

explication à donner de cet écart entre les capacités intellectuelles des acteurs et leur conscience. Est-ce que les règles sont (tacitement) connues des acteurs, ou non ?

On a cherché à rendre compte de cet écart, donc à développer une philosophie de l'esprit objectif, de deux façons. Une première réponse est de poser un esprit objectif naturel à titre d'entité théorique : les règles et les représentations sont en réalité des structures mentales, physiquement présentes dans l'organisme. Cette réponse est celle de la nouvelle philosophie mentale qui inspire le programme contemporain du cognitivisme (cf. *infra* chap. 4-12).

L'autre réponse est de définir l'esprit objectif comme un esprit social. L'esprit objectif, celui qui mesure l'écart entre la capacité intellectuelle de l'acteur et sa conscience, consiste dans un *ordre de sens* présupposé par toute manifestation d'intelligence de la part d'un sujet. C'est la voie intentionaliste : l'esprit est présent dans ses phénomènes, donc dans le monde, dans les pratiques symboliques et les institutions, et il n'y a littéralement, dans les têtes des gens, que les conditions personnelles, donc physiques (physiologiques), d'une participation à ces pratiques et à ces institutions. Le mental, lui, est là où il peut se présenter, donc dans le discours et dans l'action, dont les conditions d'existence ont un caractère holiste : c'est là le holisme anthropologique du mental, dont les thèses et les difficultés seront examinées dans la deuxième partie de cet exposé.

## 4. La nouvelle philosophie mentale

### 4.1. – PROSPECTUS DE LA NOUVELLE SCIENCE DE L'ESPRIT

Nous allons bientôt assister à la constitution, sur des bases enfin solides, d'une psychologie scientifique. L'événement de cette naissance est imminent. La preuve en est que nous commençons à comprendre comment on pourrait donner un modèle mécanique de la façon dont les choses se passent dans l'esprit de quelqu'un.

Tel est le communiqué que la philosophie mentale se prépare à publier depuis qu'elle travaille à égaler dans le domaine de l'esprit les succès de la philosophie naturelle dans le domaine des phénomènes physiques. On a assisté dans le passé à plusieurs épisodes d'annonce prématurée de l'émergence d'une psychologie scientifique. Il a fallu avouer ensuite que la théorie avait mis au monde une pseudo-explication (comme ce fut le cas avec la théorie de l'association des idées), ou que la doctrine d'abord présentée comme une grande découverte scientifique devait se prévaloir d'une autre autorité que de celle de la science (comme il est arrivé avec la psychanalyse). Aujourd'hui, le programme d'une psychologie scientifique porte le nom de cognitivisme. Dans ce chapitre, je ne considérerai qu'une seule chose : les raisons qui nous sont données de dire que ce programme renouvelle toutes nos perspectives philosophiques sur l'esprit.

Avant d'aller plus loin, il est nécessaire de distinguer entre le programme scientifique et la philosophie qui, selon les uns, soutient, ou, selon les autres, accompagne seulement ce pro-

gramme. Je suivrai l'usage recommandé par plusieurs auteurs en marquant une différence entre les *sciences cognitives* et le *cognitivisme*. Les sciences cognitives sont tout simplement les sciences de l'esprit. Je veux dire que les « sciences cognitives » qui figurent sur la liste du programme sont celles dont les recherches portent sur les conduites dans lesquelles les gens font preuve d'esprit. Mais le fait de donner le nom de « science cognitive » à une discipline, par exemple à la linguistique ou à l'anthropologie, signifie qu'on prend parti sur son objet et sa méthode : c'est ce parti pris qui est le cognitivisme. S'il suffisait de dire qu'une science va être qualifiée de « cognitive » parce qu'elle porte sur un objet intellectuel, par exemple le langage ou la culture, le mot de « cognition » serait entendu dans le sens ordinaire, pré-cognitivistique, étranger aux notions de traitement de l'information et d'intelligence artificielle<sup>2</sup>. En fait, lorsque le théoricien cognitivistique enregistre diverses disciplines déjà existantes dans cette rubrique des sciences cognitives, c'est pour proposer que ces disciplines se redéfinissent à la lumière d'un nouveau concept d'esprit. Ainsi, le cognitivisme est le programme visant à changer les sciences de l'esprit en sciences cognitives. Par exemple, la linguistique n'est pas citée parmi les sciences cognitives parce qu'elle étudie les langues, qui sont des systèmes intellectuels, mais parce qu'elle pourrait, nous dit-on, être redéfinie ou reconstruite comme une psycholinguistique : l'étude des capacités linguistiques d'un système mental doté de « organes » nécessaires à l'intelligence des phrases.

On pourrait croire que le nouveau concept d'esprit est demandé à la psychologie cognitive, celle qui a décidé de transgresser les interdits lancés par le béhaviourisme, en cherchant à savoir ce qui se passait dans la fameuse « boîte noire ». L'émergence de la psychologie cognitive est liée, comme on sait, à l'invention de l'ordinateur. Au fondement du cognitivisme, il y aurait donc une nouvelle psychologie, elle-même issue d'une innovation technologique. En réalité, le concept cognitivistique d'esprit est demandé, comme on va le voir, à une philosophie. Dès 1976, le psychologue

1. Voir par exemple la présentation de Daniel Andler à l'ouvrage publié sous sa direction : *Introduction aux sciences cognitives*, p. 12.

2. Dans le vocabulaire philosophique de la langue anglaise, on opposait souvent le cognitif et le conatif, ce qui correspond à notre distinction de la faculté de représentation et de la faculté de vouloir. En français, où le mot était inusité jusqu'à une date récente, « cognition » est ordinairement pris dans le sens cognitivistique.

Neisser mettait en garde son lecteur contre un enthousiasme excessif pour la comparaison entre nos capacités cognitives et celles de l'ordinateur. Il notait d'abord que le rôle de cette analogie avec la machine avait été de rassurer le psychologue sur l'existence d'une vie mentale à étudier. Qu'est-ce que le psychologue demande vraiment à l'ordinateur ? Ce qui compte, c'est que l'ordinateur travaille avec des informations, qu'il manipule des symboles, qu'il mette des éléments dans sa « mémoire », ou qu'il les y cherche, qu'il classe des messages, reconnaisse des formes physiques, etc. Peu importe d'abord si l'ordinateur procède comme nous lorsqu'il effectue de telles opérations. Pour le psychologue, le point décisif est que l'ordinateur soit en mesure de le faire. Comme l'écrit Neisser<sup>3</sup>, les psychologues ont été impressionnés par les performances de l'ordinateur parce qu'ils n'étaient pas sûrs d'avoir le droit de croire à la réalité des procès cognitifs (au sens pré-cognitivistique de ces mots). Les objections béhaviouristes avaient semé le doute. Mais, si un ordinateur a des activités de type cognitif, alors qu'est-ce qui empêche un être humain d'en avoir aussi ? L'ordinateur offre, comme on dit, un modèle. Or le fait que l'ordinateur soit une machine, un système physique, semble être la vertu essentielle de ce modèle : il n'est pas nécessaire à l'hypothèse cognitivistique que la psychologie humaine soit, dans ses démarches spécifiques, la même que celle de la machine à penser. L'important est qu'on soit assuré de ne pas quitter un seul instant le monde physique.

Qu'est-ce que le modèle a réellement apporté ?, se demande Neisser dans le même texte. Il répond : des techniques et des modèles dont le sens psychologique n'est pas évident, car elles concernent des systèmes étudiés en laboratoire. « On n'a toujours pas rendu compte de la façon dont les gens agissent ou de leur interaction avec le monde ordinaire. En fait, les présupposés qui soutiennent la plus grande partie de la recherche contemporaine sur le traitement de l'information ressemblent étonnamment à ceux de la psychologie introspective du XIX<sup>e</sup> siècle, toutefois sans l'introspection elle-même<sup>4</sup>. » Nous retrouvons l'idée plusieurs fois exprimée par Fodor : le cognitivisme est la réhabilitation de la théorie représentationniste de l'esprit. La différence est que la philosophie mentale renonce à la perspective de la première

3. *Cognition and Reality : Principles and Implications of Cognitive Psychology*, p. 6.

4. *Ibid.*, p. 7.

personne, ce qui lui permet de se donner pour une science théorique du comportement.

S'il en est ainsi, ce n'est pas d'abord aux sciences enregistrées comme cognitives qu'il faut demander de nous expliquer la portée de l'adjectif « cognitif » qui leur est appliqué, c'est aux philosophes que la question doit être posée. Si « révolution cognitive »<sup>5</sup> il y a, elle est surtout, à mes yeux, dans ce renversement : il appartiendrait à la philosophie de fournir elle-même, à partir de ses propres exigences, un concept d'esprit qui donnerait aux sciences de l'esprit le « paradigme » qu'elles doivent suivre pour se constituer<sup>6</sup>. La démarche de la philosophie, dans la discussion sur le statut des *Geisteswissenschaften*, était inverse : dans une philosophie des sciences morales telles qu'elles sont, le philosophe cherche à articuler un concept d'esprit qui soit capable d'embrasser les phénomènes de l'esprit tels qu'ils sont déjà décrits, classés, systématisés et ordonnés dans les sciences morales. Mais il est vrai qu'en demandant le concept d'esprit à la philosophie mentale, et pas aux sciences morales telles qu'elles sont effectivement pratiquées dès aujourd'hui, le cognitiviste ne fait que reproduire la démarche de la philosophie à l'époque conquérante de la « méthode scientifique », avant qu'il ne soit question d'un dualisme méthodologique.

Les introductions et initiations à la philosophie cognitiviste de l'esprit ne manquent pas. La partie destinée à éveiller notre intérêt pour le cognitivisme est en général rédigée selon un schéma en trois points. Il s'agit en effet de développer les trois idées suivantes :

1) Il est de nouveau permis d'avoir une vie mentale, c'en est fini du béhaviorisme.

2) Pourquoi le mentalisme est-il de nouveau respectable ? Parce qu'il ne peut plus être soupçonné de faiblesse pour le spiritualisme. Il se veut vigoureusement matérialiste et insiste sur le fait que son ontologie est moniste.

3) Comment a-t-on réussi à rendre son crédit au mentalisme ? C'est qu'on a enfin mis la main sur le bon modèle. Les philosophes mentaux d'hier avaient comparé par exemple l'esprit à une

5. Pour reprendre le titre de l'ouvrage d'introduction d'Howard Gardner, *The Mind's New Science : A History of The Cognitive Revolution*.

6. De façon plus discrète, le cognitiviste pourra se contenter de faire le tri entre les travaux de linguistique, de psychologie ou de sociologie qui entrent dans son « paradigme », et les autres qui ne sont pas assez solides pour pouvoir y prétendre.

horloge délicate, ou à une machine à vapeur. Mais il paraît difficile de parler de la psychologie d'une horloge ou d'un système thermodynamique. En revanche, on peut dire que la machine informatique fonctionne physiquement, mais on peut dire aussi qu'elle exécute un programme de calcul.

Bref, la philosophie mentale est devenue respectable depuis qu'elle peut donner des gages de son naturalisme et de son matérialisme. De ce point de vue, le cognitivisme se propose de réussir là où le béhaviorisme avait échoué, dans la tâche de naturaliser l'intentionnel. Les théoriciens béhavioristes n'avaient pas réussi à traduire les concepts psychologiques en simples types de comportement physique manifeste, donc *externe*. Mais le programme cognitiviste propose de les traduire en événements internes d'ordre physique (dont la théorie suppose la présence pour rendre compte du comportement visible)<sup>7</sup>.

Pour comprendre tout ce débat, il faut saisir le sens de la querelle sur le matérialisme en philosophie de l'esprit. Cette querelle est aussi vieille que la philosophie mentale, et elle en exprime l'aporie principale.

#### 4.2. – REMARQUE SUR LE MATÉRIALISME

En philosophie de l'esprit, le matérialisme est opposé au spiritualisme. Le spiritualisme soutient que les capacités mentales doivent être attribuées à une partie immatérielle de la personne (âme). La doctrine opposée soutient que toutes les parties de la personne sont matérielles.

Cette doctrine opposée est-elle le matérialisme ? On pourrait le croire puisque tous les manuels nous parlent d'une opposition entre ces deux grandes philosophies de l'esprit. Pourtant, les définitions qu'on donne souvent du matérialisme passent à côté du point le plus important : avant de savoir si les capacités mentales d'une personne doivent être attribuées à une partie immatérielle ou à une partie matérielle de cette personne, il faudrait savoir si ce n'est pas déjà une erreur d'attribuer les capacités du tout à une partie (matérielle ou non) de ce tout.

Si nous suivions la définition courante du conflit des deux philosophies de l'esprit, nous devrions caractériser ainsi les deux

7. Voir D. Dennett, *Content and Consciousness*, p. 39.

**positions** : le spiritualisme soutient la distinction réelle de l'âme et du corps, le matérialisme refuse cette dualité. Il est facile de voir que cette explication est inadéquate. Que nous dit en effet la définition ? Qu'il y a deux conceptions du *sujet* de la pensée : pour le spiritualiste, le sujet de la pensée est immatériel, pour le matérialiste, ce sujet est physique. Appliquons cette définition à l'épisode inaugural du *cogito*. Qui est le sujet pensant dans les *Méditations* de Descartes ? D'après le texte, ce n'est pas un sujet matériel : la réponse est donc spiritualiste. Par conséquent, ce sujet pensant n'est pas non plus Descartes, car l'individu nommé Descartes est certainement un sujet physique. Cet individu a une histoire physique. Le nom « René Descartes » a été donné à un nouveau-né, dans une cérémonie de baptême. En suivant la définition ci-dessus, nous arrivons donc à ceci : le spiritualisme soutient que ce n'est pas la personne qui pense, mais une partie immatérielle de cette personne (l'âme). Le matérialisme serait-il seulement la négation du spiritualisme ? On le croirait puisque, comme on nous le répète, la thèse matérialiste est la suivante : *la matière peut penser*. Or il est certain que Descartes, qui est assurément un être vivant, donc matériel, peut penser.

Pourtant, on découvre vite que le matérialiste ne se satisfait pas de la réponse : c'est Descartes qui pense, or Descartes est un sujet physique (pas un ange, une substance immatérielle), donc le morceau de matière qui s'appelle Descartes peut penser<sup>8</sup>. En fait, le matérialisme n'est pas de soutenir que Descartes pense, il est de soutenir que l'*esprit* de Descartes est matériel. Le sujet qui pense, une fois de plus, n'est pas Descartes, c'est l'esprit matériel de Descartes (son cerveau).

Comme ce point est le plus souvent passé sous silence, je crois qu'il n'est pas inutile de l'articuler plus en détail. On nous parle d'un conflit du « dualisme » et du « matérialisme ». S'il suffisait

8. Comme l'écrit Wittgenstein : « Une opinion est un état – Un état de quoi ? De l'âme ? De l'esprit ? Eh bien, de quoi dit-on qu'il a une opinion ? On le dit, par exemple, de Monsieur N.N. Et c'est cette réponse qui est la bonne » (*Recherches philosophiques*, § 573). Il ajoute : on ne doit pas s'attendre à être éclairé par cette réponse, même si elle est irréprochable, à la question. C'est justement l'opinion du matérialiste : par une telle réponse, dira-t-il, on n'a encore rien expliqué. Oui, mais le but n'était pas d'expliquer quoi que ce soit, il était d'indiquer, par une remarque grammaticale, que les questions à poser sur l'opinion sont des questions à poser sur Monsieur N.N., ses sources d'informations, la qualité de son jugement, etc., mais non pas des questions sur quelque chose dont on devrait inspecter l'état, comme on inspecte l'état de la toiture de sa maison.

de s'opposer au dualisme pour être matérialiste, le conflit serait entre les deux conceptions que voici :

(I) Lorsque Descartes écrit qu'il va supposer que toutes les choses qu'il voit sont fausses<sup>9</sup>, c'est Descartes qui accomplit l'acte de faire cette supposition ; de façon générale, lorsque Descartes rapporte dans son texte qu'il accomplit un acte mental, le sujet de cet acte est Descartes.

(II) Lorsque Descartes écrit qu'il va supposer que toutes les choses qu'il voit sont fausses, c'est Descartes qui l'écrit (de sa main sur le papier), mais c'est l'esprit de Descartes qui fait la supposition et c'est l'esprit de Descartes qui voit toutes les choses dont il est dit, sous la plume de Descartes, qu'elles sont vues par lui ; de façon générale, nous donnons le nom de Descartes à l'homme, donc à un sujet physique, sans nous occuper de la distinction faite par le sujet pensant entre ce qu'il juge maintenant être (*res cogitans, id est, mens, sive animus, sive intellectus, sive ratio*) et ce qu'il croyait être auparavant (quelqu'un qui a un visage, des mains, des bras, etc., et qui se nourrit, marche, et ainsi de suite).

Or l'opposition des conceptions (I) et (II) ne restitue pas le point du conflit entre les matérialistes et les spiritualistes. Les théoriciens cognitivistes ne nous décerneront pas un certificat d'orthodoxie matérialiste si nous nous contentons de dire qu'en effet toutes les *parties* de la personne de Descartes sont matérielles. En rejetant ainsi l'idée d'une « union de l'âme et du corps », nous pourrions bien réserver un rôle à la façon dont les parties sont intégrées, comme des parties vivantes, dans l'unité d'un tout. Autrement dit, nous pourrions bien avoir une arrière-pensée aristotélicienne, et vouloir distinguer, non pas une partie matérielle et une partie immatérielle de la personne, mais des parties matérielles de la personne et la forme selon laquelle ces parties sont les parties d'un tout. Or le philosophe matérialiste de l'esprit tient à pouvoir parler d'une interaction de l'esprit et du corps. Sa position est une *réponse* à la question de la dualité de l'âme et du corps, nullement un *refus* des termes de la question classique. Pour exprimer la position du matérialiste, il faudrait donc dire :

(III) Lorsque Descartes écrit qu'il va supposer que toutes les choses qu'il voit sont fausses, c'est la main de Descartes qui l'écrit,

9. « Je suppose donc que toutes les choses que je vois sont fausses » (*Méditation Seconde*, p. 71).

mais c'est le cerveau de Descartes qui le pense et qui par là cause cette activité d'écriture ; c'est donc au cerveau de Descartes qu'il faut attribuer cette pensée, et c'est d'ailleurs aussi au cerveau de Descartes qu'il faut attribuer la vision des choses, car c'est le cerveau de Descartes qui construit les représentations mentales de l'expérience visuelle.

Il y a donc en réalité deux étapes dans la querelle du dualisme. Dans la première étape, on répond à la question : Y a-t-il lieu de distinguer le sujet des actions physiques (je me nourris, je marche) du sujet des opérations mentales (je vois, je pense) ; autrement dit, y a-t-il lieu de distinguer le sujet mental de la personne ? Si vous répondez par la négative, vous considérez que le vocabulaire psychologique inclut les activités intentionnelles immanentes (comme de réfléchir à quelque chose) et aussi les actions intentionnelles transcendant (comme de chercher à trouver quelqu'un dans une foule). Vous avez donc déjà refusé les perspectives de la philosophie mentale. En revanche, si vous répondez par l'affirmative, vous devez passer à la deuxième étape du problème, celle dans laquelle vous avez à vous prononcer sur la nature de la dualité de l'esprit et du corps : la dualité de l'esprit et du corps doit-elle être comprise à l'aide d'un dualisme des genres de substance (l'âme et le corps) ou doit-elle trouver une formulation dans un monisme ? Et c'est donc seulement à cette deuxième étape que s'opposent le mentaliste cartésien, qui veut distinguer le sujet immatériel des pensées du sujet matériel des mouvements physiques, et le mentaliste matérialiste, qui veut distinguer le système matériel de l'appareil mental (cerveau) et le système matériel du comportement extérieur (corps).

En ce point, il est vraisemblable qu'une objection sera faite, peut-être sur le ton lassé de quelqu'un qui n'en finit pas d'avoir à rappeler des évidences : il faut bien, dira-t-on, passer à la deuxième étape, celle de savoir si l'organe de la pensée est matériel ou immatériel. Tout le monde doit se déclarer matérialiste ou spiritualiste : si vous prétendez qu'on peut penser sans un cerveau, vous êtes spiritualiste, si vous reconnaissez qu'on ne le peut pas, vous êtes matérialiste comme nous.

Poser cette question, c'est une fois de plus nous demander de confondre les capacités de Pierre ou de Paul (les personnes) et les conditions (externes et internes) dans lesquelles Pierre ou Paul possèdent ces capacités. Qu'il s'agisse des capacités mentales de quelqu'un ne change rien à l'affaire. Paul peut-il penser sans un

cerveau ? La question porte sur les circonstances dans lesquelles il est permis de dire que quelqu'un possède actuellement telle ou telle capacité. Si le cerveau de Paul est détruit, Paul perd définitivement la capacité de penser, en même temps que la capacité d'être Paul. Si le cerveau de Paul est temporairement mis dans un état d'hibernation, Paul, de nouveau, perd la capacité de penser : par exemple, il n'est plus capable de répondre à son courrier. Mais les conditions dans lesquelles nous pouvons faire quelque chose ne sont pas ces capacités : le fait que Paul cesse d'être Paul si on le prive de son cerveau ne prouve pas que Paul soit un cerveau, ni que ce soit en réalité son cerveau, et non lui, qui soit le sujet d'attribution des prédicats psychologiques. Car c'est bien Paul, et non son cerveau, qui répond au courrier que reçoit Paul<sup>10</sup>.

Dans une page enjouée<sup>11</sup>, Peirce explique qu'une des raisons qui font qu'on loge l'esprit dans la tête est qu'on identifie l'esprit à la conscience. Or la conscience, au sens le plus ordinaire du mot, se manifeste dans le fait de ressentir ou d'éprouver en soi des impressions, lesquelles varient selon les changements du milieu externe. C'est pourquoi on croit que la place de l'esprit est dans le système nerveux. En revanche, pour Peirce, l'esprit ne doit pas être cherché dedans, dans des événements intérieurs du genre expérientiel : il est dehors. A ceux qui font de la conscience, prise au sens d'un « sentiment » (*feeling*), un attribut essentiel de l'esprit, Peirce répond ceci : « le sentiment n'est rien que l'aspect intérieur des choses, alors que l'esprit, au contraire, est essentiellement un phénomène extérieur ». Peirce veut dire que les phénomènes mentaux ne doivent pas être cherchés dans la tête, mais à l'extérieur, par exemple dans la conduite intelligente ou dans la communication par signes. Les psychologues auxquels pense ici Peirce sont ceux qui ont tenté de localiser les facultés mentales dans telle ou telle partie du cerveau. Ils disent par exemple qu'on a prouvé que la faculté du langage était précisément localisée dans un certain lobe cérébral. La preuve : un psychologue coupera un lobe de

10. Ce point a été notamment mis en lumière par Anthony Kenny (cf. « The Homunculus Fallacy » et « Language and the Mind » dans : *The Legacy of Wittgenstein*). Il sera repris plus loin (cf. *infra*, chap. 8).

11. *Collected Papers*, t. 7, § 364. Voir sur cette page une étude de Christiane Chauviré qui compare Peirce et Wittgenstein : « Quand savoir, c'est (savoir) faire », *Critique*, LXV, n° 503, avril 1989, p. 282-299.

mon cerveau et, observant que je me découvre incapable de m'exprimer, me dira : « Vous voyez bien que votre faculté du langage était localisée dans ce lobe. » Cela revient à exiger que la capacité de faire quelque chose, ici de parler, soit localisée dans une partie du corps dont on montre qu'elle est indispensable à l'exercice de cette capacité. Peirce refuse tout simplement la preuve. Car, dit-il, si le psychologue lui avait dérobé son encrier, il se serait, là encore, trouvé incapable de s'exprimer et de poursuivre la discussion avant d'en avoir acquis un autre. *Yea, the very thoughts would not come to me*<sup>12</sup>. Ainsi, la faculté du langage était également présente dans l'encrier, de telle sorte qu'elle apparaît sous un signalement ontologique remarquable : le pouvoir du langage serait capable de se trouver en même temps à deux endroits différents.

Il vaut mieux dire, explique encore Peirce, que le langage réside dans la langue (au sens de l'organe), et que les pensées d'un auteur vivant sont dans les exemplaires de ses livres, plutôt que de dire que tout cela a son siège dans le cerveau.

Le psychologue auquel s'adresse Peirce répondra peut-être qu'il est plus facile de remplacer un encrier qu'un cerveau. Ce dernier a donc une place plus centrale. Cela va de soi : mais en disant cela, on admet qu'on est en train de ranger les conditions empiriques de la possession de capacités dans un ordre croissant de nécessité (au sens de besoin). Les conditions physiologiques des phénomènes psychologiques sont dans le cerveau : cela, c'est la recherche scientifique qui nous l'a appris, ce n'est pas la « neurophilosophie », c'est-à-dire la prétention à étudier le cerveau par les voies de la spéculation philosophique<sup>13</sup>. Mais ces phénomènes de l'esprit ne sont des phénomènes *de l'esprit* qu'au-dehors, dans le monde public.

12. « Je dirais même plus, il ne me viendrait pas la moindre idée » (*Collected Papers*, t. 7, § 367). Dans le même sens, Wittgenstein écrit qu'il est égarant de dire qu'on pense « dans sa tête » ou « avec sa tête ». Tout ce que cela signifie est qu'il y a un rapport ou une connexion entre la pensée et la tête : mais en même temps, l'expression employée suggère que la tête abrite des processus invisibles du dehors, et toute une philosophie de la pensée se met en place à partir de cette image (*Grammaire philosophique*, § 64).

13. On se gardera de confondre la « neurophilosophie » et la réflexion philosophique du neurologue sur sa discipline.

## 4.3. – LA CHARTE DE LA NOUVELLE PHILOSOPHIE MENTALE

Qu'est-ce qui peut bien rendre plausible, aux yeux d'un contemporain, la conception mentaliste ? Pour le comprendre, il convient de voir comment les mentalistes eux-mêmes s'expriment. On peut partir d'un texte de Fodor qui reproche aux mentalistes la timidité dont ils feraient preuve dans l'affirmation de leur propre doctrine. Fodor est à bien des égards le représentant d'un mentalisme *ultra* faisant campagne parmi des mentalistes *juste milieu*. Comme toujours, l'*ultra* offre une version plus cohérente de la position adoptée (puisque'il ne se soucie pas de faire des concessions et de passer des compromis avec la réalité, ou avec les apparences enregistrées par le sens commun).

S'adressant à ses collègues mentaliste<sup>14</sup>, Fodor résume leur dogme en trois points (avant de les inviter à y ajouter la thèse proprement fodorienne du « langage de la pensée »). Voici les trois articles du dogme mentaliste.

1) Le premier article pose la réalité du mental : « Il y a des croyances et des désirs, et il y a un point de fait touchant leur contenu intentionnel ; c'est-à-dire qu'il y a un point de fait touchant la question de savoir quelle est la proposition qui est l'objet intentionnel d'une croyance ou d'un désir. »

La première partie de l'article premier est une assertion d'existence par laquelle la psychologie reçoit un domaine propre. Cette thèse ontologique semble aller de soi : si la psychologie a une réalité à étudier, c'est qu'en effet les gens ont des croyances et des désirs.

La deuxième partie de ce premier article est exprimée de façon plus hermétique, et demande à être quelque peu développée. Fodor veut dire qu'on ne se contente pas d'attribuer aux gens de la croyance et du désir, mais qu'on leur attribue des croyances particulières, identifiables, et des désirs particuliers, susceptibles d'être distingués. Comment fait-on pour identifier des désirs et des croyances ? La réponse semble, ici aussi, triviale. Pour distinguer les croyances particulières de quelqu'un, nous faisons comme s'il les exprimait dans un *credo*, c'est-à-dire que nous lui donnons une attitude mentale de croyance à un certain nombre d'objets de croyance, appelés « objets intention-

14. « Why There Still Has to Be a Language of Thought », appendice à *Psychosemantics : The Problem of Meaning in the Philosophy of Mind*, p. 135.

nels » de l'attitude d'esprit consistant à croire. Ces objets intentionnels sont tout simplement les différents articles du *credo* auquel la personne considérée souscrit ou pourrait souscrire. Ainsi, cette question de savoir quel est le contenu de la croyance de quelqu'un est en principe une question de fait : il suffit qu'on ait l'occasion de demander au sujet considéré s'il approuve telle proposition.

Bref, l'article premier part d'un postulat difficile à refuser : *les gens ont des croyances et des désirs*, il en tire une assertion d'existence : *il y a des croyances et des désirs*, et il infère de là une philosophie de l'esprit qui tient dans la thèse : *les gens ont des relations à des phrases exprimant ce qu'ils croient et ce qu'ils désirent*.

2) Le deuxième article nous dit qu'une théorie matérialiste de l'esprit est concevable : « Il n'est pas exclu en principe que croire et désirer se découvrent être des états du cerveau. » Autrement dit, l'hypothèse matérialiste est intelligible. Nous comprenons ce qu'elle dit.

L'article ajoute que la philosophie ne peut pas se substituer à la recherche empirique (ici, à la neurologie). « C'est une question empirique de savoir comment travaille le système cérébral de quelqu'un lorsqu'il passe par tel ou tel état mental. » Autrement dit, le philosophe n'a pas à préférer une hypothèse empirique à une autre, car une argumentation philosophique ne peut pas nous dispenser d'aller voir sur le terrain comment les choses se passent en fait <sup>15</sup>.

3) Le troisième article pose que les entités mentales, dont la réalité vient d'être posée, ont une fonction causale : « Les croyances et les désirs ont un rôle causal, ou encore, le comportement manifeste est l'effet d'interactions complexes entre des causes mentales. » C'est la thèse à laquelle on peut plus spécialement réserver la qualité de *mentaliste*. Il va de soi que c'est cette thèse qui était visée depuis le début et que les deux premières ont préparé.

On observe dans cette page quelque chose qui relève de la

15. On ne peut qu'approuver cette sage restriction de la compétence du philosophe aux questions de sens et d'intelligibilité (tout en doutant peut-être que la restriction soit entièrement respectée dans un programme qui, tel qu'il est le plus souvent présenté, inclut la philosophie parmi les sciences cognitives et demande aux philosophes de construire des hypothèses et des modèles empiriques).

manœuvre persuasive, voire de l'art du bonimenteur <sup>16</sup>. D'abord, le dogme mentaliste nous est présenté comme étant d'une extrême banalité : comme s'il équivalait à la simple reconnaissance de l'existence d'une dimension psychologique des affaires humaines. Qui irait nier que les gens aient des opinions et des désirs (article 1), sinon le personnage démodé du béhavioriste borné dont tout le monde se moque ? Quel obscurantiste irait refuser l'intérêt pour la psychologie des recherches neurologiques sur les troubles du comportement ou de l'activité mentale (article 2) ? Qui refuserait la platitude : les gens agissent en fonction de ce qu'ils croient savoir et de ce qu'ils veulent obtenir (article 3) ? Mais, au bout du compte, le lecteur a la surprise d'apprendre qu'en accordant ces vérités peu contestables il a accepté les uns après les autres les éléments d'une métaphysique de l'esprit <sup>17</sup>, et d'une métaphysique qui va lui imposer des révisions radicales de ses conceptions.

Considérons les trois articles tour à tour. Loin de se borner à exprimer des platitudes sans grande portée, chacun d'eux nous demande de remplacer un énoncé banal par la formule mentaliste canonique grâce à laquelle la psychologie pourra accomplir le prodige annoncé : *naturaliser l'intentionnalité*. Or cette formule canonique, puisqu'elle s'écarte nettement de nos façons ordinaires de parler de la psychologie des gens, nous avons le droit de demander qu'on nous l'explique et qu'on nous la justifie.

*Article premier* (sur la réalité du mental). Nous partons d'une proposition qui, semble-t-il, n'engage à rien du point de vue de la philosophie. Nous affirmons que les gens ont des opinions et des désirs. Cette proposition initiale pourrait être une réaction exprimant notre surprise et notre incrédulité devant ce que disaient les béhavioristes. C'est ensuite seulement que commence la mise en forme métaphysique. L'article mentaliste nous demande ici

16. « D'où vient-il qu'on se pose le problème philosophique des procès et des états mentaux, et du béhaviorisme ? Le premier pas est celui qui passe entièrement inaperçu. Nous parlons de procès et d'états, et laissons leur nature indécidée ! Un jour, peut-être, nous en saurons plus à leur sujet – pensons-nous. Mais, par là même, nous nous sommes engagés dans une façon déterminée de traiter le sujet. En effet nous avons un concept déterminé de ce que cela veut dire que d'apprendre à mieux connaître un procès. (Le pas décisif du tour de passe-passe a déjà été fait, et c'était justement celui qui nous a paru innocent.) » (Wittgenstein, *Recherches philosophiques*, § 308).

17. Dans le présent contexte, le mot « métaphysique » peut donner lieu à un malentendu. Voir la note en fin de chapitre sur l'usage qui est fait ici de ce terme.

d'adopter la tournure canonique suivante : quand quelqu'un a une opinion, par exemple quand il croit que le gouvernement perdra les élections, il est dans un état qui consiste à avoir une attitude à l'égard d'un « objet intentionnel », à savoir la représentation du gouvernement comme étant destiné à perdre les élections. Cet objet avec lequel il entre en rapport est une proposition. Autrement dit, à la fin du premier article, nous avons souscrit, sans peut-être l'avoir remarqué, à une doctrine des « attitudes propositionnelles » : croire quelque chose, c'est être dans un état consistant à avoir une certaine *relation* (celle du « croire que », plutôt par exemple que celle du « douter que » ou celle du « se demander si ») envers une certaine entité, la proposition donnant l'objet représenté. Or le point stratégique de toute la discussion à venir va justement être la notion d'*état d'esprit*.

*Article second* (sur la possibilité d'assimiler des états d'esprit à des états du cerveau). La mise en forme mentaliste de l'expression a mobilisé la catégorie des états pour redécrire les croyances. La traduction du mental au physique semble donc en bonne voie. Croire quelque chose est un état de l'esprit. Croire quelque chose *pourrait* bien être un état du cerveau. Ici, le matérialisme de la thèse est peut-être ce qui attire aussitôt l'attention, tel un chiffon rouge agité devant des philosophes peu désireux d'être soupçonnés de spiritualisme. Mais le point décisif est ailleurs, il est de nous faire prendre l'attribution à quelqu'un d'une croyance et d'un désir (ce monsieur qui court vers le quai de la gare croit être en retard, il désire prendre son train) pour l'attribution d'états (il est dans l'état de croire être en retard, il est dans l'état de désirer prendre son train). La notion d'état mental est ainsi introduite – sans avoir été annoncée comme telle – *dans un sens comparable à celle des états du cerveau* : le but était justement de préparer l'idée d'une psychologie qui découvre des lois de l'esprit, donc des séquences régulières dans la succession des états de l'esprit.

On remarquera qu'une fois accompli ce travail de préparation métaphysique du terrain, la doctrine du « langage de la pensée » trouve un rôle à remplir, et reçoit un air de vraisemblance. Cette doctrine répond à la question de savoir ce que pourrait bien être un « état d'esprit », en ajoutant au dogme mentaliste ordinaire l'idée que le cerveau de quelqu'un qui a une opinion, dont le contenu peut être énoncé par une phrase, va contenir à son tour

un symbole matériel correspondant, constitué comme une phrase selon les règles d'écriture du « code cérébral ».

*Article troisième* (sur la fonction causale des croyances et des désirs). Ici encore, il y a un saut entre le point de départ qui paraît n'engager à rien et la conclusion. Si ce voyageur est en train de courir vers le quai de la gare, c'est *parce qu'il* craint d'être en retard pour prendre son train : tel est le point de départ, que personne ne contestera vraiment, à moins d'avoir un système idéologique à défendre. Qui dirait que rien ne serait changé au comportement de ce voyageur si nous l'informions que son train est déjà parti, ou bien que son train, à la suite d'un retard, n'est pas encore arrivé en gare ? Il faut donc bien que les croyances fassent une différence. Ici, la formulation conforme au dogme mentaliste donne : si les jambes du voyageur se meuvent rapidement de façon à lui donner le pas de course, c'est parce qu'il s'est produit dans la tête du voyageur une interaction complexe entre ses croyances relatives à l'horaire du train, l'heure qu'il est, le temps qu'il faut pour franchir la distance au but, et son désir prédominant de ne pas rater le train. (Ici, la doctrine du langage de la pensée ajoute que le rôle causal d'une réalité mentale, par exemple de la présence d'une croyance dans l'esprit de quelqu'un, doit être assigné à ce symbole physique inscrit dans le cerveau.)

L'ensemble des trois articles donne la substance de la philosophie mentaliste. Fodor conclut : « une explication psychologique doit postuler un réseau de relations causales entre des états intentionnels ». C'est ainsi que nous sommes passés des platitudes du sens commun à l'hypothèse théorique d'une machinerie mentale. La première thèse est ontologique : il y a des croyances et des désirs. Mais dire cela ne sert qu'à repousser le tigre de papier d'une doctrine qui dirait qu'il n'y en a pas. En réalité, l'important est dans la thèse métaphysique associée : une croyance porte sur une proposition, donc finalement sur une phrase. La deuxième thèse relève en apparence de la prospective scientifique, mais vise en réalité à modeler la métaphysique de l'esprit sur la métaphysique des systèmes physiques qui passent par des états. La troisième thèse est directement psychologique : pour qu'il y ait psychologie dans le sens du mentaliste, il faut qu'on accepte de diviser le phénomène psychologique en une partie purement

extérieure et une partie purement intérieure (celle qu'on doit postuler pour expliquer la première).

En ce point, il convient de faire le partage entre le classique et le nouveau. Une bonne façon de le faire est de souligner la parenté de la « révolution cognitive » d'aujourd'hui et des idées des philosophes du XVII<sup>e</sup> siècle. Chomsky, dans un exposé introductif<sup>18</sup>, va jusqu'à parler de deux révolutions cognitives, l'une chez Descartes et ses successeurs (elle met fin au règne de la conception scolastique), l'autre de nos jours (elle abolit l'hégémonie du béhaviourisme). La première de ces révolutions met en place la « théorie représentationniste de l'esprit »<sup>19</sup>. Ainsi, Descartes explique que le sujet de la perception visuelle est comme un aveugle muni d'un bâton. Lorsque cet aveugle touche par exemple un cube avec sa canne, il se construit une représentation mentale de ce cube. La perception du cube résulte du procès cognitif de traitement des données sensibles qui aboutit à l'image mentale. La deuxième révolution cognitive est contemporaine du développement de l'informatique : elle permet de proposer des modèles plausibles des activités mentales et de leurs structures, en les présentant comme des « procès cognitifs », c'est-à-dire des procès de calcul sur des symboles internes au système.

Mais qu'est-ce qui permet à Chomsky de dire qu'en étudiant les capacités linguistiques de la personne il étudie son *cerveau* ? Un point est sûr : il étudie les capacités d'un système physique, car il faut être un système physique pour parler et pour entendre. Mais est-ce que les capacités linguistiques dépendent de structures cérébrales ? C'est bien possible, mais le linguiste, considéré comme linguiste, n'en sait rien. Lorsque Chomsky parle donc d'étudier le cerveau du point de vue des capacités cognitives responsables de l'exercice du langage, il se place, comme il le dit lui-même, à un haut niveau d'abstraction. En fait, ce niveau d'abstraction est si élevé qu'il faudrait comprendre par le mot « cerveau » quelque chose comme : la partie de l'organisme, quelle qu'elle soit, qui est le support des capacités considérées. La neurologie est donc ici saluée de très loin. En fait, rien ne serait changé à la grammaire générative si l'on apprenait que, tout

18. « On the Nature, Use and Acquisition of Language » (conférence de 1987).

19. *Op. cit.*, p. 631.

compte fait, ce n'est pas le cerveau qui est en cause dans l'activité linguistique, mais plutôt les reins et le cœur.

Est-ce que la philosophie représentationniste de l'esprit a été réhabilitée ? C'est la question qu'il faut maintenant poser. Il convient pour cela de prendre les articles de la nouvelle philosophie mentale dans un ordre analytique.

Le but de toute l'opération est de permettre à la psychologie de formuler des explications causales. Mais l'explication psychologique, au sens courant du terme, a-t-elle besoin d'états mentaux et de représentations internes ? C'est le problème de la *causalité psychique* (cf. *infra*, chap. 5).

La seconde révolution cognitive se flatte de rendre intelligible l'idée que les états mentaux sont des états physiques, et plus précisément des états du cerveau de l'organisme. Lorsque nous pensons, nous faisons ce que fait l'ordinateur quand il calcule, et personne ne doute que l'ordinateur soit une machine entièrement physique. Ici se pose le problème de la *psychologie des ordinateurs* (cf. *infra*, chap. 6-8).

Les deux révolutions cognitives ont un point commun : l'idée que croire et désirer soient des états mentaux internes. Si l'on adoptait cette hypothèse, on devrait, au moins en principe, s'y prendre tout autrement pour déterminer ce que les gens croient et désirent (cf. *infra*, chap. 9). Or c'est la métaphysique des états d'esprit qui impose à la philosophie mentale, l'ancienne comme la nouvelle, son orientation solipsiste (cf. *infra*, chap. 10-11). Mais on ne saurait se défaire de ce solipsisme qu'en formulant une version satisfaisante du « holisme du mental » (cf. *infra*, chap. 12), et c'est ce qui sera tenté dans la deuxième partie de cet ouvrage.

#### NOTE SUR LE CONCEPT DE MÉTAPHYSIQUE

Dans le contexte d'une philosophie de l'esprit, le mot « métaphysique » peut prêter à confusion. En effet, on pourrait croire qu'une discussion a plus de chance de devenir « métaphysique » si elle porte sur l'esprit – réalité dont le statut semble problématique, peut-être suspect, puisque intangible ou caché – que s'il s'agissait de choses terre à terre comme les choux-fleurs, la fonction glyco-génique du foie, ou les règles du tennis de table (pour prendre des exemples dans des catégories variées). Ce serait

là une erreur. Le sens utile du mot « métaphysique » est celui dans lequel la réflexion porte, non pas sur des entités d'ordre spéculatif, mais sur le statut, dans notre système conceptuel, de classes d'entités parfaitement familières. Il n'est d'ailleurs pas nécessaire, pour poser ces questions, d'user du mot même de « métaphysique ». C'est ainsi que Wittgenstein, qui fait rarement appel au vocabulaire traditionnel des philosophes, et parle plus volontiers de « grammaire », illustre pourtant fort bien la teneur d'une enquête métaphysique par un exemple qu'il donne d'une différence entre des catégories. L'occasion de sa comparaison est fournie par la confusion que favorise le fait de parler de différents genres d'objet (*kinds of objects*), là où il faudrait introduire une différence plus radicale que la simple différence générique. Par exemple, lorsque le théoricien des données sensibles distingue entre des objets de genres différents, l'arbre physique et l'arbre phénoménal. Wittgenstein explique ainsi la confusion : « Ils pensent qu'ils font une déclaration du genre "Un train de chemin de fer, une gare de chemin de fer et une voiture de chemin de fer sont différents genres d'objets", alors que leur déclaration est analogue à "Un train de chemin de fer, un accident de chemin de fer et un règlement de chemin de fer sont différents genres d'objets" »<sup>20</sup>.

Il est vrai que le mot de « métaphysique » est pris en plusieurs sens chez les philosophes, et ce pourrait être une raison pour lui préférer des termes moins chargés d'histoire. Toutefois, il est difficile de s'en passer, et d'ailleurs, l'« éthique de la terminologie » (Peirce) nous prescrit de nous en tenir, dans la mesure du possible, au vocabulaire le plus propre à faciliter la communication dans la profession, et donc, chaque fois que la chose est possible, au sens traditionnel des termes.

Dans un sens (historiquement dérivé) du terme, le mot « métaphysique » s'applique à l'immatériel, au supra-sensible, à une réalité dont l'existence n'aurait pas de conditions physiques. Si le mot était toujours pris en ce sens, on devrait dire que le spiritualiste a une métaphysique, mais que le matérialiste n'en a pas. On voit aussitôt l'inconvénient du point de vue de la terminologie : comment qualifier les déclarations du matérialiste qui nous parle de son « ontologie physicaliste » ? C'est pourquoi je crois qu'il

20. *The Blue and Brown Books*, p. 64.

faut éviter tout emploi du mot dans une acception *dogmatique*, et ne retenir que l'acception *opératoire*.

Cette acception opératoire doit être neutre, ne privilégier aucune école, et n'en exclure aucune. Il me semble qu'on la trouve dans l'une des acceptions dont nous héritons. Toute philosophie doit traiter, comme le fait le traité d'Aristote qui a reçu le titre de « métaphysique », de grandes distinctions conceptuelles comme celles de l'actuel et du potentiel, ou de l'existence indépendante (substance) et de l'existence dépendante (accident). Le fait pour un philosophe de réfléchir sur ces distinctions ne lui donne encore aucune affiliation doctrinale : le sens du mot qui qualifie sa réflexion n'est donc pas dogmatique. En revanche, un théoricien qui éviterait soigneusement d'avoir à entrer dans l'examen de ce genre de distinctions montrerait par là qu'il ne veut pas s'engager sérieusement sur les chemins de la philosophie. Il importe peu que cet examen soit *appelé* « métaphysique » ou autrement : ce qui compte est qu'il y ait place dans une pensée pour une mise en lumière des différences de catégorie sur lesquelles repose notre intelligence d'un domaine d'enquête ou de discussion.

Quelques références suffiront à montrer que l'acception opératoire est tout aussi vivante en philosophie que l'acception dogmatique. Le lexique de Descartes offre ici un exemple privilégié. En principe, c'est l'acception dogmatique qui domine chez lui : les *Méditations* sont métaphysiques parce qu'elles portent sur des choses immatérielles (Dieu et l'âme). Ici, on pourrait pourtant faire observer que l'objet des *Méditations* n'exclut pas le physique, puisque, s'il y est question de l'âme, c'est pour établir la *distinction réelle* de l'âme et du corps. Et s'il y est question de Dieu, c'est pour démontrer qu'il y a une et une seule Cause universelle des choses, ce qui fait que la métaphysique cartésienne pose les principes de la *physique*. Du point de vue opératoire, ce qu'il y a de métaphysique dans les *Méditations* de Descartes, c'est qu'elles parviennent à leurs conclusions par des raisons qui roulent entièrement sur l'emploi de concepts tels que ceux de « distinction réelle » (par opposition à « distinction de raison » et à « distinction modale ») et ceux de « réalité formelle » (par opposition à « réalité matérielle » et à « réalité objective »). Dans la Préface des *Principes de la philosophie*, Descartes use bien du terme de « métaphysique » dans le sens d'« immatériel », mais il en fait aussi un usage plus proche de ce que j'ai appelé l'acception

opérateur. En effet, il distingue une partie de la philosophie qui contient « les principes de la connaissance » d'avec une autre partie qui contient « les principes des choses matérielles ». La première est la *métaphysique*, la deuxième est la *physique*. Mais les principes de la connaissance dont il s'agit ici sont les « notions premières » et les « vérités éternelles ». Or nous savons que les notions premières (notions qui entrent dans la composition de toutes nos pensées, quel que soit leur objet) se rangent pour Descartes en quatre classes<sup>21</sup> : 1) Notions générales (être, nombre, durée, substance, accident, mode) ; 2) Notions particulières à la considération des corps (extension, figure, mouvement, disposition des parties, divisibilité) ; 3) Notions particulières à la considération des esprits (pensée, perceptions de l'entendement, inclinations de la volonté) ; 4) Notions particulières à la considération du composé de l'âme et du corps (appétits, passions, sentiments). Parmi les « notions communes » entrent aussi les vérités du genre « l'effet est semblable à la cause » ou « il est impossible qu'une chose soit et ne soit pas ». On retrouve donc bien l'idée selon laquelle l'enquête métaphysique porte sur les grandes distinctions conceptuelles (plutôt que sur un domaine particulier de la réalité).

En somme, l'acception opératoire du terme technique de « métaphysique » se retrouve chaque fois qu'un philosophe doit fixer d'avance les possibilités de classification ou d'explication qui sont offertes par un système conceptuel général. Par exemple, Kant, dans son « exposition métaphysique » des concepts d'espace et de temps<sup>22</sup>, fait appel à une division classique des catégories (avant d'innover en ajoutant une quatrième possibilité imprévue), lorsqu'il se demande si l'espace et le temps sont des choses (substances), ou s'ils sont des déterminations des choses (qualités), ou s'ils sont des rapports entre les choses. Plus près de nous, Russell indique quelles sont les « convictions métaphysiques » qui s'expriment dans le symbolisme proposé dans les *Principia Mathematica* : « Ce symbolisme suppose qu'il y a des "choses", qu'elles ont des propriétés, et qu'elles ont aussi des relations à d'autres "choses" (...) J'ai employé des lettres latines minuscules pour les "choses", des lettres grecques minuscules pour les classes, et des lettres latines majuscules pour les rela-

21. Cf. Lettre à Elizabeth, 21 mai 1643. Ces notions font l'objet des articles 48 à 67 dans la 1<sup>re</sup> partie des *Principes de la philosophie*.

22. *Critique de la raison pure, Esthétique transcendentale*, § 2.

tions »<sup>23</sup>. Le sens que j'ai appelé opératoire est celui dans lequel le mot est aujourd'hui couramment utilisé dans les écrits philosophiques de langue anglaise.

Au XIX<sup>e</sup> siècle, les questions relatives au statut des sciences de l'esprit étaient tenues pour des questions de méthode. On en discutait dans ce qu'on appelait « la logique des sciences », en entendant par logique l'étude des formes d'explication. C'est ainsi que le positivisme a été contesté, sur le terrain méthodologique, par l'herméneutique. Mais cette vue du problème est trop courte : elle appartient elle-même à une conception positiviste de la science. Derrière cette réduction de la question de l'esprit à des questions portant sur la méthode, on trouve le préjugé selon lequel l'activité de la science consisterait à confectionner des théories générales compatibles avec les faits observés. Cette philosophie de la science est insuffisante, car il est bien connu qu'on peut toujours offrir plusieurs théories qui rendent compte des mêmes faits, et qui sont donc équivalentes du point de vue logique. Comment départager ces théories ? L'épistémologie positiviste est obligée d'ajouter des « contraintes » supplémentaires, par exemple d'ordre esthétique, voire économique (simplicité de l'explication, petit nombre des principes et des postulations). En fait, les théories se distinguent entre elles par leur *ontologie* : par exemple, la psychologie mentaliste donne à l'esprit la même structure ontologique qu'à la nature, car elle parle de changement d'état d'esprit comme elle parlerait d'un changement d'état de l'eau contenue dans la bouilloire. Autrement dit, ce n'est pas la logique qui tranche, mais la métaphysique. Le fait qu'une théorie soit satisfaisante du point de vue de la forme logique (hypothético-déductive) ne suffit pas à l'imposer, ni même à la rendre plausible. Le problème est plutôt de savoir si les modèles proposés par une théorie respectent les conditions ontologiques d'une description du mental : et ces conditions ontologiques doivent être fixées à titre de principe, donc par une démarche philosophique de la part du psychologue ou de l'anthropologue. Ce dépassement des dogmes du positivisme classique est aujourd'hui largement acquis, du moins, me semble-t-il, dans les écrits de langue anglaise. On observe qu'en général les philosophes contemporains de l'esprit – qu'ils soient matérialistes ou non – prennent soin d'explicitement la métaphysique des modèles conceptuels qu'ils

23. *My Philosophical Development*, p. 117.

proposent, signe qu'ils aperçoivent le besoin de justifier ceux-ci du point de vue ontologique, donc du point de vue des catégories d'être auxquelles il est fait appel. D'ailleurs, s'ils ne l'apercevaient pas d'eux-mêmes, leurs critiques ne manqueraient pas de repérer cette faiblesse dans la présentation de leur thèse.

Ces éclaircissements permettent de revenir sur la question confuse du « monisme » opposé au « dualisme » cartésien. En réalité, il y a ici deux questions confondues en une seule. La plus importante des deux n'est peut-être pas la plus spectaculaire, celle qui consiste à se demander (comme Mersenne le demande déjà à Descartes) si un corps peut penser. Nous parlons d'identifier le mental au physique : c'est la position du « monisme matérialiste » opposé au « dualisme ». D'un autre côté, nous parlons d'états mentaux et d'états physiques. Est-ce le même sens du mot « état » dans les deux cas ? Pour qu'il y ait le dualisme des substances, ou « dualisme cartésien », il faut en effet que ce soit dans le même sens qu'on parle de substance pour l'esprit et pour le corps, faute de quoi on ne pourrait pas procéder à un recensement des types de substance. De sorte que ce sera aussi dans le même sens du mot « état » (et avec les mêmes conséquences) qu'on attribuera des états à la chose étendue et à la chose pensante. Derrière le dualisme cartésien il y a donc nécessairement, du point de vue des catégories utilisées, un *monisme métaphysique*.

Le monisme proprement métaphysique n'est pas de compter le nombre de substances ou d'existences, et de trouver qu'il n'y en a qu'une (de déterminer que le Tout est Un et non pas Plusieurs, pour parler comme Russell à la fin du siècle dernier). Le monisme est de dire qu'il y a une seule classe ontologique, à l'intérieur de laquelle se fera l'inventaire. Je parlerai donc de monisme métaphysique pour désigner le présupposé suivant : un seul système de concepts métaphysiques suffit à tous les besoins. Par concepts métaphysiques nous entendons justement les notions générales de *substance*, d'*état*, de *procès*, de *capacité*, d'*opération*, de *mode*, d'*attribut*, de *relation*, de *propriété intrinsèque*, notions qui sont étudiées depuis Aristote dans les traités de métaphysique. Ces concepts sont ceux qu'on doit mettre en œuvre dans les distinctions, classifications, énumérations, corrélations, etc., auxquelles on procédera dans le domaine d'enquête qu'on s'est fixé (par exemple, celui des phénomènes de l'esprit).

Chez Descartes, le dualisme dont on parle tant suppose évi-

demment qu'on puisse compter les substances : le terme « substance » est donc doté d'un sens unique, un sens qui rend possible le dénombrement. Si Descartes parle d'une différence réelle entre l'âme et le corps, c'est parce qu'il suppose qu'on peut concevoir ces deux choses comme séparées : l'une pourrait exister sans l'autre. C'est pourquoi il conclut que le sujet d'attribution des opérations mentales doit être considéré comme une substance à part entière. Aussi n'y a-t-il aucun paradoxe à faire remarquer que le dualisme cartésien de l'âme et du corps suppose manifestement un monisme métaphysique, puisqu'on peut dénombrer les substances, qu'elles soient matérielles ou spirituelles. Un seul système de catégories suffit pour dire tout ce qu'il y a, dans quelque domaine que ce soit.

Ce point est d'ailleurs reconnu par l'historien. Guérout fait remarquer que chez ce philosophe, célèbre par son « dualisme », il y a en fait le plus complet parallélisme entre la science des choses matérielles et celle des esprits. La même philosophie fonde la physique rationnelle et la psychologie rationnelle. Autrement dit, il y a une structure métaphysique neutre, qui se retrouve dans le discours sur les corps étendus et dans le discours sur les esprits. Comme l'écrit Guérout :

Les mêmes concepts président à la vraie science des choses physiques et à celle de la pensée claire et distincte. Les *Regulae* semblent témoigner que la pensée est, tout autant que le mouvement, justiciable de notions s'inspirant de la géométrie analytique, qui font contraste avec nos perceptions courantes des choses, mais sont précisément destinées à en rendre compte. Or, de même que les instants indivisibles de mouvement élémentaire, dont la sommation donne le mouvement réel, c'est-à-dire temporel, ne comportent chacun aucun parcours assignable et sont finalement des repos, de même les intuitions instantanées dont la somme constitue ma pensée dans le temps sont des repos intemporels qui contrastent avec « le mouvement continu de la pensée » [*Regulae*, Regula 3, A.T., X, p. 370] laquelle est « toujours temporelle comme mes actions »<sup>24</sup>.

Ce monisme du système conceptuel général est plus décisif que les divisions ultérieures entre les propriétés qui relèvent du corps et celles qui relèvent de l'âme. Déjà le fait d'appliquer partout le même schéma du sujet et de ses attributs (propriétés) témoigne

24. Descartes selon l'ordre des raisons, t. I, L'âme et Dieu, p. 102.

d'une tendance à soumettre toute description à une seule et même métaphysique générale.

Chez les philosophes du XX<sup>e</sup> siècle, le monisme métaphysique domine largement (que ces philosophes soient ou non partisans d'une réduction de l'esprit à la nature). On peut citer : Husserl avec l'idée d'une « ontologie formelle » coiffant des « ontologies régionales », les positivistes logiques (avec l'idée de l'« unité de la science ») et, aujourd'hui, tous ceux qui estiment que toute description doit pouvoir se faire en termes de théorie des ensembles ou de calcul des prédicats du premier ordre.

Parmi les philosophes qui ont contesté le monisme métaphysique, on citera en particulier Heidegger (avec la distinction des « prédicamentaux », ou catégories, et des « existentiels ») et Wittgenstein (dont la « philosophie de la psychologie » repose sur la critique des notions courantes de « procès mental » et d'« état mental »). Plus près de nous, plusieurs philosophes qui ont repris ce motif wittgensteinien, certains dans un contexte aristotélicien<sup>25</sup>, d'autres dans le contexte d'une philosophie des sciences<sup>26</sup>.

En France, le poids de l'héritage cartésien, et le prestige du formalisme mathématique, ne favorisent guère la critique du monisme métaphysique. On peut citer pourtant la critique faite par Cornelius Castoriadis de l'illusion d'une métaphysique générale. Chez lui, cette critique des prétentions du philosophe à dégager un système des catégories qui aurait valeur universelle est conjuguée avec une critique de l'idée que la réalité authentique est la réalité déterminée. C'est justement parce qu'on se figure que la théorie des ensembles nous donne une ontologie générale qu'on juge irréel tout ce qui ne se plie pas bien aux conditions d'une description extensionnelle (ou « ensembliste-identitaire »), tout ce qui, du coup, paraît vague ou indéterminé<sup>27</sup>.

Dans le contexte de l'étude de l'esprit, la question métaphysique n'est donc pas d'abord de compter des genres de substance (au sein d'une unique catégorie de la substance). Elle est de savoir

25. Voir : Peter Geach, *Mental Acts : Their content and their objects*, p. 3 ; Anthony Kenny, *The Legacy of Wittgenstein*, p. 116.

26. Voir : Rom Harré, *Social Being : a Theory for Social Psychology*, ainsi que *Personal Being : a Theory for Individual Psychology*.

27. « Il n'existe pas de catégories transrégionales » (*L'institution imaginaire de la société*, p. 253). J'ai commenté quelques-uns des textes dans lesquels cette critique est exposée dans mon étude sur la pensée de Castoriadis, à laquelle je me permets ici de renvoyer : « Un renouveau philosophique », *Revue européenne des sciences sociales*, XXVII, 1989, n° 86, p. 69-85.

si les sciences de l'esprit doivent travailler avec le même système de catégories que les sciences de la nature, ou que certaines d'entre elles. Ou bien si elles ne doivent pas plutôt partager leur système métaphysique avec des disciplines traitant des affaires humaines, telles que le droit, la rhétorique, ou l'histoire.

## 10. La métaphysique des états d'esprit

*Penser, se représenter, opiner, projeter, vouloir, désirer* sont communément tenus pour des états de l'esprit de quelqu'un. Mais il n'y a pas encore grand-chose à tirer de cette vague classification car la notion ordinaire d'état n'est pas prise dans une acception bien déterminée. Un simple coup d'œil sur le lexique montre qu'elle est des plus élastiques. L'état d'esprit, au gré des auteurs, peut être l'humeur (gaieté, mélancolie), l'attitude (résolution, indécision), le niveau de l'activité (torpeur, surexcitation), et aussi l'« état mental » au sens de l'état de santé ou de trouble.

Chez les philosophes, on peut, je crois, distinguer déjà deux acceptions. Dans un usage neutre, la notion d'état d'esprit signale seulement qu'on est passé dans l'ordre de la description psychologique : elle n'introduit comme telle aucune classification du vocabulaire de cette description. D'autre part, la notion est également utilisée dans une acception qu'on pourrait dire prédicamentale, dans la mesure où elle vise à introduire des distinctions de catégorie, principalement l'opposition des procès et des états.

Le premier sens à enregistrer est le sens neutre. Il peut arriver que le fait de parler d'un état mental dans cette acception ne signale encore aucune philosophie de l'esprit particulière, sinon celle-ci : quelqu'un est dans un état mental s'il y a des faits psychologiques qui le concernent, s'il y a des descriptions contenant des verbes psychologiques qui s'appliquent à cette personne. On ne s'est pas encore prononcé sur la réalité psychologique qui faisait la différence entre une personne à laquelle convient la description et une autre personne à laquelle elle ne conviendrait pas. On a seulement exclu que les phrases qui se donnent pour

des descriptions intentionnelles soient privées de toute valeur descriptive. La seule philosophie de l'esprit ici exclue est donc celle pour laquelle les gens n'ont pas d'opinions, pas d'inclinations, pas de préférences, pas d'habitudes, pas de projets. Jusqu'ici, la notion d'état est philosophiquement vide. Comme le note Putnam<sup>1</sup>, on peut décider de poser qu'un état psychologique est un état qu'étudie la psychologie, et, dans ce cas, toute étude psychologique de votre personne devient une étude de vos états psychologiques.

Supposons qu'on dise : la différence entre deux personnes A et B, c'est que A désire gagner, alors que B ne le désire pas. On a posé par là une différence psychologique entre A et B. Un philosophe insiste-t-il pour qu'on exprime cela dans la forme suivante : la différence est entre un *état mental* de A et un *état mental* de B ? S'il le fait, nous lui demanderons en quoi cette tournure constitue plus qu'une variante stylistique de la déclaration précédente. Est-ce que l'introduction d'un idiome dans lequel les propriétés psychologiques des gens sont rebaptisées « états » offre un intérêt philosophique ? Est-ce qu'il nous éclaire sur le mental en nous permettant d'indiquer par ce mot le mode d'être de ce qu'on attribue aux gens ? Si c'est le cas, nous passons à la signification prédicamentale, de la notion d'état : en effet, une classification est introduite, en vertu de laquelle ce qui est conçu comme un état d'un sujet du même coup ne l'est pas comme un acte de ce sujet, ni comme une relation de ce sujet, etc.

Wittgenstein<sup>2</sup> souligne ceci : dans cette affaire, les choses se décident bien avant qu'on n'en vienne à se prononcer pour telle ou telle doctrine. Elles se décident dans le simple fait d'accepter de parler de procès mental ou d'événement mental. On peut ajouter : et encore d'état mental. Le tour de passe-passe du mentaliste est de nous imposer une métaphysique des états et des procès (calquée sur la philosophie naturelle des procès physiques et des choses qui sont dans tel ou tel état).

On lit souvent chez Wittgenstein des observations comme celle-ci : « Voir n'est pas une action, mais un état (remarque grammaticale »<sup>3</sup>. Quel est l'intérêt d'une remarque « grammati-

1. *Mind, Language and Reality, Philosophical Papers*, vol. II, p. 220.

2. *Recherches philosophiques*, § 308 (cité *supra*, chap. 4, n. 16).

3. « *Sehen ist keine Handlung, sondern ein Zustand (Grammatische Bemerkung)* » (*Remarques sur la philosophie de la psychologie*, t. I, § 1). Wittgenstein parle aussi de *Seelenzustand* (t. II, § 722) et de *Geisteszustand* (t. II, § 247).

cale» de ce genre? Il est d'attirer notre attention sur une différence de catégorie. Si un verbe signifie un procès ou une opération, on peut demander quand cela se fait, selon quel rythme et si nous en avons le contrôle. Mais, s'il signifie un « état », les questions à poser sont différentes. On opposera ici les états mentaux de quelqu'un aux épisodes de sa vie mentale.

Toutefois, cette première division ne suffit pas, car, parmi les états, c'est-à-dire les formes de vie mentale qui ne sont pas des actes ou des procès actuels, on doit de nouveau diviser. L'opposition de épisodes et des états – qui est en somme celle dont use Gilbert Ryle dans *Le concept d'esprit* – sert surtout à refuser que différents verbes psychologiques signifient un acte du sujet ou un épisode de sa vie mentale. L'autre classe, celle des états, ressemble jusqu'ici à un fourre-tout. Pour introduire un début de classification dans tout ce qui n'est pas épisode, Wittgenstein use du critère de la temporalité propre à chaque verbe psychologique. Correspondent à des verbes d'état psychologique : la crainte, la tristesse, la joie, la colère<sup>4</sup>. On aurait parlé, dans la rhétorique antique, de passions. Ce qui permet de parler d'état, ce sont les traits de la temporalité de ces passions. L'un de ces traits est le mode de présence du pathos considéré, du point de vue d'une chronologie : nous pouvons ne pas avoir décoléré de toute une journée, entre le moment où une certaine nouvelle irritante nous a été communiquée et un autre incident qui a fait que notre attention s'est portée sur un autre sujet de préoccupation. Il est donc naturel de parler d'état : il faut qu'un état commence, qu'il s'établisse<sup>5</sup>, qu'il dure (de façon continue ou par intermittence) et qu'il prenne fin. Il s'ensuit que les verbes intellectuels « se proposer », « croire », « penser », ne sont pas en général des verbes d'état. On ne peut pas être *sans interruption* dans un état qui serait de croyance, ou dans un état de projeter quelque chose. Non pas qu'on cesse de croire une chose quand on se met à penser à une autre : mais la continuité de la croyance n'est pas celle d'un état.

Aucun des deux sens qui viennent d'être relevés ne nous donne encore l'acceptation requise pour une psychologie conforme au programme du matérialisme psychique. Et cela pour une raison

4. *Remarques sur la philosophie de la psychologie*, t. II, § 177.

5. *Ibid.*, t. II, § 388.

qui a été bien mise en lumière par un article important de Putnam sur le « sens du sens »<sup>6</sup>.

L'article de Putnam porte avant tout sur la sémantique. Mais, traditionnellement, les questions sémantiques (et plus généralement logiques) et les questions psychologiques ont souvent été confondues, de sorte qu'il est normal qu'une discussion sémantique ait des conséquences en philosophie de l'esprit.

Putnam croit pouvoir déceler deux préjugés gouvernant les théories classiques de la signification. Premier préjugé : connaître le sens d'une expression, c'est se trouver dans un état psychologique. Deuxième préjugé : le sens d'un mot (qu'on pourrait trouver dans un dictionnaire) détermine son extension (c'est-à-dire l'ensemble des choses auxquelles ce mot s'applique).

Dans les textes qu'il écrit à cette époque, Putnam vise surtout à réfuter le deuxième préjugé : ce n'est pas le sens qui détermine l'extension, c'est même l'inverse, en vertu de sa « théorie causale » de la signification des noms propres. Les noms de genre trouvent un usage en se voyant conférés à des échantillons prélevés chacun dans la nature : quant au sens du terme, il ne peut qu'évoluer avec les progrès dans la connaissance des propriétés des choses du genre correspondant à l'échantillon qui a reçu le premier le nom en question.

C'est pour renverser le premier des deux préjugés mentionnés ci-dessus que Putnam est amené à présenter quelques idées sur la psychologie. Il fait remarquer que la psychologie de la philosophie traditionnelle est en général solipsiste. Une psychologie est solipsiste si les propriétés mentales dont elle traite ont le caractère suivant : il ne serait pas contradictoire d'attribuer ces propriétés à un individu quelconque Paul, même si Paul était la seule créature existante. Il va de soi que l'immense majorité des attributions psychologiques ordinaires ne sont pas de cette nature. Si Paul voit un chien, ou s'il se souvient de son chien, il faut qu'il y ait quelqu'un (Paul) pour voir, ou se souvenir, et quelqu'un (le chien) pour servir d'objet à ces actes mentaux. Du point de vue grammatical, on dira que les seuls verbes psychologiques susceptibles de présenter un caractère solipsiste sont les verbes monovalents (ou intransitifs) : ainsi, « souffrir », « être songeur », « s'ennuyer » (dans le sens intransitif). La psychologie solipsiste nous demande donc de *reconstruire* tous les verbes psychologiques

6. « The "meaning" of meaning » (1975), repris dans *Mind, Language and Reality*.

comme des verbes monovalents, ou verbes d'état : par exemple, le verbe « voir » est à reconstruire comme verbe attribuant à un sujet l'état d'avoir une représentation visuelle de quelque chose (sans que l'objet vu doive exister ailleurs que dans cette représentation). L'attribution d'un état doit respecter le présupposé d'un isolement ou d'une solitude du sujet psychologique. Nous reconnaissons cette psychologie : elle est tout simplement la philosophie mentale, celle qui veut que l'esprit soit isolé ou détaché (pour que son contenu puisse être pris en considération dans l'explication du cours des choses).

Putnam donne l'illustration suivante<sup>7</sup>. Soit l'état psychologique qui consiste à être jaloux. Dans l'emploi ordinaire, dit-il, on usera d'une expression qui nous impose d'admettre l'existence de plusieurs personnes. Le schéma *x est jaloux de y au sujet de z* est de forme triadique : ce drame psychologique familial réclame trois personnages. Or la psychologie traditionnelle pratique ce que Putnam appelle, reprenant l'expression de Carnap, un « solipsisme méthodologique » : l'état attribué par un prédicat qui aurait le schéma ci-dessus n'est pas considéré comme un état psychologique dans un sens acceptable par une philosophie mentaliste. Cet état met en effet en cause plusieurs personnes. L'état du jaloux (*x*) n'est pas un état de jalousie, au sens ordinaire du mot, s'il n'y a pas rival (*y*) dont il est jaloux, et quelqu'un (*z*) qui fournit le motif de cette jalousie. Du coup, l'état de *x* dépend de ce qui se passe autour de lui, et nous sortons donc de la sphère mentale qu'il s'agissait d'isoler.

Putnam propose de dire que les états psychologiques, dans l'acception ordinaire de ces termes, sont pris « au sens large », et que la théorie psychologique des classiques exige, pour des raisons de méthode, de les reformuler comme des états « au sens étroit ». La jalousie, prise au sens étroit (ou, comme on dira ensuite, prise comme un *état étroit*), ne doit concerner qu'un seul individu. Dans ce cas précis, il faudrait imaginer quelqu'un qui passerait par une crise de jalousie à l'égard d'un rival peut-être imaginaire et à propos d'un objet d'amour qui, lui aussi, pourrait ne pas exister.

Les procédés par lequel les philosophes de la tradition mentale détachent l'esprit – hallucination, suspension de la croyance – réduction phénoménologique – ont en commun de substituer un

7. *Ibid.*, p. 220.

expérience à une relation. Normalement, l'attribution de la jalousie se fait par une relation : qui est jaloux, de qui, au sujet de qui ? Mais la forme logique de l'attribution d'un sentiment de jalousie permet de distinguer les relations entre les personnages du drame et le vécu pénible du jaloux.

Ayant distingué les états mentaux *lato sensu* des états mentaux *stricto sensu*, Putnam va s'employer à montrer que la connaissance du sens d'un mot ne peut pas être un état mental *stricto sensu*. L'argument consiste à faire remarquer qu'un esprit détaché du monde (un esprit dont la description doit se faire selon la « méthode solipsiste ») *ne sait pas de quoi il parle*, comme on peut le faire voir par divers exemples. Et, s'il ne sait pas de quoi il parle, il ne connaît pas non plus la valeur sémantique des mots qu'il emploie. Donc la connaissance du sens des mots ne peut pas être un « état psychologique » dans le sens revendiqué par la psychologie traditionnelle, à savoir un état pris au sens étroit. La sémantique du langage commun doit être « externaliste ».

Cet argument a été illustré par diverses fictions qui ont été cent fois reprises dans la littérature philosophique contemporaine. Putnam lui-même a proposé une fantaisie sur le cas de deux planètes jumelles qui seraient indiscernables du point de vue des premières apparences, mais qui, à l'insu des habitants de l'une et l'autre planètes, seraient différentes du point de vue de l'essence réelle de telle ou telle substance. Par exemple, tout serait phénoménalement semblable sur la terre et sur sa jumelle, mais il y aurait une différence physique qui échapperait à tout le monde (Putnam situe son histoire avant la naissance de notre chimie) : l'eau serait H<sub>2</sub>O sur notre terre et autre chose sur l'autre planète, de sorte que les gens feraient référence à des réalités différentes en parlant ici et là de ce qu'ils appellent « eau ». Sur terre, le mot « eau » sert à parler de l'eau (H<sub>2</sub>O), tandis que sur la planète jumelle le sens d'un mot indiscernable « eau » est tout différent, puisque le mot est appliqué à ce qui, par hypothèse, n'est pas chimiquement de l'eau.

Ces fictions font appel à des cas fantastiques de dédoublement ou de gémellité, parce qu'il s'agit de construire une situation du type que voici : deux personnes sont, d'après les données de l'histoire inventée, dans le même état mental étroit, alors qu'elles vivent dans des mondes qui sont réellement différents. Ce qui nous intéresse ici est le principe de l'argument. Un esprit détaché du monde est un esprit dont les états sont étroits : il ne sait donc

pas de quoi il parle. En effet, cet esprit tient pour équivalentes des situations qui *paraissent* identiques, ou, si l'on préfère, cet esprit tire les mêmes conclusions lorsqu'il fait les mêmes expériences, et cela même si ces expériences pour lui identiques sont faites dans des conditions qui, sans qu'il s'en aperçoive, sont différentes. Il suffit donc de construire une situation logiquement concevable dans laquelle deux réalités différentes A et B donnent lieu aux mêmes épiphénomènes. L'esprit isolé ne sait pas s'il parle de A ou s'il parle de B. Il n'a aucun moyen de le savoir et n'a d'ailleurs aucune raison particulière de soupçonner que deux réalités sont en cause là où il ne voit qu'une seule et même chose.

Si savoir le sens des mots, c'est savoir de quoi l'on parle, il n'y a que deux façons de connaître le sens des mots. La première est d'avoir la connaissance complète et ultime de toute chose. La deuxième est d'avoir des rapports de types physiques avec les choses dont on parle, de telle sorte qu'on puisse pointer vers un exemplaire du genre de choses et dire que, quel que soit le verdict ultime de la science sur l'essence réelle de ce qu'on montre, c'est de cela que l'on entend parler quand on emploie tel ou tel mot. D'où la conclusion de Putnam, tournée en un slogan et devenue célèbre sous cette forme : les significations ne sont pas dans la tête (*meanings ain't in the head*)<sup>8</sup>.

Qu'est-ce que Putnam a montré ? Ceci : contrairement à ce que laissait entendre une certaine propagande cognitiviste, il n'est pas nécessaire d'être immatérialiste en philosophie de l'esprit pour récuser l'idée d'une science des états psychologiques internes. Le point en question n'a rien à faire avec une dispute portant sur le matériau dont sont faites les entités mentales : cette façon de présenter les choses est une simple manœuvre de diversion. De quoi s'agit-il donc ? L'argument montre donc qu'un état ne peut pas être à la fois intentionnel et interne : s'il est interne, il n'est pas évaluable sémantiquement, mais, s'il est évaluable sémantiquement, il n'est pas interne. (Putnam entend ici l'intentionnalité dans un sens sémantique : une pensée est au sujet de l'eau si la chose sur laquelle elle porte est de l'eau.)

L'argument de Putnam a suscité une vaste discussion, parce que son importance a été aussitôt aperçue. Si cet argument est probant, alors la « percée conceptuelle » et « théorique » de la

8. *Ibid.*, p. 227.

nouvelle philosophie mentale de l'esprit doit être réduite aux proportions d'une mode passagère. Il est piquant de se souvenir que c'est justement Putnam qui avait lancé plusieurs de ses élèves dans la voie de la psychologie des états fonctionnels d'une machine logique (fondée sur la distinction du *hardware* et du *software*, du matériel et du logiciel)<sup>9</sup>.

L'intérêt de l'argument de Putnam est qu'il nous permet de reconnaître, par-delà les engouements pour les « paradigmes » tirés de tel ou tel dispositif technique, les motifs philosophiques du solipsisme dit « méthodologique ». On comprend que ce solipsisme n'est pas du tout *methodologique*. Bien entendu, il ne s'agit pas d'un solipsisme ontologique, car le psychologue ne doute pas un seul instant de l'existence du monde extérieur au sujet qu'il étudie, et il ne suppose pas non plus que son sujet éprouve des doutes à ce propos. Pourtant, l'appeler « méthodologique », c'est donner à penser que les études psychologiques seront effectivement facilitées, ou même rendues possibles, si l'on fait comme si le sujet à étudier est le seul individu au monde. Mais on s'aperçoit rapidement qu'il ne s'agit pas ici de fixer une méthode pour la recherche en psychologie telle qu'elle se pratique en fait. Il s'agit de prescrire une méthode pour la recherche psychologique telle qu'elle *devrait* se pratiquer, de façon à se conformer aux canons de la philosophie mentaliste.

Le solipsisme que l'on dit méthodologique est imposé par une métaphysique. Les états psychologiques doivent être pris au sens étroit, non pas pour être des états *psychologiques*, mais pour être des états *internes* du sujet ou de l'organisme ou du système étudiés. L'exigence de rétrécir la réalité psychologique de la jalousie ou de la perception à ce qui resterait donné même si le milieu extérieur était différent est une exigence dogmatique, un « il faut » imposé par le système, non par la chose même.

C'est ce que Putnam nous dit fort bien dans son rappel de ce qui doit être ici entendu par le terme technique « état ». Dans le sens de ce terme qui est ici à retenir, un état est « un prédicat à deux places qui prend pour arguments un individu et une date »<sup>10</sup>. Putnam donne des exemples d'états : être grand de cinq pieds, avoir mal quelque part, savoir l'alphabet, être à cent kilomètres de Paris. Une différence saute ainsi aux yeux : parmi

9. Putnam s'explique sur son passé fonctionnaliste et la critique qu'il en fait aujourd'hui dans *Representation and Reality* (1988).

10. *Mind, Language and Reality*, p. 219.

les propriétés qu'on attribue à une chose sous le nom d'états, certaines peuvent jouer un rôle dans une explication causale et d'autres non. On peut dire pourquoi un meuble ne peut pas passer par la cage d'escalier : il est trop grand. On peut dire pourquoi l'animal recule lorsqu'il touche la clôture électrique : il ressent la douleur de la décharge électrique. Mais la propriété d'être éloigné de Paris n'explique pas *causalement* pourquoi, à cet instant, les rues sont mouillées ici, à cent kilomètres de Paris, alors qu'elles sont sèches à Paris (où il n'a pas plu). Les relations spatiales, par elles-mêmes, ne jouent aucun rôle causal. Tout cela revient à distinguer les états internes, ou intrinsèques, des états qui constituent ce que l'on appelle des « dénominations extrinsèques », justement parce qu'elles font seulement allusion à la réalité de la chose en donnant ses relations à des choses extérieures.

Dans la série ci-dessus des exemples donnés par Putnam, j'ai laissé de côté l'exemple psychologique. En quel sens savoir l'alphabet est-il un état ? Du point de vue de la distinction entre l'intrinsèque et l'extrinsèque, il s'agit d'un état mental intrinsèque : si quelqu'un sait l'alphabet, il le sait partout où il est. Savoir l'alphabet, c'est pouvoir déchiffrer des lettres (par exemple chez l'oculiste). Lorsqu'on analyse ce pouvoir, on doit distinguer les conditions à remplir par le sujet lui-même (il faut qu'il sache reconnaître les lettres) d'avec les conditions à remplir du côté des circonstances (il faut qu'il y ait des lettres présentées, qu'elles soient assez grandes pour être reconnaissables, qu'il y ait de la lumière, et ainsi de suite). Il semble donc que la connaissance de l'alphabet soit plutôt un état qu'une relation. Oui, mais s'agit-il d'un état dont l'expression aura pour forme logique « un prédicat à deux places qui prend pour arguments un individu et une date » ? Quelle sorte d'explication pourrions-nous fournir en invoquant la connaissance de l'alphabet ? On pourrait dire par exemple : ce touriste arrive à se retrouver dans Athènes parce qu'il sait l'alphabet grec, tandis que cet autre touriste n'arrive pas à déchiffrer les noms des rues. Nous faisons bien appel à une différence entre les deux touristes, qui n'est pas à chercher du côté des circonstances, mais de leur « condition interne ». Faut-il parler d'une explication par un état intrinsèque ? Ne serait-ce pas une explication par la vertu dormitive ? Une différence de catégorie se dessine ici entre les deux exemples psychologiques : la connaissance de l'alphabet ne peut pas être un état dans le sens

où la douleur est un état. Je ne peux pas être dans l'état de savoir l'alphabet grec dans le sens où l'on parle d'un état de ressentir une douleur dans la jambe gauche.

Même si la connaissance de l'alphabet grec est classée parmi les états intrinsèques de la personne, elle n'est pas un état dans le sens où l'état du corps est un état (mon corps manque d'eau), ni un état dans le sens où l'état vécu, expérientiel, de la personne est un état (je ressens une soif). Une différence majeure est celle du profil temporel des états. Les états physiques (intrinsèques et extrinsèques) sont à comprendre comme des états à *tel instant*. Quelle est la taille du meuble à l'instant présent ? Il y a une réponse. Le meuble peut changer de taille en très peu de temps. De même, la douleur peut se déclarer ou disparaître en un clin d'œil. Mais puis-je dire qu'à l'instant *t* je sais l'alphabet grec ? Certainement, si je sais l'alphabet grec, je le sais maintenant, alors que je l'ignorais avant de l'avoir appris. Toutefois, il saute aux yeux que la temporalité propre à un savoir n'est pas la même que celle de la taille d'un corps (qui peut être raccourcie en moins de temps qu'il n'en faut pour le dire), et qu'elle ne ressemble pas non plus à celle d'une douleur (qui apparaît, disparaît, s'interrompt, varie en intensité). Il n'y a pas un instant que l'on puisse pointer sur la ligne biographique du sujet, qui séparerait l'époque où il ne savait pas l'alphabet de l'époque où il le sait. Cela ne veut pas dire que le savoir de l'alphabet soit intemporel : ici, comme dans tous les apprentissages, il y a bien un moment où l'élève peut dire « maintenant, je sais », et en donner la preuve en récitant convenablement les lettres. La récitation de l'alphabet s'accomplit dans le temps naturel, celui que nous pouvons mesurer par le mouvement circulaire de l'aiguille d'une horloge. A tel instant donné, l'élève a récité l'alphabet jusqu'à la lettre *lambda*. Mais l'acquisition et l'appropriation personnelle de ce savoir n'est pas mesurable de la même façon. On peut demander : à l'instant d'avant sa profération de la lettre *alpha*, l'élève était-il en train de réciter l'alphabet ? La réponse est non : il n'avait pas commencé. Mais quel sens y aurait-il à demander : l'élève savait-il l'alphabet à l'instant précédent celui où il a déclaré « maintenant je sais l'alphabet » ?

Que retenir de la démonstration de Putnam ? Je crois qu'on doit considérer comme acquise la thèse elle-même (« les significations ne sont pas dans la tête »), mais qu'on doit noter en même

temps la limite de cet externalisme, pour reprendre le terme qui a été forgé pour désigner cette doctrine.

Les fictions de Putnam (terres jumelles, sosies, etc.) ont mis en lumière une chose : au fond de la philosophie classique de l'esprit, il y a une volonté de calquer la forme des attributions psychologiques sur celle des attributions physiques. L'attribution physique passe par la détermination d'un état, ce qui veut dire d'un état actuel (déterminable à tel instant). De même, la philosophie traditionnelle de l'esprit cherche à concevoir le mental sous la forme d'états d'une chose pensante, états qui doivent être intrinséquement déterminables à chaque instant.

Remplaçons les exemples de science-fiction par des cas plus familiers, et donc moins contestables. Prenons un cas de duplication industrielle : deux voitures de même modèle, sorties le même jour de l'usine. Du point de vue de la forme industrielle, elles sont par hypothèse identiques. Le principe d'individuation est clairement fourni par la matière. Sans doute les matériaux sont-ils seulement semblables, et pas exactement équivalents. Ils ne seront jamais entièrement équivalents, mais nous pouvons supposer que les différences soient aussi petites qu'on voudra : dans ce cas, les deux voitures neuves sont équivalentes du point de vue pratique. Un an après, au marché de l'occasion, elles ne sont plus équivalentes de ce même point de vue pratique. Elles ont un passé, qui intéresse vivement l'acheteur éventuel. L'une des deux a eu un accident et pas l'autre : laquelle est-ce ?

Nous nous intéressons à l'histoire des machines. Mais l'explication historique des différences entre les deux véhicules (entre les performances dont elles sont capables, entre leurs dispositions : tenue de route, tendance à caler, etc.) n'est pas le dernier mot sur la réalité de ces voitures. Nous nous intéressons, comme futurs conducteurs, à l'histoire en tant qu'elle affecte la constitution mécanique des véhicules. (C'est-à-dire que nous laissons de côté d'autres aspects historiques possibles : par exemple, le fait qu'une des voitures ait une grande valeur symbolique pour avoir appartenu à une vedette de la chanson, car cela n'affecte pas sa tenue de route.)

L'explication historique de la différence dans le comportement manifesté est une explication causale authentique, mais elle n'est pas une explication ultime. On peut dire que la voiture n° 1 a acquis une tendance à se déporter sur la droite à la suite d'un accident. Mais les tendances et les dispositions et les capacités

physiques sont fondées sur l'état *actuel*, à l'instant présent, du système physique de la voiture. L'histoire ne peut rien nous apprendre qu'un examen mécanique suffisamment attentif ne pourrait suffire à nous découvrir. Mieux, l'examen peut nous dire jusqu'à quel point l'histoire a affecté la réalité physique de la machine. Tout ce qu'on peut dire, c'est que, si nous connaissons l'histoire passée, notre curiosité sera éveillée : nous chercherons si la voiture a souffert de ses mésaventures passées, s'il en reste des traces durables. Les chocs subis par la voiture sont d'authentiques traumatismes physiques. Ce ne sont pas des épisodes psychologiques.

Considérons maintenant le rôle donné à l'histoire passée dans le cas de deux chiens de même espèce et de même âge : nous voulons savoir comment ils ont été dressés. La psychologie de l'animal-machine espère calquer la psychologie du chien sur la mécanique de la voiture : les tendances acquises sont conçues comme des mécanismes. Cette psychologie a-t-elle une chance de découvrir des mécanismes ? *Distinguo* : une tendance n'est certainement pas un mécanisme, mais on peut concevoir que la science explique la formation des tendances d'un animal par la construction d'un montage neurophysiologique. C'est d'ailleurs une question purement empirique que de savoir comment la psycho-physiologie animale va rendre compte de la formation des habitudes et des aptitudes acquises d'un animal.

Soit enfin le cas de deux jumeaux élevés par des parents différents : nous voulons savoir ce qui leur est arrivé pour apprécier leur « personnalité ». Le sens commun se contente (dans la rhétorique quotidienne) d'une reconstitution du passé des êtres humains. Dans quelles conditions ont-ils été éduqués ? Le passé fournit un motif par lequel on éclaire en termes généraux la conduite : par exemple, l'un des adolescents estime que « tout lui est dû », l'autre qu'il a une revanche à prendre. Nous n'en demandons pas plus. Nous n'attendons pas de savoir comment ces « traits de caractère » sont « inscrits » dans une machinerie mentale des sujets. La charge de la preuve revient donc à la psychologie mécaniste. Il n'est pas vrai que nous tenions un trait de la personnalité pour un « état à l'instant *t* ». On dira que la psychologie a bien le droit de faire l'hypothèse que certains traits aient un support, et soient donc liés à un état de la personne. Pourtant, même dans l'hypothèse où de tels états auraient été découverts, on devrait continuer à distinguer la personnalité, dont la temporalité requiert une multiplicité propre à nourrir une

narration en plusieurs épisodes, d'avec les états qui sont des états à tel ou tel instant du temps mesuré par la chronologie des mouvements naturels.

l'externalisme de P  
: réf. Putnam  
et mais en  
jeu des  
signif.  
l'histoire

Cependant, il convient de marquer la limite de la démonstration externaliste de Putnam. Son externalisme ne concerne que la fixation de la référence des termes d'une nomenclature pour les genres naturels (l'eau, le platane, l'orme, etc.). Les significations ne sont pas dans la tête : cela veut dire qu'il faut étudier les choses (les échantillons) et non les représentations (les définitions courantes) pour savoir de quoi nous parlons quand nous employons ces mots. Autrement dit, les significations ne doivent pas être confondues avec des représentations dans la tête : l'argument ne porte pas sur ces représentations dans la tête. Il y aurait donc bien des états mentaux étroits.

En somme, Putnam a transposé en termes modernes l'argument nominaliste tiré de l'omnipotence divine. Cet argument a pour but de pousser aussi loin que possible la dissociation de la réalité et des apparences. Il est possible, *de potentia Dei absoluta*, que je fasse l'expérience de percevoir intuitivement devant moi un cheval, et que pourtant il n'y ait pas de cheval devant moi : cela ne comporte pas de contradiction logique, et rien ne permet donc de dire que cela ne pourrait pas arriver si la Toute-puissance le voulait. Aujourd'hui, la toute-puissance appartient, au moins dans les exemples des philosophes, à la science des hommes de demain. Il est donc concevable qu'un jour des savants fous dérobent le cerveau d'un homme (sans qu'il s'en aperçoive), le conservent en laboratoire dans une cuve où ils puissent le stimuler grâce à un équipement sophistiqué, et qu'ils lui donnent, par exemple, le même état cérébral que celui d'un vacancier sur une plage de Californie. Dans ces conditions, le Cerveau de la cuve croit qu'il est en train de prendre un bain de soleil et de regarder les vagues de l'Océan. En réalité, il est dans la cuve d'un laboratoire, et ne voit donc ni vagues du Pacifique, ni ciel bleu<sup>11</sup>.

Autrement dit, les Apparences sont les mêmes pour le sujet dans la cuve et pour le sujet sur la plage : mais ce dernier seul a des pensées qui se rapportent à une plage et à l'océan Pacifique. Dans cette version de la fiction, nous continuons peut-être à

11. Hilary Putnam, *Reason, Truth and History*, chapitre 1 (trad. fr. sous le titre *Raison, vérité et histoire*).

parler du Cerveau dans la cuve comme d'un sujet psychologique (qui possède un « vécu »), car l'histoire nous demande de considérer que l'homme possesseur de ce cerveau a été réduit, par amputation de ses autres membres, à son cerveau. Il y a donc une continuité, au moins dans la fiction. Pourtant, l'argument est aussi appliqué à un cerveau dont toute la vie aurait été passée dans la cuve (et qui n'aurait jamais été une partie du système nerveux d'un organisme humain). Dans ce cas encore, l'argument accorde une vie mentale au Cerveau. Le Cerveau ne parle pas d'un arbre quand il croit voir un arbre devant lui, parce qu'il manque une liaison historique (causale) entre les arbres et lui. Donc – c'est la conclusion à laquelle Putnam voulait arriver –, le Cerveau ne peut pas non plus penser et dire qu'il est un Cerveau dans la cuve, car le mot « cuve », lorsque c'est lui qui l'emploie, ne s'applique pas à ce que nous appelons des cuves.

Putnam ne dit pas que le Cerveau n'a pas d'états mentaux. Il lui accorde une vie mentale *étroite*. Du point de vue des apparences déliées de toute attache à une réalité, ou, ce qui revient au même, du point de vue d'une psychologie des états mentaux étroits, nous pourrions être des cerveaux dans une cuve.

Putnam laisse donc en place la possibilité d'un Monde des Apparences. Le Cerveau dans la cuve est comme un Ego husserlien qui aurait mis entre parenthèses le monde monde naturel<sup>12</sup>. Ce sujet aurait le même vécu que nous. Du coup, l'argument a été mobilisé au service de la philosophie de la conscience. C'est ainsi que Searle peut écrire :

Le cerveau est tout ce que nous avons pour ce qui est de nous représenter le monde, et tout ce que nous pouvons utiliser dans ce but doit être dans le cerveau. Chacune de nos croyances doit être possible pour un être qui serait un cerveau dans une cuve, car chacun de nous est justement cela, un cerveau dans une cuve ; la cuve est le crâne, et les « messages » qui arrivent dedans arrivent par la voie de chocs sur le système nerveux<sup>13</sup>.

On voit que l'argument de Putnam a pour résultat d'humilier le solipsisme méthodologique, mais n'en fait pas ressortir le non-sens. Tout ce qu'il établit, c'est que ce solipsisme doit renoncer à parler sémantique, et qu'il doit renoncer à parler

12. *Reason, Truth and History*, p. 28-29.

13. *Intentionality, An Essay in the Philosophy of Mind*, p. 230 (trad. fr. sous le titre *L'intentionnalité, Essai de philosophie des états mentaux*).

psychologie, au sens ordinaire du mot. Mais il pourrait encore parler des apparences auxquelles serait livré le sujet, encore qu'il soit impossible de dire quoi que ce soit de ces apparences (car il faudrait user de notre langue, dont la sémantique est publique ou externaliste).

La conclusion de Putnam est externaliste. Du même coup, il dénonce l'illusion inhérente aux programmes mentalistes. Pour qu'une théorie puisse formuler des lois psychologiques, dans le sens où une science naturelle formule des lois, il faudrait qu'elle puisse se donner un système psychique clos, de façon à définir des états internes à ce système. Les lois consisteraient à fixer les conditions de l'obtention de ces états.

Mais, si les états psychologiques ne deviennent des états intrinsèques qu'à la condition de traiter le psychisme comme un système clos (solipsiste), alors l'essentiel nous échappe. Depuis trois siècles, on s'est essayé à faire de la psychologie « mentaliste » de ce genre : « trois siècles d'échec », conclut Putnam.

Fodor, non sans panache, va relever le défi. Les trois siècles d'échec dont parle Putnam sont pour lui trois siècles de grande fécondité théorique. Si Putnam, en parlant des penseurs qui ont échoué, vise « tout le monde de Locke et Kant jusqu'à Freud et Chomsky »<sup>14</sup>, alors, nous dit Fodor, je ne serais pas mécontent d'échouer comme eux. Le point ici disputé est en effet considérable s'il engage un jugement sur toute la philosophie de l'esprit depuis Locke jusqu'à Freud et Chomsky, et cela surtout si l'on tient compte des différences considérables qui séparent les penseurs cités. C'est un grand mérite de la discussion présente d'avoir révélé comment, derrière ce qui est trop souvent présenté comme un programme scientifique lié à des développements récents de notre neurologie ou de notre technique informatique, il y a en réalité une décision philosophique. La philosophie mentale a choisi de soutenir que l'esprit n'aurait pas de réalité s'il n'avait pas la réalité d'un système clos. L'esprit de quelqu'un fait sentir ses effets quand on se l'imagine détaché des circonstances.

14. *Representations*, p. 247.

## 11. Le détachement de l'esprit

Une psychologie des états mentaux est une psychologie solipsiste. En élevant cette objection contre la psychologie de l'esprit-ordinateur, Putnam signalait qu'elle valait aussi contre la psychologie des philosophes classiques, de Locke à Freud. Si une théorie exige que les significations soient dans la tête, elle sera solipsiste, et elle ne pourra donc prétendre restituer les descriptions psychologiques, au sens ordinaire du mot.

Vaillamment, Jerry Fodor va relever le défi lancé par Putnam à son école. Loin d'être un défaut de la psychologie cognitive, le fait d'embrasser une méthode « solipsiste » serait son point fort. Seule une psychologie solipsiste peut faire acte de candidature à un statut scientifique à part entière parmi les « sciences spéciales ». L'objection de Putnam devient chez Fodor une condition de la psychologie. En bref, une explication psychologique n'est pas une explication scientifique et causale si elle n'explique pas par des états étroits de la personne. Son article sur « le solipsisme méthodologique considéré comme une stratégie de recherche en psychologie cognitive »<sup>1</sup> est un document précieux pour notre discussion, puisqu'il revendique explicitement l'héritage de ce que j'ai proposé d'appeler la philosophie mentale.

Une fois encore, le but visé est de trouver le moyen de justifier l'idée d'une science qui donnerait des explications psychologiques, causales et naturelles. Il s'agit donc pour Fodor de montrer

1. Le texte de cet article est d'abord paru dans la revue *The Behavioral and Brain Sciences* (1980), 3, p. 63-73, accompagné, comme c'est l'usage dans cette revue, des commentaires d'une vingtaine de psychologues et de philosophes, avec les réponses de l'auteur. Il est repris par Fodor dans son livre intitulé *Representations*, auquel vont mes références.