosophique à l'idée qu'une théorie de l'esprit doive être d'une manière ou d'une autre une théorie du cerveau s'appuie principalement sur deux types d'arguments, qui ont tous deux leur source

---

<sup>17</sup> Sober, E. 1999 "The Multiple Realizability Argument Against Reductionism" *Philosophy of Science*, , 66: 542-564).

<sup>18</sup> Kim, J. 1998 *Mind in a physical World*, Cambridge Mass, MIT Press, tr. fr. *L'esprit dans un monde physique*, Paris, Syllepse, 2005

chez Wittgenstein. Le premier est avancé par des philosophes critiques des sciences cognitives<sup>19</sup>. « Que le cerveau, nous disent-ils, soit l'organe de la pensée, le substrat de la conscience, des croyances et des émotions est une chose sur lequel on peut s'accorder. Ce n'est pas, pour paraphraser Aristote au sujet de l'âme (*De Anima*, I, 4) le cerveau qui pense, mais l'homme au moyen de son cerveau.» De même les philosophes qui critiquent les explications des phénomènes psychologiques des sciences cognitives aiment à dénoncer le «sophisme de l'homoncule» : ce n'est pas l'œil qui voit, ce n'est pas un homoncule en nous qui calcule, comprend ou infère, mais c'est l'individu tout entier. Mais cet argument est un *non sequitur* : le fait, indéniable, que les critères (nos concepts, la «grammaire») par lesquels nous reconnaissons la pensée, la conscience, ou la vision diffèrent de ceux par lesquels nous reconnaissons les mécanismes cérébraux n'implique pas que la pensée n'ait rien à voir avec les mécanismes en question. Le fait que nous appelions «rêve» le type d'événements dont nous sommes conscients pendant notre sommeil et dont nous nous souvenons au réveil n'implique pas que la nature des rêves puisse être constituée par des processus cérébraux dont nous ne sommes pas conscients.

Le second type d'argument provient des philosophes qui défendent une conception «externaliste» de l'intentionnalité et des contenus mentaux : ce que nous pensons ou percevons est déterminé dans une large mesure par la nature de l'environnement auquel nous appartenons, et un individu qui serait notre réplique physique mais qui vivrait dans un environnement différent du nôtre n'aurait pas les mêmes contenus mentaux. Il y a diverses variétés d'externalisme, selon qu'on «étend» l'esprit à l'action et aux relations à l'environnement physique et biologique, ou selon qu'on l'étend

---

<sup>19</sup> Descombes, V. 1995 *La denrée mentale*, Paris, Minuit

aux relations sociales et linguistiques. Selon l'externalisme social, qui retrouve en cela la critique d'Auguste Comte contre l'assimilation de l'esprit au cerveau, l'esprit n'est pas au dedans, mais au dehors, dans ses œuvres, dans les institutions humaines. Tout dépend ici de la question de savoir si l'on entend garder une conception causale de l'esprit. Si les croyances, désirs, et autres états mentaux, y compris les états conscients ou *qualia* sont supposés jouer un rôle causal dans l'action, l'externalisme des contenus mentaux ne permet de l'expliquer, car les processus causaux doivent être *locaux*, c'est-à-dire situés là où se trouve le pilotage du système nerveux, et non pas *distaux* ou extérieurs à l'organisme. Il viole donc le principe de survenance du mental sur le physique. Quand je prends un coup de soleil, c'est certainement une cause extérieure à mon organisme qui agit, mais la cause est bien sur la surface de ma peau. L'externalisme a raison en ceci que l'esprit est très largement *constitué* par des relations externes (sociales, linguistiques, institutionnelles, historiques) à la seule boîte crânienne. Mais il a tort s'il confond cette constitution avec la causalité qu'exercent les états mentaux, et les états cérébraux, si les premiers surviennent sur les seconds. Si l'on veut maintenir la place de l'esprit dans la nature, le cerveau doit avoir un pouvoir causal. Il n'est peut être pas le seul pilote du navire de l'esprit, mais c'est sur son navire que nous sommes embarqués.

Les techniques de l'imagerie cérébrale et la compréhension des bases neurales de l'esprit ont changé la manière dont nous comprenons des questions telles que celle de l'identité humaine. Ainsi la médecine ne considère la mort cérébrale, et non plus l'arrêt du système cardiaque comme le signe de la mort clinique. L'époque semble lointaine où l'on se demandait si l'homme a une essence ou s'il constitué par la culture ou l'histoire. Plus personne ne doute que la compréhension de la relation de

l'inné et de l'acquis passe par la compréhension des bases génétiques de la nature humaine, et de la question de savoir comment les structures du cerveau ont évolué. Loin aussi semblent les discussions sur la maladie mentale et la psychopathologie qui opposaient les tenants de l'autonomie de la *psychè* et les « organicistes ». Non pas que les seconds aient gagné, mais le fait est que la question de la liberté et de la personne passe par la compréhension du degré de plasticité des structures neuronales et de leur relation à l'environnement. Ce que l'on appelait encore il y a un siècle le problème philosophique « de la connaissance d'autrui » ne peut plus être posé sans tenir compte ce que nous savons des structures de la « théorie de l'esprit » que notre espèce partage en partie avec les primates. Un grand nombre de nos maladies sont des maladies cérébrales : prions, Alzheimer. Cela ne change, à mon sens, rien à la nature des problèmes métaphysiques classiques, comme celui de la liberté et du déterminisme, qui ne sont nullement réglés, mais le cerveau est devenu l'organe métaphysique par excellence.

