

L'esprit

Le « Mind-Body problem »

Jean-Gérard Rossi

Philopsis : Revue numérique
<http://www.philopsis.fr>

Les articles publiés sur Philopsis sont protégés par le droit d'auteur. Toute reproduction intégrale ou partielle doit faire l'objet d'une demande d'autorisation auprès des éditeurs et des auteurs. Vous pouvez *citer* librement cet article en en mentionnant l'auteur et la provenance.

I. Considérations générales

1. Héritier du problème traditionnel de l'âme et du corps, le « Mind-body problem » se pose aujourd'hui en des termes nouveaux dans le cadre d'une tentative de « naturalisation » inspirée du modèle scientifique.

La plupart des philosophes contemporains versés dans les questions de philosophie de l'esprit se situent par rapport à la thèse cartésienne. Le dualisme de l'auteur des *Méditations Métaphysiques* est toujours invoqué pour être, la plupart du temps, récusé. On peut même dire qu'il est présenté de manière caricaturale pour mieux servir de repoussoir.

Dans une large mesure, il semble que ce soit G. Ryle le responsable de cette situation. C'est lui en effet qui a combattu le premier ce qu'il appelait « le dogme du fantôme dans la machine » en quoi lui semblait se résumer la position cartésienne.

On s'explique dans ces conditions que le problème de la causation mentale, principale difficulté rencontrée par Descartes soit au cœur des débats et des controverses et que la plupart des solutions proposées s'efforce en premier lieu de le résoudre.

La tonalité générale des études contemporaines sur la question est moniste. Le dualisme cartésien impliquait une référence à l'esprit ; le monisme contemporain prend une forme matérialiste. Ceci s'inscrit dans une

tentative de naturalisation des problèmes philosophiques, avec l'ambition plus ou moins avouée de garantir la scientificité des résultats de l'enquête.

Le monisme matérialiste peut revêtir plusieurs formes. Il peut reconnaître la spécificité du mental, au niveau des propriétés. On parle alors d'un dualisme des propriétés, pour l'opposer au dualisme des substances, mais l'expression peut se révéler égarante car en fait il reste au niveau du monisme. Mais le monisme matérialiste peut se révéler radical et se proposer alors d'éliminer toute terminologie mentaliste. Entre ces deux positions se situe toute la gamme des positions intermédiaires envisageables.

Ce qui est certain, c'est que la très grande majorité des auteurs anglo-saxons contemporains travaille dans le cadre d'un paradigme ne mettant pas en cause le monisme matérialiste et que toutes les controverses portent sur la meilleure façon d'accommoder ce choix théorique à nos intuitions courantes concernant le mental et notamment à ce qu'il est convenu d'appeler « la psychologie populaire ».

Ce paradigme vise explicitement à supplanter le paradigme dualiste élaboré par Descartes. Il se recommande d'une conception anti-métaphysique. La science ayant réussi à substituer une explication naturaliste à toutes les explications théologiques ou métaphysiques dans le domaine de la nature, pourquoi n'en serait-il pas de même dans le domaine de l'esprit ? Bien plus, ne doit-elle pas s'efforcer d'éliminer l'esprit et le résorber dans la Nature ? Sans doute y a-t-il là une pétition de principe car ce n'est que si on a décidé que l'esprit n'a pas de spécificité qu'on peut trouver aberrant de chercher pour lui un traitement différent de celui qui vaut pour le physique. Mais tel est le présupposé du nouveau paradigme. Présenté souvent comme relevant d'un pur choix méthodologique en faveur du physicalisme il n'est pas en réalité sans implication métaphysique, et son refus de reconnaître à l'esprit une réalité en fait, qu'on le veuille ou non, un matérialisme.

Ce paradigme, largement hérité du positivisme logique et des oukases du cercle de Vienne à l'encontre de la métaphysique s'alimente de la critique du langage développée par la philosophie analytique et il n'est pas étonnant, dans ces conditions, que ce soit dans le cadre de celle-ci qu'il se soit d'abord développé – notamment, on l'a vu, avec G. Ryle (pour ne pas parler de la position extrêmement complexe de B. Russell sur la question). Il n'est donc pas étonnant que beaucoup de philosophes analystes se soient « reconvertis », si l'on peut dire, dans la philosophie de l'esprit au moment même où la philosophie analytique, engagée sur la désastreuse voie de la sémantique formelle des langues naturelles, marquait des signes certains d'essoufflement. C'est le cas notamment de Searle, de Davidson et de Putnam pour ne citer que les plus marquants.

Dans ces conditions, ce qu'il est convenu d'appeler le « mind-body problem » va être posé et résolu dans le cadre d'une philosophie marquée par un présupposé matérialiste et une méthode analytique. Comme le dit Shoemaker, « le mind-body problem concerne moins la manière dont un esprit non physique pourrait interagir avec un corps physique que la manière dont les esprits peuvent être une partie d'une réalité fondamentalement

physique ». C'est là la conséquence de l'option matérialiste. En même temps, héritage analytique oblige, la recherche d'une solution aux problèmes posés par les rapports entre le corps et l'esprit empruntera souvent la voie d'une analyse du discours. La question de savoir ce qu'on entend par « mental », quels types de phénomènes ou de processus peuvent être qualifiés de mentaux, et la question de savoir ce que recouvre le terme « esprit » vont se poser de manière particulièrement aiguë. Au bout du compte il apparaîtra à beaucoup que le « mind-body problem » est un pseudo problème, ou qu'il est mal posé, en tous cas qu'il faut le reformuler, sinon l'éliminer.

2. On justifie généralement le dualisme entre le mental et le physique en faisant appel à notre expérience ordinaire. Nous faisons l'expérience de sensations fournies par les cinq sens, de sentiments et d'émotions, d'états tels que la croyance, le désir, l'attente, le vouloir ; nous avons des représentations, des souvenirs, nous imaginons ce qui est absent ou ce qui n'est pas encore, nous manifestons des traits de caractère, des tendances ou des habitudes. Tous ces phénomènes sont classés comme mentaux. Ils se ramènent en fait à ce que Descartes appelait la pensée puisque, comme chacun le sait, pour lui « une chose qui pense... c'est une chose qui doute, qui conçoit, qui affirme, qui nie, qui veut, qui ne veut pas, qui imagine aussi et qui sent ». Tous ces phénomènes ont pour caractéristique commune d'être directement saisis comme nôtres, en bref d'être subjectifs et privés. De plus ils ne sont pas situés dans l'espace – nous avons l'impression d'avoir affaire avec eux à quelque chose d'intérieur et d'immatériel.

Tout autre est l'expérience que nous faisons du monde des choses, dont notre corps précisément. Nous avons affaire là à des choses tangibles, localisables dans l'espace, observables par tout un chacun et susceptibles de faire l'objet d'un consensus dans la description que nous faisons d'elles.

C'est ce sentiment qui fournit la principale justification du dualisme entre l'esprit et le corps. C'est ce dualisme qui est conceptualisé chez Descartes dans le cadre d'une ontologie substantialiste. La pensée et l'étendue y apparaissent comme deux substances irréductibles l'une à l'autre – tellement irréductibles que tout rapport entre les deux devient problématique, toute tentative d'explication d'une interaction entre elles impossible.

Or Descartes insiste également sur le fait que la réalité vécue par l'être humain, en tant qu'il est composé d'une âme et d'un corps, est celle d'une union entre ces deux composantes. Est-ce à dire que chez lui l'anthropologie vient mettre en cause l'ontologie ? En tout état de cause l'auteur des *Méditations Métaphysiques* a bien vu la difficulté. En soulignant qu'il n'était pas question de voir l'âme « comme un pilote en son navire », en insistant sur le caractère substantiel de l'union de l'âme et du corps, en s'efforçant de déterminer la localisation d'un lien entre les deux composantes et en la situant dans la glande pinéale, il ne vise à rien de moins qu'à tenter de résoudre le problème de l'interaction causale. Il est d'ailleurs significatif que les auteurs anglo-saxons contemporains ne s'intéressent pas à cet aspect de la philosophie cartésienne – preuve s'il en est qu'ils ont choisi d'écarter

l'hypothèse de l'existence d'une entité distincte du corps. Aussi bien préférèrent-ils s'en tenir à la mise en évidence des difficultés du dualisme substantialiste.

La postérité immédiate de Descartes se situant au contraire dans le cadre de l'hypothèse de l'existence de l'esprit aura eu à cœur de trouver des solutions aux difficultés soulevées par la thèse dualiste. On sait que l'occasionalisme de Malebranche aussi bien que l'harmonie préétablie de Leibniz visent à surmonter l'apparente impossibilité de l'interaction. Très logiquement Spinoza tente d'écarter la difficulté dans le cadre d'un monisme où la pensée et l'étendue sont réduites au statut de modes et il ouvre la voie à une théorie du double aspect.

II. La critique ryléenne du dualisme et le behaviorisme

1. C'est dans la voie du monisme que vont s'engager les recherches contemporaines. Déjà au début du siècle, W. James, puis B. Russell avaient soutenu la thèse d'un monisme neutre destiné à renvoyer dos à dos le matérialisme et l'idéalisme et affirmant que la distinction entre le physique et le mental ne renvoie pas à une différence de propriétés intrinsèques car « les choses et les pensées sont faites d'une même étoffe, l'étoffe de l'expérience ». Elle renvoie à une distinction qui n'est nullement ontologique comme le pense le dualisme classique, mais qui est d'ordre fonctionnel seulement. Elle relève d'une « simple question de groupement des entités ». Mais ce n'est pas cette voie philosophiquement stimulante qui sera empruntée. C'est G. Ryle plutôt que Russell qui sera invoqué et l'ouvrage qui fera date sera *La notion d'esprit*, publié en 1949.

La critique ryléenne va constituer un élément décisif dans la mise en cause du dualisme cartésien. C'est une critique très souvent proche du behaviorisme, bien que son auteur s'en soit vigoureusement défendu. En admettant seulement ce qui est publiquement observable, elle vise à court-circuiter toute référence au privé et dénonce l'intériorité comme un mythe. Le mental n'a pas de lieu spécifique tout simplement parce qu'il n'a pas d'autre réalité que celle des comportements sur lesquels nous prétendons le voir s'exprimer. Ce faisant, nous sommes victimes d'une « erreur de catégorie ». C'est là l'autre aspect de l'analyse ryléenne, elle relève de l'analyse conceptuelle. Le problème du mental est appréhendé comme un problème de langage. La question du mental est traitée d'abord comme question du langage du mental. La question de l'esprit est traitée par le biais d'une analyse des concepts mentaux tels que l'imagination, la volonté, l'intellect, etc... et on a bien l'impression qu'elle s'épuise dans cette analyse.

Pour Ryle, il s'agit d'éliminer ce qu'il appelle « la doctrine officielle » laquelle remonte à Descartes et affirme que « tout être humain est à la fois un esprit et un corps ». Cette doctrine est absurde à ses yeux parce qu'elle conduit à l'idée que « chaque individu vit deux vies parallèles, celle de son corps et celle de son esprit ». Elle conduit de plus à des difficultés

inextricables et Ryle pointe du doigt la question de la causalité : « comment l'esprit peut-il influencer le corps et être influencé par lui ? ».

A l'origine de cette doctrine Ryle voit « le dogme du fantôme dans la machine », lequel relève d'une « erreur de catégorie ». Celle-ci consiste à placer le corps et l'esprit dans la même catégorie et à penser que l'esprit est sujet lui aussi aux lois de la mécanique. L'hypothèse cartésienne est une hypothèse para-mécanique. Elle fait de l'esprit une machine mais comme cette machine est décrite de manière négative comme n'ayant pas les propriétés des corps, c'est une « machine fantomatique ». D'où la fiction d'un agent intérieur qui gouvernerait nos comportements. Il faut bien comprendre que l'habileté d'un conducteur n'est rien d'autre que l'ensemble des activités déployées dans sa conduite, qu'elle ne renvoie pas à une qualité occulte attribuée à un agent fantomatique. Comme le précise Ryle, « l'esprit n'est pas un théâtre second, mais l'ensemble des façons dont s'enracinent les épisodes de la vie des êtres humains ». Les problèmes des relations entre un individu et un esprit ou entre l'esprit et le corps ont été mal posés, « aussi mal que la question : quelles sont les transactions passées entre la chambre des communes et la constitution britannique ? »

Les termes psychologiques ne renvoient pas à des états ou à des processus cachés, ils rendent compte d'aspects du comportement, de manière plus précise, de dispositions au comportement. Ce sont des termes dispositionnels, comme on en trouve en physique. De même qu'attribuer la fragilité au verre ce n'est pas se référer à une qualité intrinsèque du verre mais simplement signaler que si on le cogne on le brisera, de même dire de quelqu'un « il a soif », ce n'est pas se référer à un état mental qu'il garderait jalousement caché mais simplement signaler qu'il prendra le verre qu'on lui tend si on lui offre à boire. Aux propositions catégoriques censées décrire des processus cachés, il faut substituer des propositions hypothétiques de la forme « si... alors » ; propositions fondées sur des observations de comportement et établissant, sur la base de régularités observées, des dispositions à agir.

Cette procédure, décrite dans *La Notion d'esprit*, constitue un des aspects les plus importants du behaviorisme logique. La reformulation des énoncés permet d'évacuer de faux problèmes, d'éliminer des croyances erronées, et ceci sans perte de sens. Elle permet de fournir une interprétation aux explications psychologiques courantes. Là où la psychologie populaire attribue aux comportements des causes mentales, le behaviorisme logique préfère parler de dispositions aux comportements. Le mental est réduit au comportemental, donc à quelque chose de physique. L'esprit apparaît alors comme une construction logique à partir des comportements. Le problème des rapports entre le corps et l'esprit n'a plus lieu d'être.

On trouve, dans cette position, l'ambiguïté qui caractérise nombre d'analyses relevant de la philosophie analytique. A-t-on affaire à une réduction ayant une portée ontologique (évacuation du mental au profit du physique) ou à une position méthodologique choisissant de laisser de côté la question de la nature des états mentaux, pour s'en tenir aux seules observations sur lesquelles tout le monde peut tomber d'accord ? Ryle, quant

à lui, affirme qu'il n'est pas question de réduire le monde matériel à des états et à des processus mentaux, pas plus qu'il n'est question de réduire les états et les processus mentaux à des états et à des processus physiques. Il semble donc s'en tenir à une version méthodologique et vouloir se maintenir dans une perspective philosophique. Dans le dernier chapitre de *La Notion d'esprit*, il souligne que son but n'est pas de faire avancer la méthodologie de la psychologie mais simplement de montrer comment les philosophes avaient décrit de manière erronée le comportement logique de plusieurs concepts mentaux fondamentaux. Il conclut en indiquant que ce mauvais usage des concepts n'est pas l'apanage des dualistes. Il est tout aussi fréquent chez les tenants du mécanisme hobbesien, entendons par là du mécanisme matérialiste, dont à ses yeux le behaviorisme fait partie.

Quoiqu'il en soit de ses réserves à l'égard du béhaviorisme Ryle a tout de même contribué à orienter la réflexion philosophique dans la voie d'une recherche du mental du côté de ses inscriptions physiques. Ses successeurs ne s'en sont pas tenus, c'est le moins que l'on puisse dire, à son agnosticisme métaphysique.

L'œuvre de Ryle marque un tournant dans la formulation du « mind-body problem » et inaugure un nouveau paradigme. Le behaviorisme en constitue une des premières versions.

Du point de vue chronologique il convient d'examiner le behaviorisme avant la théorie de l'identité (dite aussi matérialisme de l'état central), le fonctionnalisme, les diverses tentatives de matérialisme non réductif et le matérialisme éliminationniste.

2. Le behaviorisme s'efforce de rendre compte des états et des processus qu'on a l'habitude de qualifier de mentaux en les réduisant aux seuls comportements observables ; il opère en mettant en évidence des séquences stimuli-réponses. Il a pour conséquences de faire l'économie d'états intermédiaires et de ne pas faire appel à la notion de causation mentale.

On a coutume de distinguer entre le behaviorisme scientifique et le behaviorisme logique. Le premier nierait tout simplement les états mentaux et du coup viderait de son contenu le problème de la causation mentale. Le second, très largement inspiré des analyses ryléennes, reconnaîtrait grâce à la théorie des dispositions une certaine réalité au mental, saisi au niveau du comportemental. Alors que le premier serait éliminationniste, le second serait réductionniste. Or, comme nous l'avons vu, la position de Ryle était elle-même ambiguë. L'auteur de *La Notion d'esprit* donne les clés pour traduire le langage de la psychologie populaire dans le langage des dispositions, donc pour opérer une réduction, mais il n'en considère pas moins les intuitions courantes concernant le mental comme des leurres à éliminer. Aussi bien ne retiendrons-nous pas cette distinction comme pertinente pour la résolution de la question des rapports entre le corps et l'esprit. Le behaviorisme évacue celle-ci, quelle que soit la forme qu'il adopte.

Le behaviorisme est largement tributaire d'une conception vérificationniste de la science : seuls des énoncés susceptibles d'être vérifiés

expérimentalement sont dotés de sens, les autres sont exclus, rejetés comme absurdes. A ce titre, il est philosophiquement caduc, comme le positivisme logique qui le sous-tend.

De plus, il ne fournit pas les moyens de rendre compte de nos intuitions courantes concernant des états et des processus mentaux dont nous faisons l'expérience, ce qui le rend insuffisant. L'évacuation de la causation mentale est contre intuitive.

Enfin, il ne peut tenir ses promesses même lorsqu'il adopte une théorie des dispositions. Si l'on reprend l'exemple de quelqu'un qui a soif, le behavioriste soutiendra que dire « Pierre a soif », c'est dire « si on tend un verre à Pierre, celui-ci le prendra », mais cela n'est vrai qu'à condition que Pierre n'ait pas peur d'être empoisonné, qu'il croie que le liquide contenu dans le verre étanchera sa soif, qu'il n'entretienne pas avec la personne qui lui tend le verre des relations telles qu'il préférerait mourir de soif plutôt que d'accepter quelque chose d'elle, etc... L'analyse qui prétend évacuer l'état mental « avoir soif » n'a de sens que si on postule un nombre indéterminé d'autres états mentaux. Ce faisant le behaviorisme dessine en creux une caractéristique du mental que le fonctionnalisme mettra en évidence, à savoir son caractère relationnel.

III. La théorie de l'identité (le matérialisme de l'état central)

Cette théorie a été défendue par D. M. Armstrong et par J.J. L. Smart dont l'article intitulé « Sensations and Brain Processes » publié en 1959 dans la *Philosophical Review* constitue une des formulations les plus claires.

Pour les tenants de cette théorie, l'esprit c'est le cerveau (d'où le qualificatif de matérialisme de l'état central pour la distinguer des théories associant l'esprit à l'ensemble du système nerveux). Cette théorie se fonde sur une opinion largement répandue selon laquelle le cerveau est le siège de la vie mentale. Mais elle va beaucoup plus loin puisqu'elle affirme que les états mentaux sont des états physiques du cerveau, que le mental est identique au neuronal. Si nous considérons par exemple la douleur, nous dirons qu'elle est l'activation des fibres C. Smart précise bien que lorsqu'il parle d'identité, c'est au sens strict du terme : « quand je dis qu'une sensation est un processus du cerveau ou que la lumière est une décharge électrique, j'emploie « est » dans le sens de l'identité stricte... les processus rapportés dans la sensation sont en fait des processus dans le cerveau ». Il y a donc un seul et unique phénomène et deux manières d'en parler. Il y a une possibilité de traduire le langage phénoménologique de la douleur dans le langage scientifique des fibres, de réduire l'un à l'autre.

L'analogie avec les sciences est souvent invoquée à l'appui de cette thèse. La science ne nous a-t-elle pas appris que la température d'un corps c'est son énergie cinétique moléculaire moyenne, que la lumière est une onde électromagnétique ou encore que l'éclair est une soudaine décharge d'électricité à large échelle au voisinage des nuages ? Ce faisant ne nous a-t-elle pas fourni les moyens de donner un compte-rendu objectif et précis des

faits que nous décrivions auparavant « avec les moyens du bord » ? Pourquoi n'en serait-il pas de même lorsqu'il s'agit des faits mentaux ?

La théorie de l'identité fait donc le pari qu'il peut y avoir une traduction inter-théorique réussie entre le discours de la psychologie populaire et le discours scientifique. Toute traduction étant une réduction, elle fournit le cadre conceptuel idéal pour un matérialisme réductionniste.

On a objecté que les sensations et les processus neuronaux ne pouvaient pas être identiques parce que les énoncés rapportant des sensations et les énoncés rapportant des processus neuronaux ne signifient pas la même chose. A cela Smart a répondu en rappelant la distinction frégréenne entre le sens et la dénotation. De même que « Etoile du matin » et « Etoile du soir » dénotent le même corps céleste tout en ayant un sens différent, les expressions « j'ai telle sensation » et « mon cerveau est dans tel état » peuvent très bien dénoter un seul et même évènement tout en ayant des sens différents.

Une autre objection se fonde sur notre expérience introspective. Nous avons affaire à des sensations, des émotions, des représentations, pas à des impulsions électrochimiques dans un réseau neuronal. Les états et les processus auxquels nous avons accès grâce à l'introspection n'ont rien à voir avec ceux dont il est question dans les neurosciences. Aussi, dira-t-on, dans le meilleur des cas peut-on parler d'une corrélation entre deux domaines hétérogènes, mais certainement pas d'identité.

Peut-on effectivement parler d'identité au sens strict comme le veut Smart ? Un paysan illettré peut très bien parler de sa douleur ou de sa sensation de rouge sans ne rien connaître à la neurophysiologie. Smart prend en considération l'objection, mais pour l'évacuer. Il note qu'elle permet seulement de montrer que ces énoncés d'identité sont des énoncés d'identité contingents et pas nécessaires comme on pourrait s'y attendre dans le cadre d'une conception logique de l'identité stricte. Mais il en est de même de la relation entre « eau » et « H₂O » qui est une relation d'identité, mais qui a été découverte à un moment donné de l'histoire des sciences, qui a besoin d'être apprise, maîtrisée pour être utilisée. S'il en est ainsi, la relation entre les états mentaux et les états du cerveau aurait tout aussi bien pu être différente.

On peut imaginer des mondes possibles dans lesquels elle n'est pas la même, tout comme on peut imaginer des mondes possibles dans lesquels l'eau n'a pas la formule H₂O. Ceci n'est évidemment pas sans poser des problèmes. Kripke dans *La logique des noms propres* a montré les enjeux philosophiques majeurs de cette problématique. Ce qui est certain, c'est que, si la douleur = l'activation des fibres C chez l'homme, on peut très bien imaginer que la douleur = l'activation des fibres D chez le dauphin ; ce qui, en stricte logique de l'identité, nous conduit à affirmer que l'activation des fibres C = l'activation des fibres D. C'est cette difficulté notamment qui a conduit certains tenants de la théorie de l'identité à défendre une version restreinte aux cas particuliers. On peut en effet reconnaître à côté d'une identité entre types une identité entre cas particuliers ou instanciations (« tokens »). En se restreignant à une identité entre cas spécifiques, on évite la difficulté signalée.

Mais l'objection la plus redoutable est celle qui consiste à dire que la théorie de l'identité viole la loi de Leibniz qui veut que deux entités x et y ne peuvent être dites identiques que si elles partagent toutes leurs propriétés, que tout ce qui peut être attribué à x peut être attribué à y . Or les sensations et les états mentaux en général ont des propriétés qualitatives qui ne peuvent être attribuées à des états physiques du cerveau et ceux-ci ont des propriétés physiques qui ne peuvent être attribuées à des émotions ou à des croyances. On ne peut parler d'états sémantiques d'un neurone pas plus qu'on ne peut parler des propriétés spatiales d'une émotion, et la liste pourrait être longue des asymétries entre propriétés du mental et propriétés du neuronal.

IV. Le fonctionnalisme

C'est pour répondre à ces objections et pour essayer d'éviter les difficultés du behaviorisme et de la théorie de l'identité (dans la version du matérialisme réductionniste défendue par Armstrong et par Smart) que Putnam a proposé la théorie fonctionnaliste. Il s'agit en effet, pour lui, de préserver la réalité des états mentaux (à la différence du behaviorisme), sans les assimiler pour autant à des états physiques du cerveau (comme le fait la théorie de l'identité).

L'idée centrale du fonctionnalisme, c'est que l'esprit est un ordinateur ou, du moins, peut être envisagé par analogie avec un ordinateur. Dès 1960, dans un article intitulé « Minds and Machines », Putnam propose de traiter la question des rapports entre le corps et l'esprit dans les termes de la machine de Turing (en laquelle on peut voir l'ancêtre des ordinateurs).

Ce qui est remarquable, c'est que la description logique de la machine de Turing ne spécifie en rien sa nature physique. « La machine de Turing est une machine abstraite qui peut être réalisée selon un nombre infini de manières différentes »¹. Une fois réalisée de manière spécifique, la machine présente des états qu'on peut qualifier de structuraux, états qui intéressent l'ingénieur ou le physicien, mais pas le logicien. Une machine de Turing peut donc être décrite du point de vue de l'ingénieur et du point de vue du logicien – en termes informatiques on distinguera entre le *hardware* et le *software*, entre le programme et son implémentation physique. De même, affirme Putnam, il y a deux descriptions possibles de la psychologie humaine : la description behavioriste des comportements humains (qui correspond à la description que pourrait faire d'une machine de Turing un ingénieur) et la description abstraite des processus mentaux humains (qui correspond à la description que pourrait faire de la machine un logicien). Or cette description c'est, en fait, le programme de la psychologie populaire. Si ce programme a échoué ce n'est pas pour des raisons méthodologiques mais pour des raisons empiriques, parce que les états mentaux ne constituent pas un système causalement clos comme c'est le cas pour les différentes configurations de la machine de Turing.

¹ « Minds and Machines », p. 371.

Du fait de l'analogie entre l'esprit et la machine, on peut formuler le « mind-body problem » dans les termes de la machine de Turing : les états logiques de la machine peuvent-ils être identifiés à ses états structuraux ? La réponse étant évidemment négative et, pourrait-on dire sans objet, on peut dire que le « mind-body problem » est sans objet et reste purement verbal . Putnam présente sa théorie comme une version de la théorie de l'identité qui, au lieu d'identifier les états mentaux à des états neuronaux, les identifie à des états computationnels du cerveau. Mais d'autres versions du fonctionnalisme peuvent avoir sur le « mind-body problem » des positions très différentes.

L'esprit est donc envisagé comme un ordinateur digital traitant de l'information ; le programme est indépendant du matériel, le mental du physique ; le fait mental est d'abord défini par sa fonction. Ce qui fait qu'un état peut être qualifié de mental plutôt que de physique c'est ce qu'il accomplit dans le cadre d'un réseau de relations causales avec les stimuli, les réponses comportementales et les autres états mentaux. Pour reprendre l'exemple de la douleur, elle va être définie, non par des expériences phénoménales ou par un élément physique ayant lieu dans le cerveau, mais par son rôle fonctionnel (causal) abstrait. Elle se distingue de la démangeaison par sa cause (un dommage dans les tissus par exemple et non une irritation de la peau), par ses effets (elle cause des cris, des grimaces et non des grattements) et par ses rapports avec d'autres états mentaux.

On voit par là toute la différence entre le fonctionnalisme et le behaviorisme qui assimile la douleur à des comportements dans le cadre d'une séquence simple. Ici, au contraire, la fonction implique la prise en considération d'une économie complexe. Le fait mental est relationnel.

A la différence de la théorie de l'identité, le fonctionnalisme n'assimile pas chaque état mental à un état neurophysiologique du cerveau. Bien plus, du fait de l'indépendance du programme par rapport au matériel, il y a possibilité d'une pluralité de réalisations physiques. C'est là la thèse de la multiréalisabilité. Pas question dans ces conditions d'identifier la douleur avec un quelconque état neuronal. On peut imaginer une créature d'une autre planète, dont le corps serait à base de silicone et non de carbone et qui éprouverait de la douleur, tout comme un Terrien. Dans ce cas, la douleur ne serait pas liée aux processus physiologiques auxquels elle est liée chez le Terrien – mais elle serait tout de même de la douleur par son rôle fonctionnel.

Le fonctionnalisme peut dans ces conditions être tout aussi compatible avec le dualisme qu'avec le monisme (dans ce cas il affirmera l'identité au niveau des cas particuliers mais pas au niveau des types). Comme il peut très bien convenir également à ceux qui profèrent un agnosticisme en la matière ; on s'explique aisément qu'il ait longtemps constitué la théorie la plus répandue en philosophie de l'esprit. Il présente également l'avantage de rendre compte des intuitions de la psychologie populaire en même temps qu'il en sauvegarde l'autonomie par rapport à la biologie et aux neurosciences. Par ailleurs les positions fonctionnalistes ont été largement confortées par le développement de l'intelligence artificielle et des sciences cognitives.

Malgré tous ces avantages, le fonctionnalisme se heurte à de sérieuses objections.

En faisant des propriétés relationnelles, saisissables à un haut niveau d'abstraction, la marque du mental, le fonctionnalisme fait l'impasse sur l'aspect qualitatif des états mentaux. Ici encore, les objections fondées sur les *qualia* vont se révéler redoutables. De nombreuses expériences de pensées ont été invoquées pour ébranler la position fonctionnaliste : celle du spectre inversé, celle de Mary la neurophysiologiste (encore appelée argument de la connaissance), celle des *qualia* absents (encore appelée argument du Cerveau Chine).

L'argument du spectre inversé part d'une expérience de pensée qui consiste à imaginer deux jumeaux, James et John, qui voient des couleurs complémentaires. Lorsque James voit une tomate rouge, John la voit verte et vice versa. Cela ne se manifeste pas dans leur comportement. Leurs états mentaux – ici des sensations de couleurs – ont exactement les mêmes relations causales à leurs comportements. Ils sont donc fonctionnellement équivalents. Peut-on pour autant faire abstraction du fait que leurs *qualia* sont différents ?

L'argument de Mary la scientifique part d'un autre type d'expérience de pensée. On imagine Mary, brillante spécialiste en neurophysiologie qui vit, depuis sa naissance, dans un environnement blanc et noir. Elle sait tout des couleurs et de la manière de les discriminer. Elle agit comme tout un chacun devant les discriminations de couleurs, mais elle n'a pas la sensation du rouge ou de toute couleur autre que le blanc et le noir. Un jour elle est brutalement introduite dans le monde coloré qui est le nôtre. Du point de vue fonctionnel rien n'est changé puisqu'elle avait la capacité de discrimination entre les couleurs et se comportait en conséquence. Peut-on néanmoins raisonnablement affirmer qu'elle ne connaît pas quelque chose de nouveau lorsqu'elle voit du rouge ? L'expérience phénoménologique n'enrichit-elle pas la connaissance ? Dans le même ordre d'idées, Thomas Nagel s'est demandé dans un article célèbre (repris par Hofstadter et Dennett dans *Vues de l'esprit*) « quel effet cela fait d'être une chauve-souris ? »...

Toujours dans le même souci de mettre l'accent sur l'importance des *qualia*, Ned Block a proposé l'expérience de pensée dite du Cerveau Chine. On imagine que le gouvernement chinois, adepte du fonctionnalisme, a décidé d'utiliser toute sa population pour simuler, pendant une heure, le fonctionnement du cerveau humain. Chaque Chinois doit accomplir pendant une heure une fonction déterminée, comme un neurone dans le cerveau, et il doit être en relation avec d'autres Chinois, dans le cadre d'un gigantesque réseau. Il est muni d'un émetteur-récepteur, d'un téléviseur et d'instructions. Le Cerveau Chine fonctionne comme le cerveau humain. Mais, pendant cette heure de simulation, un Chinois ressent une douleur à l'orteil. Le Cerveau Chine la ressent-il ? Non, car un tel système ne possède pas les propriétés associées à la douleur. Les *qualia* sont absents.

Putnam s'est révélé un des critiques les plus lucides du fonctionnalisme après en avoir été l'inspirateur, mais c'est à Searle que l'on doit une des critiques les plus pertinentes de cette théorie. Pour mettre en

cause l'idée que le cerveau n'est qu'un ordinateur digital et l'esprit un programme d'ordinateur, Searle a proposé l'expérience de pensée dite de la Chambre Chinoise. On introduit dans une pièce un individu qui ne comprend pas le chinois. Il dispose d'un stock d'étiquettes sur lesquelles sont inscrits des signes chinois et d'un livre d'instructions. On lui fait passer en entrée des étiquettes portant des signes chinois et il doit choisir des étiquettes dans le stock en fonction de celles qu'il a reçues et des instructions dans le livre pour les distribuer en sorties. Le dispositif est bien celui d'un ordinateur avec entrées, sorties et programme. La personne qui est dans la chambre va fournir des réponses appropriées aux questions qui lui ont été posées sans comprendre elle-même le chinois. La thèse de Searle c'est qu'il en va de même pour l'ordinateur. Il manipule en fonction d'un programme de données sans manifester la moindre compréhension. Les ordinateurs calculent, mais l'esprit a quelque chose de plus, un contenu.

A partir de cette expérience de pensée, Searle formalise son argument. Le comportement de celui qui est dans la Chambre Chinoise, tout comme le comportement des ordinateurs, est computationnel – purement syntaxique. Les esprits ont une sémantique : un des traits essentiels des états mentaux c'est qu'ils sont à propos de, qu'ils sont intentionnels. La syntaxe ne rend pas compte de la sémantique. Les programmes ne sont pas des esprits.

L'argument de Searle n'établit pas qu'aucun ordinateur ne peut penser, il réfute la thèse selon laquelle seuls les ordinateurs peuvent penser.

V. Le matérialisme éliminationniste

Le matérialisme réductionniste – dont la théorie de l'identité constitue l'expression la plus cohérente – espère donner de ce dont traite la psychologie populaire une version scientifique permettant une traduction inter-théorique réussie, du type de celle qu'on trouve dans les sciences de la nature lorsqu'on affirme que eau = H₂O. Cet espoir n'est pas partagé par tous les matérialistes. Beaucoup d'entre eux émettent de sérieuses raisons quant au statut théorique de la psychologie populaire, d'autres mettent en cause ses analyses. C'est notamment le cas de Paul et Patricia Churchland qui se sont faits les défenseurs d'un matérialisme éliminationniste.

L'idée de base de cette version du matérialisme, c'est que le cadre conceptuel de la psychologie populaire est faux et que l'ensemble des platitudes du sens commun concernant notre activité mentale (en quoi elle consiste) se révèlent égarantes. Aux yeux des Churchland, il n'y a là que préjugés et intuitions plus ou moins vagues mais, contrairement à Stich par exemple, ils considèrent la psychologie populaire comme une théorie. C'est une théorie comme l'étaient également la cosmologie populaire, la physique populaire ou la biologie populaire qui ont toutes fini par céder la place à des théories scientifiques. La psychologie populaire est la seule conception populaire qui résiste, les autres ont été éliminées.

D'une portée restreinte – elle n'a pu rendre compte ni du sommeil, ni de la mémoire, ni des différences d'intelligence ni de la maladie mentale –,

n'ayant pas fait de progrès depuis deux mille ans, fondée sur des fabulations, la psychologie populaire doit à son tour être éliminée. Le problème n'est donc pas de la traduire, c'est-à-dire de la réduire à une théorie scientifique, c'est de l'éliminer au profit d'une théorie scientifique. Aux yeux des Churchland, ce sont les neurosciences qui peuvent fournir le cadre scientifique approprié. C'est notamment la thèse développée par Paul Churchland dans *Matière et Conscience* (1984) et par Patricia Churchland dans *Neurophylosophy : Toward a Unified Science of the Mind-Brain* (1986).

La conséquence de cette décision théorique c'est évidemment l'élimination pure et simple de la notion de mental. Pour les Churchland, l'avenir est dans les neurosciences. Paul Churchland va jusqu'à suggérer dans *Matière et Conscience* que le développement de celles-ci sera de nature à affiner nos capacités introspectives en développant nos facultés discriminatoires dans un cadre conceptuel maîtrisé. De même, soutient-il, que le musicien professionnel a acquis grâce à un savoir et de la pratique des capacités discriminatoires sans commune mesure avec celles qu'il avait enfant, que l'œnologue peut, grâce à ses connaissances, mieux saisir les différences entre les vins, nous pourrions, grâce aux neurosciences, mieux maîtriser ce que nous sentons. « La consommation de glucose dans le cerveau antérieur, les niveaux de dopamine dans le thalamus, les vecteurs de codage dans les voies neuronales spécifiques... pourraient être transposés dans le foyer objectif de notre pouvoir de discrimination introspective et de reconnaissance conceptuelle de la même manière qu'un accord en sol mineur 7 et un accord en la 9 sont transposés en une discrimination et une reconnaissance dans l'audition d'un musicien entraîné »².

Le trait dominant de la position des Churchland, c'est qu'elle évacue le discours sur le mental au profit du discours sur le cerveau. Il est significatif que leurs travaux soient du coup des travaux scientifiques, la seule manière de faire avancer la résolution des problèmes philosophiques étant, à leurs yeux, de fournir des connaissances scientifiques indubitables.

Contre le fonctionnalisme de la machine de Turing par exemple, la seule objection pertinente, de leur point de vue, ne peut venir que d'une meilleure connaissance du fonctionnement du cerveau. Lorsqu'on a compris que le cerveau a une architecture physique et computationnelle différente de l'ordinateur dans le traitement de l'information et lui permettant d'effectuer simultanément des millions d'opérations simples, qu'il a un traitement parallèle au lieu du traitement en série des machines informatiques, on pose le problème du rapport entre esprit et machine différemment.

Pour les Churchland, le problème traditionnel de l'esprit et du corps est désormais une question exclusivement scientifique, c'est « la question de savoir comment le cadre théorique de la psychologie populaire finira par être relié à toute théorie neuropsychologique destinée à la remplacer ». D'ailleurs, les différentes positions que l'on peut avoir à l'égard du « mind-body problem » se ramènent pour eux aux différentes conceptions que l'on

² *Matière et Conscience*, p. 222.

peut avoir du destin de la psychologie populaire : « Si la psychologie populaire est réduite en douceur à une théorie matérialiste de substitution, alors la théorie de l'identité sera défendue. Si elle apparaît disjonctivement réductible alors le fonctionnalisme sera défendu. Si on prouve qu'elle n'est pas réductible parce qu'on ne peut pas trouver de théorie matérialiste pour lui succéder, alors c'est une forme de dualisme qu'il faudra défendre. Et si on prouve qu'elle n'est pas réductible pour la simple raison qu'on ne peut pas la projeter sur la théorie matérialiste destinée à lui succéder alors la position appelée matérialisme éliminationniste sera défendue ». C'est évidemment cette dernière hypothèse que les Churchland retiennent.

Contre le matérialisme éliminationniste, on a objecté qu'il est faux parce qu'on voit mal comment considérer comme erronées, voire égarantes, les données que nous révèle l'introspection ; on n'a pas manqué naturellement de lui reprocher d'ignorer les *qualia*. Mais il semble que ce soit surtout la volonté de ramener toute la vie mentale, faite de relations aux multiples facettes de l'environnement – dont autrui n'est pas l'élément le moins négligeable – à des microprocessus se déroulant dans le cerveau qui constitue la difficulté majeure de cette position qui représente, rappelons-le, la forme la plus radicale du matérialisme. Peut-on envisager alors une forme de matérialisme qui reconnaisse tout de même au mental sa spécificité ? Les diverses tentatives qui ont été faites en ce sens, sont souvent groupées sous la dénomination de matérialisme non réductionniste.

VI. Le matérialisme non réductionniste.

Une des versions les plus célèbres de ce type de matérialisme est le monisme anomal introduit par Davidson en 1970 dans son article intitulé « Mental Events ».

Le propos de Davidson est de montrer qu'il est possible de concilier trois thèses en apparence incompatibles :

La thèse de la dépendance causale qui soutient que les événements mentaux sont la cause de certains événements physiques ;

La thèse du caractère nomologique de la causalité qui soutient que là où il y a causalité, il y a des régularités exprimables sous forme de lois ;

La thèse de l'anomie du mental qui soutient qu'il n'existe pas de régularités formulables sous forme de lois liant le mental au physique.

Ces thèses ne se contredisent pas dès lors qu'on professe un monisme matérialiste permettant d'identifier les événements mentaux à des événements physiques et que l'on admet qu'en classant des événements comme mentaux, on les fait entrer dans une trame explicative ne permettant pas l'énoncé de lois. A cette double condition satisfait précisément le monisme anomal.

Davidson présente souvent sa théorie comme une version de la théorie de l'identité. Nous avons effectivement affaire à une identification du mental au physique. Il n'y a que des événements spatiotemporels, susceptibles d'être décrits comme événements physiques. Mais certains d'entre eux sont

susceptibles aussi d'être décrits comme événements mentaux. Le fait de donner d'un événement une description physique implique qu'on l'insère dans une trame d'événements dans laquelle les lois sont strictes ; le fait qu'on le décrive comme mental requiert une interprétation parmi d'autres, ce qui rend impossible la formulation de lois. La thèse centrale de Davidson c'est que les événements mentaux tels que percevoir, se souvenir, prendre des décisions, accomplir des actions, ne se laissent pas intégrer dans le réseau nomologique de la théorie physique.

On peut objecter à cette théorie que, si le mental, c'est ce qui est décrit comme mental, il ne joue alors aucun rôle causal ; donc la version davidsonienne de la théorie de l'identité ne parvient pas à surmonter une des difficultés majeures de celle-ci.

Dans ces conditions la question se pose de savoir si, d'une manière générale, le matérialisme non réductionniste n'est pas un mythe. Kim a soutenu que c'est une position condamnée à verser soit dans l'élimination du mental soit dans un dualisme explicite. Cela est vrai pour le dualisme des propriétés aussi bien que pour le physicalisme survenant – que nous ne pouvons malheureusement que citer dans le cadre restreint de cet article. Tant il apparaît que défendre une position matérialiste tout en essayant de sauvegarder la spécificité du mental est une gageure.

Conclusion

Si l'effondrement des ontologies substantialistes a vidé de son contenu la problématique de l'interaction entre substances hétérogènes et exclusives l'une de l'autre, la question n'en demeure pas moins lancinante de savoir quels types de rapports peuvent entretenir le mental et le physique. Le parti pris du naturalisme conduit à s'interroger sur la possibilité même de parler d'esprit dans le cadre d'une explication physicaliste de tous les phénomènes possibles. Les progrès enregistrés dans la connaissance des mécanismes du fonctionnement cérébral tendent à poser le « mind-body problem » dans les termes d'un « mind-brain problem », sans que l'on puisse d'ailleurs voir là le fruit d'une évolution nécessaire car la question du rôle central du cerveau dans le système nerveux peut toujours être posée.

Ce qui émerge des différentes positions que nous avons passées en revue c'est, dans un premier temps, l'opposition entre une approche computationnelle du mental et une approche qu'on pourrait qualifier de « biologique » – la sous-tendant la question de savoir si le médium est ou non indifférent. A Putnam qui, dans un de ses tout premiers textes, écrivait qu'après tout l'esprit pourrait être tout aussi bien fait de gruyère, Dennett répond dans *La diversité des esprits* qu'on ne pourrait jamais avoir affaire à un esprit doué d'appréhension avec des puces de silicone, ou bien des fils ou du verre, voire des canettes de bière attachées par des ficelles. « L'esprit, machine de Turing ou machine organique ? » fut une des questions les plus débattues à l'époque où le fonctionnalisme était la position dominante.

Le fonctionnement de l'ordinateur nous a sans doute beaucoup appris sur le fonctionnement du mental – ne fût-ce qu'en en dessinant en creux les traits irréductibles au calcul. Aujourd'hui, c'est de l'étude du cerveau qu'on semble attendre la résolution de l'énigme. Le fonctionnement du cerveau peut effectivement nous apprendre beaucoup sur les conditions physiques de l'activité mentale, il peut nous aider à mieux cerner les problèmes et à écarter définitivement de pseudo solutions, mais aussi loin que l'on puisse aller dans ce domaine, aussi loin qu'on en repousse les limites et qu'on en diffère la réponse, la question se pose de savoir comment rendre compte de la spécificité du mental – même si la pensée n'était qu'une sécrétion du cerveau, encore faudrait-il expliquer ce qui la fait être pensée et rendre compte des quatre caractéristiques principales que Searle reconnaît au mental : la conscience, l'intentionnalité, la subjectivité et la causalité.

BIBLIOGRAPHIE (ouvrages accessibles en français) :

- Churchland Paul (1988), *Matière et Conscience*, Champ Vallon, 1999.
Davidson Donald (1980), *Actions et événements*, PUF, 1993.
Fisette Denis et Poirier Pierre (2000), *Philosophie de l'esprit – Etat des lieux*, Vrin, 2000.
Hofstadter D. et Dennett D. (1984), *Vues de l'esprit*, Interéditeurs, 1987.
Putnam H. (1988), *Représentation et Réalité*, Gallimard, 1990.
Ryle G. (1949), *La notion d'esprit*, Payot, 1976.
Searle J. (1984), *Du cerveau au savoir*, Hermann, 1985.