

Swiss Philosophical Preprint Series

10

Olivier Massin

**Le Corps et L'Ésprit
Part 2**

added 10/11/2008

ISSN 1662-937X

© Olivier Massin

Le corps et l'esprit

Deuxième partie

Nous avons montré en première partie que la question du corps et de l'esprit posait trois problèmes distincts : le problème de la corrélation (comment expliquer la corrélation du corps et de l'esprit sans renoncer à leur différence de nature ?), le problème du corps propre (notre corps est-il un simple objet pour notre esprit ?), et le problème de la prévention de l'esprit à l'égard des corps (pourquoi considérons-nous plus volontiers que le monde se compose de corps que d'évènements ou de processus ?). Il ne faut donc pas réduire la question du corps et de l'esprit au seul problème de la corrélation. Cela dit, il faut maintenant souligner que le problème de la corrélation est de loin de plus important des trois, qu'il est en substance celui qui, des présocratiques à nos jours, a fait l'objet de l'attention philosophique la plus grande. La deuxième partie de ce cours lui sera donc consacrée.

2. Deuxième partie : le problème de la corrélation

Le corps et l'esprit sont donc nettement distincts. Mais ils sont également étroitement corrélés. D'où naît un problème : comment expliquer une corrélation aussi étroite si la distinction est si nette ? Avant d'aborder les différentes solutions qui peuvent être apportées à ce problème majeur, il convient de s'assurer qu'une telle corrélation a bien lieu.

2.1. La corrélation du corps et de l'esprit

Pour peu qu'ils admettent qu'une telle proposition ait un sens, tous les philosophes semblent admettre qu'il y a une corrélation entre le corps et l'esprit. Le véritable débat sur ce point, oppose ceux qui pensent que cette corrélation est totale, en ce qu'à tout phénomène mental, on peut associer un phénomène physique, à ceux qui soutiennent que cette corrélation n'est que partielle, dans la mesure où certains phénomènes mentaux n'ont aucun corrélat corporel. Il est clair que les partisans d'une théorie moniste d'identification de l'esprit au corps (que ce monisme soit matérialiste, neutre ou spiritualiste) doivent nécessairement

soutenir que la corrélation du corps et de l'esprit est totale : si l'esprit est le corps, il ne peut y avoir de phénomènes mentaux sans phénomènes physiques. Les partisans d'une théorie dualiste en revanche peuvent aussi bien admettre une corrélation partielle, qu'une corrélation totale. Une des principales lignes d'argumentation des philosophes dualistes consiste à chercher à mettre en évidence des phénomènes mentaux (la conscience, les *qualia*) qui n'ont pas de corollaires physiques.

2.1.1. La corrélation chez Descartes

Il convient d'insister sur le fait que l'admission d'une corrélation forte entre le corps et l'esprit n'est pas le privilège des monistes. Ainsi Descartes souligne-t-il l'importance d'une telle corrélation. Celle-ci, selon lui, n'est que partielle. Il y a pour Descartes deux types de phénomènes mentaux (de « fonctions de l'âme ») : les actions et les passions. Les actions s'identifient à nos volontés, qui viennent directement de notre âme, les passions s'identifient à nos perceptions, que notre âme trouve en elle mais qu'elle ne fait pas¹. Les volontés et les actions sont respectivement de deux sortes : il y a les volontés qui sont des actions de l'âme qui se terminent dans l'âme et les volontés qui sont des actions de l'âme et qui se terminent dans le corps ; de même, il y a les perceptions qui ont l'âme pour cause et celles qui ont le corps pour cause².

Il y a dès lors deux types de phénomènes mentaux qui n'ont pas de corrélats physiques : les volontés qui sont des actions de l'âme qui se terminent dans l'âme (« comme lorsque nous voulons aimer Dieu ou généralement appliquer notre pensée à quelque objet qui n'est point matériel ») et les perceptions qui ont l'âme pour cause (les perceptions de nos volontés). Ces phénomènes se déroulent de manière purement interne à l'âme et sont indifférents à ce qui se passe dans le corps. Les deux autres types de phénomènes mentaux en revanche ont *nécessairement* des corrélats corporels. Il s'agit des volontés qui se terminent dans le corps (« comme lorsque de cela seul que nous avons la volonté de nous promener, il suit que nos jambes se remuent et que nous marchons ») et des perceptions qui ont le corps pour cause. Pour mieux s'y retrouver, on peut présenter le tableau suivant, représentant des fonctions de l'âme dans *Les passions de l'âme* (Les « ↓ » symbolisent les fonctions ayant un corrélat corporel, les « x » les fonctions n'en n'ayant pas).

¹ *Passions de l'âme*, art. 17.

² Art. 18, 19.

Fonctions de l'âme						
Actions (volontés)		Passions (perceptions)				
Qui se terminent dans l'âme	Qui se terminent dans le corps	Qui ont l'âme pour cause (Perceptions de nos volontés, des imaginations formées par l'âme, de nos pensées)	Qui ont le corps pour cause			
			Qui ont les esprits animaux pour cause (imaginations formées par le corps)	Qui ont les nerfs pour cause		
			Rapportées aux objets du dehors (entendre une cloche)	Rapportées à notre corps (avoir soif)	Rapportées à notre âme (la joie, la colère)	
x	↓	x	↓	↓	↓	↓

Les phénomènes mentaux qui sont corrélés à des phénomènes physiques sont donc légion. Un point essentiel est que les corrélats des phénomènes mentaux ne sont pas nécessairement les phénomènes physiques *voulus* ou *perçus* par ces phénomènes mentaux. Pour le dire autrement, il y a toujours un corrélat corporel à ce type de phénomène, mais *ce corrélat n'est pas toujours l'objet intentionnel de ce phénomène*. Si tel était le cas, il faudrait admettre que seules les perceptions véridiques, qui perçoivent l'objet intentionnel tel qu'il est, auraient un corrélat physique. Les hallucinations et les rêves en seraient dépourvus. Telle n'est certainement pas la position de Descartes³. S'il affirme que les illusions de nos songes n'ont pas leur cause dans nos nerfs, c'est pour préciser aussitôt qu'elles sont causées par les esprits animaux, qui se trouvent accidentellement dans tel état d'agitation⁴. Descartes à clairement précisé auparavant que ces esprit animaux sont des corps⁵. Il écrit ainsi :

Il reste ici à remarquer, que toutes les mêmes choses que l'âme aperçoit par l'entremise des nerfs, lui peuvent aussi être représentées par le cours fortuit des esprits ; sans qu'il y ait autre différence, sinon que les impressions qui viennent dans le cerveau par les nerfs ont coutume d'être plus vives et plus expresses que celles que les esprits y excitent [...]⁶.

Il y a donc un corrélat corporel à tous nos rêves, à toutes nos illusions perceptives. Ce corrélat en l'occurrence n'est pas ce à quoi nous rêvons, ni ce que nous percevons, il n'est pas l'objet intentionnel de nos rêves et de nos perceptions, mais il est la cause physique de ces

³ Ainsi ne faut-il pas se laisser abuser lorsque Descartes écrit que les actions de l'âme qui se terminent dans l'âme incluent les cas où nous pensons à des objets immatériels. Cela concerne les pensées de Dieu, des vérités mathématiques, de la nature de l'âme elle-même, ou l'imagination *active*, comme lorsque nous nous appliquons à imaginer un palais enchanté (art. 20) mais en aucun cas les rêves ou les hallucinations perceptives, dans lesquels nous demeurons *passifs*.

⁴ Art. 21.

⁵ Art. 10.

⁶ Art. 26.

actes, l'agitation cérébrale des esprits animaux ou l'excitation des nerfs. Cependant, les esprits animaux censés être les corrélats des rêves relèvent plus de la conjecture que de l'observation. En outre, la corrélation que Descartes introduit entre le corps et l'esprit est peu quantifiée. Nous voudrions savoir non seulement si à tout phénomène mental (ou à certains types de phénomènes mentaux seulement) correspond un phénomène physique, mais également s'il y a possibilité d'établir un lien quantitatif entre le phénomène physique et le phénomène mental, entre le corps et l'esprit. Il faudra attendre deux siècles avant que Weber et Fechner ne parviennent à la formulation d'une telle loi psychophysique.

2.1.2. La corrélation chez E.H. Weber et G.T. Fechner

2.1.2.1. Naissance de la psychophysique

C'est à travers l'étude du toucher que la première véritable loi de corrélation psychophysique a été découverte. Elle est le fait de deux hommes : Ernst Heinrich Weber tout d'abord, puis Gustav Theodor Fechner qui formula la loi sur la base des travaux de Weber.

Dans les années 1820, Weber était professeur d'anatomie puis de physiologie à Leipzig. Dans son ouvrage *De tactu : annotationes anatomicæ et physiologicæ* (1834), Weber présente son expérience devenue célèbre du « test du compas », sur le seuil de discrimination entre deux points stimulés à la surface de la peau⁷. Le résultat de cette expérience est que ce seuil varie énormément selon les endroits stimulés sur la peau : le bout des doigts peut distinguer les deux pointes rapprochées du compas, alors que pour ce même écartement, le haut du dos ne sent qu'une seule pointe. On trouve déjà là une quantification très nette des seuils de discrimination sensorielle, qui seront au cœur de la psychophysique contemporaine. Mais l'expérience qui fut au fondement de cette psychophysique, bien qu'elle relève également le toucher, concerne la discrimination entre les *poids*. C'est à l'occasion de l'étude de nos capacités à distinguer deux poids que Weber formule ce que Fechner appellera « la loi de Weber ». L'idée fondamentale de Weber est que la différence perceptible entre deux poids n'est pas une différence absolue, mais une différence relative, un *ratio* entre ces deux poids.

⁷ Ces expériences seront précisées dans une œuvre ultérieure, considérée comme sa plus importante, *Der Tastinn und das Gemeingefühl* (1846). Nous nous référons ici à la traduction anglaise de ces deux œuvres, parue dans *E. H. Weber on the Tactile Sense*, éd. et trad. Helen E. Ros et David J. Murray, Erlbaum (GB), Taylor & Francis, 1996.

Nous ne trouvons pas de discrimination du poids *par le toucher* à moins que la différence entre les poids soit au moins d'une quantité d'un quinzième ou d'un trentième.

Nous ne trouvons pas de discrimination entre des lignes *par la vue* à moins que la différence entre elles soit d'une quantité d'un centième.

Enfin, nous ne trouvons pas de discrimination entre les sons *par l'ouïe* à moins qu'il y ait une différence tonale d'une quantité d'un sur trois cent vingt.

Lorsque nous notons une différence entre des choses qui ont été comparées, nous ne percevons pas la différence entre ces choses, mais le ratio de leur différence par leur magnitude.

Si deux poids de 30 et 29 onces [une once = 28,35 grammes] sont comparés par le toucher, la différence n'est pas perçue plus facilement que pour des poids de 30 et 29 drachmes [1 drachme = 3,24 grammes]⁸.

Dans ses *Eléments de Psychophysique* (1860), ouvrage fondateur de la psychophysique, Gustav Theodor Fechner, repart de ce résultat de Weber selon lequel nous percevons non pas les différences absolues entre les choses (qu'il appelle stimuli — terme quasiment absent chez Weber) mais seulement les différences relatives entre ces choses. Il en tire des conséquences tout à fait originales, qu'il continue pourtant à attribuer, trop modestement, à Weber. Le projet de Fechner est de fonder une psychophysique scientifique⁹. Celle-ci se donne pour tâche de dégager les relations entre la quantité de stimulus (ou de processus corporels pour la psychophysique interne) et la quantité de sensation. Un tel projet se heurte à un obstacle évident : si l'on peut concevoir des moyens de déterminer la quantité (ou la magnitude) d'un stimulus, l'idée d'une quantité de sensation semble très étrange : nous pouvons mesurer les stimuli, mais il ne semble pas possible de mesurer les sensations. Quel serait en effet l'étalon d'une telle mesure ? Fechner reconnaît que nous ne pouvons pas parvenir à déterminer *directement* la magnitude absolue des sensations, à leur attribuer une valeur cardinale. A partir de là, sa réponse se fait en deux temps : (i) dans le premier, Fechner montre, derrière Weber, que bien que nous n'ayons pas de mesure directe de la sensation, nous pouvons établir une relation mathématique entre la quantité d'énergie du stimulus d'une part et l'apparition et la variation des sensations de l'autre. (ii) Dans le second, Weber montre

⁸ *De tactu*, à partir de l'édition anglaise, p. 125-126.

⁹ Ce projet s'intègre dans un projet plus vaste, qui tire argument du parallélisme psycho-physique dégagé par la psychophysique, en faveur d'une thèse d'identité du corps et de l'esprit (l'un et l'autre n'étant que deux points de vue, internes et externes sur une même chose) et en faveur d'une ontologie panpsychiste. Ce vaste projet est le plus important aux yeux de Fechner, qui se concevait comme un philosophe, mais il n'est pas celui qu'a retenu l'histoire, qui s'est presque exclusivement intéressée au Fechner psychophysicien. Nous la suivons ici en ne nous appuyant sur Fechner que pour présenter le problème de la corrélation et non les solutions qui lui ont été proposées. Pour une présentation de la solution de Fechner au problème de la corrélation et de sa *philosophie*, voir l'ouvrage d'Isabelle Dupéron, *G. T. Fechner, le parallélisme psychophysique*.

qu'en intégrant cette formule mathématique et faisant l'hypothèse que l'unité de mesure de la sensation est la plus petite variation notable, nous pouvons parvenir, via la mesure de la quantité de stimulus, à une mesure *indirecte* de la sensation.

(i) Fechner, suivant Weber, souligne d'abord qu'en dépit de l'impossibilité de mesurer directement les sensations, deux choses restent possibles : premièrement, nous pouvons dire si une sensation est présente ou absente ; deuxièmement, nous pouvons déterminer la valeur relative des sensations, leur attribuer une valeur ordinale, en disant qu'une sensation est plus ou moins intense qu'une autre. Je ne peux pas dire d'une sensation qu'elle est d'intensité 12, mais je peux dire de deux sensations que l'une est plus forte que l'autre. Puisque nous pouvons par ailleurs mesurer sans trop de problème les phénomènes physiques, il est possible de mesurer la quantité de stimulus (physique) nécessaire pour donner naissance à une sensation particulière ou pour qu'apparaisse une différence entre deux sensations. C'est là précisément la loi de Weber, que Fechner formule en termes mathématiques. R étant le stimulus, dR étant une (petite) variation de ce stimulus, D'après Weber, on sait que pour toute plus petite variation notable de la sensation :

$$dR/R = \text{une constante}$$

Comme le remarque Edwin G. Boring, c'est à strictement parler cette formule qu'il convient d'appeler loi de Weber (et non, comme le fait Fechner, le résultat auquel lui-même parviendra)¹⁰. Elle signifie que la plus petite différence entre deux poids que nous sommes en mesure de percevoir correspond à une différence relative constante de ces deux poids. A l'issue de ce premier temps, Fechner n'a donc fait « que » donner une formulation mathématique à la loi de Weber, et n'a pas encore proposé de moyen de quantifier les sensations. C'est pourquoi les critiques que William James et Henri Bergson adresseront à l'idée que les sensations peuvent être mesurées épargneront cette première étape du raisonnement.

(ii) La deuxième étape du raisonnement de Fechner introduit alors son apport central par rapport à Weber, outre la formulation mathématique : Fechner propose d'utiliser la plus petite différence notable entre deux sensations comme unité de mesure des sensations. Certes,

¹⁰ Boring, *A History of Experimental Psychology*, , p. 278.

c'est déjà implicitement sur les plus petites différences sensibles que s'était appuyé Weber pour formuler son résultat. Mais l'hypothèse forte que fait Fechner, que certains rejettent, est que ces plus petites différences sensibles sont *égales* entre elles, et peuvent donc servir d'unité de mesure pour l'intensité des sensations. Fechner peut alors introduire une nouvelle formule, qu'il appelle la *formule fondamentale* (S symbolise la sensation, dS la plus petite différence sensible de la sensation et c est une constante).

$$dS = c \, dR/R$$

Il s'agit là d'une première formulation de ce qu'il appelle la « loi de Weber », selon laquelle « des incréments relatifs égaux de stimuli sont proportionnels à des incréments égaux de sensations ». c est donc ici une constante de *proportionnalité*. Bien sûr, Weber n'a jamais formulé une telle loi, qui repose sur l'hypothèse selon laquelle la plus petite différence remarquable est l'unité de mesure des sensations. Cette formule fondamentale est en effet l'étape décisive du raisonnement de Fechner, parce qu'il n'y a plus qu'à l'intégrer pour trouver la formule qui permet d'attribuer une quantité aux sensations. Cette intégration fait apparaître un logarithme, et Fechner aboutit à formule suivante, qu'il appelle *formule de la mesure (Massformel)* :

$$S = k \log R/r$$

Où k est une nouvelle constante issue de la transformation de la constante c , où R est la valeur du stimulus et r la valeur liminale du stimulus, c'est-à-dire à dire la valeur de seuil du stimulus à laquelle la sensation commence et disparaît (en r , $S=0$). Ce qu'il faut remarquer est que l'intégration de la formule fondamentale a permis de faire apparaître la sensation, et non plus la variation de la sensation, à gauche de l'égalité. Autrement dit, cette formule nous donne le moyen de trouver la quantité d'une sensation uniquement à partir de la mesure de la quantité du stimulus, qui ne fait pas problème pour Fechner. Il conclut alors :

Dans la formule de la mesure, on a une relation de dépendance générale entre la taille du stimulus fondamental et la taille de la sensation correspondante, et cette relation n'est pas valable seulement pour les cas de sensations égales. Cela permet de calculer la quantité de sensation à partir des quantités relatives du stimulus fondamental, et d'avoir ainsi une mesure de la sensation¹¹.

¹¹ Fechner, 1912, Section XIV.

C'est ainsi que l'on parvient à une mesure indirecte de la quantité de sensation à partir d'une mesure directe de la quantité de stimulus. La quantité d'une sensation est la somme des différences sensibles minimales traversées avant de l'atteindre ($S = \sum dS$). Il ne reste à Fechner qu'à introduire une dernière simplification pour aboutir à la formule ultime de la « loi de Weber », que Boring suggère l'appeler plus justement « loi de Fechner », ou, à la rigueur « loi de Weber-Fechner ». Cette simplification consiste à prendre r , la valeur liminale du stimulus, comme unité de mesure de R , la valeur du stimulus. Par exemple, la plus petite énergie nécessaire pour produire la sensation d'un son sera, par convention, considérée comme l'unité de mesure du son. Dès lors que r est l'unité de mesure de R , on peut écrire :

$$S = k \log R$$

En se rappelant bien que cette formule n'est vraie que si l'unité de R est la valeur liminale du stimulus r , c'est-à-dire la plus petite valeur en deçà de laquelle elle n'éveille aucune sensation (Fechner appelle aussi cette valeur de seuil du stimulus le *limen*). La présence d'un logarithme dans cette loi finale de Fechner signifie que pour augmenter d'une unité la sensation (unité de sensation = la plus petite différence sensible) il ne faut pas toujours augmenter d'autant la quantité du stimulus. Plus précisément, pour que la sensation augmente d'une unité, il faut ajouter de plus en plus de stimulus à mesure que la quantité de celui-ci s'accroît. Ainsi il faudra ajouter d'autant plus de poids au stimulus (pour que cet ajout soit sensible) que le stimulus est d'un poids élevé.

Tels sont donc les deux temps à l'issue desquels Fechner parvient à donner une formule de la mesure de la sensation. Cette formule est-elle strictement exacte ? Fechner remarque que bien qu'étant d'une importance fondamentale à ses yeux, la formule de la mesure n'est pourtant pas d'une application universelle. Elle donne seulement une bonne *approximation* de la relation entre le stimulus et la sensation, et est comparable à ce titre à la loi de Kepler en astronomie, qui fait abstraction des déviations, ou aux lois de la réfraction simple des instruments dioptriques qui font abstraction des aberrations optiques. Toutes ces lois perdent leur validité dès lors que les hypothèses dans le cadre desquelles elles sont vraies,

les conditions *normales* de leur application, ne sont pas remplies¹². A quoi réfèrent alors ces « conditions normales » dans le cas de la loi de Weber ?

En réponse à cette question, Fechner introduit une distinction importante entre *psychophysique externe* et *psychophysique interne*. La psychophysique externe étudie les relations entre le stimulus et la sensation, le stimulus étant appliqué sur la peau, ou présenté à l'œil : il est *extérieur* au corps. Dans le cadre de cette psychophysique externe, la loi de Weber est une bonne approximation, mais sa validité n'est pas absolument générale car les processus corporels qui se produisent entre la présentation du stimulus et la sensation peuvent interférer. En revanche, la loi de Weber décrit de façon exacte la corrélation entre la sensation et, non pas le stimulus, mais les processus corporels internes et la sensation. Cela signifie que la loi de Weber vaut de façon seulement approximative pour les relations entre les stimuli – disons les corps extérieurs qui nous sont présentés– et l'esprit, alors qu'elle vaut de façon stricte pour les relations entre les processus cérébraux ultimes et l'esprit. Autrement dit, la corrélation corps/esprit n'est parfaite que si on la considère comme une corrélation esprit/cerveau. Tout ceci rappelle évidemment ce que nous venons de voir sur Descartes et les esprits animaux : comme Descartes, Fechner en arrive à la conclusion que la corrélation du corps et de l'esprit ne vaut de manière stricte que par rapport à des phénomènes corporels dont nous ne sommes pas conscients, tels que les esprits animaux ou les processus corporels (rappelons toutefois que pour Descartes seuls *certain*s phénomènes mentaux sont corrélés à des phénomènes corporels, alors que *tous* le sont pour Fechner). Ce n'est pas avec les objets intentionnels (ici les stimuli) que notre esprit est corrélé de la façon la plus forte. A mesure que nous voulons établir cette corrélation, il semble que nous devons quitter le domaine de la psychologie (ou de la phénoménologie) purement descriptive et accepter de prendre en compte des hypothèses et observations de nature biologique. C'est ce que fait la psychophysique interne, contrairement à la psychophysique externe selon Fechner.

Cette loi de Fechner a été d'une très grande importance historique. D'aucuns la considèrent comme l'acte fondateur de la psychologie scientifique, faisant valoir qu'il n'y pas de science sans mesure. La psychophysique n'a cessé de se développer depuis, et s'intéresse aujourd'hui à quatre types de mesures : (i) la détection : on cherche à mesurer l'intensité en

¹² « La loi de Weber, écrit Fechner, perd entièrement sa validité aussitôt que les conditions normales ou moyennes sous lesquelles le stimulus produit la sensation ne sont pas réalisées. Elle sera toujours décisive cependant pour ces conditions particulières » (Fechner, 1912, Section VII).

deçà de laquelle le stimulus n'est pas décelé. (ii) la discrimination : on cherche à mesurer la différence détectable entre deux intensités de stimulus. (iii) l'échelonnage : on cherche à mettre en correspondance des échelles de stimuli avec les échelles de sensations du sujet. (iv) les temps de réaction : on mesure la durée en le début de la stimulation et la réponse du sujet (ainsi la loi de Piéron selon laquelle il y a une relation hyperbolique entre l'intensité de la stimulation et le temps de réaction).

2.1.2.2. Critiques de W. James et H. Bergson

Cependant, très tôt un certain nombre de voies se sont élevées contre la loi de Fechner. Deux d'entre elles sont le fait de personnalités importantes : Henri Bergson et William James. Dans son *Essai sur les données immédiates de la conscience*, Bergson commence par admettre la phase (i) du raisonnement de Fechner, la formalisation des résultats de Weber. A cette étape, il ne s'agit pas encore de mesurer la sensation, « mais seulement de déterminer le moment précis où un accroissement d'excitation la fait changer ». Cela ne pose pas de problème de principe. C'est à la seconde étape du raisonnement de Fechner que s'en prend Bergson, lorsque celui-ci tente de donner une formule pour la mesure des sensations. La psychophysique fechnerienne qui en découle repose, selon Bergson, sur une confusion entre la qualité et la quantité. Quand nous passons d'une sensation à une autre, il n'est pas vrai selon lui que nous ayons conscience d'une différence quantifiable, d'un intervalle numérique. Les sensations sont des états simples, il n'y a entre elles qu'une différence qualitative : une différence d'intension et non d'extension. On ne peut faire par exemple la soustraction entre deux sensations S et S' pour déterminer leur différence dS . Il y a bien un contraste (ou un passage), grâce auquel nous distinguons deux sensations, mais ce contraste n'a rien d'une différence arithmétique. Si nous l'associons à la différence arithmétique qui existe entre les causes des sensations (les stimuli), ce n'est que par convention : un tel rapprochement étant arbitraire, le fait que les sensations soient mesurables n'est pas un résultat, mais une simple décision. Fechner n'a rien démontré, il a postulé.

Mais si S et S' [deux sensations] sont des états simples, en quoi consistera l'intervalle qui les sépare ? Et que sera donc le passage du premier état au second, sinon un acte de votre pensée, qui assimile arbitrairement, et pour le besoin de la cause, une succession de deux états à une différenciation de deux grandeurs ?

Ou vous vous en tenez à ce que la conscience vous donne, ou vous usez d'un mode de représentation conventionnel. Dans le premier cas, vous trouverez entre S et

S' une différence analogue à celle des nuances de l'arc-en-ciel, et point du tout un intervalle de grandeur. Dans le second, vous pourrez introduire le symbole dS , si vous voulez, mais c'est par convention que vous parlerez de différence arithmétique, par convention aussi que vous assimilerez une sensation donnée à une somme¹³.

James oppose pour sa part quatre objections importantes à la théorie de Fechner¹⁴. La première est tout à fait analogue à celle de Bergson, que l'on peut appeler, suivant Boring, « l'objection de la quantité ». Comme Bergson, James admet la première étape du raisonnement de Fechner et reconnaît la validité de la loi de Weber, mais refuse la seconde étape de son raisonnement. Sa première objection à son encontre est que les sensations sont simples, elles ne sont pas composées d'unités.

A l'introspection, notre sensation (*feeling*) de rouge n'est certainement pas une portion de notre sensation d'écarlate ; pas plus que la lumière d'un arc électrique ne semble contenir celle d'une chandelle. Les choses composées contiennent des parties et une chose peut avoir deux ou trois fois plus de parties qu'une autre. Mais quand nous prenons une qualité sensible simple comme la lumière ou le son, et disons qu'il y en a alors deux ou trois fois plus qu'au moment précédent, bien que nous semblions dire la même chose que pour les objets différents, nous disons en fait autre chose¹⁵.

Ce que nous mesurons alors n'est pas la quantité d'unités qui composent intrinsèquement la sensation, mais sa différence, sa distance avec d'autres sensation. Quand nous disons d'une chose composée qu'elle est de mesure n , nous faisons selon James référence à une caractéristique *intrinsèque* de la chose : le nombre de ses parties. En revanche, nous ne pouvons dire d'une sensation qu'elle est de n , mais seulement qu'elle est n fois plus rouge par exemple qu'une autre : il s'agit là seulement d'une propriété *extrinsèque* de cette sensation, c'est-à-dire d'une propriété que la sensation ne posséderait pas si s'il n'y avait pas d'autres sensations. Pour James comme pour Bergson, les plus petites différences notables ne peuvent donc constituer l'unité de mesure des sensations car celles-ci sont simples et qualitatives.

La seconde raison que James invoque à l'encontre de l'étape (ii) du raisonnement de Fechner, « l'objection de l'égalité », est absente chez Bergson. Elle consiste à soutenir que ces plus petites différences notables non seulement ne sont pas des parties de sensation, mais ne sont de surcroît pas nécessairement égales entre elles. Cette hypothèse de Fechner est

¹³ *Essai sur les données immédiates de la conscience*, p. 49-50.

¹⁴ *The principles of psychology*, Chapitre XXIII, « The measurement of discriminative sensibility : Weber's law » sqq.

¹⁵ *ibid.*, p. 546.

arbitraire selon James : ce n'est pas parce que les variations de sensations sont perçues comme égales qu'elles le sont en fait. « Pourquoi un petite addition à une petite sensation ne pourrait-elle pas être aussi *perceptible* qu'une grande addition à une grande sensation ? » demande-t-il¹⁶.

La troisième objection de James, « l'objection physiologique », consiste à montrer que la loi de Weber, bien que probablement valide, n'apporte rien à la psychologie mais relève seulement de la physiologie. Pour lui, ni Weber ni Fechner n'ont mis en évidence de relation psychophysique. Leurs lois n'expriment en fait que des relations nomologiques entre le stimulus et le système nerveux. Ce ne sont pas des lois corps-esprit, mais des lois corps-corps, ou plus précisément, des lois stimuli-cerveau, qui aussi importantes qu'elles soient pour la physiologie, n'intéressent pas la psychologie. Si nous remarquons moins ce que nous ajoutons à un grand stimulus qu'à un petit c'est parce qu'il y doit y avoir une résistance, une friction du système nerveux (James reconnaît le caractère spéculatif d'une telle supposition). La seule corrélation que la loi de Weber exprime est donc une corrélation entre le stimulus et la « machine neurale ». Qu'il y ait ensuite une corrélation entre cette machine neurale et la sensation est hors de doute, mais elle n'est pas exprimée par une loi mathématique.

Enfin, il existe une critique plus fondamentale à l'encontre de la loi Fechner, qu'il convient de bien distinguer de l'objection précédente. Elle fait valoir que cette loi exprime une relation non pas entre les stimuli physiques et les phénomènes mentaux, mais entre les stimuli physiques et les objets *intentionnels*, c'est-à-dire les objets qui nous apparaissent, indépendamment de la question de savoir s'ils existent ou non en dehors de nous. Ainsi James écrit-il :

Ce que Fechner appelle la « sensation » est ce qui apparaît à l'esprit comme le *phénomène objectif* de la lumière, de la chaleur, du poids, du son, de la partie sensible du corps, etc¹⁷.

Brentano avait déjà soulevé cette difficulté :

¹⁶ *ibid*, vol. 1, p. 547. Brentano propose une critique similaire dans la *Psychologie d'un point de vue empirique*, I, i 1, p. 7-8 ; I, ii, 4, , p. 66-67.

¹⁷ *The Principles of Psychology*, p. 546.

On peut dire pour finir qu'une compréhension claire de ce qui est effectivement mesuré par la méthode de Fechner nous montrerait que l'objet de la mesure n'est pas tant un phénomène mental que physique. Que sont les phénomènes physiques, sinon les couleurs, les sons, le froid et le chaud, etc. qui se manifestent dans nos sensations ? –Donc quand nous mesurons les intensités de couleurs, de sons, etc., comme le faisait Fechner, nous mesurons en fait les intensités des phénomènes physiques. La couleur n'est pas voir, le son n'est pas entendre, le chaud n'est pas sentir du chaud.–¹⁸

Autrement dit, la loi de Fechner établirait une relation quantitative entre le stimulus externe et, non nos *actes* mentaux, mais les objets intentionnels de ces actes. Ou encore : la loi de Fechner relie la cause physique à ce que nous avons l'impression de percevoir et non à ce qui perçoit. On retombe, là encore, sur la confusion liée au terme de « sensation », qui est la même que celle qui vaut pour le terme « idée » : désigne-t-elle un acte ou un objet de l'esprit ? Dans le cas des lois de Weber et Fechner, cette confusion a probablement été encouragée par le fait que ces lois trouvent leur origine dans l'étude du toucher, et en particulier de la perception des poids. Cette perception s'accompagne d'un sentiment d'effort qui pose les mêmes problèmes que la douleur au regard de la théorie intentionnelle selon laquelle tout phénomène mental porte sur un objet : il n'est pas aisé de distinguer entre l'acte et l'objet dans le cas d'un effort musculaire. Comme la douleur, le sentiment d'effort physique tend à être considéré comme un acte purement subjectif, sans objet. On peut choisir de le baptiser « sensation d'effort », le terme de sensation désignant alors l'acte mental. Mais si l'on garde le même terme pour désigner la « sensation de rouge » ou la « sensation de chaud », on commet alors une confusion car le terme désigne cette fois l'objet physique perçu.

Brentano ne considère pas cette objection comme définitive, car bien qu'il souligne que la loi de Fechner ne corrèle que les stimuli et les objets intentionnels, il admet par ailleurs que l'intensité de l'acte est corrélée, en partie, à l'intensité de ces objets. On peut donc dire que pour Brentano, la loi de Fechner continue à valoir, de façon approximative, entre le stimulus et l'acte mental. La loi de Fechner peut donc être sauvée en tant que loi *psychophysique*. James en revanche refuse d'admettre une telle corrélation entre l'intensité de l'acte et celle de son objet, et nie donc que la loi de Fechner établisse une quelconque relation quantifiée entre le corps et l'esprit.

Historiquement, l'objection la plus importante a été celle de la quantité. Les plus petites différences notables ne sont pas des unités de sensations. Cette objection pourtant n'est

¹⁸ *Psychologie d'un point de vue empirique*, p. 69.

pas irrésistible. James et Bergson partagent tous les deux l'hypothèse selon laquelle une chose n'est mesurable que si elle a des parties, c'est-à-dire si elle n'est pas simple et contient d'autres choses¹⁹. Mesurer une chose, c'est dire le nombre de x qu'elle contient. Or c'est peut-être sur une théorie erronée de la mesure que repose cette objection. Il convient en effet de distinguer entre les parties naturelles d'un objet, et ses parties arbitraires ou conventionnelles. Les parties naturelles d'un objet sont par exemple les atomes et molécules qui le composent. Nous pouvons certes dénombrer ces parties et prétendre donner ainsi une mesure de la quantité de particules qui le composent. Mais nous pouvons aussi mesurer la hauteur de l'objet : ce que nous dénombrons alors, disons, des centimètres, ne sont pas des parties naturelles, mais des parties conventionnelles de l'objet : en l'absence d'un étalon rapporté arbitrairement à l'objet, aucune inspection de l'objet au microscope n'aurait pu nous révéler qu'il mesurait 15 centimètres. Or la plupart des mesures que nous faisons reposent sur la mise en rapport de l'objet physique mesuré à un étalon ou une unité qui est déterminée de façon purement conventionnelle. Cela signifie que pour l'objet en question, mesurer 15 centimètres n'est pas une propriété intrinsèque, mais une propriété extrinsèque que l'objet ne posséderait pas en l'absence de l'étalon du centimètre. Mesurer, c'est le plus souvent dénombrer autre chose que des parties naturelles. C'est découper arbitrairement un objet en parties égales, que l'on dénombre ensuite.

Il est aisé dès lors de répondre à l'objection de la quantité : certes le découpage des sensations en sommes de plus petites différences notables est arbitraire, certes leur mesure ne dénombre pas leurs parties naturelles, *mais il en va exactement de même pour la mesure du stimulus*. Nous ne procédons pas moins arbitrairement quand nous disons d'un objet qu'il pèse 100 grammes. Si Fechner s'est trompé, c'est non pas sur la mesure de la sensation, mais sur celle du stimulus, en disant que celui-ci était mesuré directement. Telle est la réponse que Boring apporte à l'objection de la quantité :

La bonne réponse à l'objection de la quantité [...] n'a pas été vue du tout jusqu'à maintenant. Elle réside dans le fait que la mesure est toujours indirecte, aussi bien pour le stimulus que pour la sensation. Bien sûr, du point de vue de l'observation

¹⁹ Bergson ne cesse de rappeler qu'il n'y a pas de quantité sans rapports de « contenant à contenu », James que la mesure quantitative porte sur la *composition* de l'entité mesurée, qui doit donc comporter des *parties*. Dans l'article « Kritik der hergebrachten Intensitätslehre » (« Critique de la doctrine reçue de l'intensité ») de ses *Untersuchungen zur Sinnespsychologie*, Brentano attribue à Gauss et Euler l'idée qu'une entité n'a une grandeur que si elle a des parties. Il propose une critique de la quantification de l'intensité qui rejoint en partie celles de Bergson et James.

immédiate, une écarlate n'est pas un certain nombre de roses ; mais dans l'observation immédiate un yard n'est pas un amas de pouces. Une grande magnitude n'est pas plus complexe qu'une petite. La mesure est une méthode qui établit une relation et qui n'est jamais directe et immédiate. Un poids de 100 grammes n'est pas du tout 100 poids d'un gramme,[...]. La véritable erreur a été celle de Fechner, quand il a dit que les stimuli pouvaient être mesurés directement, car ils ne le peuvent pas plus que les sensations. L'observation n'est pas la mesure²⁰.

Notons que cette réponse à la première objection de James (et Bergson) vaut également pour sa deuxième objection : à partir du moment où toute mesure implique un étalon arbitraire, l'égalité des unités de cet étalon ne fait plus problème.

Enfin, même si cette réponse à l'objection de James et Bergson à l'encontre de Fechner devait ne pas convaincre, et que les deux autres objections devaient se révéler décisives, notre affirmation initiale selon laquelle tous les philosophes admettent une corrélation (au moins partielle) entre le corps et l'esprit ne serait pas invalidée pour autant. Ce que récuse James et Bergson n'est pas l'existence d'une telle corrélation (puisqu'ils admettent la plausibilité de la loi initiale de Weber), mais seulement le fait qu'elle soit quantifiable. La seule restriction qu'il convient d'apporter à notre affirmation est que certains philosophes nient simplement qu'elle ait un sens. Avant d'aborder les réponses au problème de la corrélation proprement dit, il convient donc d'aborder les solutions qui tendent à dissoudre ce problème plutôt qu'à le résoudre.

2.2. Le problème dissolu

Il y a deux façons (au moins) de ne pas prendre au sérieux le problème de l'interaction : la première consiste à nier qu'il y ait une référence quelconque au terme « esprit », et à affirmer que les seules corrélations constatées relient des phénomènes corporels. La seconde consiste à admettre cette corrélation comme un fait ultime qui n'a pas besoin d'autre explication. La première approche est défendue par les psychologues behavioristes, ainsi que par certains philosophes, comme Gilbert Ryle et, selon certaines interprétations, Ludwig Wittgenstein. La seconde approche quant à elle, a été fortement défendue par David Hume²¹.

²⁰ Boring, *A History of experimental psychology*, p. 282.

²¹ En toute rigueur, il faudrait ajouter à la liste de ceux qui ne prennent pas au sérieux le problème de la corrélation les philosophes dits « éliminativistes », qui soutiennent que la psychologie du sens commun, qui fait référence à des états mentaux, est fautive et appelée à être remplacée par une autre théorie, fondée notamment sur

2.2.1 Le béhaviorisme

2.2.1.1. Le béhaviorisme psychologique : J. B. Watson et B. F. Skinner

On associe d'ordinaire le béhaviorisme aux noms de John B. Watson et B. F. Skinner. Cela n'est pas faux, bien qu'il convienne d'insister sur le véritable renversement que constitue la position de Skinner par rapport à celle de Watson. Reste que tous deux s'accordent sur l'article de base du béhaviorisme : *le rejet de l'introspection comme méthode d'investigation psychologique*. Watson entame ainsi son article « La psychologie telle qu'un béhavioriste la voit » (1913), manifeste fondateur du béhaviorisme :

La psychologie telle qu'un béhavioriste la voit est une branche expérimentale purement objective de la science naturelle. Son but théorique est la prédiction et le contrôle du comportement. L'introspection ne forme aucune partie essentielle de ses méthodes, pas plus que la valeur scientifique de ses résultats ne dépendent de la facilité avec laquelle ils se laissent interpréter en termes de conscience.

Le grief central de Watson à l'encontre fait à l'introspection est qu'elle ne permet pas à la psychologie d'obtenir des résultats *reproductibles*. Si un psychologue échoue à reproduire les résultats d'un autre, cela ne sera pas dû au fait qu'il contrôle mal le stimulus ou que ses instruments sont défectueux, mais seulement au fait que son introspection, ou celle de son sujet, est mal exercée. « L'attaque porte alors sur l'observateur et non sur le dispositif expérimental ». Dès lors, la psychologie introspectionniste pourrait bien n'être qu'une discipline où règne un malentendu constant, où personne ne donne la même signification aux termes employés, et où les questions de savoir si les sensations auditives ont une étendue ou si la couleur a une intensité, sont condamnées à rester éternellement en débat. Que propose alors Watson en lieu et place de l'introspection ?

Je crois que nous pouvons écrire une psychologie [qui] n'use jamais des termes conscience, état mental, esprit, contenu, vérifiable par l'introspection, images, et autres termes de ce genre. [...] Elle peut être écrite en termes de stimulus et de réponse, de formation d'habitude, d'intégration d'habitude et d'autres termes de ce type.

les neurosciences. Pour une défense d'une théorie de ce type voir Paul Churchland, « Le matérialisme éliminativiste et les attitudes propositionnelles ».

Comme son nom l'indique, le béhaviorisme (aussi appelé comportementalisme) refuse donc de prendre pour objet autre chose que ce qui relève du comportement observable. Toute référence à des états mentaux internes est soigneusement évitée : l'esprit est considéré comme une boîte noire impénétrable, seule ce qui y entre (les stimuli) et en sort (le comportement) peut faire l'objet d'une étude scientifique rigoureuse. Dans ce cadre, l'interprétation des lois de Fechner et Weber doit être purement comportementale : aucune n'établit de relation entre le stimulus et l'esprit. La seule relation que nous pouvons établir est celle qui relie le stimulus à la réaction comportementale. Il n'y a donc aucun moyen de décider s'il y a une corrélation entre le corps et l'esprit, et Watson laisse au métaphysicien le soin de s'occuper du problème corps-esprit, de décider entre interaction cartésienne et parallélisme leibnizien. Il faut bien préciser que Watson n'entend pas nier qu'il y ait des états mentaux. Simplement, il s'agit là d'une question métaphysique dans laquelle il refuse de s'engager. Tout ce qu'il affirme est que ces états ne peuvent être les objets de la psychologie scientifique. En ce sens, son béhaviorisme est *épistémologique* et non *ontologique* : la psychologie (comme toutes les autres *sciences* finit par dire Watson) ne peut rien *savoir* des états mentaux, elle ne peut donc ni nier, ni affirmer qu'ils existent. Dès lors, le problème de la corrélation n'est pas à proprement parlé dissolu, mais plutôt suspendu. Il cesse d'occuper les psychologues et autres scientifiques et demeure un problème exclusivement philosophique. La psychologie s'inscrit alors parmi les autres sciences naturelles, aux côtés de la physique et de la chimie. Son objet, le comportement, est lui-même explicable ensuite en termes physico-chimiques.

Il importe de présenter rapidement le type de loi que la psychologie béhavioriste entend formuler, notamment parce que c'est sur ce point que la différence entre Watson et son disciple Skinner est la plus importante. Watson est issu de la psychologie animale, où il a été marqué par le physiologiste russe I. P. Pavlov (1849-1936). Dans cette discipline, le recours à la notion d'état mental ou de conscience devient très tôt suspect d'anthropomorphisme²². Watson ne cesse d'y insister : il n'y a pas de différence de principe entre la psychologie animale et la psychologie humaine. Attribuer aux humains une conscience n'est pas plus fondé scientifiquement qu'en attribuer une aux animaux. Les humains doivent être étudiés comme les animaux : par la méthode pavlovienne du *conditionnement*. Qu'est-ce à dire ? La psychologie béhavioriste est l'étude des relations entre les stimuli et les réponses : elle

²² Voir notamment la mise en garde fameuse du zoologue et psychologue anglais C. Lloyd Morgan : « Nous ne devons en aucun cas interpréter une action comme le produit d'une faculté psychologique de haut niveau si elle peut être interprétée comme le produit d'une faculté de plus bas niveau sur l'échelle psychologique », *An Introduction to Comparative Psychology*, 1894.

cherche à établir des liens fonctionnels entre les variations de l'environnement et les réponses du sujet. Or Watson adopte une forme d'empirisme radical (très proche du sensualisme de Condillac) : à la naissance, l'esprit est une véritable *tabula rasa*. Cela signifie que la quasi-totalité des réponses que l'organisme va adopter vis-à-vis de son environnement sont des réponses acquises, ou, dans les termes que Watson emprunte à Pavlov, *conditionnées*²³. Ainsi en va-t-il de la fameuse expérience du « chien de Pavlov » : lorsque l'on présente un morceau de viande à un chien, il salive. Admettons qu'il s'agisse là d'une des rares réponses inconditionnées du chien : il est inné, ou naturel pour un chien de saliver quand on lui présente de la viande. L'expérience Pavlov consiste alors à présenter de manière répétée un signal sonore juste avant la présentation du morceau de viande. Au début, le chien tourne la tête vers le signal, mais au bout d'un certain temps, la seule présentation du signal sonore sans présentation de la viande suffit à enclencher la réponse de salivation chez le chien : il s'agit là alors d'une réponse conditionnée, dans la mesure où le chien l'a acquise. Ainsi le monde se construit-il pour Pavlov et Watson par un élargissement extraordinaire du répertoire de stimuli pertinents. Le monde de notre chien vient de s'enrichir du son. Si l'on peut prédire le comportement, c'est donc parce que l'on peut renforcer certains stimuli.

C'est sur point que la différence entre Watson et Skinner est la plus importante. Pour ce dernier, il ne va plus s'agir de renforcer des stimuli (en répétant plusieurs fois le son avant la présentation de la viande), mais de renforcer certaines *réponses*. On passe du conditionnement répondant (l'association, par présentation répétée, d'un stimulus neutre à un stimulus inconditionné, qui devient ainsi conditionné) au conditionnement opérant (le renforcement positif ou négatif du comportement du sujet, qui conduit à l'association d'une réponse comportementale à un stimulus). Si l'influence centrale de Watson était Pavlov, l'influence déterminante pour Skinner est celle d'Edward Thorndike (1874-1949) qui formula la célèbre « loi de l'effet » selon laquelle une action renforcée positivement (c'est-à-dire une action dont les conséquences sont bénéfiques pour l'animal ou le sujet humain en question) tend à être reproduite. Ce dernier mis au point un paradigme expérimental qui allait fortement marquer Skinner : un animal (un chat) était placé dans une cage qui contenait différents leviers. Un seul levier (ou une seule combinaison de leviers) ouvrait la porte de la cage. En répétant l'expérience, il était alors possible de mesurer le temps que le chat mettait pour ouvrir la cage, quantifiant ainsi le degré de son apprentissage. Il y a, derrière ce paradigme que Skinner

²³ Watson admet toutefois un répertoire très restreint d'actions inconditionnées : le cri provoque le retrait, la chute provoque la peur, cela, de manière innée.

poursuivra en parlant de conditionnement opérant, une hypothèse darwinienne selon laquelle notre comportement fonctionne de la même façon que l'évolution des espèces. Skinner mit même en évidence des comportement « magiques » de certains animaux : lorsqu'une souris appuie par hasard sur un levier qui déclenche l'arrivée de nourriture, et qu'elle a fait auparavant un tour sur elle-même, elle va avoir tendance, les fois suivantes, à tourner une fois sur elle-même avant d'aller appuyer sur le levier.

En dépit de leurs différences, Watson comme Skinner sont absolument convaincus de la nécessité d'appliquer aux comportements humains la même méthodologie behavioriste que celle qui est en jeu dans l'étude du comportement animal. Aucun des deux en outre, ne considère cette position comme un pis-aller face au rejet de l'introspection. Tous deux se montrent au contraire résolument optimistes quant aux perspectives d'application d'une telle méthode aux comportements humains les plus complexes, tel le langage. C'est sur ce point justement, que le behaviorisme va rencontrer des difficultés que beaucoup considèrent aujourd'hui comme insurmontables.

L'objection la plus décisive à l'encontre du behaviorisme a été formulée en 1959 par Noam Chomsky, dans une critique restée célèbre du livre de Skinner *Verbal Behavior*. Skinner cherche à montrer dans ce livre que le comportement verbal peut être prédit sur la seule base d'observation (et de manipulations) de l'environnement physique du locuteur. Chomsky commence par remarquer que si les notions de stimulus et de réponse sont clairement déterminées dans le contexte du laboratoire (l'émission d'un son dans la boîte, l'actionnement du levier par la souris) elles deviennent nettement plus floues dans les contextes de vie réelle. Il y a dans ces contextes un nombre très élevé d'événements physiques qui peuvent prétendre au statut de stimulus (tout événement auquel l'organisme est capable de réagir) ainsi qu'une pléiade de sorties comportementales qui peuvent prétendre au statut de réponse. Le problème, pour le behavioriste, est que s'il définit de manière aussi large les stimuli et les réponses, il ne peut mettre en évidence aucune relation nomique entre eux, et que la seule façon d'expliquer le comportement est alors de recourir à des phénomènes mentaux que le behavioriste exclut par hypothèse. Il semble donc nécessaire de restreindre la classe des stimuli et des réponses. Mais la seule façon de le faire paraît être de définir les stimuli comme les événements nomiquement connectés (c'est-à-dire connecté par des lois) à

des réponses et les réponses comme des événements nomiquement connectés à des stimuli. Il devient alors clairement trivial de dire qu'il y a des lois reliant les stimuli aux réponses. En toute rigueur, le béhavioriste semble donc devoir choisir entre le renoncement et la trivialité. Skinner, selon Chomsky, ne parvient à échapper à ce dilemme qu'en apparence, en exportant les termes rigoureusement définis en laboratoire à des situations complexes où ils ne sont plus applicables. Ce qu'est un stimulus est clair et contrôlable dans un laboratoire, mais devient totalement flou dans un contexte naturel. Il vaut la peine de citer longuement le passage où Chomsky montre les difficultés que pose l'exportation des notions de stimulus et de réponse dans le contexte de la vraie vie.

Un exemple typique de *contrôle du stimulus* pour Skinner serait la réponse « Mozart » à un morceau de musique ou la réponse « Hollandais » à une peinture. Ces réponses sont dites être « sous le contrôle de propriétés extrêmement subtiles » de l'objet ou de l'événement physique. Supposez qu'au lieu de « Hollandais », nous ayons dit « Ca jure avec le papier peint », « Je pensais que tu aimais l'art abstrait », « Jamais vu avant », « Penché », « Accroché trop bas », « Superbe », « Hideux », « Tu te souviens de notre séjour en camping l'été dernier ? », ou quoi que ce soit d'autre qui nous vienne à l'esprit quand nous regardons une image (dans les termes skinneriens, n'importe quelle réponse de force suffisante). Skinner pourrait seulement dire que chacune de ces réponses est sous le contrôle de quelque autre propriété-stimulus de l'objet physique. Si nous regardons une chaise rouge et disons « Rouge », la réponse est sous le contrôle du stimulus *rougeur* ; si nous disons « chaise », elle est sous le contrôle de la collection de propriétés (pour Skinner, l'objet) *chaise*, et ainsi pour les autres réponses. Le dispositif est aussi simple que vide. Puisque les propriétés sont disponibles à la demande (nous en avons autant que nous avons d'expressions descriptives non synonymes dans le langage), nous pouvons expliquer une vaste classe de réponses dans les termes d'une analyse skinnerienne fonctionnelle en identifiant les *stimuli de contrôle*. Mais le mot *stimulus* a perdu toute son objectivité dans cet usage. Les stimuli ne sont plus des parties du monde physique extérieur ; ils sont reconduits dans l'organisme. Nous identifions le stimulus quand nous entendons une réponse. Il est clair à partir de tels exemples, qui abondent, que le fait de parler de contrôle du stimulus ne fait que couvrir une retraite au sein de la psychologie mentaliste²⁴.

Chomsky met donc le béhavioriste face à dilemme : soit la trivialité, soit la réintroduction de l'idiome psychologique. Suite notamment à cette critique, l'échec du programme béhavioriste de Skinner est aujourd'hui majoritairement admis.

2.2.1.2. Le béhaviorisme philosophique : G. Ryle

²⁴ « A Review of B. Skinner's Verbal Behavior », III.

Il est cependant une autre forme de béhaviorisme, d'origine philosophique et non psychologique, qui continue à exercer une certaine influence. Il s'agit principalement de l'analyse que le philosophe Gilbert Ryle a donné du concept d'esprit dans son livre *The concept of Mind* (1949)²⁵. Commençons par remarquer que bien qu'on le lui attribue souvent, Ryle refuse le qualificatif de béhavioriste²⁶. Le béhaviorisme est pour lui une théorie qui cherche à réformer la *méthodologie* de la psychologie. Or l'ambition de Ryle n'est pas de réformer une méthodologie, mais de remettre en cause une *ontologie*. Cette ontologie que Ryle rejette avec virulence correspond à l'ontologie dualiste cartésienne, selon laquelle il y aurait dans le monde des esprits, qui seraient non étendus, privés, et infailliblement accessibles en première personne, et des corps, étendus, publics, et à propos desquels nous pourrions nous tromper. (Notons que Ryle n'envisage pas le critère intentionnel de distinction entre le corps et l'esprit). Ryle qualifie tour à tour cette doctrine de « théorie de la double vie », de « mythe cartésien » de « légende des deux mondes » et de « dogme du fantôme dans la machine ». La critique que Ryle fait du dualisme cartésien *ne vise pas cependant à réduire le corps à l'esprit ou l'esprit au corps*. Sa critique cherche à montrer que ce dualisme, comme tout monisme physicaliste ou idéaliste qui lui seraient opposés, sont des réponses à un faux problème. Celui-ci repose sur une erreur de catégorie qui consiste à admettre que le corps et l'esprit relèvent d'un même type logique.

[...] l'idéalisme et le matérialisme sont des réponses à des questions mal posées. La « réduction » du monde matériel à des états et processus mentaux aussi bien que la « réduction » des états et processus mentaux à des états et processus physiques présuppose la légitimité de la disjonction : « ou il existe des esprits ou il existe des corps (mais pas les deux) ». Ce qui équivaut à dire : « ou elle a acheté un gant pour la main droite et un gant pour la main gauche ou elle a acheté une paire de gants (mais pas les deux) »²⁷.

Ryle ne définit pas précisément ce qu'il entend par « erreur de catégorie », il se contente de donner une série d'exemples. Une personne visite pour la première fois un campus universitaire, voit les bibliothèques, les terrains de sport et les laboratoires et demande « Mais, où est l'Université ? », supposant que l'Université est un nouveau bâtiment alors qu'elle consiste en l'ensemble de ce qu'il a vu. C'est une erreur de catégorie. Il en va de même dans le cas d'un enfant assistant au défilé d'une division qui demande, après avoir vu

²⁵ Trad. Fr. S. Stern-Gillet, *La notion d'esprit, pour une critique des concepts mentaux*, Paris, Payot, 1978.

²⁶ 1978, p. 309 ssq.

²⁷ 1978, p. 22. Nous laissons ici de côté la question de savoir si le « second » Wittgenstein soutient lui aussi une forme de béhaviorisme.

les bataillons, à quel moment il verra la division, « supposant ainsi qu'elle est analogue aux unités déjà vues ». Les erreurs de catégorie proviennent donc d'une inaptitude à utiliser certains concepts dans une langue. Mais alors, si l'esprit ne relève pas de la même catégorie que le corps, de quelle catégorie relève-t-il ?

La réponse de Ryle est qu'il relève de la catégorie des *dispositions*²⁸. Les concepts mentaux (tels que la volonté, la connaissance, la croyance, le désir) désignent des dispositions. Les dispositions s'opposent pour Ryle aux épisodes. Ainsi, quand nous disons de quelqu'un qu'il est fumeur, nous lui attribuons une disposition, c'est-à-dire une tendance, une compétence, une capacité ou une inclination. Mais quand nous disons de lui qu'il fume, nous décrivons un épisode. Les épisodes sont des faits, les dispositions ne sont rien d'autre que des possibilités. Pour Ryle, les dispositions sont dénotées par des termes ou énoncés *hypothétiques* (c'est-à-dire conditionnels), alors que l'on réfère à des faits à l'aide de termes ou d'énoncés *catégoriques*. Le point crucial est que *les termes et phrases dispositionnels (hypothétiques), ne relatent aucun fait, aucun épisode*. Quand nous disons que Paul est un fumeur, nous ne décrivons pas un fait : nous disons simplement que si Paul est dans telle ou telle condition (qu'il n'a pas fumé depuis deux heures, qu'il vient de boire son café et qu'il a un paquet de cigarette dans la poche par exemple), il fumera. Paul reste fumeur, même lorsqu'il ne fume pas, même lorsqu'aucun fait ne correspond à l'énoncé dispositionnel.

Posséder une propriété dispositionnelle ne revient pas à se trouver dans un état particulier ou à subir un changement particulier ; c'est se trouver dans un état particulier ou exposé à un changement donné lorsqu'une condition déterminée se trouve réalisée. La même chose vaut pour les dispositions spécifiquement humaines telles que les qualités de caractère. Le fait que je sois fumeur n'entraîne pas que je fume à tel ou tel moment mais plutôt que j'ai envie de fumer lorsque je ne suis pas en train de manger, de dormir, de donner un cours, d'assister à des funérailles ou que je ne viens pas de finir une pipe²⁹.

Les attributions de dispositions doivent dès lors être paraphrasées par des énoncés hypothétiques formulés au conditionnel. Par exemple « Julie croit qu'il fait chaud » peut-être paraphrasé par « Si Julie sort, elle ne prendra pas son parapluie ». A l'inverse, les descriptions de faits, « Paul fume », sont des énoncés catégoriques qui ne nécessitent aucune paraphrase.

²⁸ Nous avons déjà rencontré les dispositions en première partie, lors de l'examen de l'objection à la thèse de l'intentionnalité de Brentano : les dispositions ne sont pas toutes mentales, et pourtant elles semblent orientées vers un objet, leur actualisation. Ici ce sont les dispositions mentales qui nous intéressent.

²⁹ *La notion d'esprit*, p. 43.

Une disposition ne réfère donc pas à une propriété dans le monde, mais peut-être réduite à un énoncé conditionnel. Ryle propose donc une réduction conditionnelle des dispositions³⁰.

Dire de ce morceau de sucre qu'il est soluble revient à dire qu'en toutes circonstances, il se dissoudrait s'il était plongé dans l'eau. Dire d'un homme endormi qu'il connaît l'anglais revient à dire que, si on lui adressait la parole dans cette langue ou si on lui montrait un journal britannique, il répondrait en anglais, agirait de façon appropriée ou traduirait correctement les textes dans sa langue maternelle³¹.

Qu'est-ce que ces remarques sur le caractère purement hypothétique, non factuel des énoncés dispositionnels nous apprennent sur le statut de l'esprit ? Beaucoup de choses, dans la mesure où les concepts mentaux désignent pour Ryle des dispositions. Les termes « connaître », « croire », « vouloir », « aspirer », « intelligent », « spirituel » sont des termes dispositionnels qui ne désignent aucun fait, mais seulement des dispositions. Si l'on ne le réalise pas tout de suite, c'est parce que dans ces cas, à la différence du cas de « fumeur », les dispositions peuvent se manifester de plusieurs façons. En effet, à la disposition « fumeur » correspond un seul type d'épisode « fumer », et l'épisode et la disposition sont désignés par des termes de même racine. Mais il n'en va pas ainsi de toutes les dispositions. « Fumeur » est en effet un terme dispositionnel *déterminé*, c'est-à-dire que cette disposition ne s'actualise que dans un sens. Il y a également des termes dispositionnels *déterminables* qui peuvent s'actualiser en plusieurs sens : « avide », « élastique », « croire », « intelligent ». Plusieurs types d'épisodes peuvent correspondre à l'actualisation de ces dispositions³². C'est le cas de la plupart des dispositions mentales. La thèse centrale de Ryle est donc que *les concepts mentaux sont des concepts dispositionnels, qui contrairement ce que suggère le mythe du fantôme dans la machine, ne désignent aucun fait*. Si je dis que Janine comprend l'algèbre, je ne fais pas référence pour Ryle à un acte ou un état qui se trouverait « à l'intérieur » de Janine, ou dans son esprit, je dis simplement qu'elle est disposée, par exemple, à trouver des solutions correctes à des équations algébriques si on les lui présente.

³⁰ L'expression est proposée par Mumford, 1998, p. 16.

³¹ *La notion d'esprit*, p. 120 (trad. corrigée). Voir également p. 42 : « Lorsque nous décrivons le verre en disant qu'il est fragile et le sucre en disant qu'il est soluble, nous utilisons des concepts de disposition dont la force logique est la suivante. La fragilité du verre ne consiste pas dans le fait qu'il a été, à un moment donné, brisé, car le verre peut être fragile sans avoir été brisé. Dire du verre qu'il est fragile revient à dire que, s'il reçoit un coup ou est soumis à une tension trop forte, il se brise en morceaux. Dire du sucre qu'il est soluble, c'est dire que, dans l'eau, il se dissout ».

³² On trouve déjà une distinction voisine entre deux types de dispositions chez Aristote (*Ethique à Nicomaque*, Livre II), qui distingue *l'hexis*, ne pouvant s'actualiser que dans un sens, de la *dunamis* qui peut s'actualiser dans des sens contraires.

Demeure un problème : si les termes et énoncés dispositionnels ne désignent pas de faits, en vertu de quoi peuvent-ils être vrais ou faux ? Un énoncé catégorique est vrai si un fait lui correspond et faux si aucun fait ne lui correspond. Mais un énoncé dispositionnel ? Comment pouvons-nous dire qu'il est vrai que Paul est fumeur s'il n'y a rien dans le monde qui rend vrai cet énoncé ? C'est sur ce point que David M. Armstrong centrera sa célèbre critique de Ryle. Il convient cependant de rappeler que Ryle envisage lui-même cette objection. Quelle est sa réponse ? Ryle commence par remarquer que le langage n'a pas pour seule fonction de décrire : « Relater des faits n'est que l'une des nombreuses fonctions des phrases en général »³³. Ainsi les phrases interrogatives ou impératives ne désignent aucun fait. Reste que les énoncés dispositionnels sont formulés à l'indicatif. Ryle cherche donc à montrer dans un second temps que même parmi les phrases indicatives, certaines n'ont pas pour fonction de décrire. C'est le cas notamment des énoncés de loi. Ceux-ci sont de la forme « Tout ce qui est *a* est *b* ». Ces énoncés sont, selon Ryle, susceptibles d'être vrais ou faux bien qu'ils ne désignent aucun fait. Il ne sont pas de même niveau dit-il que les énoncés factuels qu'ils présupposent. Un énoncé de loi nous permet seulement de faire des prédictions :

On utilise une loi comme un billet d'inférence (du genre abonnement) permettant au titulaire de passer de l'affirmation de certains faits particuliers à l'affirmation d'autres faits particuliers³⁴.

Ce serait donc une erreur de penser que les lois décrivent des faits : « Autant prétendre qu'une règle de grammaire est un nom ou un verbe supplémentaire mais muet et qu'une règle du jeu d'échecs est un pion supplémentaire mais invisible ! ». Une loi n'exprime que le passage d'un fait à un autre, ce passage n'étant pas lui-même un fait. La troisième étape du raisonnement de Ryle est de rapprocher les énoncés dispositionnels des énoncés de loi. Bien qu'ils ne soient pas, à proprement parler des énoncés de loi, puisqu'ils ne s'appliquent qu'à des choses particulières, les énoncés dispositionnels fonctionnent néanmoins comme des « billets d'inférence ». Si nous savons de Paul qu'il est fumeur, nous pouvons inférer qu'en telles circonstances, il fumera. Mais on aurait tort de prétendre que cette inférence réfère elle-même à une relation entre des états de choses.

³³ P. 116. Ce point deviendra central pour J. L. Austin qui parlera d'« illusion descriptive ». Le langage n'a pas selon lui pour seule fonction de décrire, mais aussi de faire (de promettre, d'interroger, d'ordonner, de baptiser...). Voir *Quand dire c'est faire*, 1962.

³⁴ P. 118.

En résumé, Ryle soutient que l'on peut maintenir une distinction de catégorie entre les termes physiques et les termes mentaux sans pour autant la fonder sur une distinction entre deux types de choses. La distinction conceptuelle n'implique pas une distinction ontologique. Les termes mentaux ne désignent pas de faits, contrairement à de nombreux termes physiques. On remarquera qu'il y a une forme de chiasme entre la position de Ryle et celle d'un behavioriste comme Skinner. Ce dernier suspend son jugement en matière de métaphysique et ne se prononce jamais sur l'existence ou l'inexistence d'entités mentales. En revanche, il s'emploie à éliminer l'idiome mentaliste de la psychologie. C'est l'inverse que fait Ryle : il nie l'existence ontologique d'entités mentales mais s'emploie à sauver l'idiome mentaliste. Les termes mentaux, bien qu'ils n'aient pas à proprement parler de référence, ne sont pas pour autant privés d'usage. Ils appartiennent au contraire à la classe des termes et des énoncés dont la fonction n'est pas de décrire mais de prédire.

Trois objections au moins peuvent être adressées à la position de Ryle. La première tient à la relative imprécision de ce qu'il entend par « catégorie ». Ryle ne le définit jamais explicitement, mais quand on lit les pages consacrées à l'« erreur de catégorie »³⁵, il semble que Ryle ait en tête deux catégories : les termes qui désignent des faits d'un côté, et ceux qui n'en désignent pas de l'autre. Ainsi, pour reprendre son exemple, les termes « Bataillons » et « Escadrons » ne sont pas de la même catégorie que « division », parce que les premiers désignent des faits, et le dernier ne désigne aucun fait supplémentaire (parler de division n'implique aucun supplément d'être en dehors des bataillons et escadrons qui la compose). On comprend bien dès lors pourquoi Ryle interdit toute question sur la possible réduction du corps à l'esprit ou de l'esprit au corps. De telles relations de réduction ou d'irréductibilité ne peuvent valoir qu'au sein d'une même catégorie. Le problème apparaît cependant lorsque l'on considère que Ryle admet, parmi les énoncés qui ne décrivent pas de faits, bien d'autres termes que les énoncés mentaux. En particulier, certains énoncés physiques ne désignent pas de faits : ce sont les énoncés physiques qui expriment des lois ou des dispositions comme la solubilité ou la malléabilité. Le dilemme est alors le suivant : soit Ryle doit fournir un critère expliquant pourquoi les dispositions mentales et les dispositions physiques relèvent de catégories distinctes, ce qu'il ne fait pas de manière évidente ; soit il admet que dispositions physiques et mentales relèvent d'une même catégorie, auquel cas rien n'interdit plus de

³⁵ P. 17-23.

s'interroger sur la possibilité de réduire des unes aux autres. Prenons un exemple. « Courage » et « neurone » sont des termes de catégories différentes, puisque « courage » est un terme dispositionnel alors que « neurone » réfère à un fait. La question « peut-on réduire le courage aux neurones ? » est donc dénuée de sens, elle repose sur une erreur de catégorie. Mais si l'on considère les termes « Courage » et « excitabilité d'un neurone », rien n'interdit plus de se demander si le courage peut-être réduit à l'excitabilité d'un neurone, dans la mesure où l'un et l'autre sont des termes dispositionnels. La seule façon pour Ryle d'interdire une telle question serait d'expliquer en quoi ces deux dispositions relèvent de catégories distinctes, bien que toutes deux ne portent pas sur des faits, étant de simples billets d'inférence particuliers. Il ne le fait pas.

La seconde objection qui peut être adressée au « béhaviorisme » concerne la raison que nous avons d'unir différentes manifestations comportementales conditionnelles sous un même concept mental. Selon Ryle, une disposition n'est rien d'autre que la tendance d'une personne à se comporter d'une certaine manière en certaines circonstances. Dire que John croit qu'il pleut, c'est dire que s'il sort il prendra son parapluie, s'il prend la voiture, il allumera les essuie-glaces, etc. Le problème n'est pas tant le caractère infini de cette liste (que Ryle reconnaît, et admet) que son caractère arbitraire. La question est : qu'est-ce qui unifie, qui justifie le fait que l'on regroupe ensemble tous ces différents comportements ? Qu'est-ce qui nous interdit, dans l'optique béhavioriste, d'adjoindre à la liste des comportements composants la croyance qu'il pleut « sauter à cloche-pied dans les escaliers » ? Comme le remarque E. J. Lowe :

Il semble que nous devons *déjà comprendre* ce que signifie l'attribution à quelqu'un de la croyance qu'il pleut afin de pouvoir générer les items sur la liste, de telle sorte que la liste ne peut être utilisée pour *expliquer* ce que signifie l'attribution d'une telle croyance à quelqu'un³⁶.

Cette objection rejoint en partie celle que Chomsky faisait à Skinner : le béhavioriste n'a pas de moyens de spécifier les conditions de manifestations d'un comportement, ni de justifier que l'on regroupe ensemble certains comportements et pas d'autres sans présupposer un état mental. Officiellement, il ne peut pas dire que « prendre son parapluie » et « allumer ses essuie-glaces » sont à réunir parce qu'ils *résultent* de la croyance qu'il pleut, conçu comme un événement interne. Mais implicitement, il semble bien admettre quelque chose de

³⁶ *An Introduction to the Philosophy of Mind*, p. 43

ce type quand il se considère en mesure d'inclure à cette liste « annuler le pique-nique » mais pas « se gratter l'oreille ».

Le troisième type d'objection qui a été soulevé à l'encontre de la théorie de Ryle concerne précisément sa théorie des dispositions. Ryle considère que ce qui distingue une disposition d'un fait ne relève pas de l'ontologie mais de la sémantique. Contrairement aux énoncés décrivant des épisodes ou des faits, les énoncés dispositionnels (« Herbert sait que les ours sont dangereux») peuvent être traduits en des énoncés conditionnels du type « Si Herbert rencontre un ours, il fuira ». Mais si Herbert ne rencontre jamais d'ours, l'énoncé « Herbert sait que les ours sont dangereux » sera-t-il vrai ou faux ? On a vu que Ryle envisageait cette objection en principe, et y répondait en rapprochant les dispositions des lois de la nature. Reste alors la question de savoir sur quelle base nous affirmons que certaines lois sont vraies et d'autres sont fausses. En disant que les lois ne désignent aucun fait, mais ne sont que des billets d'inférence, Ryle s'inscrit implicitement dans le cadre d'une conception humienne des lois de la nature : celles-ci ne décrivent que des régularités, des conjonctions constantes, mais aucunement des forces ou des nécessités qui reliaient ces faits. (Nous reviendrons sur ce point à l'occasion de l'examen de la solution de Hume). Il s'ensuit qu'avant d'attribuer une disposition à Jean, nous devons avoir constaté que Jean a manifesté cette disposition de façon répétée. Ainsi, parlant de la compétence d'un enfant à faire des nœuds, Ryle écrit :

Qu'il sache nouer des nœuds plats, nous l'inférons à partir des points suivants : on le lui a enseigné, il a de l'expérience, il les réussit presque toujours, il remarque les nœuds mal noués et sait les corriger, il a honte de celui qu'il vient de nouer et, sans aide, le refait. [...]

Qu'il y ait ou non des épisodes occultes dans le flux de conscience de l'enfant n'a rien à voir avec la question de savoir s'il possède ou non la compétence requise pour nouer des nœuds plats³⁷.

Le problème est que certaines dispositions semblent pouvoir être attribuées bien qu'elles se manifestent pour la première fois ou qu'elles ne se soient jamais manifestées. C'est le cas notamment des dispositions qui ne se manifestent qu'une fois : ainsi, un verre peut être fragile, bien qu'il ne se soit jamais brisé (un verre brisé n'est plus fragile). Autrement dit, si l'énoncé « Ce x est fragile » doit avoir une valeur de vérité, quelque chose d'autre que les manifestations précédentes de sa fragilité (le bris) doit le rendre vrai. Ce type d'argument a été développé par le philosophe australien David M. Armstrong, dès 1968 dans

³⁷ P. 70.

son livre *A Materialist Theory of the Mind*, qui, rompant avec l'approche ryléenne, est considéré comme un des ouvrages fondateurs de la philosophie contemporaine de l'esprit. Ryle affirme à plusieurs reprises que l'on peut attribuer une disposition à un objet même si celui-ci ne la manifeste pas à l'instant de l'attribution. Armstrong est d'accord, mais considère que Ryle n'a aucun moyen de justifier de telles attributions. Il objecte que tant que nous concevons les dispositions comme de purs énoncés conditionnels sans aucun corollaire ontologique, nous n'avons aucune raison de considérer qu'une attribution de disposition non manifestée est vraie ou fausse, ou plus vraie ou plus fausse qu'une autre. Trois problèmes se posent à Ryle dans ce contexte. Le premier, que nous venons d'envisager, est l'attribution à un objet d'une disposition qu'il n'a encore jamais manifestée. Le second est l'acquisition ou la perte d'une disposition non manifestée. Supposons que je réalise soudain que les ours sont dangereux, alors que je les croyais auparavant inoffensifs. Comment rendre compte d'un tel changement sans supposer que quelque chose a changé dans la scène interne, cachée, que Ryle tient en horreur ? Il se peut que je n'aie jamais l'occasion de fuir devant un ours, et pourtant, il se sera bien passé quelque chose au moment où j'aurai réalisé que les ours étaient dangereux. En d'autres termes, sur quelle base Ryle va-t-il pouvoir justifier que jusqu'à l'instant t , la proposition « S'il voit un ours, il fuira » est fausse, mais qu'elle devient vraie à partir de t ? Le troisième problème que rencontre Ryle est développé plus longuement par Armstrong. Il concerne cette fois un objet qui a déjà manifesté plusieurs fois une disposition (par exemple, un klaxon qui a manifesté à plusieurs reprises que dans certaines conditions de pression, un son était émis). La question que pose Armstrong est la suivante : sur quelle base Ryle affirme-t-il qu'un objet qui a toujours manifesté une disposition continuera à le faire ? Quelles raisons avons-nous de penser, dans l'optique de Ryle, que le klaxon sonnera encore la prochaine fois qu'on l'actionnera ? Notre réponse naïve, qu'Armstrong s'efforce de défendre, est que cela tient au fait que le dispositif interne du klaxon ne soit pas endommagé, que l'interrupteur soit bien relié à la sirène, que les deux soient en état de marche etc. Mais Ryle ne peut accepter une telle réponse qui réduit les dispositions non plus à des énoncés hypothétiques, mais à des faits qui se situent dans une scène interne. Que propose Armstrong, une fois rejetée la réduction conditionnelle des dispositions soutenue par Ryle ?

L'approche d'Armstrong prend le contre-pied direct de celle de Ryle : pour Armstrong, les énoncés dispositionnels, qu'ils concernent les dispositions physiques ou mentales, désignent quelque chose dans le monde, ils ne sont pas de simples billets

d'inférence. La position d'Armstrong peut être présentée comme la conjonction de deux thèses : (i) Toutes les dispositions ont une base. (ii) Cette base est catégorique.

(i) La première thèse défendue par Armstrong revient à dire que toutes les dispositions sont fondées ontologiquement. Par exemple, la disposition « connaître l'anglais » réfère non plus à un énoncé conditionnel mais à un état qui persiste au travers des différentes manifestations de la connaissance de l'anglais. L'argument central en faveur d'une telle thèse est ce qu'Armstrong appellera le principe du vérifacteur (*truthmaker*) : si un énoncé est vrai, quelque chose doit le rendre vrai. Il s'agit là d'un

principe général que j'appelle, derrière C. B. Martin, le principe du truthmaker. Selon ce principe, pour toute vérité contingente au moins (et peut-être pour toute vérité contingente et nécessaire) il doit y avoir quelque chose dans le monde qui la rend vraie³⁸.

Appliquant ce principe aux dispositions, Armstrong en vient donc à l'idée que celles-ci désignent quelque chose de distinct des manifestations, qui est à leur base et qui persiste même quand celles-ci cessent :

Je pense que Ryle a eu tout à fait raison de revendiquer un rôle essentiel pour la dispositionnalité dans l'élucidation de notre notion du mental. Ce fut là une contribution importante et durable. Mais nous devons alors poursuivre en considérons la question du vérifacteur pour ces vérités dispositionnelles. Qu'y a-t-il dans le monde qui rend vraies ces vérités ? Ryle n'a pas de réponse.

Aussitôt que nous soulèverons la question du vérifacteur, notre approche de la nature de l'esprit se transformera très probablement et nous prendrons une direction tout à fait anti-ryléenne. Nous identifierons sûrement une croyance avec, disons, un état interne de l'esprit (les métaphysiciens matérialistes l'identifieront ensuite avec un état du cerveau) qui, dans les circonstances adéquates, mais seulement dans ces circonstances, se manifestera de différentes manières, dont certaines peuvent être un comportement apparent.

[...] L'intuition du vérifacteur, telle que je la comprends, empêche le métaphysicien de laisser les dispositions flotter dans l'air, telles qu'elles le font dans la philosophie de l'esprit de Ryle³⁹.

Le principe du vérifacteur implique donc que les dispositions doivent être fondées. Si la disposition ne désigne rien d'autre qu'un billet d'inférence, que l'on ne peut formuler que sur la base de ses manifestations précédentes, il n'y a rien dans le monde qui rend vrai un

³⁸ Armstrong, 1989, p. 88.

³⁹ « Truths and Truthmakers », à paraître.

énoncé dispositionnel tant qu'aucune manifestation épisodique n'a lieu. Armstrong soutient au contraire qu'à l'énoncé d'une disposition non manifestée doit correspondre une différence dans le monde, qui le rend vrai. En résumé, la première thèse d'Armstrong soutient que les énoncés dispositionnels réfèrent à une *base* persistante qui sous-tend les différentes manifestations de la disposition. Il y a dans le monde, en plus des manifestations de la croyance qu'il pleut (prendre son parapluie, allumer les essuie-glaces...) un état qui fonde toutes ces dispositions (un état interne, cérébral...). Avant de passer à la seconde thèse d'Armstrong, il est intéressant de noter pour finir que Brentano a soutenu une position tout à fait voisine de celle d'Armstrong sur ce point. Critiquant Bain, qui soutient déjà une forme de béhaviorisme dispositionnel selon lequel ce qui distingue les facultés mentales (en l'occurrence présenter et juger) est leurs conséquences comportementales, Brentano fait valoir que ces comportements doivent bien avoir une *base* interne :

Bain a l'idée malheureuse de chercher la différence entre présenter et juger non dans les activités de penser elles-mêmes, mais dans leurs conséquences. Parce que nous laissons quelque chose que nous jugeons vrai influencer nos actions d'une manière spéciale, contrairement à quelque chose qui nous est seulement présenté, il pense que la différence entre tenir quelque chose pour vrai et la simple présentation consiste simplement dans cette influence sur la volonté. [...] J'ai dit que cette théorie était malheureuse. Et, de fait, comment se fait-il qu'une présentation d'un objet ait une telle influence sur le comportement pendant que l'autre ne l'a pas. Soulever la question suffit à montrer très clairement la négligence dont Bain est coupable. *Il n'y aurait pas de conséquences spéciales, si celles-ci n'avaient pas de bases spéciales dans la nature du processus de penser.* Plutôt que de rendre futile l'hypothèse d'une différence intrinsèque entre une simple présentation et un jugement, la différence dans les conséquences fait ressortir de façon manifeste cette différence intrinsèque⁴⁰.

(ii) La seconde thèse défendue par Armstrong concerne la nature ontologique de ces bases dispositionnelles. Celles-ci ne sont selon lui pas elles-mêmes des dispositions, mais des faits catégoriques. Non seulement les dispositions ont des bases, mais ces bases sont catégoriques. Qu'est ce que cela signifie ? Cette seconde thèse d'Armstrong n'est pas tant dirigée à l'encontre de Ryle qu'à l'encontre de ceux qui seraient tentés par une « ontologisation » trop littérale des dispositions, qui conduirait à admettre qu'il existe réellement des pouvoirs dans la nature. Armstrong tente de trouver une voie médiane entre la position de Ryle, qui soutient que les énoncés dispositionnels ne désignent rien, et la position selon laquelle ils désignent littéralement. Cette dernière position est soutenue par ceux qui, à la suite de Platon, pensent que le monde est un ensemble de pouvoirs. Ainsi l'Etranger

⁴⁰ *Psychology from an empirical standpoint*, p. 202-203, nous soulignons.

affirme-t-il dans le *Sophiste* :

Je dis que ce qui possède une puissance, quelle qu'elle soit, soit d'agir sur n'importe quelle autre chose naturelle, soit de pâtir –même dans un degré minime, par l'action de l'agent le plus faible et même si cela n'arrive qu'une seule fois– tout cela, je dis, existe réellement. Et, par conséquent, je pose comme définition qui définit les êtres que ceux-ci ne sont autre chose que puissance⁴¹.

Armstrong pense au contraire que le monde est composé de particuliers et de propriétés qui existent indépendamment de leurs différents pouvoirs. Ces pouvoirs ne leur sont pas essentiels. Ainsi Armstrong distingue-t-il entre le *catégoricalisme* et le *dispositionnalisme*, se revendiquant du premier⁴². Le catégoricalisme est la thèse selon laquelle les propriétés ont leur nature propre, indépendamment de leurs pouvoirs. Le dispositionnalisme est la thèse contraire selon laquelle la nature des propriétés réside exclusivement dans leurs pouvoirs⁴³. Il est important de bien séparer les deux thèses d'Armstrong. La première, selon laquelle les dispositions à se comporter ont une base ontologique, est aujourd'hui largement admise. Elle représente, pour la philosophie de l'esprit, un tournant important : la sortie d'une forme ryléenne ou wittgensteinienne de béhaviorisme. La seconde thèse d'Armstrong en revanche est plus controversée. Si de nombreux philosophes s'accordent à reconnaître qu'une disposition doit avoir une base, certains n'acceptent pas que cette base soit catégorique et soutiennent qu'une disposition peut être fondée une autre disposition plus fondamentale etc. La conjonction des deux thèses d'Armstrong le conduit à ce qu'il appelle une théorie réaliste des dispositions.

Selon l'approche réaliste [des dispositions] parler d'un objet qui a une propriété dispositionnelle implique que l'objet soit dans un état non dispositionnel ou qu'il ait quelque propriété (qu'il existe une 'base catégorique') qui soit responsable du fait qu'il manifeste un certain comportement dans certaines circonstances [...]⁴⁴.

La thèse la plus décisive au regard du problème corps-esprit demeure la première, c'est-à-dire l'admission de bases dispositionnelles. Tout d'abord parce que c'est elle qui

⁴¹ 247 d-e.

⁴² *A World of States of Affairs*, p. 69.

⁴³ Parmi les philosophes qui ont récemment soutenu une thèse dispositionnalisme, on trouve notamment .K.R. Popper, H. Mellor, S. Shoemaker, S. Mumford, B. Ellis ou G. Molnar. Pour une synthèse engagée, voir C. Tiercelin « Sur la réalité des propriétés dispositionnelles », qui tente de trouver une voie médiane entre le catégoricalisme et un dispositionnalisme fort en imposant aux dispositions de figurer dans une loi de la nature (suivant sur ce point Kistler). D'autres philosophes enfin, tels C.B. Martin, refusent de choisir entre dispositionnalisme et catégoricalisme en soutenant que toute propriété a deux « faces », l'une dispositionnelle, l'autre catégorique.

⁴⁴ *A Materialist theory of the mind*, p. 86.

permet de sortir des trois impasses auxquelles conduisait la réduction sémantique des dispositions à des énoncés conditionnels. On peut ainsi comprendre, premièrement, comment il est possible de posséder une disposition sans la manifester : si le verre est fragile, c'est du fait de sa structure intrinsèque (le type de verre qui le compose, l'épaisseur des parois, leur forme, etc.). Cette base dispositionnelle est présente même si le verre ne se brise pas. De même, lorsque je réalise soudain que les ours sont dangereux, sans manifester nullement cette nouvelle croyance, ce qui s'est passé est un évènement interne, que les matérialistes identifieront à un processus cérébral, et que les dualistes définiront comme une modification de mon âme. Enfin, nous avons une bonne raison de penser que le klaxon qui a sonné jusque-là continuera de le faire : c'est parce que la base de sa disposition à sonner, les circuits et éléments qui le composent, perdure. Armstrong insiste sur ce dernier point. Supposons, écrit-il, qu'une bande élastique s'accroisse de 1 cm à chaque fois qu'on exerce sur elle une force F. On peut lui attribuer la disposition de s'accroître de 1 cm à chaque fois qu'on lui applique la force F. Le point essentiel concernant les dispositions, rappelle Armstrong, est que l'on peut les attribuer même lorsqu'elles ne se manifestent pas. Je peux dire : si la bande était soumise à la force F à l'instant T, elle s'étendrait de 1 cm. La question est de savoir ce qui justifie cet énoncé. Qui admet des bases dispositionnelles a une réponse : il a toutes les raisons de croire que la base dispositionnelle de la bande qui est responsable du fait qu'elle s'étire de 1 cm quand on lui applique la force F existe en T. Puisqu'elle existe en T, c'est là une question de nécessité physique, la bande doit s'étirer de 1 cm si on lui applique la force F. Mais quelle sera la réponse du phénoménaliste (de Ryle) ? La seule raison qu'il pourrait donner pour dire que la bande se serait étirée de 1 cm en T si on lui avait appliqué la force F est que la même bande numérique s'est comportée de la même manière en d'autres occasions. Soit, mais quelles raisons a-t-il de croire que ces propriétés dispositionnelles n'ont pas changé, si ce n'est qu'elles sont fondées sur des bases perdurantes ? Ainsi le phénoménaliste ne peut nullement être fondé à attribuer des dispositions qui ne sont pas manifestées :

Le phénoménaliste concernant les dispositions sera réduit à un scepticisme total concernant les dispositions, excepté dans les occasions où elles sont actuellement manifestées⁴⁵.

L'autre raison pour laquelle l'admission de bases dispositionnelles est la thèse la plus importante ici, est que c'est elle qui permet de restaurer la pertinence du problème de la

⁴⁵ *A Materialist Theory of the Mind*, p. 87.

corrélation. En effet, si l'on admet que l'esprit doit être qualifié par des concepts dispositionnels (comme la croyance), la question de la corrélation du corps et de l'esprit n'est dissolue que pour peu que nous admettions comme Ryle que les dispositions n'ont pas de base, que les termes dispositionnels ne réfèrent pas à des faits. Mais une fois réintroduites les bases dispositionnelles, la question de la corrélation de ces bases et du corps ressurgit : la base interne qui apparaît au moment de l'acquisition d'une croyance, éventuellement non manifestée, a des chances d'être corrélée à des événements cérébraux. En résumé, admettre que les concepts mentaux sont des dispositions ne nous engage pas à soutenir qu'ils sont réductibles à des comportements conditionnels⁴⁶. L'esprit est réhabilité comme fait, et l'on est dès lors autorisé à se demander pourquoi il est si fortement corrélé au corps.

2.2.2. *La dissolution humienne*

Pour le béhaviorisme, le problème de la corrélation est considéré comme une question mal posée pour la raison qu'un des termes de cette corrélation, l'esprit, ne désigne rien à proprement parler. Si Hume en vient à considérer que le problème de la corrélation est mal posé, c'est pour une tout autre raison : ce n'est pas l'esprit qui pose problème mais plutôt la *relation* de l'esprit au corps. La façon la plus intuitive d'expliquer la corrélation des phénomènes physiques et mentaux est en effet de supposer qu'ils sont reliés par une relation de causalité : la flamme cause notre douleur, et notre volonté cause le mouvement de notre bras. Ce n'est donc pas à l'esprit que s'attaque Hume pour dissoudre le problème de la corrélation, mais à la causalité.

Il nous faut pour commencer, rappeler brièvement l'analyse que Hume propose de la causalité dans la section de *l'Enquête sur l'entendement humain* consacrée à l'idée de connexion nécessaire⁴⁷. Cette analyse procède en trois temps. Dans un premier temps, Hume établit que si l'on veut connaître l'idée de pouvoir ou de connexion nécessaire, il faut examiner son impression sensible. Cela découle directement de la position empiriste de Hume, selon laquelle toute idée est la copie d'une impression. Dans un deuxième temps, Hume se demande si dans des cas isolés, nous pouvons avoir l'impression d'une connexion nécessaire. Sa réponse est négative : nous ne sentons jamais qu'une succession d'événements,

⁴⁶ Armstrong maintient que les concepts mentaux sont dispositionnels. Mais l'on peut également soutenir que la croyance n'est pas plus une disposition que le levé du soleil. Pour un rejet de l'idée que les concepts seraient des capacités, voir J. Fodor, *Concepts*.

⁴⁷ Section VII, « L'idée de connexion nécessaire », p. 125-146.

mais en aucun cas le pouvoir ou la nécessité qui les relie, qu'il s'agisse des objets extérieurs ou des opérations de notre esprit. Aussi, si l'on objecte que nous sentons l'influence de la volonté sur le corps, Hume répond que nous n'avons conscience que d'idées consécutives : celle de la volonté, puis celle du mouvement de notre corps, mais que nous n'avons aucune idée du pouvoir de l'âme sur le corps. Dans un troisième temps, Hume restaure notre idée de la causalité : celle-ci n'est pas l'idée pouvoir, mais l'habitude que nous prenons d'anticiper ou d'inférer un événement à partir d'un autre, suite à la consécution répétée de ces deux événements dans l'expérience. La causalité n'est pas un fait supplémentaire, elle se réduit à la conjonction constante de certains événements ou objets (le vocabulaire de Hume fluctue sur ce point)⁴⁸. Le sentiment que nous pouvons en avoir ne peut dès lors être le produit d'un cas isolé, mais seulement d'une pluralité de cas : la première fois que nous voyons deux boules de billards s'entrechoquer, nous ne nous attendons à rien. Nous ne voyons qu'une conjonction, et nullement une connexion. Mais si cette expérience est répétée, la conjonction du mouvement de la première boule et du mouvement de la seconde après le choc s'avérera constante, et nous tendrons à inférer qu'il en ira de même les fois suivantes. Hume conclut :

Nous pouvons donc conformément à cette expérience [de la transition coutumière], former une autre définition de la cause et l'appeler un objet suivi d'un autre et dont l'apparition conduit toujours la pensée à l'idée de cet autre objet⁴⁹.

Hume fournit un résumé assez clair des trois temps de son raisonnement (que nous numérotions entre crochets) :

Récapitulons donc les raisonnements de cette section : [1] toute idée est copiée d'une impression d'un sentiment qui la précède ; si nous ne pouvons trouver d'impression, nous pouvons être sûrs qu'il n'y a pas d'idée. [2] Dans tous les cas isolés d'opération des corps ou des esprits il n'y a rien qui produise une impression, ni, par suite, qui puisse suggérer une idée de pouvoir ou de connexion nécessaire. [3] Mais quand beaucoup de cas semblables se présentent et que le même objet est toujours suivi du même événement, nous commençons alors à concevoir la notion de cause et de connexion. Nous sentons alors un nouveau sentiment, une nouvelle impression, à savoir une connexion coutumière dans la pensée ou l'imagination entre un objet et l'événement qui l'accompagne habituellement ; ce sentiment est l'original de l'idée que nous recherchons⁵⁰.

La conséquence d'une telle approche est qu'une corrélation, ou une conjonction

⁴⁸ On voit ici combien Ryle est héritier de Hume lorsqu'il écrit : « La vieille erreur consistant à traiter le terme 'Force' comme dénotant une influence (*agency*) occulte exerçant une force a été abandonné dans les sciences physiques, mais ses corollaires survivent dans de nombreuses théories de l'esprit. (cité par Mumford, p. 27)

⁴⁹ P. 144.

⁵⁰ P. 145-146.

constante de deux séries d'événements ne requiert aucune explication supplémentaire. La causalité n'est rien d'autre que l'observation d'une uniformité, le sentiment d'une *conjonction constante*, l'habitude d'une corrélation entre des événements ou objets. L'implication immédiate, au regard du problème de la corrélation du corps et de l'esprit, est qu'une fois établie cette corrélation, il n'y a plus rien à en dire : la relation causale du corps et de l'esprit *consiste* en cette corrélation et il est absurde de se demander comment une substance spirituelle pourrait influencer sur une substance matérielle⁵¹. Il n'y a pas d'influence des phénomènes mentaux sur les phénomènes physiques, mais seulement une conjonction ; cela suffit pour que l'on soit en mesure de parler de causalité entre l'âme et le corps. Hume souligne que la causalité corps esprit n'est ni plus ni moins problématique que la causalité corps-corps :

En effet, bien qu'aucune connexion n'apparaisse entre le mouvement et la pensée, il en est de même de toutes les autres relations de causalité. [...]

Il semble qu'il ne nous reste, dans le cas présent, que cette alternative : soit nous affirmons que rien ne peut être la cause d'autre chose [...], soit nous soutenons que tous les objets que nous trouvons constamment associés doivent, pour cette raison, être considérés comme causes et effets⁵².

C'est bien entendu la seconde branche de l'alternative que retient Hume. Or, puisque nous constatons des conjonctions constantes non seulement entre les corps, mais aussi entre le corps et l'esprit (« chacun peut percevoir que les différentes dispositions de son corps changent ses pensées et ses sentiments »), il faut admettre que le corps et l'esprit sont liés par des relations de causalité. Il faut préciser enfin, Hume y insiste, que la conjonction entre le corps et l'esprit est de nature temporelle et non spatiale : l'esprit n'étant pas étendu (tout comme les idées de l'odorat, de l'ouïe et du goût⁵³), ce n'est pas la proximité dans l'espace, mais la proximité dans le temps qui permet d'établir une corrélation entre phénomènes physiques et phénomènes corporels.

Dans la mesure où la position de Ryle repose implicitement sur l'analyse humienne de la causalité, cette dernière peut à bon droit être considérée comme la tentative la plus fondamentale de dissoudre le problème de la corrélation. Le principal problème que rencontre

⁵¹ E. J. Lowe (1996) défend aujourd'hui une position de ce type.

⁵² P. 337-338.

⁵³ Cf. 1.2.

l'analyse humienne de la causalité est qu'elle conduit à admettre *trop* de relations causales⁵⁴. Deux objections complémentaires peuvent être soulevées dans cette optique. Selon la première, elle ne permet pas de distinguer les régularités accidentelles des régularités nomiques. Ainsi, comme nous l'avons remarqué en première partie (1.1.4.), le fait qu'il y ait une corrélation systématique entre les propriétés « être une boule en or » et « avoir une masse inférieure à cent tonnes » ne semble pas désigner une véritable loi de la nature, contrairement à la corrélation de même type qui existe entre le fait d'être une sphère d'uranium enrichi et celui de peser moins de cent tonnes (dans ce cas, la masse critique de l'uranium est dépassée : contrairement à la sphère d'or, une telle sphère est impossible). Deuxièmement, outre le fait que la théorie humienne de la causalité échoue à distinguer les régularités accidentelles des régularités non accidentelles, elle échoue encore à distinguer, au sein même des régularités non accidentelles celles qui sont causales de celles qui ne le sont pas. Il semble pourtant exister des corrélations qui ne sont ni accidentelles ni causales. Autrement dit, si la conjonction *temporelle* constante suffisait à la causalité, il y aurait beaucoup plus de relations causales que ce que nous serions prêts à admettre. Imaginez que des milliers d'auditeurs répartis sur tout le territoire français écoutent une même émission de radio. Chacun des récepteurs radiophoniques cause les sensations sonores des auditeurs situés dans son voisinage. Cela ne pose pas problème. Mais à suivre Hume, il y a beaucoup d'autres relations causales qui entrent en jeu ici. Un récepteur situé à Paris peut être considéré comme causant les sensations sonores d'un auditeur bordelais (car seule la conjonction temporelle entre en jeu). En outre, la théorie de Hume ne semble pas permettre de faire la différence entre les événements corrélés parce qu'ils relèvent d'une même cause et les événements corrélés parce qu'ils se causent entre eux : ainsi, il faut aussi admettre que tous les récepteurs, dans la mesure où leurs émissions sont conjointes, entrent en relation de causalité. Et il faut également admettre, que les sensations sonores de l'auditeur parisien et celles de l'auditeur bordelais se causent entre elles⁵⁵. Autrement dit, puisque la corrélation entre le corps et l'esprit n'est pas spatiale mais temporelle pour Hume, il est difficile de comprendre pourquoi *cet* esprit serait corrélé avec *ce* corps. Le deuxième problème posé à la théorie humienne de la causalité est donc qu'elle ne paraît pas capable de sélectionner, parmi un ensemble de

⁵⁴ Elle rencontre également le problème opposé : dans la mesure où elle exclut la causalité singulière, certains peuvent penser que d'un autre côté qu'elle n'admet *pas assez* de relations causales.

⁵⁵ Un partisan de l'approche humienne de la causalité pourrait rejeter ces contre-exemples pour la raison qu'ils impliquent une simultanéité de la cause et de l'effet, alors que Hume souligne qu'il doit y avoir une succession. Il n'est pas difficile cependant d'écartier une telle réponse : considérons simplement que l'émission de radio est retransmise depuis Paris, et que les temps de transmissions produisent un léger décalage entre les différentes réceptions sur le territoire.

corrélations non accidentelles celles qui sont authentiquement causales et celles qui ne le sont pas. Ainsi devrait-on dire que le jour cause la nuit et la nuit le jour. La théorie humienne de la causalité semble donc trop faible : une relation causale est plus qu'une conjonction constante. Tout le problème est que si l'on renforce la définition de la causalité, afin de soutenir qu'elle est par exemple une relation réelle dans la nature, on est immédiatement ramené au problème de la corrélation : comment le corps et l'esprit peuvent-ils interagir s'ils sont de nature différente ?

Ni la tentative béhavioriste ni la tentative humienne de dissolution du problème de la corrélation ne sont donc incontestables. Puisqu'il n'est pas certain que le problème puisse être dissolu, il est légitime de se demander s'il pourrait être résolu. Comment expliquer la corrélation entre le corps et l'esprit sans renoncer à leur différence ?

2.3. Le problème résolu

Il y a au moins quatre façons d'expliquer une corrélation. La première consiste à dire qu'il y a une relation causale entre les éléments corrélés. Cette relation causale peut être directe (Descartes), ou indirecte (médiatisée par une intervention divine comme chez Malebranche), elle peut valoir dans les deux sens (interactionisme) ou dans un seul (épiphénoménalisme⁵⁶). La deuxième est de dire que les séries corrélées, bien que n'entrant pas en relation causale, procèdent d'une même cause commune (l'harmonie préétablie de Leibniz par exemple). La troisième est d'identifier, purement et simplement, les deux éléments corrélés, ou d'identifier un élément à une partie de l'autre. Concernant le problème du corps et de l'esprit, cela aboutit soit à un monisme matérialiste (identification de l'esprit au corps), soit à un monisme spiritualiste (identification du corps à l'esprit), soit à un monisme neutre (identification du corps et de l'esprit, sans engagement sur leur nature finale). La quatrième est de renoncer à une telle identification tout en maintenant qu'il demeure néanmoins une *dépendance* d'un des éléments à l'égard de l'autre, voir de chacun vis-à-vis de l'autre⁵⁷. Ainsi les corrélations entre le trou et le fromage peuvent-elles être expliquées par le

⁵⁶ L'épiphénoménalisme est généralement entendu comme la thèse selon laquelle le corps exerce une influence causale sur l'esprit, mais nous l'inverse. L'esprit est alors causalement impotent. Rien n'interdit cependant de concevoir un épiphénoménalisme contraire selon lequel l'esprit exerce un pouvoir sur le corps, mais non l'inverse.

⁵⁷ La notion de *dépendance existentielle* (ou de *fondation*) est introduite par Husserl dans les *Recherches Logiques*. Il convient d'introduire une distinction entre la fondation forte et la fondation faible (Smith, Simons, Mulligan, 1984 ; Simons 1987 ; Simons, 1994). A est faiblement fondé sur B, si et seulement si A ne peut

fait que le trou dépend du fromage⁵⁸. On ne peut dire en effet que le trou est identique au fromage, ni qu'il est causé par lui, ni qu'il le cause. Appliquée au corps et à l'esprit, cette relation de dépendance sans réduction a conduit aux notions d'émergence et, plus récemment, de survenance.

Maintenant, si l'on considère d'un peu plus près ces quatre types d'explications de la corrélation du corps et de l'esprit, on constate qu'il est à première vue difficile d'avoir à la fois une explication informative de cette corrélation sans renoncer, ou du moins atténuer, leur différence de nature. Les deux premiers types d'explication sont de nature causale. Or il semble que des relations causales ne peuvent valoir qu'entre des termes *homogènes*. On ne peut dire par exemple que le nombre 8 a causé une avalanche. Par transitivité, cette homogénéité s'applique à tous les termes d'un réseau causal. Il suit que selon les deux premiers types d'explications, le corps et l'esprit devraient être homogènes, dans un sens qui reste à déterminer, mais qui fait craindre que l'on ne doive renoncer à leur différence de nature. Il est encore plus évident que le troisième type d'explication, l'identification, conduit à renoncer à cette différence⁵⁹. La quatrième solution explique la corrélation du corps et de l'esprit par le fait que l'esprit dépend du corps. Mais une telle réponse, si elle était tenable, serait-elle elle vraiment explicative ? Nous préservons certes ainsi une distinction importante, mais une fois la sentence prononcée, « l'esprit dépend du corps », nous ne sommes pas forcément plus avancés. Dans les trois premiers cas, nous ne parvenons (à première vue) à donner une explication de la corrélation du corps et de l'esprit qu'en renonçant à leur différence de nature. Dans le dernier, nous réussissons à sauver cette différence de nature, mais nous n'avons alors plus de véritable explication de leur corrélation. Nous envisagerons ici surtout la troisième et la quatrième solutions : peut-on soutenir qu'il y a une identité entre le corps et l'esprit ? Sinon, en quel sens peut-on dire que l'un dépend de l'autre ?⁶⁰ La

exister sans que B existe. Dans ce cas, la relation de dépendance faible inclut aussi la relation du tout à ses parties essentielles : un tout (l'homme) ne peut exister sans ses parties essentielles (son cœur ou son cerveau par exemple). Nous voulons ici exclure ce cas car il nous ramène directement à la troisième solution : l'explication de la corrélation par *identification* d'un des phénomènes corrélés à une partie de l'autre. Il est alors utile d'introduire une relation de dépendance forte : A est fortement fondé sur B si et seulement si A est faiblement fondé sur B et que B n'est pas une partie de A. Par exemple, il y a une dépendance mutuelle forte entre la couleur et l'étendue.

⁵⁸ La thèse selon laquelle les trous sont des particuliers dépendants est défendue par Achille C. Varzi et Roberto Casati, *Holes and other superficialities*, MIT Press, 1994.

⁵⁹ Nous verrons cependant que les partisans de la thèse de l'identité peuvent maintenir que les termes « corps » et « esprit », bien qu'ils aient la même référence, n'ont pas le même sens.

⁶⁰ Les autres solutions évoquées, le dualisme de Descartes, l'occasionalisme de Malebranche, l'harmonie préétablie de Leibniz ou la théorie du double aspect de Spinoza ne seront pas développées ici. La principale raison en est que ces positions sont en général plus connues et que les textes sont accessibles en français. Il existe

première position est logiquement appelée théorie de l'identité du corps et de l'esprit. Les difficultés qu'elle a rencontrées ont donné naissance à une seconde position, le fonctionnalisme. Certains la considèrent comme une sophistication de la théorie de l'identité, d'autres comme une nouvelle théorie incompatible, qui repose sur la relation de survenance et non plus d'identité. Ces derniers ont notamment tenté de rapprocher le fonctionnalisme de l'hylémorphisme défendu par Aristote.

2.3.1. La théorie de l'identité

2.3.1.1. Présentation

La façon la plus directe d'expliquer une corrélation est l'identification des termes corrélés. En ce qui concerne la corrélation du corps et de l'esprit, nous disposons aujourd'hui de données empiriques qui nous donnent de bonnes raisons de penser que si l'esprit doit être identique à une partie du corps, celle-ci sera probablement le cerveau (plutôt que le cœur ou le foie). Ainsi Fechner faisait-il l'hypothèse que la corrélation corps/esprit n'était parfaite que dans la mesure où l'on considérait la corrélation cerveau/esprit (la corrélation stimulus/esprit n'étant elle qu'approximative)⁶¹. Edwin Boring, dont on a vu qu'il proposait une réponse à l'objection de la quantité que James et Bergson adressaient à Fechner⁶², a été le premier à formuler au vingtième siècle une théorie de l'identité du corps et de l'esprit :

Selon l'auteur [Boring lui-même] une corrélation parfaite est identité. Deux événements qui se produisent toujours ensemble à la même place au même moment, sans aucune différence spatiale ou temporelle du tout, ne sont pas deux événements mais le même événement. Les corrélations du corps et l'esprit, telles qu'elles sont formulées à présent, ne prennent pas en considération la corrélation spatiale et réduisent donc la question à une simple corrélation dans le temps. Le besoin d'identification n'est pas moins urgent dans ce cas⁶³.

Cette première formulation de la théorie de l'identité passa cependant relativement inaperçue, du fait, selon Place de raisons sociologiques (Boring était psychologue, peu connu des philosophes) ainsi que de raisons historiques (les thèses de Frege concernant le sens et la

en outre de très bons commentaires qu'il serait inutile de paraphraser. Sur le dualisme de Descartes, il est notamment très utile de lire les chapitre XII, XIII et XIV de l'ouvrage de Henri Gouhier, *La pensée métaphysique de Descartes*.

⁶¹ Voir 2.1..2.

⁶² Ibid.

⁶³ *The Physical Dimensions of Consciousness*, 1933, cité par Ullin Place, « Identity Theories ».

référence qui seules permettent, toujours selon Place, de formuler une théorie viable de l'identité étaient encore en 1933 réservées à un cercle d'initiés viennois)⁶⁴. C'est seulement une vingtaine d'années plus tard que l'hypothèse d'une identité entre les états et processus mentaux et cérébraux entra au coeur des débats, à la suite de la publication de trois articles importants : l'article de U. T. Place « Is Consciousness a Brain Process ? » (1956), celui de H. Feigl « The 'Mental' and the 'Physical' » (1958) et celui de J. J. C. Smart « Sensations and Brain Processes » (1959)⁶⁵. Ni Place, ni Feigl ni Smart n'entendent à l'origine appliquer la théorie de l'identité à l'ensemble des phénomènes mentaux. Seuls les phénomènes mentaux non intentionnels tels que les sensations (la douleur notamment) ou la conscience sont concernés par cette identification au départ. Pour le reste, les états intentionnels comme les croyances ou les désirs demeurent pour Place, Feigl et Smart (à cette époque) explicables de façon ryléenne en termes de dispositions⁶⁶. (C'est Armstrong qui rompra avec définitivement avec une telle solution en étendant la théorie de l'identité aux à tous les états mentaux⁶⁷). Tous demeurent donc béhavioristes (ryléens) en ce qui concerne les états intentionnels. Ainsi Place (1956) écrit-il :

Dans le cas de concepts cognitifs tels que « savoir », « croire », « comprendre », « se souvenir », et de concepts volitifs comme « vouloir » et « avoir l'intention de », il est hors de doute, je pense, qu'une analyse en termes de dispositions à se comporter est fondamentalement bonne.

⁶⁴ U.T. Place, « Identity Theories ».

⁶⁵ Ces trois articles sont reproduits dans le recueil édité par D. Chalmers, *Philosophy of Mind*. Pour une très bonne présentation de la théorie de l'identité et de ses difficultés, voir Jaegwon Kim, *Philosophy of Mind*, chap. 3.

⁶⁶ Un certain retournement s'est produit depuis dans la philosophie de l'esprit : on estime plutôt aujourd'hui que la réduction physicaliste des phénomènes mentaux dits non intentionnels, telles les douleurs ou la conscience est plus problématique que la réduction de phénomènes intentionnels comme les croyances. Ainsi Paul Churchland écrivait-il en 1981 « On croyait il y a vingt ans que les émotions, les *qualia* et les "sensations brutes" constituaient les principales pierres d'achoppement du programme matérialiste. Mais la résistance s'est déplacée avec la disparition de ces barrières [Churchland cite ici une série d'article dans la lignée des théories de l'identité]. Désormais [1981], c'est plutôt du domaine de l'intentionnel, donc du domaine des attitudes propositionnelles dont on dit qu'il ne peut être ni réduit au domaine matériel, ni remplacé par celui-ci ». Aujourd'hui (2003), on en est plutôt revenu à l'idée que les *qualia*, les sensations étaient bien l'obstacle majeur à tout programme matérialiste, qui prévalait avant l'apparition des théories de l'identité. Cela est attesté par le fait que la stratégie matérialiste standard consiste aujourd'hui à réduire les *qualia*, c'est-à-dire les propriétés phénoménales de l'expérience, à des propriétés présentées dans l'expérience, c'est-à-dire à des objets intentionnels (voir note 27 de la première partie de ce cours).

⁶⁷ Dans la préface à la réédition de 1993 de *A materialist theory of the mind*, Armstrong écrit : « Une différence claire qu'il y avait entre ma théorie et celle de Place et Smart est que contrairement à eux, j'appliquais la théorie à tous les états, processus et événements mentaux, au lieu de la limiter aux sensations et aux processus conscients ». Smart s'est ensuite rallié à la position d'Armstrong, contrairement à Place.

La théorie de l'identité n'entre en jeu que pour la conscience, les expériences, les sensations ou les images mentales. La motivation qui conduit Place, Feigl et Smart à défendre une telle thèse semble double : d'une part, elle permet, d'un point de vue philosophique, de faire une économie ontologique en évitant de postuler des entités spécifiquement mentales ; d'autre part, elle permet de prendre en compte les progrès des sciences de la nature, en particulier de la physique, et de faire ainsi rentrer le mental dans un cadre physicaliste. Smart (1959) écrit ainsi :

Pourquoi est-ce que je souhaite [identifier les sensations avec les processus du cerveau] ? Principalement en raison du rasoir d'Occam. Il me semble que la science, de plus en plus, nous offre un point de vue d'après lequel les organismes peuvent être considérés comme des mécanismes physico-chimiques ; il semble que même le comportement de l'homme lui-même sera un jour explicable en termes mécaniques. Au regard de la science, il semble effectivement n'y avoir rien d'autre dans le monde que des arrangements de constituants physiques d'une complexité croissante. Partout sauf à un endroit : dans la conscience. Cela signifie que pour donner une description complète de ce qui se passe dans un homme, vous aurez à mentionner non seulement les processus physiques dans ses tissus, ses glandes, son système nerveux et ainsi de suite, mais aussi ses états de conscience : ses sensations visuelles, auditives et tactiles, ses maux et ses douleurs. Qu'ils soient corrélés à des processus du cerveau n'aide en rien car cela revient à dire qu'ils sont quelque chose « en plus et au dessus » [...] Donc les sensations, les états de conscience, semblent être la seule sorte de chose qui demeure en dehors de l'image physicaliste, et pour diverses raisons, je ne peux simplement pas croire qu'il en soit ainsi. Que tout soit explicable dans les termes de la physique [...] sauf l'occurrence de sensations me semble franchement incroyable.

Qui prend au sérieux les sciences de la nature doit donc chercher à identifier les sensations à des entités physiques. Comment s'y prendre ? A première vue, il peut paraître absurde de dire « Les sensations sont des processus cérébraux ». Manifestement, les concepts de sensations et de cerveau sont totalement différents, et ce serait leur faire violence que de réduire l'un à l'autre, de traduire les sensations en termes de concepts physiques. Ce n'est pas le but de la théorie de l'identité : celle-ci ne porte pas sur les concepts, mais sur leurs référents : c'est une identification ontologique et non une identification épistémique qui est cherchée. Smart écrit :

En considérant la théorie de l'identité (dans ses diverses formes) comme une espèce de physicalisme, je dois dire que c'est un physicalisme ontologique et non un physicalisme de traduction. Il serait absurde d'essayer de traduire les phrases

contenant les mots « cerveau » ou « sensation » en des phrases concernant les électrons, les protons etc⁶⁸.

« Sensation » n'est pas traductible en « processus cérébral », même si les deux expressions désignent la même chose, à savoir un processus cérébral. Mais, objectera-t-on, si deux termes réfèrent à la même chose, pourquoi ne peut-on pas traduire l'un dans l'autre, substituer l'un à l'autre ? Si « sensation » n'est qu'un autre nom désignant un processus cérébral, pourquoi maintenir qu'il ne peut être traduit par « processus cérébral » ? Pourquoi y aurait-il plus de différence entre « sensation » et « cerveau » qu'entre « brain » et « cerveau » ? C'est sur ce point que la théorie de l'identité est tributaire de l'analyse frégréenne du sens et de la référence. Dans son article « Sinn und Bedeutung » (« Sens et référence ») de 1892, considéré comme un des actes fondateurs de la philosophie analytique, Gottlob Frege a introduit cette distinction entre sens et référence⁶⁹. Le problème de Frege est précisément de comprendre les propositions d'identité : il y a deux types de propositions d'identité, celles qui sont de la forme $a = a$, qui sont analytiques et *a priori*, et celles qui sont de la forme $a = b$, qui sont synthétiques et *a posteriori*⁷⁰. Ce sont les propositions du second type que Frege cherche à comprendre : en vertu de quoi sont-elles informatives ? C'est pour répondre à cette question qu'il distingue entre sens et référence. Pour introduire cette distinction, Frege utilise un premier exemple. Considérons les trois médianes d'un triangle, a , b , et c , qui joignent ses sommets aux milieux des côtés opposés. On sait que a , b et c se coupent en un même point. Nous avons dès lors quatre façons de désigner ce point à l'aide

⁶⁸ Smart 2000, introduction.

⁶⁹ L'article, paru dans une revue de philosophie critique, est traduit en français dans le volume *Ecrits logiques et philosophiques*. La traduction du terme *Bedeutung* a posé problème. Littéralement, le terme devrait se traduire par « signification » : c'est entre le sens et la signification que distingue Frege. Bien que cette traduction soit sans doute la plus fidèle, elle est aussi trompeuse : nous considérons souvent, en français comme en anglais que ces termes sont synonymes. Pour Frege, la signification d'un mot désigne son référent dans le monde. La traduction « Sens et référence » semble alors la plus appropriée bien qu'elle s'éloigne de l'expression allemande originale. Imbert traduit par « Sens et dénotation », qui est également acceptable, mais qui historiquement peut induire en erreur étant donné l'usage que Russell fera du terme dénotation. Nous retenons donc ici la traduction de « Bedeutung » par « référence ».

⁷⁰ C'est bien sûr à Kant que Frege emprunte la distinction entre ces deux types de propositions (qu'on trouvait déjà chez Hume). L'analyticité est notamment définie chez Kant par le verbe *entfalten*, qui signifie déplier : un jugement est analytique s'il est ampliatif, s'il ne fait que déplier ou décomposer les prédicats qui étaient déjà contenus dans son sujet. Au contraire, un jugement synthétique ajoute au sujet un prédicat qui n'était pas contenu en lui. Une connaissance est *a priori* si et seulement elle est indépendante de l'expérience, et *a posteriori* ou empirique si elle trouve sa source dans l'expérience (sur tous ces points, voir l'introduction à la *Critique de la raison pure*). Pour Kant, la logique relève de l'analytique *a priori*, et l'arithmétique du synthétique *a priori*. Le synthétique *a priori* est au cœur du projet kantien : il permet à Kant de répondre au scepticisme humien en maintenant qu'il existe une connaissance féconde (parce que non analytique) et certaine (parce que non dépendante de l'expérience). En montrant que l'arithmétique est réductible à la logique, et donc que l'arithmétique relève de l'analytique *a priori*, Frege porte donc un premier coup au système de Kant. Mais Frege maintient que la géométrie est synthétique *a priori*. Russell montrera en 1903 qu'elle est également réductible à la logique. Carnap soulignera alors que le synthétique *a priori* a disparu.

des médianes : (i) le point d'intersection de a et de b , (ii) de a et de c , (iii) de b et de c , et (iv) de a , de b , et de c . Ces quatre signes ou expressions ont la même référence, mais ce qui les distingue est qu'ils ont des sens différents : ils dénotent une même chose mais de différentes *manières*.

Il est naturel d'associer à un signe (nom, groupe de mots, caractères), outre ce qu'il désigne et qu'on pourrait appeler sa dénotation, ce que je voudrais appeler le sens du signe, où est contenu le mode de dénotation de l'objet. Pour reprendre l'exemple ci-dessus, la référence des expressions « point d'intersection de a et b » et « point d'intersection de b et c » serait bien la même, mais non leur sens. La référence d'« étoile du soir » et d'« étoile du matin » serait la même, mais leurs sens seraient différents⁷¹.

Le sens d'une expression est donc distinct de sa référence : c'est la manière dont l'expression désigne sa référence, le mode de dénotation de son objet. Les partisans de la théorie de l'identité vont ainsi mettre à profit l'analyse frégréenne. L'énoncé « Les sensations sont des processus cérébraux » doit être conçu sur le modèle de l'énoncé « L'étoile du soir est l'étoile du matin ». L'identification de « L'étoile du soir » à « L'étoile du matin », consiste dans la découverte empirique que leur référence est la même (alors que l'on présupposait auparavant que chacun référerait à un objet distinct). Pour autant, cette identification ne rend pas caduque l'existence des deux expressions distinctes, parce que chacune a un sens différent. De même, ce n'est pas parce que nous identifions les sensations à des processus cérébraux que nous devons renoncer à parler de « sensations », et traduire l'énoncé « J'ai mal aux dents » par l'énoncé « Il y a une activation de fibres-C dans mon cerveau ». « Sensations » et « processus cérébraux » ont des sens distincts mais une référence commune⁷².

Une précision doit être apportée, qui est principalement présente dans la position de Place. Elle est que l'identification que soutient la théorie de l'identité revêt également le caractère d'une *réduction* : cela signifie que contrairement à ce que suggère l'exemple de l'étoile du soir et de l'étoile du matin, les deux termes identifiés, les sensations et les

⁷¹ *Ecrits logiques et philosophiques*, p. 103 (traduction corrigée).

⁷² Le fait de recourir à la notion de sens frégréen pose cependant une difficulté à la théorie de l'identité : elle concerne justement le statut ontologique de ces sens. Frege maintient que les sens, qu'il identifie aux pensées, existent dans un troisième monde, qui n'est ni l'esprit ni le monde physique, parce qu'il refuse la solution psychologue selon laquelle les sens seraient des colorations subjectives que nous donnerions aux références. On a pu en ce sens parler d'un réalisme platonicien chez Frege. Le problème est qu'une telle solution est difficilement compatible avec le physicalisme, que professe les théoriciens de l'identité. D'autres conceptions des sens sont cependant envisageables, qui seraient compatibles avec le physicalisme.

processus cérébraux, ne sont pas de même *niveau*. Les processus cérébraux semble être d'un niveau inférieur à celui des sensations. Ce qu'il faut entendre ici par « niveau » est certes flou : il y a au moins deux façons de comprendre le terme. Soit nous parlons de différents niveaux sur l'échelle microscopique/macrosopique, soit nous parlons de niveaux sur une échelle de type physique/biologie/psychologie. C'est au regard de cette seconde échelle que les sensations sont dire d'un niveau « supérieur ». Le point essentiel est donc que l'on tente d'identifier l'esprit à quelque chose de *plus fondamental* que lui. Cela est très net dans cette citation de Place (1956) :

Les objections logiques qui peuvent être soulevées à l'encontre de l'énoncé « La conscience est un processus dans le cerveau » ne sont pas plus fortes que celles qui peuvent être soulevées à l'encontre de l'énoncé « Un éclair est un mouvement de charges électriques ».

Autre exemple « Les gènes sont des molécules d'ADN », « La température est l'énergie cinétique moléculaire moyenne », « La lumière est la radiation électromagnétique ». Ces exemples ont pour mérite de mettre en évidence l'aspect réductionniste des théories de l'identité : les sensations sont identifiées à des processus conçus comme plus fondamentaux (l'étoile du soir n'est à l'inverse pas plus fondamentale que l'étoile du matin).

Place, Feigl et Smart insistent en outre sur le fait que l'identification des sensations aux processus cérébraux est contingente et empirique, tout comme l'identification de l'étoile du soir à l'étoile du matin. De même que nous avons découvert que l'étoile du soir était l'étoile du matin, tous trois pensent que nous découvrirons que les sensations sont des processus cérébraux. La proposition « Les événements mentaux sont identiques aux événements cérébraux » n'est donc vraie que de façon contingente et empirique. Ce n'est pas une vérité logiquement nécessaire. Elle n'est pas du type « $2+2=4$ » ou « Les célibataires ne sont pas mariés », mais du type « Un éclair est un mouvement de charges électriques » ou « Le gène est une molécule d'ADN ». Les partisans de la théorie de l'identité ne soutiennent pas cependant qu'ils sont actuellement en mesure de dire à quel processus cérébral doit être identifié telle sensation. L'exemple canonique « La douleur est l'activation des fibres-C » ne pas censé être pris à la lettre : la position de Place, Feigl et Smart est seulement de souligner qu'il y a de bonnes raisons d'identifier les sensations aux processus cérébraux et de montrer qu'il n'y a pas d'objection en principe insurmontable à une telle identification. Cela constituerait une explication définitive du problème de la corrélation.

Place et Smart soulignent que contrairement aux sensations, les états intentionnels, comme les croyances, ne peuvent pas être soumis à la réduction identitiste, ce qui implique que pour eux les sensations sont des états non intentionnels. Et l'emploi récurrent de l'exemple de la douleur pourrait laisser penser que dans leur optique, il n'y a pas de distinction entre l'acte et l'objet des sensations⁷³. Une telle lecture serait pourtant erronée : la douleur n'est qu'un exemple de sensations parmi d'autres, dont la perception visuelle d'un rond jaune ou la perception auditive d'un do dièse. Il ne faut pas se laisser abuser par l'usage que ces auteurs font du terme « intentionnalité », qui n'est pas l'usage phénoménal initial de Brentano mais est plutôt le sens linguistique que lui donne Chisholm à la même époque en rapprochant l'intentionnalité de d'intensionnalité (voir 1.1.4.). Ainsi les états intentionnels qu'ils donnent en exemple, comme la croyance, le souvenir ou l'intention, sont tous des états conceptuels. Donc ce n'est pas parce qu'ils soutiennent que les sensations sont des phénomènes mentaux non intentionnels, que les théoriciens de l'identité nient qu'il faille distinguer entre l'acte sentant et l'objet senti.

Au contraire : le fait qu'ils admettent cette distinction apparaît clairement lorsque Smart envisage l'objection que constitue, pour le physicaliste, la possibilité des hallucinations perceptives. Dans ces cas, ou dans le cas des images rémanentes, le sujet a l'impression de percevoir quelque chose qui n'est pas pourtant pas présent dans son environnement physique immédiat. On a souvent appelé « sensation », puis « sense-datum » la chose qu'il a l'impression de voir dans de tels cas, à savoir *l'objet* de sa perception, par exemple un cercle orange. (On retrouve ici l'ambiguïté déjà rencontrée du terme sensation, utilisé tantôt pour désigner l'acte, tantôt pour désigner l'objet de nos expériences.) L'objection poursuit : les sensations étant les objets de nos expériences, Smart prétend identifier ces objets à des processus cérébraux. Or, en vertu du principe d'indiscernabilité des identiques, si x et y sont identiques, alors toute propriété possédée par x est également une propriété possédée par y (sans cela, on devrait autoriser qu'une même chose possède et ne possède pas une propriété)⁷⁴.

⁷³ Voir l'attitude de Reid ou Hamilton à l'égard de la douleur (1.1.1.).

⁷⁴ Le principe d'indiscernabilité des identiques est probablement d'origine aristotélicienne (voir notamment *Réfutations sophistiques* 179a37-39, *Physique* 202b14-16), il ne doit pas être confondu avec le principe d'identité des indiscernables qui lui est d'origine Leibnizienne. Le premier dit que si deux choses sont les mêmes, elles ont toutes leurs propriétés en commun, ce qui est très largement accepté. Le second est plus controversé et dit que si deux choses ont toutes leurs propriétés en commun, alors ce sont les mêmes. Formellement on a : indiscernabilité des identiques (Aristote), $[(x = y) \rightarrow (f)(fx \Leftrightarrow fy)]$; identité des indiscernables (Leibniz), $[(f)(fx \Leftrightarrow fy) \rightarrow (x = y)]$. Sur tous ces points, voir S. Ferret, *L'identité*.

Il s'ensuit qu'il faut admettre que le cerveau possède les propriétés des objets hallucinés. Le problème est que ces objets ont des propriétés qu'il semble inenvisageable d'attribuer au cerveau : Smart devrait admettre qu'il y a des cercles orange dans le cerveau, ce qui est en contradiction avec les connaissances neurobiologiques les plus rudimentaires. Donc la thèse de Smart conduit à des absurdités empiriques : il y a littéralement des odeurs de figues et des roucoulements de tourterelles dans le cerveau.

La réponse de Smart à cette objection est sans ambiguïté : il ne prétend nullement que ce sont les sensations comme *objet*, (c'est-à-dire les *sense data* ou données sensorielles), qui sont identiques à des processus cérébraux. Ce sont les expériences de ces *sense data* qui sont identifiables des événements du cerveau, c'est-à-dire les sensations comme *actes* de l'esprit. Il y a dans le cerveau des *expériences* de cercles orange, mais pas de cercles orange. La théorie de l'identité ne concerne directement que les premières. Demeure un problème cependant : si la théorie de l'identité ne concerne pas directement le statut ontologique des objets de nos perceptions (ou de nos sensations), elle doit néanmoins, dans la mesure où son adoption est motivée par le physicalisme, expliquer en quoi ceux-ci sont compatibles avec une image physique du monde. Si la théorie de l'identité réussit à réduire les actes perceptifs à des processus cérébraux, il lui reste encore à réduire les objets de ces expériences, en cas d'hallucination notamment : dans quelle mesure la tâche orange que nous hallucinons est-elle physique ? C'est là un autre problème, très délicat, auquel nous allons le voir, Smart répond indirectement lors de l'examen d'une autre objection. Ce qu'il faut donc retenir est que Place et Smart entendent appliquer la théorie de l'identité non à l'ensemble des phénomènes mentaux, c'est-à-dire des actes intentionnels, mais seulement aux actes intentionnels non conceptuels qui entrent en jeu dans l'expérience perceptive et l'expérience de la douleur.

Trois objections majeures ont été adressées à la théorie de l'identité : l'objection de la réalisabilité multiple, l'objection des propriétés phénoménales, et celle de la nécessité de l'identité. Nous ne présenterons ici que les deux premières⁷⁵.

2.3.1.2. L'objection de la réalisabilité multiple

⁷⁵ L'objection fondée sur la nécessité de l'identité a été formulée par S. Kripke. Nous renvoyons au troisième chapitre de son livre *La logique des noms propres*. Voir également P. Engel, 1994.

L'objection de la réalisabilité multiple a été formulée par Hilary Putnam en 1967⁷⁶. La théorie de l'identité soutient selon Putnam que la douleur est un processus cérébral d'un certain *type*, disons l'activation de fibres-C. L'objection de la réalisabilité multiple consiste à demander si un être qui serait dépourvu de fibres-C ne pourrait pas cependant avoir des douleurs. Il ne semble a priori pas absurde de penser que l'état mental « douleur » peut être réalisé par des fibres-C chez l'homme, et par un autre type de processus biologique chez le poulpe. Après avoir remarqué que les pieuvres et les humains se situaient sur des branches distinctes de l'évolution, Putnam écrit :

Si donc nous pouvons trouver un seul prédicat psychologique qui peut être clairement appliqué à la fois à un mammifère et à une pieuvre (par exemple « avoir faim »), mais dont le "corrélat" physico-chimique est différent dans les deux cas, la théorie de l'état cérébral s'effondre. Il me paraît plus que probable que nous en soyons capables.

L'objection s'appuie donc sur le fait qu'il n'est pas nécessaire qu'un même type d'état mental soit toujours identique à un même type d'état physique. Un même type d'état mental paraît pouvoir être réalisé de multiples façons, selon les espèces vivantes (et peut-être même les artefacts comme les ordinateurs ou les robots), voir les individus ou les différentes périodes de la vie d'un individu. Autrement dit, il est faux qu'à chaque type (c'est-à-dire genre, ou espèce) d'état mental corresponde un seul et même type d'état physique. Dès lors qu'à un type d'état mental correspond plus d'un type d'état physique, il est impossible d'identifier chaque type d'état mental à un et un seul type d'état physique.

2.3.1.3. L'objection des propriétés phénoménales

L'objection des propriétés phénoménales procède pour sa part en deux étapes : elle souligne dans un premier temps que puisque l'identité entre les sensations et les états du cerveau est contingente, nous devons avoir un moyen d'individualiser une sensation *indépendamment* de ses propriétés neuronales⁷⁷. Dans un deuxième temps, l'objection soutient

⁷⁶ L'article de Putnam s'intitulait originellement « Mental predicates » et a été réédité sous le titre « The Nature of Mental States » (trad. fr. in D. Fisette et P. Poirier, *Philosophie de l'esprit*, 2002).

⁷⁷ La question est bien ici celle de *l'individualisation* et non celle de *l'individuation* des sensations. L'individuation est une notion purement *métaphysique* : le principe d'individuation cherche à exprimer ce qui fait qu'une chose est ce qu'elle est. L'individualisation a en revanche une portée *cognitive*, et ne concerne pas (ou pas directement) la question de la nature de la chose : la question centrale posée par l'individuation est celle de la façon dont *nous* recrutons, identifions puis réidentifions une chose. Vous pouvez vous demander quelle est la nature d'un morceau de basalte, qu'est-ce qui fait qu'il est ce qu'il est : votre problème sera alors celui de

que les seules propriétés d'une sensation qui nous permettent de la connaître indépendamment de connaissances neurobiologiques sont ses propriétés phénoménales qui sont nécessairement mentales et irréductibles aux propriétés du cerveau. Reprenons ces deux étapes.

(i) Que l'identité esprit/cerveau soit contingente et empirique implique qu'avant d'identifier les références de ces deux expressions « sensations » et « processus cérébraux », nous avons été (et demeurons) en mesure d'appliquer l'une et l'autre indépendamment. A cette condition seulement l'identité revendiquée entre sensation et cerveau n'est pas triviale (analytique dans le sens kantien). Par exemple, si on a pu découvrir que les gènes étaient des molécules d'ADN (en gros), c'est parce qu'avant cette identification, nous avions des idées distinctes et indépendantes de ce qu'était les gènes et de ce qu'était l'ADN. Voici ce qu'écrit Armstrong, lorsqu'il présente cette difficulté :

Si « L'esprit est le cerveau » est un énoncé contingent, alors il doit être possible de donner des explications (ou, alternativement des « définitions ostensives ») de la signification des deux mots « esprit » et « cerveau ». [...] « Le gène est la molécule d'ADN » est un énoncé contingent. Nous pouvons expliquer le sens du mot « gène » ainsi : c'est cette chose ou ce principe en nous qui est responsable de la transmission de caractéristiques héréditaires, telles que la couleur des yeux. Nous pouvons expliquer le sens de l'expression « molécule d'ADN » de la façon suivante : c'est une molécule d'une constitution chimique très complexe qui forme le noyau de la cellule. Nous pouvons donner des explications logiquement indépendantes des mots « gène » et « molécule d'ADN ».

Dès lors, s'il y a un sens à dire que « L'esprit est le cerveau » il doit être possible de traiter les mots « esprit » et « cerveau » de la même manière⁷⁸.

Donner une description du cerveau, ou plus précisément du système nerveux central ne semble pas insurmontable. En revanche, un des principaux problèmes que rencontre la théorie de l'identité est de donner un sens au mot « esprit », ou plus précisément « sensation », qui soit compatible avec le physicalisme. Si nous voulons identifier la sensation d'un rond rouge à un processus cérébral, nous devons au préalable avoir un accès à cette sensation indépendamment du processus cérébral. Effectivement, il semble que nous sachions très bien ce qu'est une sensation de rouge, bien que nous ne sachions absolument pas quels

l'individuation. Mais vous pouvez également vous demander comment vous êtes parvenu à recruter ce morceau de basalte, à le distinguer de son environnement, à le réidentifier à travers les différentes expériences visuelles et tactiles que vous avez eu de lui. En ce cas, vous vous intéressez à la question de l'individualisation. L'objection des propriétés phénoménales part du problème de l'individualisation des sensations, la question de leur individuation se réduisant par définition, selon la théorie de l'identité corps-esprit, à celle de l'individuation des processus cérébraux. (Sur la distinction individuation/individualisation, voir Ferret, 1998).

⁷⁸ *A materialist theory of the mind*, p. 77.

mécanismes neuronaux la sous-tendent, ou, en l'occurrence, la constituent. Autrement dit, il doit y avoir quelque chose de caractéristique qui permet à la sensation d'être individualisée. Bien que l'étoile du matin soit identique à l'étoile du soir, l'une a la propriété d'apparaître le matin, l'autre d'apparaître le soir. Dès lors que nous identifions l'une à l'autre, nous réalisons qu'une même chose, Venus, possède les deux propriétés d'apparaître le matin et d'apparaître le soir : ces deux propriétés ne sont ni contradictoires entre elles, ni contraires à une thèse plus générale que nous aurions adopté, tel le physicalisme.

(ii) La seconde étape de l'objection fait valoir que les seules choses que nous pouvons connaître de la sensation, si ce ne sont ses propriétés physiques, sont ses propriétés phénoménales. Nous savons « ce que ça fait » que d'avoir une sensation de rouge⁷⁹. Autrement dit, puisque selon (i), l'individualisation des douleurs doit être tout à fait distincte de celle des fibres-C (ce qui semble plus que plausible), la douleur (ou tout acte perceptif) doit avoir des propriétés distinctes en vertu desquelles nous parvenons à l'identifier, indépendamment de toute connaissance neurobiologique⁸⁰. Smart (2000) formule ainsi l'objection :

On peut se demander : même si les sensations sont identiques aux processus cérébraux, n'y a-t-il pas des propriétés introspectées non physiques des sensations qui ne sont pas identiques aux propriétés des processus cérébraux ?

Tout le problème est que puisque ces propriétés introspectibles en vertu desquelles nous individualisons les douleurs doivent être distinctes des propriétés physiques du cerveau, on ne voit pas ce qu'elles pourraient être sinon des propriétés mentales de nature ontologiques distinctes de celles du cerveau ! Pour répéter : si nous distinguons ordinairement entre une douleur et une démangeaison, ce n'est pas parce que qu'elles ont des propriétés neuronales différentes. Mais pourquoi alors ? Parce qu'elles ont des propriétés phénoménales différentes

⁷⁹ Cette objection a été initialement soulevée par Jérôme Shaffer in « Mental Events and the Brain », *Journal of philosophy* 60 (1963), p. 160-166. Elle est en partie à l'origine des nombreuses discussions autour de la notion de *qualia* ou de propriétés phénoménales de l'expérience qui peuvent être en première approche expliquée par l'expression de Thomas Nagel « l'impression que cela fait de » (voir du rouge par exemple). Cette impression est le *quale* de l'expérience du rouge, qui est donc à distinguer du rouge expérimenté.

⁸⁰ Il faut bien faire la distinction entre cette objection, qui porte sur les propriétés phénoménales des expériences (= des actes), et celle qu'on vient de voir qui porte sur les propriétés phénoménales des objets des expériences. Dans le premier cas, on s'intéresse aux propriétés « être conscient », « être introspectible », « faire un certain effet » ; dans le second, on s'intéresse aux propriétés « être rouge », « être rond », « être acide » etc. Pour une présentation claire de ces deux objections, souvent confondues, parfois même par Smart lui-même, voir Kim, *Philosophy of Mind*, p. 63-68.

soutient l'objecteur. Donc, conclut-il, les sensations ont des propriétés phénoménales irréductibles à des processus cérébraux.

La réponse officielle de Smart à cette objection consiste à introduire la notion de « neutralité thématique » (« *topic neutrality* »). L'idée de Smart est que les propriétés « mentales » que nous attribuons aux sensations ne sont ni physiques, ni mentales. Les énoncés qui font apparemment référence à des propriétés mentales peuvent être paraphrasés par des énoncés thématiquement neutre :

Quand une personne dit « Je vois l'image rémanente d'une orange jaune », il dit quelque chose de ce type : « *Il se passe quelque de similaire à ce qui se passe quand j'ai les yeux ouverts que je suis réveillé, et qu'il y a une orange illuminée sous une bonne lumière en façon de moi* »⁸¹.

Cette réponse paraît étrange : la paraphrase proposée semble répondre plutôt à l'objection des objets hallucinés envisagée plus haut qu'à l'objection des propriétés phénoménales (il faut garder en tête la distinction entre propriétés des objets hallucinés et propriétés introspectibles des actes de perception de ces objets : l'objection des propriétés phénoménales concerne ces dernières). En effet, le problème posé par les images rémanentes est qu'en tant qu'objet halluciné on est tenté de leur attribuer un statut purement mental. La solution proposée par Smart est que tout ce que nous pouvons dire de l'orange hallucinée est qu'elle est similaire à l'orange perçue véritablement (celle qui cause effectivement notre expérience), mais qu'il est impossible de dire si cette similitude est une similitude entre les propriétés physiques ou les propriétés purement phénoménales *de l'orange*. Autrement dit, le physicaliste peut selon Smart faire une place aux objets mentaux (*sense data*) de nos expériences s'il soutient que ceux-ci sont similaires, aux objets physiques qui causent nos expériences véritiques. Il est ainsi possible de caractériser les objets « mentaux » de manière neutre, sans référer à leurs propriétés phénoménales. Il semble donc que la neutralité thématique, officiellement introduite pour expliquer les propriétés de l'expérience n'explique directement que les propriétés de ses objets. L'objet concerne les propriétés de la vue, la réponse porte sur les propriétés de l'objet vu. Cette confusion est très nette dans l'article de Smart « *The Identity theory of Mind* », 2000, où Smart présente la neutralité thématique comme une réponse à l'objection des objets hallucinés :

⁸¹ Smart, 1959.

Ainsi est-il important pour les théoriciens de l'identité de dire (comme ils l'ont fait) que les *sense data* et les images ne font pas partie de l'ameublement du monde. « J'ai un *sense datum* vert » est réellement juste une façon de dire que je vois ou semble voir quelque chose qui est réellement vert ».

[...] Revenons à la question d'avoir un *sense datum* ou une image mentale rayée jaune, pourpre et vert alors qu'il n'y a aucune chose rayée jaune pourpre et vert dans le cerveau. Le théoricien de l'identité peut dire que les *sense data* et les images ne sont pas de vraies choses dans le monde : ils sont comme le plombier moyen. Les phrases qui réfèrent apparemment au plombier moyen peuvent être traduites en, ou élucidées en termes de phrases qui réfèrent aux plombiers. De la même façon, bien qu'il existe « le fait d'avoir un *sense datum* vert », il n'existe pas de *sense data* ou d'images, et le fait d'avoir un *sense datum* vert ou une image n'est pas lui-même vert.

Ces passages manifestent clairement que la neutralité thématique résout le problème des *sense data* et non des actes des expériences. La position de Smart à l'égard des *sense data* est proche de celle de Ryle à l'égard des pouvoirs : ils n'existent pas, les propositions qui les mentionnent ne réfèrent à aucun fait, mais peuvent être traduites en d'autres propositions référentiellement moins suspectes. Il est clair alors les *sense data* sont à l'opposé des sensations (qui sont des actes mentaux) : ces dernières existent selon Smart qui insiste par ailleurs sur le fait que sa théorie de l'identité ne consiste pas à traduire les énoncés référant à des sensations en des énoncés n'y référant plus, mais à identifier ontologiquement les sensations à des processus cérébraux. Au contraire, lorsqu'il est question des *sense data*, la neutralité thématique consiste à éviter la question de leur statut ontologique. La question ne concerne plus les choses, mais le langage, elle devient celle de l'usage sémantique des termes « *sense data* », « données sensorielles », « impressions », « images rémanente » etc⁸². Quand je dis « J'ai l'impression de voir une orange », je ne prétends pas qu'il existe une orange mentale que je vois, je dis seulement « J'ai la même impression que quand je vois véridiquement une orange ». Remarquez que cet argument ne peut valoir que si l'on admet que quand je vois une orange véridiquement, je ne vois pas une idée, une sensation, ou un *sense data* qui représente cette orange : comme le remarque Place dans « Identity Theories », le développement des théories de l'identité doit non seulement à l'analyse frégéenne du sens

⁸² Ce type d'analyse langagière des *sense data* est notamment développé par Wittgenstein dans *Notes sur l'expérience privée et les « sense data »*, ainsi que par Austin dans *Sense and Sensibilia*. Notez le trait commun entre le traitement que Chisholm fait de l'intentionnalité en termes d'intensionnalité, la théorie ryléenne des dispositions et la théorie de Smart concernant les *sense data* : dans tous les cas, le problème est ramené au niveau du langage et l'on minimise, quand on ne la nie pas, sa portée ontologique (cependant pour Chisholm, Place et Smart il ne s'agit absolument pas d'une attitude générale, mais seulement d'une position relative à certaines questions restreintes : aucun ne soutient que toutes les questions ontologiques peuvent être dissoutes par une analyse méticuleuse du langage : cette position est revanche plus proche de celle de Wittgenstein, d'Austin et de Ryle).

et de la référence, mais aussi au rejet des théories réalistes ou phénoménalistes au profit de théories réalistes directes de la perception⁸³.

La neutralité thématique ne répond donc pas directement à l'objection des propriétés phénoménales. Elle permet par contre d'apporter une réponse subtile à la question de la nature des objets de l'expérience, posée par les cas d'illusions et d'hallucination. Smart n'a peut-être pas tort de penser que la neutralité thématique permet de répondre à l'objection des propriétés phénoménales de l'expérience : son erreur est plutôt d'avoir présenté cette neutralité comme une réponse *directe* à l'objection, alors qu'elle ne constitue au mieux qu'une réponse *indirecte*. Voyons comment. L'objection des propriétés phénoménales procède en deux étapes (i) nous sommes capables d'individualiser les sensations (de les distinguer entre elles etc...) indépendamment de toute référence à des propriétés neuronales. (ii) La seule façon que nous avons de les individualiser est alors de recourir à leurs propriétés non neuronales, purement phénoménales. La réponse du partisan de l'identité corps-esprit consiste à admettre (i) mais à rejeter (ii) : nous avons besoin concède-t-il, d'un critère d'individualisation des sensations indépendant de leurs propriétés cérébrales (ce sans quoi la théorie de l'identité serait inintelligible), mais il est faux que le seul critère d'individualisation disponible soit celui des propriétés phénoménales. Nous pouvons individualiser une sensation autrement qu'en faisant appel à l'impression subjective qu'elle nous fait, si nous faisons référence à son *objet*, qui est physique. Ce qui distingue une sensation de rouge d'une sensation de vert n'est pas que l'une a une qualité subjective que l'autre n'a pas, mais qu'elles ont pour objets des propriétés physiques distinctes : le rouge et le vert. Autrement dit, pour reprendre le mot d'ordre de nombreux matérialistes contemporains : *les propriétés phénoménales ne sont pas des propriétés introspectibles des expériences mais des propriétés représentées dans l'expérience*. L'hypothèse du réalisme direct nous assure que les propriétés présentées dans l'expérience sont bien physiques : ce sont les stimuli qui causent nos expériences. Les sensations, actes perceptifs, ou expériences sont donc individuées par leurs causes respectives (leurs objets),

⁸³ Un bref rappel : la théorie réaliste indirecte correspond à ce que nous avons appelé en première partie (1.1.2.) « théorie moderne des idées » (Descartes, surtout Locke, et même Brentano ou l'a vu) : les objets immédiats de la perception sont des entités dépendantes du mental qui représentent (en général par ressemblance) des entités indépendante du mental (c'est-à-dire du monde extérieur). La théorie phénoménaliste, d'origine berkeleyenne, consiste à nier qu'il y ait quoique ce soit « derrière » ce « voile des idées » et que les entités directement senties, dépendantes de l'esprit, constitue l'ameublement ultime du monde. La théorie réaliste directe, défendue par Reid, soutient que les objets immédiats de la perception sont les entités du monde existant indépendamment de l'esprit. Précisons que si le réalisme direct a facilité historiquement l'apparition des théories de l'identité, il ne s'ensuit pas qu'il leur soit nécessairement lié. Il est peut-être possible de soutenir conjointement un réalisme indirect et une théorie de l'identité corps-esprit.

qui sont parfaitement physiques : il est alors faux de penser, comme le font les partisans de l'objection des propriétés phénoménales, que seules des propriétés mentales introspectibles nous permettent de distinguer une sensation d'une autre. Les propriétés physiques qui causent les sensations ont aussi bien l'affaire, et même mieux, dans la mesure où elles sont compatibles avec le physicalisme. L'acte de perception d'une orange n'est pas défini par ses propriétés intrinsèques, mais par le fait qu'il soit causé par une orange.

« Soit, rétorque le partisan de l'objection des propriétés phénoménales au théoricien de l'identité, mais vous n'avez ainsi qu'un moyen d'individualiser la perception *véridique* d'une orange. Si l'orange est hallucinée, vous chercherez en vain dans le monde l'objet physique qui cause cette sensation illusoire ! Le seul moyen vous aurez d'individualiser cette sensation illusoire sera de recourir à ses propriétés phénoménales : vous devrez donc admettre qu'il y a des propriétés des actes mentaux qui ne sont pas les propriétés physiques du cerveau ». C'est ici qu'intervient la *neutralité thématique* : si nous pouvons détecter des sensations illusoires indépendamment de leurs propriétés phénoménales, c'est parce que ces sensations sont similaires à d'autres, qui ont pour leur part des causes physiques définies. Ainsi, la perception illusoire d'une orange est-elle similaire à la perception véridique d'une orange, qui elle est individualisée par sa cause physique : l'orange. Le partisan de l'identité corps-esprit nous enjoint donc de procéder ainsi pour identifier une sensation : cherchez sa cause physique. Si vous n'en trouvez pas, cherchez la cause physique des sensations similaires. En résumé, si on conjoint la neutralité thématique avec la thèse selon laquelle les expériences sont individualisées par leurs objets (qui sont aussi leurs causes), on parvient à individualiser les sensations sans recourir à leurs propriétés neuronales ni à leurs hypothétiques propriétés phénoménales : on s'appuie uniquement sur leurs causes-objets, disons leurs stimuli. Comme le résume David Chalmers : « Etre l'expérience d'une orange c'est en gros être le type d'expérience qui est généralement causé par une orange »⁸⁴. Telle est donc la réponse de Smart à l'objection des propriétés phénoménales : ce qui nous permet d'identifier un état mental préalablement à toute connaissance neurobiologique est le stimulus qui le cause.

2.3.2. *Le fonctionnalisme*

2.3.2.1. Présentation

⁸⁴ *The Conscious Mind*, p. 360, note 13.

Il n'est pas exagéré de dire que le fonctionnalisme est depuis plus de trente ans la position orthodoxe en philosophie de l'esprit. Il le demeure peut-être, bien qu'il fasse aujourd'hui l'objet de critiques grandissantes. Son apparition, en tant que thèse philosophique, est associée à trois noms : David Lewis, David Armstrong et Hilary Putnam. La première formulation du fonctionnalisme se trouve dans un article de Lewis (1966) intitulé « An Argument for the Identity theory »⁸⁵. De façon indépendante, Armstrong parvient dans son livre *A materialist theory of the mind* (1968⁸⁶) à une position très proche. Putnam formule une version différente du fonctionnalisme dans « La nature des états mentaux » (1967), où est présentée l'objection de la réalisabilité multiple. La relation entre la théorie de l'identité et le fonctionnalisme est loin de faire l'unanimité : comme l'indique le titre de son article, Lewis considère le fonctionnalisme comme un prolongement de la théorie de l'identité alors que Putnam le considère comme une nouvelle théorie incompatible avec l'identité des propriétés physiques et mentales. Mais avant d'aborder les relations complexes entre théorie de l'identité et fonctionnalisme, il convient de définir ce dernier. Quelles que soient leurs divergences quant à la théorie de l'identité, tous les philosophes rangés, avec ou sans leur accord, sous la bannière « fonctionnalistes » admettent au moins deux thèses. Premièrement, alors que la théorie de l'identité était restreinte aux seules sensations, le fonctionnalisme prétend s'appliquer à *tous* les états mentaux, y compris les états dits intentionnels tels que les croyances ou les intentions. Deuxièmement, le fonctionnalisme, qu'on le conçoive comme une théorie de l'identité améliorée ou comme une nouvelle théorie, cherche à caractériser les états mentaux à l'aide de leur *rôle causal* (Lewis) ou *fonctionnel* (Putnam). Qu'est-ce que cela signifie ? Le point nodal du fonctionnalisme consiste à caractériser les états mentaux non pas directement à l'aide de leurs propriétés intrinsèques (c'est-à-dire les propriétés qu'une entité possède indépendamment de ses relations avec les autres entités), mais à l'aide de certaines propriétés extrinsèques : les propriétés d'un état mental d'avoir un certain type de cause et d'avoir un certain type d'effet. *Un état mental est donc défini d'abord, indépendamment de sa réalisation biologique, par ce qui le cause et ce qu'il cause.* Il est conçu comme un médiateur causal. Ces causes et ses effets peuvent être de deux types : les causes de l'état mental peuvent être soit des stimuli sensoriels soit d'autres états mentaux, ses effets peuvent être soit des comportements physiques soit d'autres états mentaux. En définitive, pour individualiser un état mental, nous n'avons pas à regarder « à l'intérieur » de lui, de la boîte noire qu'il

⁸⁵ *Journal of Philosophy*, 63, 17-25.

⁸⁶ Voir en particulier le chapitre 6.

constitue : il nous suffit de considérer que quel qu'il soit, il médiatise des relations d'input et d'output. David Lewis formule ainsi cette thèse :

Un état mental M (disons, une expérience) est définissable comme l'occupant d'un certain rôle causal R –c'est-à-dire comme l'état, de quelque sorte qu'il soit, qui est causalement connecté de manière spécifique aux stimuli sensoriels, aux réponses motrices et aux autres états mentaux⁸⁷.

Cela vaut comme une définition minimale du fonctionnalisme : un état mental est défini par son *rôle causal*, il est ce qui est causé par certains stimuli ou états mentaux et ce qui cause certains comportements ou états mentaux. Par exemple, le rôle fonctionnel de la douleur sera d'être causée par un dommage corporel et de provoquer le retrait.

2.3.2.2. Fonctionnalisme et béhaviorisme

Nous avons parlé de stimulus, de comportement, et même de boîte noire : tout ceci n'est pas sans rappeler le béhaviorisme. Sous prétexte de résoudre les difficultés de la théorie de l'identité, le fonctionnalisme ne serait-il finalement qu'un retour au béhaviorisme ? Il n'en est rien. Tout d'abord, le fonctionnalisme considère clairement les états mentaux, ou ce qui les réalise, comme des états *internes* ayant certains pouvoirs causaux. S'il admet des états internes, c'est parce que les pouvoirs causaux à l'aide desquels il définit les états mentaux ne sont pas conçus comme de purs pouvoirs flottant en l'air : ces pouvoirs sont fondés, ont une base dans la structure biologique du cerveau. Le rejet d'une conception ryléenne des dispositions est très net de la part du fonctionnaliste : la base des dispositions causales d'un état mental est l'état physique qui le réalise. L'individualisation d'un état mental à l'aide de ses propriétés causales ne signifie nullement que cet état s'identifie à des dispositions comportementales infondées, au contraire. La seconde raison pour laquelle le fonctionnalisme est incompatible avec le béhaviorisme est qu'il admet des relations causales *entre états mentaux*. Un état perceptif peut causer une croyance, une croyance peut causer un désir etc... Bien entendu, ces deux divergences avec le béhaviorisme sont liées : dès lors que l'on admet que des états mentaux existent réellement, indépendamment des stimuli et des réponses, il leur est possible d'entretenir entre eux des relations causales.

⁸⁷ Lewis, « Psychophysical and theoretical identifications », 1972.

2.3.2.3. Fonctionnalisme et théorie de l'identité : fonctionnalisme « australien » vs fonctionnalisme « non—réductionniste »

Le fonctionnalisme australien

Si l'antinomie entre le fonctionnalisme et le béhaviorisme est claire, les relations entre le fonctionnalisme et la théorie de l'identité sont plus ambiguës. Il y a une différence assez forte entre la façon dont Armstrong et Lewis d'un côté, et Putnam et Fodor de l'autre considèrent cette relation⁸⁸. Alors que ces derniers prétendent proposer une alternative à la théorie de l'identité, Armstrong et Lewis présentent leurs arguments comme précisant, développant et modifiant sur certains points la théorie de l'identité de Smart, qu'ils considèrent comme fondamentalement valide. Nous commençons par présenter cette version du fonctionnalisme, qu'on peut appeler fonctionnalisme « australien » ou « de l'identité ». Aux yeux d'Armstrong et Lewis le fonctionnalisme (terme qu'ils n'emploient pas à l'époque) constitue donc une version améliorée de la théorie de l'identité. Cela apparaît clairement dans ce qu'écrivit Lewis en 1966⁸⁹ :

Mon argument est le suivant : la caractéristique définitive de toute (sorte d') expérience en tant que telle est son rôle causal, ses syndromes de causes et d'effets les plus typiques. Nous autres matérialistes croyons que ces rôles causaux qui appartiennent par nécessité analytique aux expériences appartiennent en fait à certains états physiques. Puisque ces états physiques possèdent le caractère définitif des expériences, ils doivent *être* les expériences.

Lewis considère qu'il ne fait que compléter la théorie de la neutralité thématique de Smart, qui contenait déjà en germe l'idée de rôle causal : « Quand nous décrivons un état mental *M* comme l'occupant d'un rôle *M*, nous faisons ce que Smart appelle une description thématiquement neutre (topic-neutral). Elle ne dit rien de la sorte d'état qui occupe ce rôle. »⁹⁰ Lewis se range explicitement aux côtés de Place, Smart et Armstrong au sein de ce qu'il

⁸⁸ Sur la distinction entre le fonctionnalisme de Lewis et Armstrong d'un côté, et celui de Putnam et Fodor de l'autre, voir Ned Block, « What is fonctionnalism ? », et Jaegwon Kim, *Philosophy of Mind*.

⁸⁹ Nous soulignons.

⁹⁰ « Reduction of Mind », 1994.

appelle le « matérialisme australien »⁹¹. De fait, dans son article récapitulatif « The Identity Theory of Mind », Smart souligne également qu'il n'a jamais considéré le fonctionnalisme d'Armstrong ou de Lewis comme une théorie concurrente de la sienne, mais plutôt comme une heureuse sophistication. La principale difficulté, aux yeux de Smart, Armstrong et Lewis, n'est pas tant la réalisabilité multiple que la première étape de l'objection des propriétés phénoménales : l'individualisation des états mentaux (*id est* : pour identifier les états mentaux aux processus cérébraux, nous devons être capables de les décrire sans faire référence à leurs propriétés biologiques). Et dans une large mesure, leur attitude à son égard est la même : ils admettent qu'un état mental doit pouvoir être individualisé par d'autres propriétés que ses propriétés physiques intrinsèques, mais nient que le seul recours possible soit d'invoquer ses propriétés introspectibles. Tous soutiennent que les états mentaux peuvent être individualisés par leurs propriétés extrinsèques ou relationnelles : leurs pouvoirs causaux⁹². Ce que Lewis et Armstrong ajoutent à la solution de Place et de Smart est « seulement » un élargissement de ce qui compte comme pouvoir causal d'un état mental : il faut prendre en compte non seulement les causes distinctives des états mentaux, mais aussi leurs effets propres ; et les causes des états mentaux ne sont pas seulement les stimuli, mais aussi les autres états mentaux. Cet élargissement du nombre de pouvoirs causaux pris en considération permet à la version fonctionnaliste de la théorie de l'identité d'offrir une réponse plus satisfaisante à l'étape 1 de l'objection des propriétés phénoménales, l'individualisation des états mentaux. Voyons comment dans le détail.

La première raison pour laquelle le fonctionnalisme est moins embarrassé que la théorie initiale de l'identité par l'objection de l'individualisation des états mentaux est donc qu'il prend *aussi* en compte des effets comportementaux de ces états. Armstrong a fortement insisté sur cet avantage. Il est très net qu'à ses yeux, cette objection de l'individualisation est la principale difficulté que rencontre la théorie de l'identité. Armstrong s'accorde avec Smart sur le rejet de la seconde étape de l'objection des propriétés phénoménales : une fois admis que le partisan de l'identité corps-esprit ne pouvait individualiser les états mentaux par leurs propriétés physiques intrinsèques (première étape), il est faux qu'il ne nous reste comme seule solution que d'attribuer aux états mentaux des propriétés mentales (seconde étape). Notre

⁹¹ Ibid. Contrairement à Place, Smart, et Armstrong, Lewis n'était pas australien mais américain. Il a toujours reconnu sa dette envers la philosophie australienne.

⁹² Il y a bien un sens cependant dans lequel les pouvoirs (les dispositions) sont des propriétés intrinsèques, qu'un objet possède indépendamment de tout autre objet. C'est la manifestation de ces pouvoirs qui nécessite l'intervention d'autres entités.

choix n'est pas aussi restreint : outre les propriétés mentales et les propriétés physiques, il existe des propriétés physiques extrinsèques, ou relationnelles. Smart recourt à la propriété *d'être causé par certaines causes typiques*⁹³. C'est sur ce point qu'Armstrong vient compléter la théorie de Smart : il soutient que pour individualiser un état mental sans recourir à ses propriétés physiques intrinsèques, il ne suffit pas de prendre en compte ses causes, il faut également, et surtout, prendre en compte, ses effets :

Les difficultés dans les positions de Place et de Smart me conduisent à chercher du côté de la réponse plutôt que du stimulus pour trouver une description générale des concepts mentaux. Le concept d'un état mental est avant tout le concept d'un état de la personne *capable de causer une certaine sorte de comportement*. Sacrifiant la précision à la brièveté, on peut dire que bien que l'esprit ne soit pas le comportement, il est la cause du comportement. Dans le cas de certains états mentaux seulement, ils sont aussi des états de la personne *aptés à être causés par une certaine sorte de stimulus*. Mais cette formule est secondaire⁹⁴.

Armstrong donne deux arguments pour justifier sa thèse selon laquelle les états mentaux doivent être individualisés par leurs effets comportementaux et non seulement par leurs causes. Le premier est précisément qu'il entend donner une théorie qui vaut pour *tous* les états mentaux, y compris les croyances et les désirs. Tant que l'on en restait à des états perceptifs, la référence aux stimuli pouvait sembler suffisante. Mais il est désespéré de tenter de caractériser une croyance ou un désir uniquement à l'aide de certains stimuli : si l'on veut aussi les identifier à des processus cérébraux, il devient nécessaire de prendre en compte leurs effets comportementaux. Le second argument avancé par Armstrong en faveur de l'introduction de ces effets est que selon lui, même pour les états purement perceptifs comme les sensations, la seule prise en compte du stimulus ne suffit pas : si nous attribuons à autrui une certaine sensation, c'est rarement parce que nous percevons le stimulus de sa sensation, c'est plutôt parce que nous percevons certains traits de son comportement.

La seconde raison pour laquelle le fonctionnalisme est moins embarrassé que la théorie initiale de l'identité face à l'objection de l'individualisation est qu'il peut, pour individualiser les états mentaux faire appel à leurs relations causales avec d'autres états mentaux. Cela lui permet d'admettre des états mentaux qui n'ont ni causes proches ni effets immédiats observables, comme lorsque l'on effectue une addition « de tête » pour reprendre

⁹³ A vrai dire, Smart ne recourt pas aussi explicitement à la notion de causalité, et c'est là un des mérites d'Armstrong que d'avoir formulé le rôle implicite que joue la causalité dans la position de Smart.

⁹⁴ *A Materialist Theory of the Mind*, p. 82.

l'exemple d'Armstrong. Cependant, l'admission d'états mentaux dont les causes et effets ne sont pas directement observables pose un nouveau problème au fonctionnaliste : comment allons-nous caractériser ces rôles causaux internes, et donc, inobservables ? La réponse d'Armstrong est tout à fait voisine de celle que donne Smart à la question de savoir comment nous allons définir l'objet halluciné : l'un et l'autre traitent des objets physiquement douteux tels que les états purement internes et les objets hallucinés en les rapprochant d'objets physiques non douteux, tels que des états ayant des effets comportementaux et des objets physiques véridiquement perçus. Ainsi Armstrong écrit-il :

De fait, il apparaîtra parfois que certaines sortes d'états mentaux peuvent seulement être décrits en termes de *ressemblance* à d'autres états mentaux qui entrent en relations causale avec le comportement. [...]

Une intention d'effectuer une somme « de tête » serait un exemple. L'intention est une cause mentale capable de causer les pensées qui sont les étapes successives du calcul. Donc tout ce qui est demandé est que notre analyse atteigne *ultimement* des états mentaux qui sont descriptibles en termes du comportement qu'ils peuvent causer⁹⁵.

En résumé, les fonctionnalistes « australiens » ou « de l'identité » considèrent la théorie de l'identité de Smart comme valide, mais pensent qu'elle doit renforcer sa réponse à l'objection de l'individualisation des états mentaux. C'est à cette fin qu'ils introduisent les rôles causaux : pour eux, les rôles causaux sont des moyens d'individualiser les états mentaux, mais ne sont pas les états mentaux eux-mêmes. L'état mental est ce qui satisfait ce rôle causal. Notre concept courant de douleur est dans leur optique le concept d'un état qui est causé par des dommages corporels et qui cause des dommages corporels : mais tout ceci n'est pas la douleur, ce sont seulement ses symptômes extérieurs. La douleur, à proprement parler, est un processus cérébral, l'activation des fibres-C.

Le fonctionnalisme non-réductionniste

Alors que les fonctionnalistes de l'identité (Lewis, Armstrong) introduisent avant tout la notion de rôle causal ou fonctionnel pour répondre à la question de l'individualisation des états en vue de leur identification aux états physiques, les fonctionnalistes non-réductionnistes (Putnam, Fodor) introduisent ces rôles afin de répondre à l'objection de la réalisabilité multiple. La réalisabilité multiple d'un même état mental est expliquée par le fait qu'un même

⁹⁵ *A materialist theory of the mind*, p. 83.

rôle causal peut-être occupé par différents états physiques : ainsi le rôle causal de la douleur (être causé par un dommage corporel et provoquer le retrait) pour être *réalisé* par l'activation de fibres-C chez l'homme, et par un tout autre mécanisme biologique chez le poulpe. Puisqu'ils ont à cœur de montrer qu'un même type d'état mental peut avoir différentes réalisations physiques, les fonctionnalistes non-réductionnistes tendent à identifier les états mentaux non pas aux états physiques qui occupent un rôle fonctionnel, mais au rôle fonctionnel lui-même. Ainsi Putnam soutient-il que la douleur est un état fonctionnel et non un état cérébral. C'est parce qu'ils refusent d'identifier purement et simplement les états mentaux à des processus neuronaux que ces fonctionnalistes sont dits « non réductionnistes ». Pour Armstrong et Lewis, le recours au rôle causal n'est qu'une étape préalable à la réduction des états mentaux. Pour Putnam et Fodor, il constitue le point final du fonctionnalisme. Les premiers identifient les états mentaux aux processus neuronaux. Les seconds les identifient à des rôles fonctionnels, qui de façon somme toute annexe, sont réalisés par des processus biologiques. Cette relation de réalisation n'est pas une relation d'identité. Tout état mental est certes réalisé par un état physique, et en ce sens, le fonctionnalisme de Putnam et Fodor demeure un matérialisme (ou un physicalisme, nous ne distinguons pas ici ces deux termes). Il n'y a en effet pas d'états mentaux qui « flottent » immatériellement. Mais d'un autre côté, ils insistent sur le fait que les états mentaux ne sont pas réductibles aux processus qui les réalisent. Ceci pour le versant « non réductionniste » de ce fonctionnalisme.

Cette position découle donc de la réalisation multiple : si nous devons identifier les états mentaux aux processus qui les réalisent, alors ces états mentaux s'éparpilleraient en autant de réalisateurs physiques possibles. Dès lors, dans l'optique de Putnam, si nous voulons lutter contre l'éclatement d'un état mental comme la douleur en de multiples processus physiques, nous devons soutenir que la douleur est ce qu'elle est indépendamment des processus qui la réalise. Cela implique que la douleur soit quelque chose de plus que l'activation des fibres-C, ce qui conduit à une forme de dualisme des propriétés : la douleur n'est pas qu'un processus neuronal. Nous ne serions autorisés à adopter une telle théorie de l'identité que s'il y avait une correspondance biunivoque (terme à terme) entre les types d'états physiques et les types d'états mentaux. Puisque la réalisabilité multiple montre que ce n'est pas le cas, la théorie de l'identité est prise en défaut et il faut admettre des propriétés mentales non identiques aux propriétés physiques.

Cet argument selon lequel la conjonction d'une théorie de l'identité et de la réalisabilité multiple conduit à un éparpillement des phénomènes mentaux est cependant peu développé par Putnam. Est-il si évident pourtant que la réalisabilité multiple soit une objection décisive à la théorie de l'identité ? Pourquoi le tenant de la théorie de l'identité ne peut-il pas proposer de réduire les états mentaux du type « douleur » *tantôt* à des états physiques de type « activation des fibres-C », *tantôt* à des états de type (imaginons) « activation des fibres-Z » ? Nous aurions *des identités multiples et relativisées à des espèces*, voir à des individus : chez certaines espèces, la douleur est identique à un processus physique de tel type, chez d'autres, elle est identique à un processus physique de tel autre type⁹⁶. Autrement dit, la théorie de l'identité, dont le fonctionnalisme d'Armstrong et Lewis, peut répondre qu'une même sorte d'états mentaux peut être identifiée à plusieurs (sortes d') états physiques, c'est-à-dire à une *disjonction* (de sortes) d'états physiques : une douleur est identique à *a*, ou *b*, ou *c*... où *a*, *b* et *c* désignent des processus physiques de genres différents. Putnam semble considérer une telle solution comme manifestement absurde, si bien qu'il attribue d'emblée à la théorie de l'identité la thèse selon laquelle chaque type d'état mental doit être identique à *un et un seul* type d'état physique :

il est vrai que dans ce cas, le partisan de la théorie de l'état cérébral [= la théorie de l'identité] pourrait trouver une échappatoire en recourant à des hypothèses *ad hoc* (par exemple en définissant la disjonction de deux états comme un seul « état physico-chimique »), mais une telle objection ne mérite pas d'être prise au sérieux⁹⁷.

Le problème est que Putnam n'explique nulle part pourquoi ces disjonctions sont si absurdes⁹⁸. La difficulté qu'il semble avoir en tête est celle que l'on vient de mentionner : si les identités entre les sensations et les processus cérébraux sont relativisées (à des espèces vivantes, des individus, des individus à un certain temps...), le risque est d'assister à un éparpillement des phénomènes mentaux. Admettons que la douleur soit identique à la disjonction de phénomènes physiques '*a* ou *b* ou *c*' : on est alors en droit de se demander *ce que ces processus physiques ont en commun* qui font d'eux une douleur. La seule solution

⁹⁶ C'est notamment la réponse que fait David Lewis dans son article de 1980 « Douleur de fou et douleur de martien ». Franck Jackson (1996) adopte une position semblable. Jaegwon Kim va en partie dans ce sens dans son article de 1992. Pourtant, il ne prétend pas défendre une version de la théorie de l'identité, mais une forme de réductionnisme distinct. Son rejet de la théorie de l'identité est explicité dans « L'émergence, les modèles de réduction et le mental », 2000.

⁹⁷ « La nature des états mentaux », 2002 (1967).

⁹⁸ Kim (1992) critique l'étonnante rapidité avec laquelle Putnam rejette une telle hypothèse. Il remarque : « S'il y a quelque chose de profondément faux avec les disjonctions de ce type, cela n'est en tout cas pas évident ».

pour unifier ces différents états physiques est de postuler une propriété fonctionnelle que tous partagent, par définition distincte de leurs propriétés physiques. C'est ce fait Putnam.

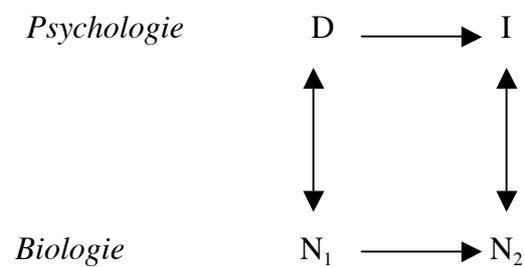
Dans son article de 1974 « Special Sciences (or : The disunity of science as a Working Hypothesis) », Jerry Fodor développe de façon plus détaillée un argument de ce type à l'encontre des propriétés disjonctives admises par le théoricien de l'identité en réponse à l'objection de la réalisation multiple. Fodor fait valoir qu'une telle identification d'un type d'état mental à une disjonction de types d'états ou de processus physiques ne satisfait pas les réquisits de la réduction interthéorique. Notons que la question qui intéresse Fodor résulte d'une forme d'ascension sémantique à l'égard du problème initial de la théorie de l'identité⁹⁹. Il ne cherche pas tant à savoir si le mental est identique au physique qu'à déterminer si la science du mental (la psychologie) est réductible à la science des événements physiques (la physique), et plus généralement, si toutes les sciences sont réductibles à la physique. Son principal souci est de montrer que la psychologie demeure une science autonome, qui n'a pas à se réduire à la science physique. Les termes de la relation de réduction ne sont plus les états, physiques ou mentaux, mais les théories du physique et du mental. Ce que Fodor entend par réductionnisme n'est pas la thèse selon laquelle tout ce qui existe est de nature physique, mais la thèse selon laquelle toutes les sciences sont en principe réductibles à la seule physique fondamentale. Ce passage au niveau théorique semble cependant ici relativement innocent : que la réalisabilité multiple concerne les états mentaux eux-mêmes ou les concepts d'états mentaux, le problème demeure le même : accepte-t-on de les réduire à des états physique ou à des concepts d'états physiques ?

L'idée de fond de la critique de Fodor est donc que la réalisabilité multiple conduit le partisan de la théorie de l'identité à admettre des propriétés disjonctives qui ne cadrent pas avec le modèle de réduction interthéorique. De quoi s'agit-il ? Selon Fodor, deux conditions doivent être satisfaites pour qu'une théorie puisse être réduite à une autre. La première condition est qu'il doit exister des lois-ponts (*bridge laws*) entre les sortes naturelles d'entités postulées par la théorie à réduire et les sortes naturelles d'entités qui entrent dans la théorie de base¹⁰⁰. Imaginons que nous disposions d'une théorie rudimentaire du mental qui contienne deux sortes d'états mentaux, les douleurs (D) et les intentions de fuite (I). Selon la première

⁹⁹ Une ascension analogue a été abordée en première partie quand nous avons traité du rapprochement opéré par Chisholm entre intensionnalité et intentionnalité.

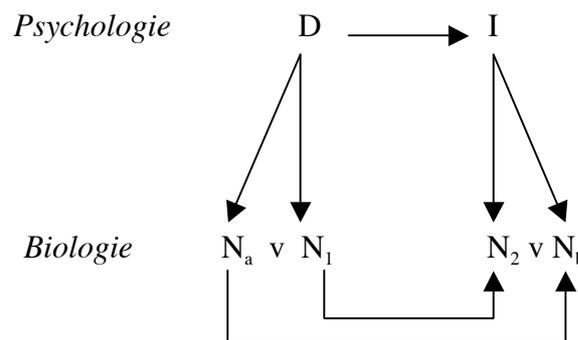
¹⁰⁰ Bien qu'il ne le cite pas, Fodor emprunte certainement à Ernest Nagel sa conception de la réduction interthéorique, et l'expression de *bridge laws*.

condition, il doit exister, dans la théorie biologique, deux sortes de processus distincts, disons, N_1 et N_2 , qui peuvent être mis en bijection avec les douleurs et les intentions. Par exemple, à D correspond N_1 (et inversement) et à I correspond N_2 . D'autre part, la seconde condition de réduction précise qu'il doit exister les mêmes rapports nomologiques entre les processus neurobiologiques qu'entre les états mentaux auxquels ils correspondent. Si les douleurs causent les intentions de fuite, les processus biologiques de types N_1 doivent causer les processus de type N_2 . Les conditions idéales de réduction sont donc satisfaites dans le cas suivant :



Les doubles flèches symbolisent les lois-pont interthéoriques, les flèches simples les lois causales, qui valent à l'intérieur des théories. La première condition n'est qu'une reformulation, au niveau des théories, de la remarque de Putnam selon laquelle le partisan de l'identité doit soutenir qu'à chaque type d'état mental correspond un et un seul type d'état physique. La seconde condition est plus nouvelle : s'il existe une loi qui relie certains types d'états mentaux à d'autres, alors il doit exister une loi analogue qui relie les types de processus biologiques qui correspondent à ces états mentaux. C'est grâce à l'introduction de ces lois intrathéoriques que Fodor pense pouvoir montrer en quoi les propriétés disjonctives aboutissent à des conséquences inacceptables. Admettons maintenant que la psychologie (D, I) s'applique à deux espèces d'êtres vivants, chez qui la réalisation de ces deux états mentaux est distincte. Pour la première espèce, les états (D, I) correspondent processus biologiques N_1 et N_2 , pour la seconde aux processus N_a , N_b . Cela signifie que les deux espèces ont les mêmes états psychologiques (les poulpes et les pangolins ont des douleurs et des intentions de fuite)

mais biologiquement réalisés de deux façons différentes. Nous avons alors le schéma suivant¹⁰¹ :



Le même type d'état mental peut être réalisé différemment selon les espèces. Quel problème cela pose-t-il ? Selon ce schéma, la biologie nous donne deux lois chacune étant relative à une espèce distincte. « Les processus biologiques de type N_1 causent des processus biologiques de type N_2 ($N_1 \rightarrow N_2$), et « Les processus biologiques de type N_a causent des processus biologiques de type N_b ($N_a \rightarrow N_b$)¹⁰². Logiquement, la conjonction de ces deux lois est équivalente à une seule loi : $(N_1 \vee N_a) \rightarrow (N_2 \vee N_b)$. Qui veut accepter la réalisation multiple sans renoncer à la théorie de l'identité doit donc accepter de telles lois, qui ont pour antécédent et pour conséquent des propriétés disjonctives. Or cette position, selon Fodor, est intenable : les lois, à ses yeux, ne peuvent avoir pour termes que des sortes naturelles. L'argument qu'il donne en faveur de cette thèse est que l'unique façon grâce à laquelle nous pouvons espérer distinguer les authentiques propriétés naturelles des propriétés arbitraire (disjonctives) est précisément que seules les propriétés naturelles figurent dans les lois théoriques. Si nous autorisons des propriétés disjonctives à y figurer, nous n'avons donc plus aucun critère de définition des propriétés naturelles. Or les propriétés disjonctives ne sont pas des sortes naturelles. Ce sont des artefacts théoriques. La solution du théoricien de

¹⁰¹ « \vee » signifie « ou » en logique. $P \vee Q$ est vraie dans trois cas : si P est vraie, si Q est vraie, et si P et Q sont toutes les deux vraies. Le « ou » logique (\vee) signifie donc l'un ou l'autre ou les deux.

¹⁰² On admet ici, ce qui peut être discuté, que l'implication logique permet de traduire les énoncés de loi.

l'identité relativisée conduit donc à admettre que les lois peuvent porter sur des sortes non naturelles. Fodor exprime ainsi son idée :

Je pense, par exemple, que c'est une loi que l'irradiation des plantes vertes par le soleil cause la synthèse d'hydrate de carbone. Je crois aussi que c'est une loi que la friction cause la chaleur. Mais je ne crois pas que c'est une loi que : (L'irradiation des plantes vertes par le soleil ou la friction) cause (la synthèse d'hydrate de carbone ou la chaleur). De façon correspondante, je doute que « est une synthèse d'hydrate de carbone ou est chaud » doive être considéré comme un prédicat de sorte naturelle plausible¹⁰³.

Ainsi formulée, la position de Fodor semble intuitive : si on admet des propriétés disjonctives, on peut construire des prétendues lois tout à fait inutiles du type « Le froid ou l'étincelle causent la gelée ou l'explosion ». Donc on ne peut admettre de propriétés disjonctives : elles sont purement arbitraires et ad hoc. Puisqu'elles étaient la réponse du théoricien de l'identité à l'objection de la réalisabilité multiple, Fodor en conclut que la théorie de l'identité de l'esprit et du corps doit être rejetée. Le mental ne doit pas être identifié aux propriétés biologiques, mais aux propriétés fonctionnelles.

Un tel argument cependant n'est pas forcé de convaincre le théoricien de l'identité. Il fera valoir que l'exemple de Fodor est mal choisi : l'irradiation des plantes vertes par le soleil et la friction, de même que le froid et l'étincelle, *n'ont pas le même rôle causal*. Il ne fait donc pas de doute que la disjonction est tout à fait arbitraire. En revanche, N_a et N_1 , ou si on veut, l'activation des fibres-C et l'activation des fibres-Z sont des processus biologiques qui par hypothèse *ont le même rôle causal*. Le théoricien de l'identité peut donc tenter de montrer que les classes de propriétés physiques disjointes auxquelles il identifie les états mentaux ne sont pas arbitraires. Pourquoi, demande-t-il, des processus biologiques qui ont les mêmes causes et les mêmes effets devraient-ils être considérés comme relevant de sortes biologiques totalement hétérogènes ? La disjonction « irradiation des plantes vertes par le soleil ou friction » est arbitraire. La disjonction « activation des fibres-C ou activation des fibres-Z » semble bien plus naturelle. Le fonctionnaliste de l'identité refuse donc l'alternative à laquelle l'acculent Fodor et Putnam : l'éparpillement ou le dualisme des propriétés.

¹⁰³ « Special sciences », 1974.

Un tel dialogue fait fortement penser à la controverse qui, en métaphysique, oppose nominalistes et réalistes. Pour les premiers, il n'y a dans le monde que des entités particulières, qui ne partagent rien où n'ont rien en commun indépendamment des rapprochement que *nous* pouvons faire entre eux. Pour les seconds, il existe dans le monde, outre des entités particulières, des entités universelles qui demeurent numériquement les mêmes aux travers de leur instanciations ou réalisation dans des entités particulières. Elle explique ainsi ce que différents particuliers ont en commun, indépendamment de nous. Considérez deux disques rouges parfaitement ressemblants. Le nominaliste dira qu'ils ne sont similaires *qu'à nos yeux*. Le réaliste dira que l'un et l'autre partagent la *même* propriété de rondeur, ainsi que la *même* propriété de rougeur¹⁰⁴. Cela même si nous ne les avons jamais regardés. Ces clés en main, on peut éclaircir l'opposition entre les deux types de fonctionnalisme. Le fonctionnalisme non-réductionniste est un réalisme concernant les propriétés mentales/fonctionnelles. Le fonctionnalisme de l'identité est un nominalisme concernant ces propriétés, qui sont réduites, par exemple, à des classes de ressemblance.

Le fonctionnalisme non-réductionniste comme réalisme des propriétés fonctionnelles

La pomme de discorde entre fonctionnalistes de l'identité (Armstrong, Lewis) et fonctionnalistes non-réductionnistes (Putnam, Fodor) est donc la question des propriétés disjonctives : ont-elles ou non quelque chose en commun qui font d'elles des sortes naturelles ? Autrement dit, les différentes réalisations physiques de la douleur ont-elles véritablement une propriété fonctionnelle commune ? Les fonctionnalistes non-réductionnistes répondent oui, et adoptent ainsi à l'égard de ces propriétés une attitude réaliste. Elles existent indépendamment de notre activité de perception ou de théorisation et sont ce qui unit les différents types de processus physiques qui réalisent un même état mental. De même que le réaliste des universaux soutient que ce que deux ronds ont en commun est d'instancier une même propriété de rondeur, le fonctionnaliste non-réductionniste soutient que ce que les activations de fibres-C et de fibres-Z ont en commun est de réaliser un *même* état fonctionnel de douleur. Cette propriété fonctionnelle est alors conçue comme *irréductible* aux propriétés physiques qui la réalisent. Cela implique un dualisme des propriétés : les

¹⁰⁴ Pour le réaliste, les universaux sont des entités qui peuvent exister en plusieurs endroits au même moment : la rondeur existe à la fois dans le rond de gauche et dans celui de droite.

propriétés fonctionnelles sont distinctes des propriétés physiques¹⁰⁵. Les fonctionnalistes non-réductionnistes ont généralement à cœur de donner un tour physicaliste à leur position, malgré ce dualisme des propriétés. Ils adoptent pour cela une thèse additionnelle : les propriétés fonctionnelles, bien que distinctes des propriétés physiques, *dépendent* d'elles. Cette dépendance prend deux formes. Il s'agit d'abord d'une dépendance existentielle : les propriétés fonctionnelles ne peuvent exister sans réalisateurs physiques. Cela équivaut à un rejet du dualisme des substances cartésien en faveur d'un dualisme des propriétés : bien qu'il ne se réduise pas au corps, l'esprit ne peut exister indépendamment de lui. Il s'agit ensuite d'une dépendance générique : si deux propriétés fonctionnelles sont réalisées par un même type d'état physique, alors elles sont elles-mêmes des propriétés fonctionnelles de même type¹⁰⁶. Par exemple, si deux états mentaux sont réalisés par l'activation des fibres-C, alors ce sont des douleurs. On parle alors de « survenance » ou d'« émergence » de l'esprit sur le corps : bien qu'il puisse y avoir une différence physique sans différence mentale (deux types d'états physiques pour un même type d'état mental), il ne peut y avoir de différence mentale sans différence physique (il est impossible qu'un même type d'état physique réalise des états mentaux de types différents). Les fonctionnalistes non-réductionnistes rachètent en quelque sorte leur dualisme en faisant valoir que certes, le corps peut varier indépendamment de l'esprit, mais que l'esprit ne peut varier indépendamment du corps. Autrement, bien que vos états mentaux ne soient pas identiques à vos états physiques, il est impossible que vous changiez d'états mentaux sans changer d'états physiques. Toute variation mentale est corrélée à une variation physique, mais toute variation physique n'est pas corrélée à une variation mentale. Bref, les propriétés mentales (fonctionnelles) sont distinctes des propriétés physiques, mais elles sont fermement ancrées à elles.

Le fonctionnalisme de l'identité comme nominalisme des propriétés fonctionnelles

¹⁰⁵ De même que le réaliste des universaux envisagé ici adopte une ontologie à deux catégories : les particuliers et les universaux.

¹⁰⁶ Il nous semble que ces deux formes de dépendance ne sont pas toujours distinguées : tantôt la survenance la dépendance d'une entité A sur une entité B est interprétée en termes de dépendance existentielle de A par rapport à B (A ne peut exister sans B —Armstrong, 1997, p. 11) ; tantôt cette survenance est interprétée en termes de dépendance générique ou de variation (A ne peut changer sans que B change —Lewis 1983). Pourtant, ces deux formes de dépendance sont indépendantes. Il semble qu'une entité puisse dépendre d'une autre pour son existence sans pour autant varier avec elle et inversement, que les variations d'une entité A puisse dépendre des variations d'une entité B sans que A dépende existentiellement de B (on a une proposition du type : *Si A existe, elle covarie avec B.*). Les théoriciens de l'identité acceptent la survenance générique, mais comme une thèse triviale, dans la mesure où il est évident qu'une chose covarie nécessairement avec elle-même.

A l'inverse, les fonctionnalistes de l'identité adoptent à l'égard des propriétés mentales une attitude nominaliste : en eux-mêmes, les différents types de processus physiques auxquels s'identifient les états mentaux n'ont rien en commun, leur communauté n'est que le produit de notre activité cognitive. La seule chose qui rassemble ces différents types de processus physiques est qu'ils tombent sous un même *concept*, le concept de rôle causal, les concepts étant les produits de notre esprit¹⁰⁷. Pour le fonctionnaliste de l'identité, notre *concept* de douleur est « un concept d'état jouant un certain rôle causal, état dont certains effets et causes sont typiques »¹⁰⁸, mais la douleur *elle-même* est un état neuronal. Deux types d'états neuronaux ne sont donc des états de douleur que si nous pouvons leur appliquer notre concept causal de douleur. Le fonctionnaliste de l'identité soutient donc que si nous n'étions pas là, l'activation des fibres-C et l'activation des fibres-Z n'auraient rien en commun. Mais, il faut y insister, il ne soutient nullement que le fait de regrouper sous un même concept les fibres-C et les fibres-Z soit purement arbitraire. Il y a bien une raison pour laquelle nous considérons l'activation des unes et des autres comme appartenant à une même classe, et que nous excluons par exemple de cette classe les processus neuronaux corollaires à la perception des couleurs. Cette raison est que les causes et effets de ces différents processus physiques ont, au regard de notre théorie psychologique ordinaire, une certaine similitude. Autrement dit, les rôles causaux s'ils n'existent pas en nature, sont néanmoins définis par les termes théoriques de notre psychologie de sens commun. En l'absence d'une telle psychologie, la douleur du poulpe et celle de l'homme n'auraient rien en commun, mais puisque nous en disposons comme d'une donnée, nous ne sommes pas livrés à l'arbitraire lorsque nous déterminons quels processus physiques sont les réalisations d'un même type d'état mental. Autrement dit, une des façons d'expliquer que notre concept de douleur s'applique à certaines disjonctions d'états physiques et pas à d'autres est de soutenir que bien que l'activation des fibres-C et des fibres-Z n'aient intrinsèquement rien en commun, leurs causes et effets tombent sous nos mêmes concepts ordinaires. L'appel à la psychologie populaire pour défendre le fonctionnalisme de l'identité est implicite chez Armstrong qui se propose de montrer que *nos concepts* d'état mentaux sont des concepts causaux. Lewis thématise ce recours à la psychologie ordinaire dans « Psychophysical and theoretical identifications ». Il faut noter par ailleurs que le fonctionnalisme de l'identité s'il admet que le concept de

¹⁰⁷ Bien entendu, si les concepts sont conçus de manière frégréenne, comme existant dans un autre monde, nous sortons du fonctionnalisme de l'identité pour aller vers un fonctionnalisme non-réductionniste, puisque le physicalisme est abandonné au profit d'une conception réaliste des propriétés mentales communes à différents états physiques.

¹⁰⁸ Lewis, « Douleur de fou et douleur de martien », 1978.

douleur, appliqué aux poulpes comme aux humains, ne désigne pas une sorte naturelle, maintient toutefois que le concept de douleur appliqué seulement aux humains, désigne une telle sorte. En effet, chez tous les humains, la douleur est (par hypothèse) identique à un même type de processus physique –l'activation des fibres-C. Donc, si le fonctionnaliste de l'identité soutient que la douleur du poulpe et celle de l'homme n'ont pas de propriété réelle en commun, il maintient que la douleur de Jacques et celle de Claudine ont une propriété physique commune.

Il est intéressant de remarquer pour finir qu'après avoir défendu une théorie réaliste des types d'états mentaux, selon laquelle il existe un état fonctionnel –computationnel– commun à différents types de réalisations physiques, Putnam a évolué dans un sens nominaliste en introduisant la notion d'isomorphie fonctionnelle dans son article « Philosophy and our mental life » (1973). Il s'agit là en substance d'un nominalisme de la ressemblance concernant les types d'états mentaux, selon lequel ce que la douleur du poulpe et de l'humain ont en commun est d'avoir des causes et des effets semblables. Dans *Représentation et réalité* (ou il dresse une utile présentation de son évolution sur la question du fonctionnalisme), Putnam fini également par rejeter cette seconde position nominaliste, ce qui (entre autres choses) le conduit à abandonner le fonctionnalisme :

Nous devons limiter la classe des réalisateurs autorisés aux disjonctions d'états physiques fondamentaux qui ont réellement (en un sens intuitif) « quelque chose en commun ». En outre, ce « quelque chose en commun » doit lui-même pouvoir être décrit à un niveau physique, ou au pire computationnel : si les membres disjoints d'une disjonction d'états physiques maximaux n'ont rien en commun qui puisse se voir au niveau physique, et rien en commun qui puisse se voir au niveau computationnel, alors dire qu'ils « ont en commun le fait d'être tous des réalisations de l'attitude propositionnelle A » quand A est une attitude propositionnelle que nous souhaitons réduire serait pure et simple tricherie.

Résumons. La douleur du poulpe est réalisée par l'activation des fibres-Z, celle de l'homme par l'activation des fibres-C. La question est de savoir si la douleur est identique à ces différentes réalisations physiques, ici l'activation des fibres-C ou l'activation des fibres-Z. Le fonctionnaliste de l'identité répond oui, ce qui lui permet de maintenir un monisme physicaliste, mais le conduit à soutenir que le terme « douleur » ne désigne rien d'indépendant de notre activité cognitive. Le fonctionnaliste non-réductionniste répond non, ce qui lui permet de maintenir que la douleur est une sorte naturelle indépendante, mais le contraint à

admettre une forme de dualisme entre propriétés fonctionnelles et propriétés physiques. Tous les fonctionnalistes s'accordent sur la place centrale qu'il convient d'accorder à la description des états mentaux en termes de rôle causal. Mais alors que les fonctionnalistes non-réductionnistes (Putnam, Fodor) prennent cette description au sérieux en pensant qu'elle réfère à un type spécifique d'entités irréductibles (les rôles causaux), les fonctionnalistes de l'identité (Lewis et Armstrong) considèrent cette description comme une simple analyse de nos concepts mentaux, préalable à la réduction physicaliste.

2.3.2.4. Fonctionnalisme et hylémorphisme

Lorsque l'on considère l'approche fonctionnaliste en ayant en le traité *De l'âme* d'Aristote, il est difficile de s'empêcher de penser à un rapprochement entre le fonctionnalisme contemporain de l'hylémorphisme aristotélicien. Dans la mesure où Aristote définit l'âme comme la forme du corps matériel, et que les fonctionnalistes non-réductionnistes définissent le mental comme le rôle causal occupé par des processus physiques, il tentant de se demander si l'analogie forme/matière et rôle causal/occupant physique est purement fortuite.

Au début du livre II du *De Anima*, Aristote propose cette définition de l'âme :

Il faut donc nécessairement que l'âme soit substance comme forme d'un corps naturel qui a potentiellement la vie. Or cette substance est réalisation (*entelecheia*). Donc elle est la réalisation d'un tel corps.

L'âme est la forme du corps. Les états mentaux sont des fonctions réalisées par des processus biologiques. Aristote est-il le premier fonctionnaliste ? Cette question s'est rapidement imposée après les premières formulations du fonctionnalisme. Dès 1974, R. Sorabji esquisse une interprétation des rapports entre l'âme et le corps chez Aristote tout à fait voisine de celle qu'en donnent les fonctionnalistes non réductionnistes. L'année suivante, Putnam écrit :

Ce qui nous intéresse vraiment, nous autres fonctionnalistes est, telle qu'Aristote le voyait, la forme et non la matière. Quelle est notre forme intellectuelle ? est la question, et non pas quelle est la matière.

Depuis lors, question de la proximité de l'hylémorphisme et du fonctionnalisme divise les commentateurs. Nous présentons d'abord les arguments qui vont dans le sens d'un tel rapprochement, puis les difficultés qu'il rencontre.

Le premier argument en faveur du rapprochement de la forme et de la fonction est que la forme n'est reliée que de façon contingente à la matière : une même forme, la rondeur, peut informer plusieurs matières différentes : de l'airain, de la pierre ou du bois :

Dans les cas où la forme apparaît s'imposer à des matières spécifiquement différentes, comme un cercle peut être d'airain, de pierre ou de bois, dans tous ces cas il semble manifeste que ne font partie de l'essence du cercle ni l'airain, la pierre, puisque le cercle a une existence séparée de la leur¹⁰⁹.

Cela évoque bien entendu la réalisabilité multiple : un même rôle causal, peut être réalisé par différents processus physiques. En outre, le fait qu'Aristote soutienne que la forme existe séparément de la matière permet de déterminer que le seul fonctionnalisme auquel on puisse prétendre rapprocher sa position le fonctionnalisme non-réductionniste, mais en aucun cas le fonctionnalisme de l'identité. Pour ce dernier, il n'existe rien de tel que des rôles causaux irréductibles à des processus physiques. Seul le fonctionnaliste non-réductionniste, qui adopte une attitude réaliste à l'égard des rôles causaux, peut donc se prétendre aristotélicien sur ce problème. L'existence séparée de la forme par rapport à la matière répond tout à fait à son souci de ménager à la psychologie un objet d'étude distinct de celui de la physique.

En outre, bien que la forme soit distincte de la matière, Aristote souligne, contre Platon, que la forme se trouve toujours réalisée dans la matière. « [la forme d'une chose] doit être nécessairement inhérente à telle sorte précise de matière »¹¹⁰. On a vu que les fonctionnalistes non-réductionnistes, soucieux de rester le plus proche possible du physicalisme, soutenaient que le mental « survenait » sur le physique en deux sens : le mental dépend existentiellement du physique, et les variations du mental dépendent des variations physiques. Ici, de même que le rôle fonctionnel dépend existentiellement de ses réalisateurs, la forme dépend existentiellement de la matière.

¹⁰⁹ *Métaphysique* Z11, 1036a33.

¹¹⁰ *De l'âme*, 403b3.

Dernier point important de parenté entre Aristote et les fonctionnalistes : Aristote donne la priorité à la forme sur la matière. Dans le livre Z de la métaphysique, après s'être demandé si la substance est la forme, la matière ou le composé des deux, Aristote conclut nettement en faveur de la forme. Les fonctionnalistes considèrent également que puisque un même état fonctionnel peut être réalisé par différents processus physiques, la question de ces processus, bien qu'importante, est secondaire par rapport à la question des états fonctionnels.

Il y a cependant un certain nombre d'objections à l'assimilation de l'hylémorphisme au fonctionnalisme. La première a trait à la thèse de survenance que le fonctionnaliste non-réductionniste adopte traditionnellement. Selon cette thèse, un même type d'état physique ne peut réaliser différents types d'état mentaux. Or Aristote soutient non seulement qu'une même forme peut être réalisée dans plusieurs matières (le bois, l'airain) *mais également qu'une même matière peut être actualisée par plusieurs formes*. Le bois peut être rond, mais aussi carré. Pour le fonctionnaliste à l'opposé, l'activation des fibres-C ne peut pas réaliser parfois la douleur, parfois les chatouilles. Admettre qu'un même état du corps peut correspondre à différents états mentaux, reviendrait à abandonner la thèse de la survenance du mental sur le physique, qui, bien qu'elle ne soit pas essentielle au fonctionnalisme non réductionniste, constitue son ancrage physicaliste. Face cette difficulté, Martha C. Nussbaum et Hilary Putnam (1992) qui défendent une lecture fonctionnaliste d'Aristote, finissent par abandonner la survenance du mental sur le physique, bien qu'ils continuent de soutenir que le mental dépend existentiellement du physique. Simplement, il peut y avoir des différences mentales sans différences physiques. Le physique n'est plus une condition suffisante du mental. Ce retranchement est possible, puisque la thèse de la survenance n'est pas essentielle au fonctionnalisme non réductionniste. Mais elle en diminue l'attrait, pour qui partage le souci du physicaliste : la seule dépendance qu'il reste entre le corps et l'esprit est une relation de dépendance existentielle, l'un et l'autre peuvent pour le reste varier indépendamment.

Une autre difficulté touche cette fois directement le cœur de la thèse du fonctionnalisme non réductionniste. Nous avons vu que la différence fondamentale entre celui-ci et le fonctionnalisme de l'identité était qu'il admettait que les types d'état mentaux existaient indépendamment de notre conceptualisation. Dans un monde sans hommes, la douleur du poulpe et celle du pangolin partageraient encore quelque chose : l'une et l'autre

instancieraient une propriété fonctionnelle. Les propriétés mentales (fonctionnelles) pouvant être réalisées par plusieurs états physiques, il est donc essentiel que le fonctionnaliste non-réductionniste les conçoive comme des universaux. Le problème est qu'au chapitre 13 du livre Z de la métaphysique, après avoir clairement identifié la substance à la forme (livre 7-9, puis 17) Aristote affirme que les substances ne sont pas des universaux :

Il nous paraît en effet impossible qu'aucun terme universel, quel qu'il soit, soit une substance [=une forme]. D'abord, la substance d'un individu est celle qui lui est propre et qui n'appartient pas à un autre ; l'universel est au contraire commun, puisqu'on nomme universel ce qui appartient naturellement à une multiplicité ; de quoi sera-t-il donc la substance ? Il devra l'être de toutes les choses auxquelles on l'attribue, ou il ne le sera d'aucune ; mais qu'il le soit de toutes, ce n'est pas possible, et s'il est la substance d'un seul individu, cet individu sera tous les autres aussi, car les êtres dont la substance est une et la quiddité [l'essence] une sont aussi un seul et même être.

Donc, l'âme, qui est la substance (c'est-à-dire la forme) du corps est un particulier. Ceci interdit donc tout rapprochement de l'âme aristotélicienne et du rôle fonctionnel, qui pour le fonctionnaliste non-réductionniste doit être un universel, qui demeure le même à travers ses multiples (type de) réalisations. Certains commentateurs cependant soutiennent que ce chapitre 13 est trompeur est que la position d'Aristote est au fond que les formes sont des universaux¹¹¹. Cette position serait en partie plus compatible avec ses *Catégories*, et semble implicite en Z4 et Z11, où Aristote soutient qu'il est impossible de définir le particulier, alors même qu'il définit la forme. Mais paradoxalement, une telle interprétation ne fait pas non plus les affaires du fonctionnaliste. S'il est essentiel au fonctionnalisme qu'un même état mental fonctionnel puisse avoir différentes réalisations physiques, il faut que ces réalisations physiques aient leurs propres critères d'individuation, en vertu desquels elles sont plusieurs. Or, dans l'hylémorphisme d'Aristote, c'est la forme qui est censée jouer le rôle d'individuateur, par rapport au substrat ou à la matière première. Cette dernière, selon Aristote, n'est pas séparable et n'est pas une chose individuelle. Seule la forme lui procure son individualité. Cela constitue non seulement un argument en faveur de l'interprétation selon laquelle la forme est un particulier, mais également un argument contre le rapprochement avec le fonctionnalisme : si la forme est un universel, l'hylémorphisme ne permet pas d'individuer les processus physiques indépendamment de leurs rôles fonctionnels.

¹¹¹ Pour une très bonne présentation de ces débats, ainsi que plus généralement de la métaphysique d'Aristote, distinguant scrupuleusement la *Métaphysique* proprement dite des *Catégories*, voir S. M. Cohen (2003).

La réalisabilité multiple devient impossible, non plus faute d'universel, mais faute de particuliers. Ackrill écrit ainsi :

Le corps qu'on nous demande de distinguer en tant que « constituant » matériel de l'animal dépend pour son identité même de fait qu'il soit vivant, informé par l'âme¹¹².

Si l'identité du corps dépend du fait qu'il est informé par l'âme, il est impossible de concevoir différents types de réalisations corporelles pour une même espèce d'âme. En résumé, soit la forme est un universel et les processus physiques n'ont pas de critères d'individuation. Soit la forme est un particulier la douleur du poulpe et du pangolin n'ont rien en commun.

La dernière difficulté mentionnée ici à l'encontre de la lecture fonctionnaliste d'Aristote est soulevée par Myles Burnyeat. Nous avons déjà rencontré cette interprétation en première partie ainsi que celle, opposée, de Richard Sorabji. Pour mémoire, Sorabji, et un certain nombre d'interprètes fonctionnalistes, considèrent que la description qu'Aristote donne de la perception, « prendre la forme sans la matière », décrit un processus physique, selon lequel l'organe perceptif prend littéralement la forme de l'objet perçu. L'œil rougit. Burnyeat à l'opposée considère que la formule d'Aristote ne décrit aucun processus physique, mais un simple acte de conscientisation. La raison pour laquelle la lecture de Burnyeat contredit le fonctionnaliste est que dans ce cas, l'acte mental de perception n'est réalisé, ni même corrélé, à aucun processus physique. Selon Burnyeat, il peut y avoir conscience perceptive sans aucun changement physiologique correspondant. Si cela est vrai, alors il existe des états mentaux qui ne sont pas réalisés physiquement et Aristote ne peut être fonctionnaliste.

La discussion entre Burnyeat et ses opposants fonctionnalistes (Putnam-Nussbaum, Sorabji, S. M. Cohen et d'autres) est pointue et va jusqu'à poser des problèmes d'édition du texte. Il n'est pas essentiel ici d'y entrer dans en détail. Un point cependant mérite d'être souligné pour finir. Il est significatif que Burnyeat justifie son rejet de l'interprétation fonctionnaliste en faisant valoir que ce qu'Aristote a avant tout en vue est la *conscience*. Indépendamment de la question de savoir si cette interprétation est justifiée ou non par

¹¹² 1972, cité par S. M. Cohen, 1992.

rapport au texte d'Aristote, force est de constater que le recours à la notion de conscience met le doigt sur un point faible du fonctionnalisme. Lorsque nous avons cherché en première partie à déterminer un critère de distinction entre le mental et le physique, nous avons souligné que le critère de la conscience, selon lequel seuls les phénomènes mentaux seraient accessibles subjectivement et infailliblement, posait des problèmes méthodologiques, et qu'en un certain sens, il ne nous éclairait pas beaucoup sur ce qu'était le mental puisqu'il semblait être un terme synonyme. Nous avons finalement retenu le critère de l'intentionnalité, qui sans exclure la conscience (tout phénomène mental est conscient de lui-même) lui ajoutait un critère plus informatif : tout phénomène est conscient d'autre chose que de lui-même. La leçon que nous pouvons en tirer est que si la conscience est floue et problématique, il est difficile de définir le mental sans y faire appel. C'est pourtant ce que fait le fonctionnaliste, qu'il adopte une position favorable ou non à l'égard de la théorie de l'identité. Ce qu'il appelle esprit, et qu'il s'enquiert de réduire ou de relier aux processus physiques, est défini « de l'extérieur », *en troisième personne* : lorsqu'il se demande ce qu'est une douleur, il se demande dans quelles conditions nous attribuons une douleur à un être distinct de nous. Et il a certainement raison de dire que nous n'attribuons de tels états mentaux qu'à l'aide d'indices comportementaux : l'être en question est blessé, il effectue un mouvement de retrait, gémi etc. Pourtant, il semble restrictif de penser que l'on capture ainsi l'essence de ce que nous appelons « douleur » et plus généralement « esprit ». Une douleur n'est-elle pas avant tout une sensation que nous expérimentons en première personne avant de l'attribuer à autrui ? N'y a-t-il pas certaines douleurs qui n'ont pour corollaires ni dommages corporels, ni comportement manifestes ? Autrement dit, en caractérisant les états mentaux comme des médiateurs causaux, le fonctionnaliste n'oublie-t-il pas un aspect essentiel de ces états : la conscience ? Dès lors, l'entreprise de réduction du « mental » au physique peut sembler relativement secondaire : il ne s'agit seulement de la question de savoir comme certains états physiques (neuronaux) peuvent en fonder d'autres (comportementaux). L'esprit, a-t-on envie de dire, est déjà réduit au départ. Il serait injuste de considérer cependant que les fonctionnalistes n'ont pas pris en compte cette objection : aussitôt sa théorie fonctionnaliste esquissée, Armstrong consacre une dizaine de pages à répondre à l'objection que nous venons de formuler, selon laquelle le fonctionnalisme oublierait l'aspect essentiel du mental, la conscience. David Lewis a consacré plusieurs articles à l'objection des *qualia*, selon laquelle sa théorie ne permettrait pas de rendre compte de l'impression subjective que cela nous fait de percevoir une couleur ou de sentir une douleur. On ne peut donc accuser les fonctionnalistes de négligence. On peut

cependant constater que le temps n'a pas dissipé cette perplexité, et que depuis une dizaine d'année maintenant, le vent sembler tourner en faveur de ceux qui considèrent que la conscience demeure inexplicée par les états fonctionnels¹¹³. Le dualisme devient de moins en moins honteux, et bien qu'il soit rarement soutenu dans une version forte, du type dualisme des substances, il n'est plus absurde de dire que Descartes est d'actualité en philosophie de l'esprit.

¹¹³ Voir notamment le livre important de D. Chalmers, *The Conscious Mind*, 1996.

BIBLIOGRAPHIE

- ABBOTT, Thomas K., (1864), *Sight and Touch : An Attempt to Disprove the Received or Berkeleian Theory of vision*. London : Longman, Green.
- ARISTOTE, *De l'âme*, trad. R. Bodéüs, GF Flammarion, Paris, 1993.
- *Ethique à Nicomaque*, trad. J. Tricot, Paris, Vrin, 1990.
- *Métaphysique*, trad. J. Tricot, Paris, Vrin, 1991 (1^{ère} éd. 1933).
- ARMSTRONG, D. M. , *A World of States of Affairs*, Cambridge, Cambridge University Press, 1997.
- *Universals, An Opiniated Introduction*, London, Westview, 1989.
- *A Materialist Theory of the Mind*, Londres, Routledge, 1993 (1^{ère} éd. 1968).
- AUSTIN, J. L. , *Quand dire c'est faire*, Paris, Seuil, 1970, (1^{ère} éd. 1962).
- *Sense and Sensibilia*, Oxford, Clarendon Press, 1962, tr. Fr. P. Gochet, *Le langage de la perception*, Paris, Armand Colin, 1971.
- BAIN, A, *L'esprit et le corps, considérés au point de vue de leur relation*, Paris, Librairie Germer Baillièrre et Cie, 1878 (3^{ème} éd.).
- BERGSON, H., *Essai sur les données immédiates de la conscience*, Paris, PUF, 1927.
- BLOCK, N., « What is functionalism ? », 1980, <http://cogprints.ecs.soton.ac.uk/archive/00000235/00/199712008.html>
- BORING, Edwin G., *A History of Experimental Psychology*, New York, The Century Co, 1929.
- BOUVERESSE, J., *Langage, perception et réalité, Tome 1, La perception et le jugement*, Nîmes, Jacqueline Chambon, 1995.
- BRENTANO, F., *Psychology from an empirical standpoint*, éd. O. Kraus, trad. Anglaise Linda L. McAlister, London, Routledge, 1995 (1^{ère} éd. Allemande, 1874).
- *Untersuchungen zur Sinnespsychologie*. Leipzig: Duncker & Humblot 1907. 2^{ème} édition augmentée édité par R. Chisholm and R. Fabian, Meiner, Hamburg, 1979.
- BURNHEIM, J., « Intentionality and Materialism », article non publié lu à un séminaire du département de philosophie de l'université de Sydney, 1969.
- BURNYEAT, M. F.,, « Is an aristotelian philosophy of mind still credible ? (a draft) », in *Essays on Aristotle's De Anima*, M. C. Nussbaum & A. Oksenberg Rorty (éds), Oxford, Clarendon Press, 1992.
- CASATI, R., DOKIC J., *La philosophie du son*, Nîmes, Jacqueline Chambon, 1994.

- CHALMERS, D., *The Conscious Mind, In Search of a Fundamental Theory*, Oxford University Press, 1996.
- éd., *Philosophy of Mind, classical and contemporary readings*, Oxford University Press, 2002.
- CHISHOLM, R. M., *Perceiving : a Philosophical Study*, Ithaca, Cornell University Press, 1957.
- CHOMSKY, Noam, « A Review of B. F. Skinner's *Verbal Behavior* » in *Language*, 35, No. 1 (1959), 26-58 <http://cogprints.ecs.soton.ac.uk/archive/00001148/00/chomsky.htm>
- CHURCHLAND, Patricia Smith, 1982, « Is the Visual System as Smart as It Looks ? », *Proceedings of the Philosophy of Science Association* 2, pp. 541-552. Réimprimé in Schwartz 2004, p. 217-228.
- CHURCHLAND, Paul, « Le matérialisme éliminativiste et les attitudes propositionnelles », in Fissette, D., Poirier, P., (éds), *Philosophie de l'esprit, Psychologie du sens commun et sciences de l'esprit*, Paris, Vrin, 2002.
- COHEN, S. Marc, « Aristotle's Metaphysics », *The Stanford Encyclopedia of Philosophy* (Winter 2003 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/win2003/entries/aristotle-metaphysics/>.
- « Hylomorphism and fonctionnalism », in », in *Essays on Aristotle's De Anima*, M. C. Nussbaum & A. Oksenberg Rorty (éds), Oxford, Clarendon Press, 1992.
- CONDILLAC, E. B., *Essai sur l'origine des connaissances humaines, Ouvrage où l'on réduit à un seul principe tout ce qui concerne l'entendement humain*, Alive, 1998 (1^{ère} éd. 1746).
- CRANE, T., *Elements of Mind, An Introduction to the Philosophy of Mind*, New York, Oxford University Press, 2001.
- « Intentionality as the Mark of the Mental », in *Current Issues in Philosophy of Mind*, éd. A. O'Hear, Cambridge University Press, 1998, p. 229-251.
- DESCARTES, *Passions de l'âme*, Paris, Gallimard, 1988.
- *Méditations métaphysiques*, Paris, Flammarion, 1992.
- DIDEROT, D., *Lettre sur les Aveugles, Lettre sur les sourds et les muets*, Paris, Flammarion, 2000 (1^{ère} éd. 1749).
- DRETSKE, F., *Seeing and Knowing*, Routledge & Kegan Paul, Chicago, 1969.
- « Laws of nature », *Philosophy of science*, 44, 1977, p. 248-268.
- ENGEL, P., *Philosophie et psychologie*, Paris, Gallimard, 1996.
- *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994.

- EVANS, G., (1985), « Molyneux's question », in *Collected Papers*, Oxford, Clarendon Press.
- DUMMETT, M., *Les origines de la philosophie analytique*, trad. M.-A. Lescourret, Paris, Gallimard, 1991 (1^{ère} éd. Allemande, 1988).
- FECHNER, Gustav Theodor, *Elements of psychophysics*, Sections VII and XIV, trad., Herbert Sidney Langfeld, 1912. <http://psychclassics.yorku.ca/Fechner/>
- FEIGL, H., « The "Mental" and the "Physical" », in Feigl, Scriven et Maxwell (éd.), *Concepts, Theories, and the Mind-Body Problem*, University of Minnesota Press, 1958, reimprimé partiellement in Chalmers, D., *Philosophy of mind, Classical and contemporary readings*, Oxford University Press 2002, p. 68-72.
- FERRET, S., *L'identité*, Paris, Flammarion, Corpus, 1998.
- FISSETTE, D., POIRIER, P., (éds), *Philosophie de l'esprit, Psychologie du sens commun et sciences de l'esprit*, Paris, Vrin, 2002.
- FODOR, J., *Concepts, Where cognitive sciences went wrong*, Oxford, Clarendon Press, 1998.
— *Psychosemantics*, Cambridge, MIT Press, 1987.
— « Special Sciences (or: the disunity of science as a Working Hypothesis) », *Synthese*, 28, 1974, reimprimé in Chalmers, D., *Philosophy of mind, Classical and contemporary readings*, Oxford University Press 2002, p. 126-135.
- FREGE, G., *Ecrits logiques et philosophiques*, trad. C. Imbert, Paris, Seuil, 1971.
- GLAUSER, R., *Berkeley et les philosophes du XVII^e siècle, Perception et scepticisme*, Bruxelles, Mardaga, 1999.
- GOUHIER, H., *La pensée métaphysique de Descartes*, Paris, Vrin, 1999 (1^{ère} éd. 1962).
- HUME, D., *Enquête sur l'entendement humain*, trad. A. Leroy et M. Beyssade, Paris, Flammarion, 1983.
— *Traité de la nature humaine*, livre I, L'entendement, trad. P. Baranger et P. Saltel, Paris, Flammarion, 1995 (1^{ère} éd. 1739).
- JACKSON, F., « Mental causation », *Mind*, n°105, p. 377-413, 1996.
- JACOB, P., "Intentionality", *The Stanford Encyclopedia of Philosophy (Fall 2003 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2003/entries/intentionality>.
- JAMES, William, *The Principles of Psychology*, New-York, Dover, 1950 (1^{ère} éd. 1890). (Il existe une traduction française.)
- KANT, E., *Critique de la raison pure*, trad. A. Renaut, Paris, Flammarion, 2001.

- KIM, J., « L'émergence, les modèles de réduction et le mental », in *Philosophiques*, 27/1, Printemps 2000, p. 11-26, URL = <http://www.erudit.org/revue/philoso/2000/v27/n1/004937ar.html>
- « Multiple Realization and the Metaphysics of Reduction », *Philosophy and Phenomenological Research*, 52, 1992, reimprimé in Chalmers, D., *Philosophy of mind, Classical and contemporary readings*, Oxford University Press 2002, p. 60-68.
- *Philosophy of Mind*, London, Westview, 1996.
- KISTLER, M., *Causalité et lois de la nature*, Paris, Vrin, 1999.
- KRIPKE, S., *La logique des noms propres*, trad. P. Jacob et F. Recanati, Paris, Editions de Minuit, 1982 (1^{ère} éd. 1972).
- LEIBNIZ, G. W., *Nouveaux essais sur l'entendement humain*, Paris, GF Flammarion, 1990.
- LEWIS, D., « Reduction of Mind », in S. Guttenplan, *A Companion to Philosophy of Mind*, Blackwell, 1994, réimprimé in Lewis, D., *Papers in Metaphysics and epistemology*, Cambridge University Press, p. 291-324.
- « Douleur de fou et douleur de martien », in D. Fisette et P. Poirier, 2002 (1^{ère} éd. anglaise 1980), p. 289-306.
- « Psychophysical and theoretical identifications », *Australasian Journal of Philosophy*, 50, 1972, réimprimé in Lewis, D., *Papers in Metaphysics and epistemology*, Cambridge University Press, p. 248-261.
- « An Argument for the Identity Theory », *Journal of Philosophy*, 63, 17-25, 1966..
- LOCKE, J., *Essai sur l'entendement humain*, vol. 1, trad. J. -M. Vienne, Paris, Vrin, 2001.
- LOWE, E. J., *An introduction to the philosophy of mind*, Cambridge University Press, 2000.
- *Subjects of experience*, Cambridge, Cambridge University Press, 1996.
- MARTIN, C. B., PFEIFER, K., « Intentionality and the Non-Psychological », *Philosophy and Phenomenological Research*, 46, 531-54, 1986.
- MCCULLOCH, G., « Intentionnality and Interpretation », in *Current Issues in Philosophy of Mind*, éd. A. O'Hear, Cambridge University Press, 1998, p. 253-271.
- MOLNAR, G., *Powers, A Study in Metaphysics*, éd. Peter Mumford, Oxford University Press, 2003.
- MULLIGAN, K., « De la philosophie autrichienne et de sa place » in *La philosophie autrichienne de Bolzano à Musil, Histoire et actualité*, éd. J.-P. Cometti, K. Mulligan, Paris, Vrin, 2001, p. 7-25.

- MUMFORD, S., « Intentionality and the Physical : A New Theory of disposition Ascription », *Philosophical Quarterly*, 49, p. 215-225, 1999.
— *Dispositions*, New York, Oxford University Press, 1998.
- O'SHAUGHNESSY, B., *The Will, A dual aspect theory*, Cambridge University Press, 1980.
- PLACE, U. T., « Intentionality as the Mark of the Dispositional », *Dialectica*, 50, p. 91-120, 1996.
— « Identity Theories », draft, <http://host.uniroma3.it/progetti/kant/field/mbit.htm>
— « Is Consciousness a Brain Process ? », *British Journal of Psychology*, 47, 1956, reimprimé in Chalmers, D., *Philosophy of mind, Classical and contemporary readings*, Oxford University Press 2002, p. 55-60.
- PLATNER, Ernst, (1793), *Philosophischen Aphorismen*, 2 vols., Leipzig : im Schwickertschen Berlage.
- PLATON, *Le Sophiste*, trad. N. Cordero, Paris, Flammarion, 1993.
- PUTNAM, H., *Représentation et réalité*, trad. C. Tiercelin, Paris, Gallimard, 1990 (1^{ère} éd. anglaise 1988).
— « Philosophy and our mental life », in *Philosophical papers, vol. 2, Mind, language and reality*, p. 291-203, Cambridge University Press, 1975 (1^{ère} éd. 1973)
— « La nature des états mentaux », in Fisette et Poirier, 2002, p. 269-287 (1^{ère} éd. anglaise 1967).
- PUTNAM, H, NUSSBAUM, M. C., « Changing Aristotle's Mind », in *Essays on Aristotle's De Anima*, M. C. Nussbaum & A. Oksenberg Rorty (éds), Oxford, Clarendon Press, 1992.
- QUINE, W. V. O., *Le mot et la chose*, Trad. Fr. J. Dopp et P. Gochet, Paris, Flammarion, 1977 (1^{ère} éd. Anglaise 1960).
- REID, *An Inquiry into the Human Mind on the Principles of Common Sense*, Pennsylvania State University Press, 1997 (1^{ère} éd. 1764).
- ROCK, Irvin, 1982, « Inference in Perception », in *Proceedings of the Philosophy of Science Association* 2, p. 525-540. Réimprimé in Schwartz 2004 p. 200-216.
- RYLE, G., *La notion d'esprit, Pour une critique des concepts mentaux*, trad. S. Stern-Gillet, Paris, Payot, 1978 (1^{ère} éd. anglaise 1949).
- SCHWARTZ, R.(éd.), *Perception*, Blackwell, 2004.
- SEARLE, J., *Intentionality*, Cambridge , Cambridge University Press, 1983.
- SIMONS P., *Parts, A study in ontology*, Oxford University Press, 2000 (1^{ère} éd. 1987).

- SMART, J. J. C., "The Identity Theory of Mind", *The Stanford Encyclopedia of Philosophy (Fall 2000 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2000/entries/mind-identity/>>.
- « Sensations and brain processes », *Philosophical Review*, 68 :141-156, 1959, reimprimé in Chalmers, D., *Philosophy of mind, Classical and contemporary readings*, Oxford University Press 2002, p. 60-68.
- SMITH, B., *Austrian Philosophy, The Legacy of Franz Brentano*, Open Court, Chicago, 1994.
- SORABJI, R., « Intentionality and Physiological Processes: Aristotle's Theory of Sense Perception », in *Essays on Aristotle's De Anima*, M. C. Nussbaum & A. Oksenberg Rorty (éds), Oxford, Clarendon Press, 1992.
- « Body and soul in Aristotle », *Philosophy*, 49, 63-89, 1974.
- SKINNER, B. F., « Two types of conditioned reflex and a pseudo type », *Journal of General Psychology*, 12, 66-77, 1935. <http://psychclassics.yorku.ca/Skinner/Twotypes/twotypes.htm>
- TIERCELIN, C., « Sur la réalité des propriétés dispositionnelles », in *Le réalisme des universaux*, Cahiers de Philosophie de l'Université de Caen, dir. V. Carraud et S. Chauvier, n°38-39, Presses Universitaires de Caen, 2002, p. 127-158.
- TWARDOWSKI, K., Sur la théorie du contenu et de l'objet des représentations, une étude psychologique, 1894, in *Sur les objets intentionnels, 1893-1901*, trad. J. English, Paris, Vrin, 1993.
- VERNANT, D., *Bertrand Russell*, Paris, Flammarion, 2003.
- VON SENDEN, Max, (1960), *Space and Sight : the Perception of Space and Shape in the Congenitally Blind Before and After Operation*, trans. P. Heath, London : Methuen.
- WATSON, John B., « Psychology as the behaviorist views it », *Psychological Review*, 20, 158-177, 1913. <http://psychclassics.yorku.ca/Watson/views.htm>
- WEBER, E.H., *De tactu, : annotationes anatomicæ et physiologicæ* in *E. H. Weber on the Tactile Sense*, éd. et trad. Helen E. Ros et David J. Murray, Erlbaum (GB), Taylor & Francis, 1996. (1^{ère} éd. allemande 1834).
- *Tastinn und Gemeingefühl*, in *E. H. Weber on the Tactile Sense*, éd. et trad. Helen E. Ross et David J. Murray, Erlbaum (GB), Taylor & Francis, 1996 (1^{ère} éd. allemande 1846).
- WITTGENSTEIN, L., *Notes sur l'expérience privée et les « sense data »*, trad. E. Rigal, 2^{ème} éd., Ter Bilingue, Trans-Europ-Repress, 1989.

Plan du cours

1. Première partie : la distinction du corps et de l'esprit

1.1. Le critère intentionnel

1.1.1. L'intentionnalité

1.1.2. L'inexistence

1.1.3. Aux origines aristotéliennes de l'inexistence intentionnelle

1.1.4. De l'intentionnalité à l'intensionnalité

1.2. Le critère spatial

1.3. Le critère épistémologique

2. Deuxième partie : le problème de la corrélation

2.1. La corrélation du corps et de l'esprit

2.1.1. La corrélation chez Descartes

2.1.2. La corrélation chez E.H. Weber et G.T. Fechner

2.1.2.1. Naissance de la psychophysique

2.1.2.2. Critiques de W. James et H. Bergson

2.2. Le problème dissolu

2.2.1 Le béhaviorisme

2.2.1.1. Le béhaviorisme psychologique : J. B. Watson et B. F. Skinner

2.2.1.2. Le béhaviorisme philosophique : G. Ryle

2.2.2. La dissolution humienne

2.3. Le problème résolu

2.3.1. La théorie de l'identité.

2.3.1.1. Présentation

2.3.1.2. L'objection de la réalisabilité multiple

2.3.1.3. L'objection des propriétés phénoménales

2.3.2. Le fonctionnalisme

2.3.2.1. Présentation

2.3.2.2. Fonctionnalisme et béhaviorisme

2.3.2.3. Fonctionnalisme et théorie de l'identité : fonctionnalisme
« australien » vs fonctionnalisme « non-réductionniste »

2.3.2.4. Fonctionnalisme et hylémorphisme

Bibliographie